



NATIONAL COMPREHENSIVE CENTER
FOR **TEACHER QUALITY**

Evaluating Teacher Effectiveness: Where Do We Go From Here?

Laura Goe, Ph.D.

Presentation to Learning First Alliance

May 12, 2011 ♦ Washington, DC

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

Today's presentation available online

- To download a copy of this presentation or look at on your iPad, smart phone or computer, go to www.lauragoe.com Publications and Presentations page.
 - Today's presentation is at the bottom of the page, along with the handout "Questions to ask about measures and models."

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
 - “Last hired, first fired” and job security based on tenure are under fire in Los Angeles, New York City, and Washington, DC, Illinois, Florida, Colorado, Tennessee and other states and districts
- Impacting privacy
 - 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

Definitions in the research & policy worlds

- Much of the research on teacher effectiveness doesn't define effectiveness at all though it is often assumed to be teachers' contribution to student achievement
- Bryan C. Hassel of Public Impact stated in 2009 that "The core of a state's definition of teacher effectiveness must be student outcomes"
- Checker Finn stated in 2010 that "An effective teacher is one whose pupils learn what they should while under his/her tutelage"

Goe, Bell, & Little (2008) definition of teacher effectiveness

1. Have high expectations for all students and help students learn, as measured by value-added or alternative measures.
 2. 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.
 3. 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.
 4. Contribute to the development of classrooms and schools that value diversity and civic-mindedness.
 5. 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.
-

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

Teacher evaluation

TEACHER CLASSROOM EVALUATION CHECKLIST (LONG FORM)

Note to Evaluation: This instrument is designed for a 45 minute or longer classroom evaluation visit. If less time is allocated for the evaluation, adapt the instrument by focusing on selected items of your choosing or use the Intermediate or Short Form of this instrument. Familiarize yourself with this instrument prior to the classroom visit.

School _____ Teacher: _____ Date: _____

Subject: _____ Grade: _____ Evaluator: _____

Evaluation # 1 2 3 4 5 (Circle One) Length of Evaluation: _____ minutes

Meeting set with teacher to discuss evaluation: Date _____ Time _____

Was an evaluation conference conducted? YES NO If yes, when? _____

Directions: The following checklist fairly represents the classroom behaviors and activities common to classrooms. Mark an "X" in the column that best represents the behavior observed.

Yes _____ Definitely observed

No _____ Not observed

Uncertain _____ Not sure

Not applicable (N/A) _____ Does not apply to this observation

GENERAL CLASSROOM MANAGEMENT

THE TEACHER:	YES	NO	UNCERTAIN	N/A
Maintains a warm, comfortable learning environment				
Establishes and communicates realistic expectations and well-defined behavioral expectations				
Maintains eye contact				
Displays self-confidence and self-control				
Maintains attendance				
Uses rewards with discretion				
Minimizes wasted time				
Students display understanding and cooperation				
Classroom is orderly				
Calls students by name				
Involved students in learning tasks				



Wherein we will consider the statement “When all you have is a hammer, everything looks like a nail.”

Research behind the push for new evaluation measures and systems

- 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)

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
-

Teacher observations: strengths and weaknesses

- **Strengths**

- Great for teacher formative evaluation (if observation is followed by opportunity to discuss)
- Helps evaluator (principals or others) understand teachers' needs across school or across district

- **Weaknesses**

- Only as good as the instruments and the observers
- Considered “less objective”
- Expensive to conduct (personnel time, training, calibrating)
- Validity of observation results may vary with who is doing them, depending on how well trained and calibrated they are

Example: University of Virginia's CLASS observation tool

	Emotional Support	Classroom Organization	Instructional Support
Pre-K and K-3	<p>Positive Climate</p> <p>Negative Climate</p>	<p>Behavior Management</p> <p>Productivity</p>	<p>Concept Development</p> <p>Quality of Feedback</p> <p>Language Modeling</p>
Upper Elementary/ Secondary	<p>Teacher Sensitivity</p> <p>Regard for Student (Adolescent) Perspectives</p>	<p>Instructional Learning Formats</p>	<p>Content Understanding</p> <p>Analysis and Problem Solving</p> <p>Quality of Feedback</p>

Example: Charlotte Danielson's Framework for Teaching

Domain 1: Planning and Preparation includes comprehensive understanding of the content to be taught, knowledge of the students' backgrounds, and designing instruction and assessment.

Domain 3: Instruction is concerned with the teacher's skill in engaging students in learning the content, and includes the wide range of instructional strategies that enable students to learn.

Domain 2: The Classroom Environment addresses the teacher's skill in establishing an environment conducive to learning, including both the physical and interpersonal aspects of the environment.

Domain 4: Professional Responsibilities addresses a teacher's additional professional responsibilities, including self-assessment and reflection, communication with parents, participating in ongoing professional development, and contributing to the school and district environment.

Example: Kim Marshall's Rubric

Planning & Preparation for Learning

	Highly Effective	Effective	Improvement Necessary	Does Not Meet Standards
a. Knowledge	Is expert in the subject area and has a cutting-edge grasp of child development and how students learn.	Knows the subject matter well and has a good grasp of child development and how students learn.	Is somewhat familiar with the subject and has a few ideas of ways students develop and learn.	Has little familiarity with the subject matter and few ideas on how to teach it and how students learn.
b. Strategy	Has a well-honed game plan for the year that is tightly aligned with state standards and assessments.	Plans the year so students will meet state standards and be ready for external assessments.	Has done some thinking about how to cover high standards and test requirements this year.	Plans lesson by lesson and has little familiarity with state standards and tests.

Research on observations: Danielson Framework

- Lots of research on Danielson Framework (1996) and whether its scores correlate with student achievement growth
 - Goe (2007) reviews many studies, most finding weak correlation (likely due to lack of rater training)
 - Kane et al. (2010) describes research linking observation scores with value-added scores (found relatively small but significant correlations)
 - Sartain et al. (2010) describes challenges in implementation, differences in researcher/principal ratings
 - Principals tend to see the evidence the same as researchers, but score teachers higher

Research on observations: CLASS

- Considerable research, mostly conducted by creators of CLASS
 - Howes et al. (2008): children's relationship with teachers, not teachers' qualifications, mattered
 - Pianta et al. (2007): "Children from nonpoor families and who scored high on achievement at 54 months were most likely to experience classrooms high in positive emotional or instructional climate throughout elementary school. Poor children were highly unlikely (only 10%) to experience classrooms with high instructional climate across multiple grades."

Teacher behaviors & practices that correlate with achievement

- High ratings on learning environment (classroom observations (Kane et al., 2010)
- Positive student/teacher relationships (Howes et al., 2008)
- Parent engagement efforts by teachers and schools (Redding et al., 2004)
- Teachers' participation in intensive professional development with follow-up (Yoon et al., 2007)

**IN MANY CURRENT TEACHER EVALUATION MODELS,
THESE ARE NEVER MEASURED.**

Value-added models

- Many variations on value-added models
 - TVAAS (Sander's original model) typically uses 3+ years of prior test scores to predict the next score for a student
 - Used since the 1990's for teachers in Tennessee, but not for high-stakes evaluation purposes
 - Most states and districts that currently use VAMs use the Sanders' model, also called EVAAS
 - There are other models that use less student data to make predictions
 - Considerable variation in “controls” used

A recent description of value-added

- “There are various methods for estimating teacher value-added, but all typically entail some variant of subtracting the achievement test scores of a teacher’s students at the beginning of the year from their scores at the end of the year, and making statistical adjustments to account for differences in student learning that might result from student background or school-wide factors outside the teacher’s control.” (Glazerman et al., 2011)

Promises of value-added...

- Value-added can provide useful evidence of teacher's contribution to student growth
 - “It is not a perfect system of measurement, but it can complement observational measures, parent feedback, and personal reflections on teaching far better than any available alternative.” Glazerman et al. (2010) pg 4
 - “...value-added estimates should best be used in combination with other indicators...when there are no serious consequences for individual teachers, administrators, or students.” Braun et al. (2010) p 54

...and perils of value-added

- “Any practical application of value-added measures should make use of confidence intervals in order to avoid false precision, and should include multiple years of value-added data in combination with other sources of information to increase reliability and validity.” Glazerman et al. (2010) pg 5
- “Type I and II error rates for comparing a teacher’s performance to the average are likely to be about 25 percent with three years of data and 35 percent with one year of data.” Schochet & Chiang (2010) abstract

Other recent entrants into the field of value-added analysis (besides SAS)

- University of Wisconsin Value Added Research Center (working with numerous districts, including Los Angeles and Chicago)
- Mathematica (working with Pittsburgh and Washington, DC)
- American Institutes of Research (working with Florida Race to the Top)

Cautions about using value-added for teacher evaluation

- Braun et al. (2010) “...value-added estimates should best be used in combination with other indicators...when there are no serious consequences for individual teachers, administrators, or students.”
- Schochet & Chiang (2010) “Type I and II error rates for comparing a teacher’s performance to the average are likely to be about 25 percent with three years of data and 35 percent with one year of data.”

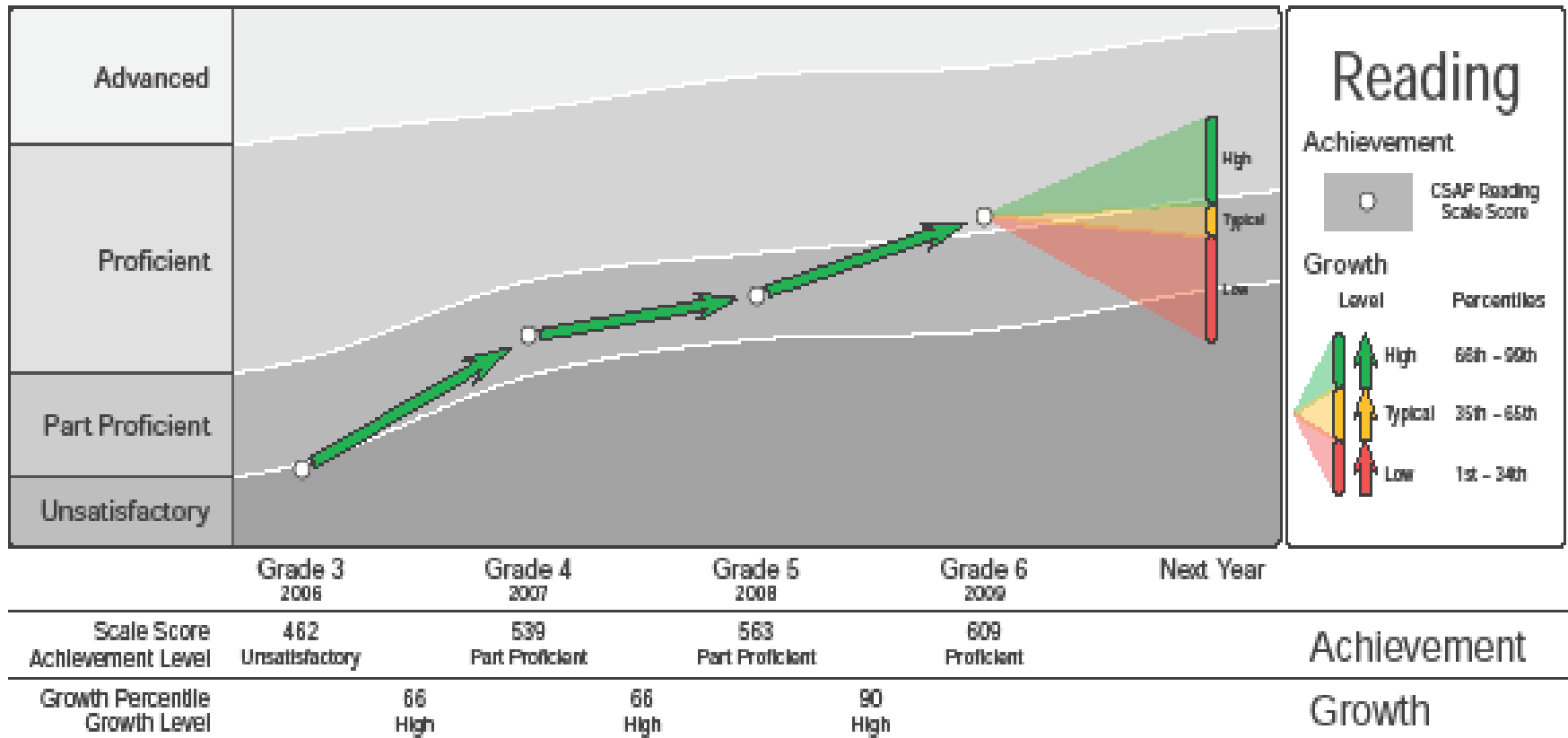
Technical considerations in using value-added for teacher evaluation

- Koedel & Betts (2009) suggest using multiple years of data for teacher evaluation to mitigate sorting bias; novice teachers cannot be evaluated under this system
- McCaffrey et al. (2009) “...there are significant gains in the stability [of teachers’ value-added scores] obtained by using two-year average performance measures rather than single-year estimates”

Most popular growth models: Colorado Growth Model

- Colorado Growth model
 - Focuses on “growth to proficiency”
 - Measures students against “academic peers”
 - Also called criterion-referenced growth-to-standard models
- The student growth percentile is “descriptive” whereas value-added seeks to determine the contribution of a school or teacher to student achievement (Betebenner 2008)

Sample student growth report: Colorado Growth Model



Slide courtesy of Damian Betebenner at www.nciea.org

What value-added and growth models cannot tell you

- Value-added models are really measuring *classroom* effects, not teacher effects
- Value-added models can't tell you why a particular teacher's students are scoring higher than expected
 - Maybe the teacher is focusing instruction narrowly on test content
 - Or maybe the teacher is offering a rich, engaging curriculum that fosters deep student learning.
- *How* the teacher is achieving results matters!

VAMs and other test-based growth models don't measure most teachers

- About 69% of teachers (Prince et al., 2006) can't be accurately assessed with VAMs/growth models
 - Teachers in subject areas that are not tested with annual standardized tests
 - Teachers in grade levels (lower elementary) where no prior test scores are available
 - Questions about the validity of measuring special education teachers and ELL teachers with VAMs

Federal priorities (August 2010)

- From “Race to the Top” and reiterated in the August 5, 2010 Federal Register (Vol. 75, No. 150) “Secretary’s Priorities for Discretionary Grant Programs”
 - Teachers should be evaluated using state standardized tests where possible
 - For non-tested subjects, other measures (including pre- and post-tests) can be used but must be “**rigorous and comparable across classrooms**” and must be “**between two points in time**”
 - Multiple measures should be used, such as multiple classroom evaluations

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	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>

Model highlight: Rigor

Austin's Reach Program includes a rubric for determining the rigor of teacher-created student learning objectives (SLOs)

Austin Reach Program: Rubric for Determining SLO Rigor (DRAFT)

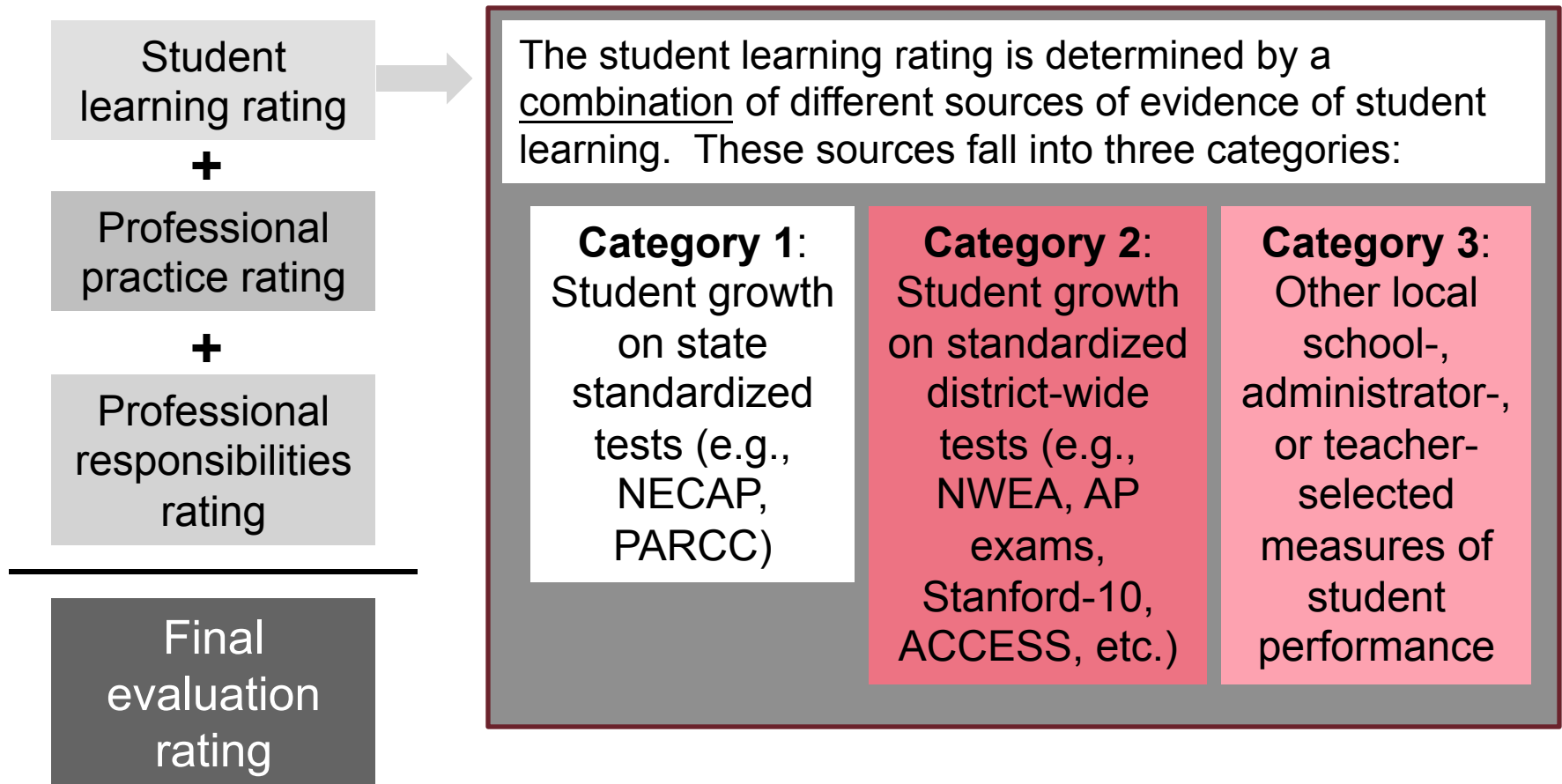
Student Learning Objective Rigor Rubric

4 Exemplary	3 Proficient	2 Progressing	1 Does not meet standard
<u>Assessment</u> <ul style="list-style-type: none"> Variety of levels of questions (Beginning, Progressing, Proficient, Advanced) At least one very challenging question Sufficient number of items Grade level appropriate Extends and deepens knowledge Measures what is intended 	<u>Assessment</u> <ul style="list-style-type: none"> Variety of levels of questions (Beginning, Progressing, Proficient, Advanced) Sufficient number of items Grade level appropriate Measures what is intended 	<u>Assessment</u> <ul style="list-style-type: none"> Addresses 2 or 3 levels of questions Spread of questions is insufficient Grade level appropriate Mostly measures what is intended 	<u>Assessment</u> <ul style="list-style-type: none"> Addresses only 1 level of questions Insufficient number of questions Not grade level appropriate Does not measure what is intended
<u>Objective</u> <ul style="list-style-type: none"> Reflects a high need Yearlong objective Grade level appropriate Deepens and extends knowledge for all students 	<u>Objective</u> <ul style="list-style-type: none"> Reflects a significant need Yearlong objective Grade level appropriate 	<u>Objective</u> <ul style="list-style-type: none"> Addresses a need Yearlong objective Grade level appropriate 	<u>Objective</u> <ul style="list-style-type: none"> Does not address a need Not a yearlong objective Not grade level appropriate
<u>Growth Target</u> <ul style="list-style-type: none"> Addresses more than 75% of students Substantial growth expected (2 or more years) Students and teachers exceeding expectations 	<u>Growth Target</u> <ul style="list-style-type: none"> Addresses 75% of students (exceptions for sped, small classes, etc) Significant individual growth (at least one year) Pushes students and teachers to exceed typical expectations 	<u>Growth Target</u> <ul style="list-style-type: none"> Addresses fewer than 75% of students Moderate individual growth (less than one year) Students and teachers barely meet expectations 	<u>Growth Target</u> <ul style="list-style-type: none"> Does not address 75% of students Minor individual student growth (less than ½ year) Students and teachers do not meet expectations

Model Highlight: Inclusive

Rhode Island's Framework ensures that all teachers, not just those in tested grades and subjects, are evaluated on their contribution to student learning growth

Rhode Island DOE Model: Framework for Applying Multiple Measures of Student Learning



Model highlight: Validity

One way New Haven, CT verifies validity of results is through placing scores on a matrix to look for mismatches that may indicate problems (with instruments, training, scoring, etc.)

New Haven “matrix”

		Student Learning Growth				
		1	2	3	4	5
Instructional Practice and Professional Values	1	1	1	2	3*	3*
	2	1	2	2	3	4*
	3	1	2	3	4	5
	4	2*	3	4	4	5
	5	3*	3*	4	5	5

Asterisks indicate a mismatch between teacher’s performance on different types of measures

Model highlight: Transparency

DC's Impact system publishes teacher handbooks that contain information about processes and scoring as well as the rubrics that will be used in all aspects of the evaluation

Washington DC IMPACT: Educator Groups

1. General Education Teachers with Individual Value-Added Student Achievement Data
2. General Education Teachers without Individual Value-Added Student Achievement Data
3. Special Education Teachers
- 3a. Special Education Teachers — Autism Program
4. Non-Itinerant English Language Learner (ELL) Teachers
5. Itinerant English Language Learner (ELL) Teachers
6. Shared Special Subject Teachers
7. Visiting Instruction Service Teachers
8. Student Support Professionals
9. Library Media Specialists
10. Counselors
11. School-Based Social Workers and Psychologists
12. Related Service Providers
13. Special Education Coordinators
14. Program Coordinators & Deans
15. Instructional Coaches
16. Mentor Teachers
17. Educational Aides
18. Office Staff
19. Custodial Staff
20. All Other School-Based Personnel

Model highlight: Training, opportunity to discuss results, growth opportunity

The Teacher Advancement Program (TAP) requires rigorous training and ongoing calibration of observers. After each observation, a 40-minute discussion between teacher and observer is scheduled, focusing on areas for teacher growth.

Teacher Advancement Program (TAP) Model

- TAP requires that teachers in tested subjects be evaluated with value-added models
- All teachers are observed in their classrooms (using a Charlotte Danielson type instrument) six times per year by different observers (usually one administrator and two teachers who have been trained as evaluators)
- Teacher effectiveness (for performance awards) determined by combination of value-added and observations
- Teachers in non-tested subjects are given the school-wide average for their value-added component, which is combined with their observation scores

Model highlight: Comparability, standards-based, growth over time

NYSUT ensures comparability by having “like” teachers examine and recommend appropriate standards-based measures. Measures must include a “pre” and “post” score to show student learning growth over time.

NYSUT Model

- Standardized test will be used as part of teachers' scores in appropriate grades/subjects
- "Group alike" teachers, meeting with facilitators, determine which assessments, rubrics, processes can be used in their subjects/grades (multiple measures)
- Assessments must focus on standards, be given in a "standardized" way, i.e., giving pre-test on same day, for same length of time, with same preparation
- Teachers recommend assessments to the district for approval
- District will consider and approve measures
- Consortium of districts share measures

Considerations

- Partner with national and regional comprehensive centers
- Engage stakeholders (teachers, administrators, parents, school board members, union representatives, business leaders, etc.) in decision-making processes early and often
- If lacking grade-level and subject standards, adopt such standards
- Conserve resources by encouraging districts to join forces with other districts or regional groups

Considerations

- Consider whether human resources and capacity are sufficient to ensure fidelity of implementation.
- Develop a communication strategy to increase awareness and buy-in (FAQs on website, public meetings, news “blasts” to email subscribers.
- Establish a plan to evaluate measures to determine if they can effectively differentiate among teacher performance
- Examine correlations among measures.
- Evaluate processes and data each year and make needed adjustments.

Observation instruments

Charlotte Danielson's Framework for Teaching

<http://www.danielsongroup.org/theframeteach.htm>

CLASS

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

Kim Marshall Rubric

[http://www.marshallmemo.com/articles/Kim%20Marshall%20Teacher%20Eval%20Rubrics%20Jan%](http://www.marshallmemo.com/articles/Kim%20Marshall%20Teacher%20Eval%20Rubrics%20Jan%20)

Growth Models

Wisconsin's Value-Added Research Center (VARC)

<http://varc.wceruw.org/>

SAS Education Value-Added Assessment System (EVAAS)

<http://www.sas.com/govedu/edu/k12/evaas/index.html>

Mathematica

[http://www.mathematica-mpr.com/education/
value_added.asp](http://www.mathematica-mpr.com/education/value_added.asp)

American Institutes of Research (AIR)

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

Colorado Growth Model

www.nciea.org

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](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>

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

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.

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