



NATIONAL COMPREHENSIVE CENTER
FOR **TEACHER QUALITY**

Models for Evaluating Teacher Effectiveness

Laura Goe, Ph.D.

California Labor Management Conference

May 5, 2011 ♦ Los Angeles, CA

Laura Goe, Ph.D.



- Former middle school teacher
- Graduate of UC Berkeley's Policy, Organizations, Measurement & Evaluation Doctoral Program
- Principal Investigator for 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

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

Questions to ask about multiple measures

Evaluating teacher effectiveness with student growth

- 1. Rigorous.** Are measures “rigorous,” focused on appropriate subject/grade standards? Measuring students’ progress towards college and career readiness?
 - 2. Comparable.** Are measures “comparable across classrooms,” ensuring that students are being measured with the same instruments and processes?
-

Questions to ask about multiple measures

- 3. Growth over time.** Do the measures enable student learning growth to be assessed “between two points in time”?
- 4. Standards-based.** Are the measures focused on assessing growth on important high-quality grade level and subject standards for students?

Questions to ask about multiple measures

Evaluating teaching practice and professional contributions

5. Teaching practice. Are measures aligned with teaching standards? Are evaluators (principals, others) properly trained to agree upon what good professional practice is? Do the measures include guidance for observing, collecting, and evaluating evidence to support decisions?

Questions to ask about multiple measures

Improving teaching and learning

- 6. Improve teaching.** Does evidence from using the measures contribute to teachers' understanding of their students' needs/progress so that instruction can be planned/adapted in a timely manner to ensure success?

Questions to ask about multiple measures

Informing professional growth opportunities

- 7. Professional growth.** Can results from the measures inform and be aligned with professional growth opportunities? Does the evaluation system ensure that teachers have an opportunity to discuss evaluation results with evaluators (principals, others)?

Questions to ask about teacher evaluation *models*

Characteristics of evaluation models

- 1. Inclusive (all teachers, subjects, grades).**
Do evaluation models allow teachers from all subjects and grades (not just 4-8 math & reading) to be evaluated with evidence of student learning growth according to standards for that subject/grade?

Questions to ask about teacher evaluation *models*

- 2. Validity.** Are mechanisms in place to evaluate the validity of the measures and the model itself and make appropriate adjustments over time?
- 3. Transparency.** Do all teachers know what is expected of them and what the measures and processes are for their evaluation?

Questions to ask about teacher evaluation *models*

Practical considerations for implementation

- 4. Capacity.** Will implementing this model create an undue burden on principals? Teachers? Evaluators? Are expectations for participants realistic given current roles, contracts, and other demands on time?
- 5. Resources.** How will states pay for the model requirements (training for evaluators, data analysis, etc? What share will districts pay for? Are federal dollars available?

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

	Model	Description
All teachers have their own growth scores	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
Only tests subjects/grades...	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>
	Tested subjects only	Teachers' contributions to student learning growth is determined with value-added; <i>other teachers are not evaluated with student outcomes</i>

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)

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)

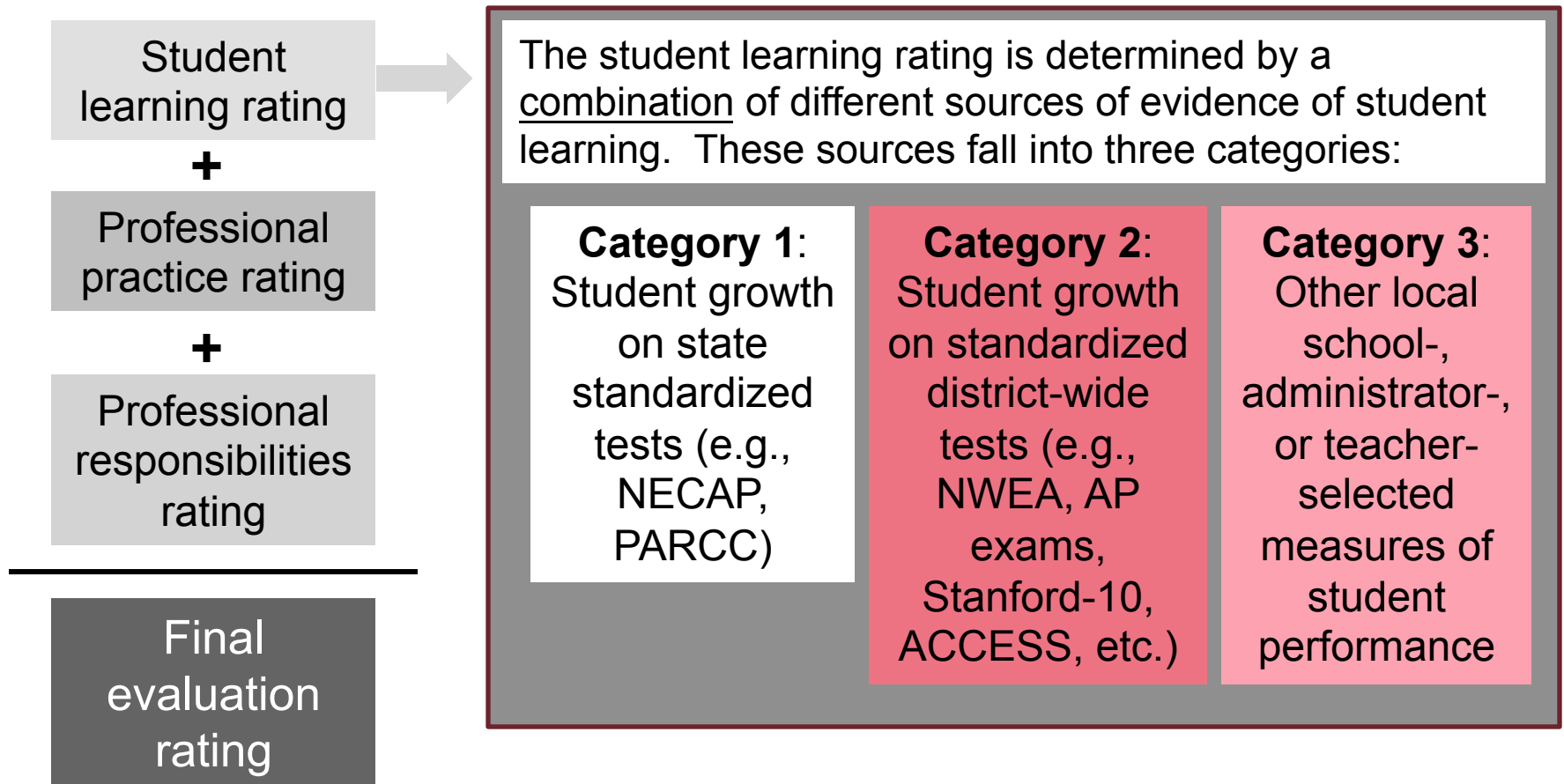
Student Learning Objective Rigor Rubric

4 Exemplary	3 Proficient	2 Progressing	1 Does not meet standard
<p><u>Assessment</u></p> <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 	<p><u>Assessment</u></p> <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 	<p><u>Assessment</u></p> <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 	<p><u>Assessment</u></p> <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
<p><u>Objective</u></p> <ul style="list-style-type: none"> Reflects a high need Yearlong objective Grade level appropriate Deepens and extends knowledge for all students 	<p><u>Objective</u></p> <ul style="list-style-type: none"> Reflects a significant need Yearlong objective Grade level appropriate 	<p><u>Objective</u></p> <ul style="list-style-type: none"> Addresses a need Yearlong objective Grade level appropriate 	<p><u>Objective</u></p> <ul style="list-style-type: none"> Does not address a need Not a yearlong objective Not grade level appropriate
<p><u>Growth Target</u></p> <ul style="list-style-type: none"> Addresses more than 75% of students Substantial growth expected (2 or more years) Students and teachers exceeding expectations 	<p><u>Growth Target</u></p> <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 	<p><u>Growth Target</u></p> <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 	<p><u>Growth Target</u></p> <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.



NATIONAL COMPREHENSIVE CENTER
FOR **TEACHER QUALITY**



NATIONAL COMPREHENSIVE CENTER
FOR **TEACHER QUALITY**

Laura Goe, Ph.D.

P: 609-734-1076

E-Mail: 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
