



# **Teacher Evaluation in Transition: Using Evaluation to Improve Teacher Effectiveness**

**Laura Goe, Ph.D.**

*New Hampshire Teacher Effectiveness Task Force*

**November 16, 2010**

Copyright © 2009 National Comprehensive Center for Teacher Quality. All rights reserved.



# Laura Goe, Ph.D.

- Former teacher in rural & urban schools
  - Special education (7<sup>th</sup> & 8<sup>th</sup> grade, Tunica, MS)
  - Language arts (7<sup>th</sup> 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 Learning & Teaching Research Center at ETS

# National Comprehensive Center for Teacher Quality (the TQ Center)

A federally-funded partnership whose mission is to help states carry out the teacher quality mandates of ESEA

- Vanderbilt University
  - Students with special needs, at-risk students
- Learning Point Associates
  - Technical assistance, research, fiscal agent
- ETS

# The goal of teacher evaluation

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

**TO IMPROVE  
TEACHING AND  
LEARNING**

# Systems, Models, & Measures for Educator Evaluation

- **System** (state level): Provides guidance to help districts carry out state and federal mandates
- **Model** (state **or** district level): Provides details
  - Which specific measures or types of measures will be used
  - In what combinations
  - What percentages of a teachers “score” various measures or combinations of measures will account for
  - Who is evaluated, how often, and by whom
  - Training for evaluators, scorers, etc.
- **Measures**: Tool, instruments, protocols, assessments, and processes used to collect evidence on teachers’ performance and effectiveness

# Decision Points for NH

- What will an evaluation score consist of?
  - Likely to be a combination of scores on different measures such as observations, growth in student achievement, professional responsibilities
- What percentage of a total score will each measure (or category of measures) be?
- What measures of student learning growth are acceptable for teachers in tested vs. non-tested subjects/grades?

# Example of a System: New York

- 40 percent of teachers' evaluation score will be based on student achievement measures
  - Year One: 20 percent student growth on state assessments or comparable measures *for teachers in the common branch subjects or ELA and Math in grades four to eight only*, and 20 percent other locally selected measures that are rigorous and comparable across classrooms
- 60 percent of the evaluation score will be based on locally negotiated processes (*e.g.*, classroom observations by trained evaluators)

# 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

# Challenges in implementation

- Context: politics and policies
- Teacher buy-in
- Valid measures & instruments
- Resources to support implementation
- Time pressure (i.e., must have system in place by a certain date)
- Research base
- Not many models of comprehensive teacher evaluation systems that are responsive to federal priorities

# 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).

# 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.

# Multiple measures

- Multiple sources of evidence of a students' learning provide...
  - The teacher with better evidence about what the student knows and is able to do, so she can adapt instructional strategies accordingly
  - The evaluator with better evidence about a teachers' contribution to student learning
    - Results from a rubric-based assessment of performance and results from a standardized test may show very different aspects of a students' knowledge and skills

# 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

## ➤ Evidence of *professional responsibility*

- Administrator/supervisor reports
- Surveys of students and/or parents
- An “evidence binder” created & presented by the teacher

# Teacher observations

## ➤ Strengths

- Great for teacher formative evaluation
- Helps evaluator 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.

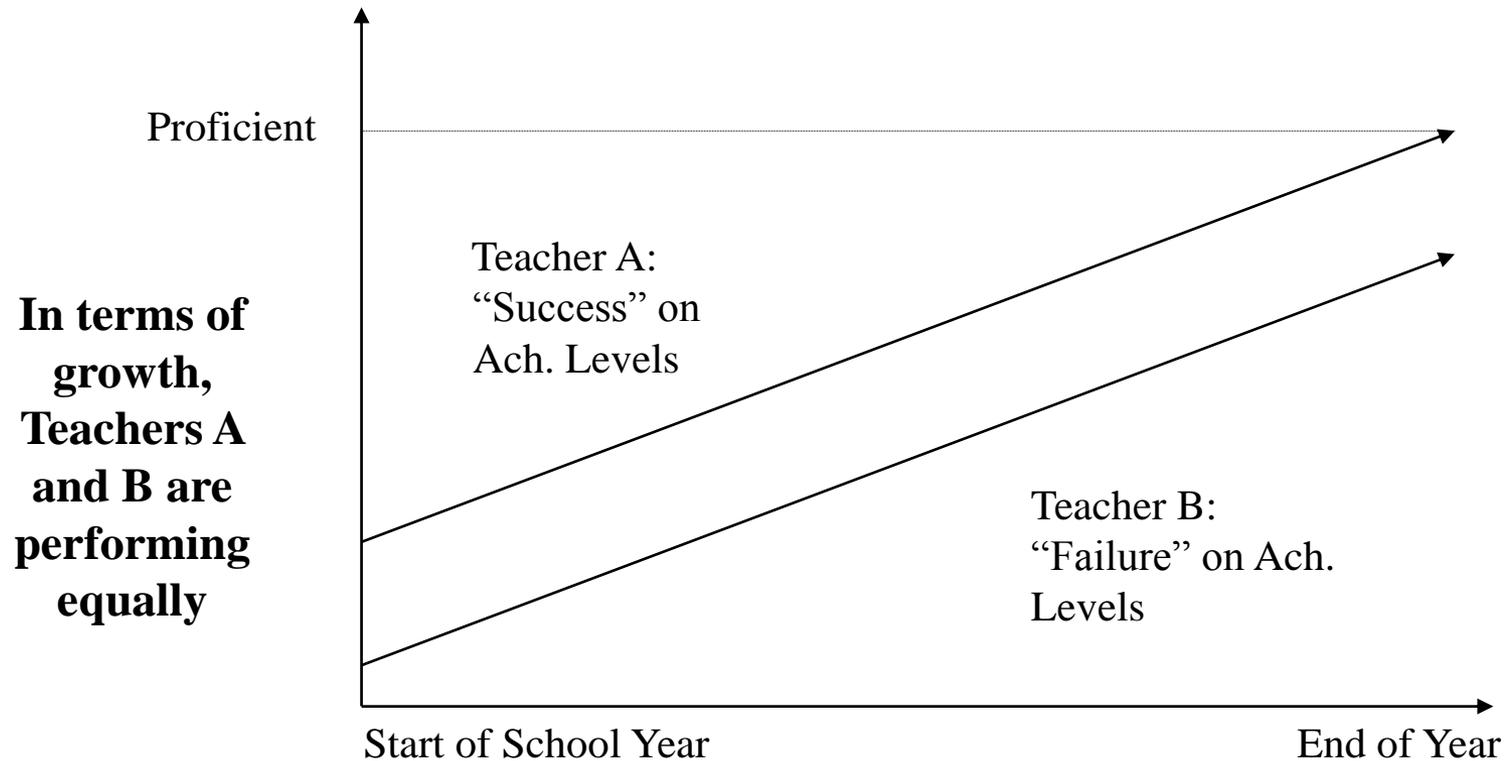
# Example: Kim Marshall's Rubric

## *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.

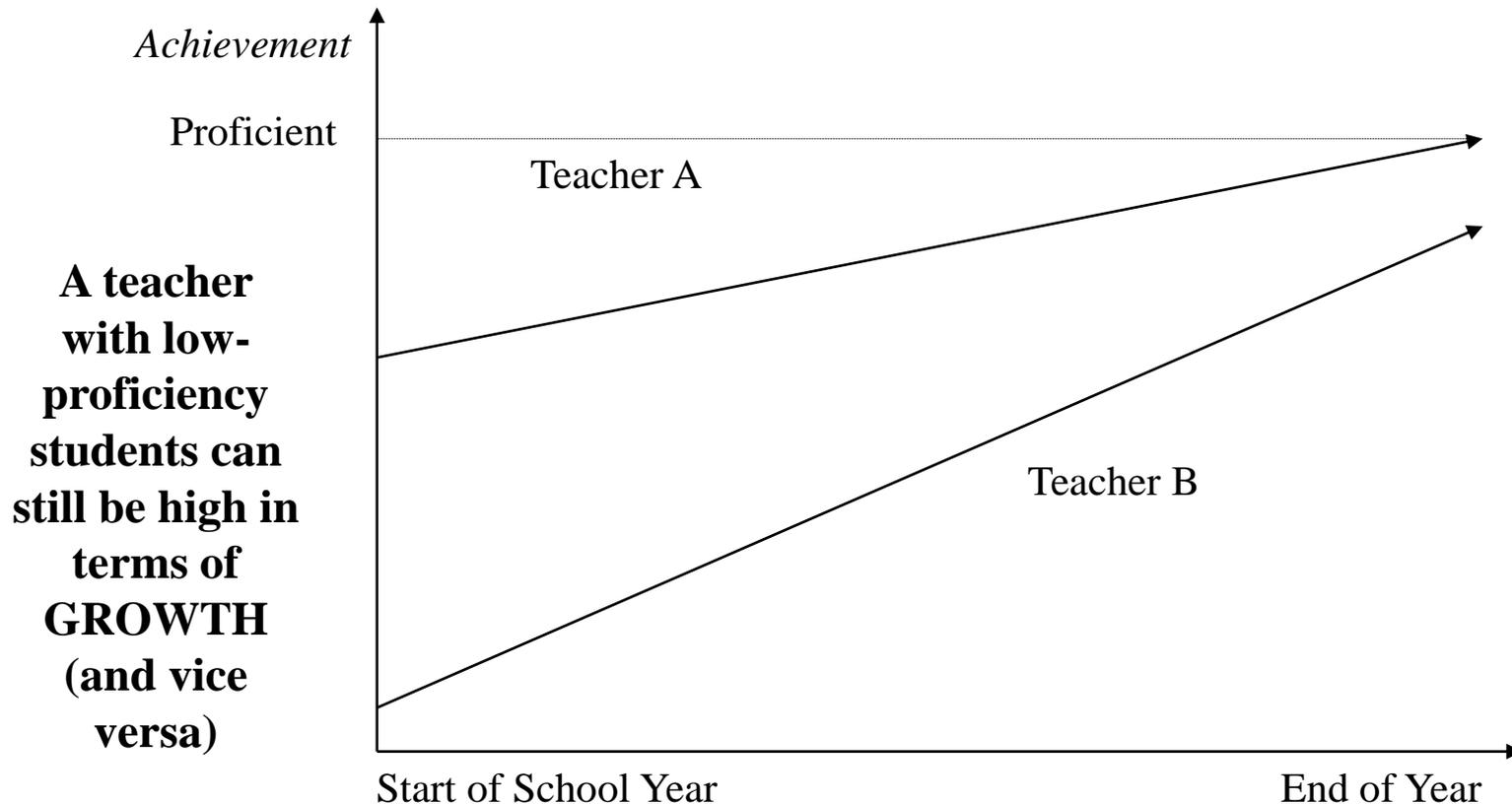
# Growth vs. Proficiency Models

*Achievement*



*Slide courtesy of Doug Harris, Ph.D, University of Wisconsin-Madison*

# Growth vs. Proficiency Models (2)



*Slide courtesy of Doug Harris, Ph.D, University of Wisconsin-Madison*

# Most popular growth models

## ➤ Value-added models

- There are many versions of value-added models (VAMs), but results from the different models are quite similar
- Most states and districts that use VAMs use the Sanders' model, also called TVAAS
- Prior test scores (3+ years in the Sanders' model) are used to predict the next test score for a student

## ➤ Colorado Growth model

- Focuses on “growth to proficiency,” i.e., whether a student is on course to reach proficiency

# What Value-Added 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!

# Other Measures

- Many types of evidence\*—including portfolios, administrator recommendations, analysis of teachers' assignments, analysis of students' work, documentation of teachers' positive contributions to the school, student and parent reports, and documentation of teacher leadership and mentoring—can be used in addition to student test scores

\* For descriptions and discussions of instruments for measuring various aspects of teacher performance, see Goe, Bell, and Little (2008).

# Measures that help teachers grow

- Not all measures are equally useful in helping teachers to grow professionally and improve their practice
- Measures that may contribute to teacher growth include
  - 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, mentors, coaches, TLCs, etc.
  - Measures that are aligned with professional development offerings
  - Measures based on protocols and processes that teachers can comprehend in terms of relationship to their own practice

# Austin Independent School District

## Student Learning Objectives:

- Teachers determine two SLOs for the semester/year
- One SLO must address all students, other may be targeted
- Use broad array of assessments
- Assess student needs more directly
- Align classroom, campus, and district expectations
- Aligned to state standards/campus improvement plans
- Based on multiple sources of student data
- Assessed with pre and post assessment
- Targets of student growth
- Peer collaboration

# Rubric for student learning objectives

APPROVAL		NEEDS REVISION	
LEVEL 4	LEVEL 3	LEVEL 2	LEVEL 1
<p><i>Indicates expert use of data, rigorous goal-setting based on student and community strengths and needs, incorporates appropriate valid/reliable assessments, considers teacher's own prior performance, and demonstrates alignment with campus improvement goals.</i></p>	<p><i>Indicates good use of data and acceptable level of rigor for students, considers teacher's prior performance, and utilizes appropriate assessments, but is not aligned with broader campus goals and does not incorporate the strengths of students or the school community.</i></p>	<p><i>Suggests superficial use of data, is minimally rigorous for students and the teacher, uses related but unproven assessments, and does not incorporate the goals of the campus or the strengths of students or the school community.</i></p>	<p><i>Represents ineffective use of data, is not rigorous for students or the teacher, does not support the campus goals, utilizes inappropriate assessment, and does not consider strengths of students or the school community.</i></p>
<p><b>Needs Assessment/Rationale: What are the needs?</b></p>			
<ul style="list-style-type: none"> <li>■ Campus data are reviewed for areas of strength and need (within subject area, within grade level, within student group, examining the Campus Improvement Plan (CIP), etc.).</li> <li>■ Classroom data are reviewed for areas of strength and need (by subject area, by student group, by concepts/skills/behavior).</li> </ul>	<ul style="list-style-type: none"> <li>■ Classroom data are reviewed for areas of strength and need (by subject area, by student group, by concepts/skills/behavior).</li> <li>■ Campus needs and strengths are not incorporated.</li> </ul>	<ul style="list-style-type: none"> <li>■ Classroom data are reviewed for areas of need, but needs of specific student groups are not examined in depth, and strengths are not identified.</li> <li>■ Campus needs and strengths are not incorporated.</li> </ul>	<ul style="list-style-type: none"> <li>■ Classroom data are not used to identify student needs.</li> <li>■ Campus needs and strengths are not incorporated.</li> </ul>
<p><b>Learning Content/Context and Student Group: What and who is targeted?</b></p>			
<ul style="list-style-type: none"> <li>■ Targets specific academic concepts, skills, or behaviors based on TEKS/TAKS Objective</li> <li>■ Targets the needs of the identified population</li> <li>■ Considers demonstrated strengths of identified population, as well as classroom &amp; school community</li> <li>■ Targets year-long (or semester-long) concepts, skills, or behaviors</li> <li>■ Supports goals of the Campus Improvement Plan (CIP)</li> </ul>	<ul style="list-style-type: none"> <li>■ Targets specific academic concepts, skills, or behaviors based on TEKS/ TAKS Objective</li> <li>■ Targets the needs of the identified population.</li> <li>■ Considers demonstrated strengths of identified population, as well as classroom &amp; school community.</li> <li>■ Targets year-long (or semester-long) concepts, skills, or behaviors.</li> <li>■ CIP Goals are not incorporated.</li> </ul>	<ul style="list-style-type: none"> <li>■ Targets specific academic concepts, skills, or behaviors based on TEKS/TAKS Objective</li> <li>■ Targets year-long (or semester-long) concepts, skills, or behaviors.</li> <li>■ Does not target the needs of all students in the identified population.</li> <li>■ Does not consider the strengths of the identified population, classroom, or school community.</li> <li>■ CIP Goals are not incorporated.</li> </ul>	<ul style="list-style-type: none"> <li>■ Does not target concepts, skills, or behaviors based on TEKS/TAKS Objective</li> <li>■ Does not target year-long (or semester-long) concepts, skills, or behaviors.</li> <li>■ Does not target the needs of all students in the identified population.</li> <li>■ Does not consider the strengths of the identified population, classroom, or school community.</li> <li>■ CIP Goals are not incorporated.</li> </ul>

# Rubric for student learning objectives (cont'd)

<i>Level 4</i>	<i>Level 3</i>	<i>Level 2</i>	<i>Level 1</i>
<b>Learning Objective:</b> What will students learn?			
<ul style="list-style-type: none"> <li>■ Based on the identified student needs.</li> <li>■ Supports goals of the CIP.</li> <li>■ Is rigorous.</li> <li>■ Is a good example of ongoing, reflexive practice.</li> <li>■ Provides clear focus for instruction and assessment.</li> <li>■ Is measurable.</li> <li>■ Reflects strengths of students and school community.</li> </ul>	<ul style="list-style-type: none"> <li>■ Based on the identified student needs.</li> <li>■ Is rigorous.</li> <li>■ Is a good example of ongoing, reflexive practice.</li> <li>■ Provides clear focus for instruction and assessment.</li> <li>■ Is measurable.</li> <li>■ Reflects strengths of students and school community.</li> <li>■ Does not relate to goals of CIP.</li> </ul>	<ul style="list-style-type: none"> <li>■ Based on the identified student needs of some students.</li> <li>■ Provides clear focus for instruction and assessment.</li> <li>■ Is measurable.</li> <li>■ Does not reflect strengths of students and school community.</li> <li>■ Does not relate to goals of CIP.</li> <li>■ Is not rigorous.</li> <li>■ Is not a good example of ongoing, reflexive practice.</li> </ul>	<ul style="list-style-type: none"> <li>■ Does not consider student needs.</li> <li>■ Does not provide a clear focus for instruction.</li> <li>■ Is not measurable.</li> <li>■ Does not reflect strengths of students and school community.</li> <li>■ Does not relate to goals of CIP.</li> <li>■ Is not rigorous.</li> <li>■ Is not a good example of ongoing, reflexive practice.</li> </ul>
<b>Outcome Assessment:</b> How will you know whether they learned it?			
<ul style="list-style-type: none"> <li>■ Aligns with the targeted learning content area.</li> <li>■ Relationship with learning objective is apparent.</li> <li>■ Has been demonstrated as reliable and valid for targeted students.</li> <li>■ Follows guidelines for appropriate assessments.</li> </ul>	<ul style="list-style-type: none"> <li>■ Aligns with the targeted learning content area.</li> <li>■ Relationship with learning objective is apparent.</li> <li>■ Follows guidelines for appropriate assessments.</li> <li>■ Has not been demonstrated as reliable and valid for targeted students.</li> </ul>	<ul style="list-style-type: none"> <li>■ Aligns with the targeted learning content area.</li> <li>■ Relationship with learning objective is minimally apparent.</li> <li>■ Does not follow guidelines for appropriate assessments.</li> <li>■ Has not been demonstrated as reliable and valid for targeted students.</li> </ul>	<ul style="list-style-type: none"> <li>■ Does not align with the targeted learning content area.</li> <li>■ Does not follow guidelines for appropriate assessments.</li> <li>■ Has not been demonstrated as reliable and valid for targeted students.</li> </ul>
<b>Performance Target:</b> What is your goal for student achievement?			
<ul style="list-style-type: none"> <li>■ Predicts performance based on past performance of students when available.</li> <li>■ Is a rigorous expectation for students.</li> <li>■ Is a rigorous expectation for teachers, based on past performance.</li> </ul>	<ul style="list-style-type: none"> <li>■ Predicts performance based on past performance of students when available.</li> <li>■ Is a rigorous expectation for students.</li> <li>■ Does not “stretch” the teacher beyond past performance history.</li> </ul>	<ul style="list-style-type: none"> <li>■ Predicts performance based on past performance of students when available.</li> <li>■ Is not a rigorous expectation for students.</li> <li>■ Does not “stretch” the teacher beyond past performance history.</li> </ul>	<ul style="list-style-type: none"> <li>■ Does not utilize past performance of students when available.</li> <li>■ Is not a rigorous expectation for students.</li> <li>■ Does not “stretch” the teacher beyond past performance history.</li> </ul>
<b>Rigor:</b> How rigorous is your SLO?			
<ul style="list-style-type: none"> <li>■ Content is challenging, complex and progressively deepens knowledge of core content</li> <li>■ Content is thought-provoking requiring high thinking demand</li> <li>■ Requires analytical thinking and active use of knowledge</li> <li>■ Content is relevant to life/ experiences.</li> </ul>	<ul style="list-style-type: none"> <li>■ Content is challenging and complex for most, but not all, students.</li> <li>■ Requires analytical thinking.</li> <li>■ Content is relevant for most, but not all, students.</li> </ul>	<ul style="list-style-type: none"> <li>■ Content is challenging for some students.</li> <li>■ Does not require analytical thinking.</li> <li>■ Content is relevant for some students.</li> </ul>	<ul style="list-style-type: none"> <li>■ Content is not challenging.</li> <li>■ Does not require analytical thinking.</li> <li>■ Is not relevant to life and learning experiences.</li> </ul>

# Austin SLO Model

## Strengths/Weaknesses

### ➤ Strengths

- Teachers take an active role in determining student learning goals
- Good professional growth opportunity for teachers
- If objectives are of high-quality and teachers plan instruction to meet them, students should benefit
- Student growth measured between **two points in time**

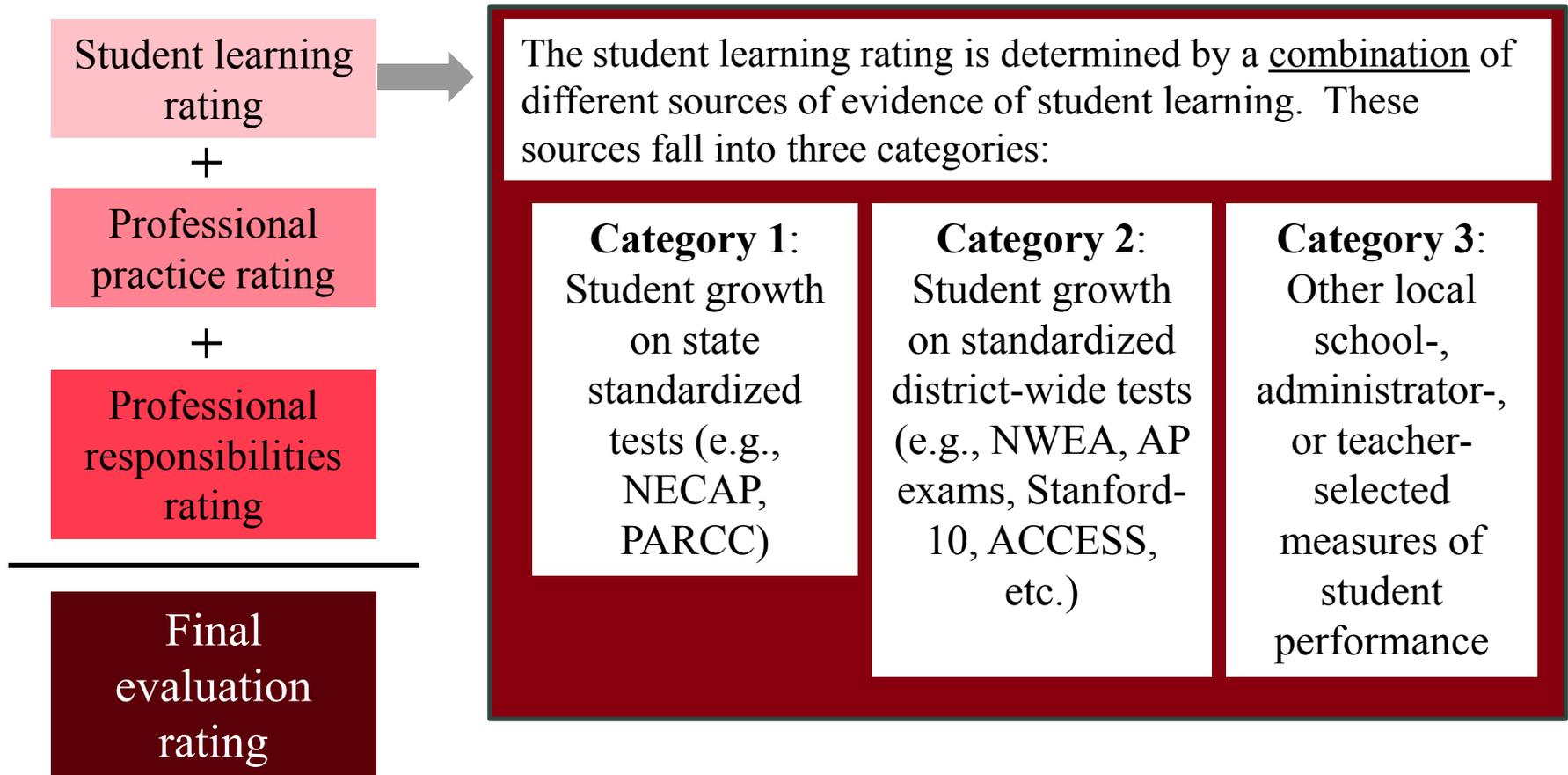
### ➤ Weaknesses

- Heavily dependent on administrator understanding and time commitment to supervision
- Not clear how well **rigor** is determined (based on rubric)
- Not **comparable across classrooms** since teachers decisions about objectives, assessments, and measures may vary widely

# “Rhode Island Model” is another example of an SLO Model

- Teachers measure student growth by setting student academic goals aligned to standards
- Principals, during the goal setting process, will confer with teachers to establish each goal’s degree of ambition and select the appropriate assessments for measuring progress against the goals
- Teacher evaluation will be based on students’ progress on the established goals, as determined by an end-of-the-year principal review of the pre-determined assessments and their results

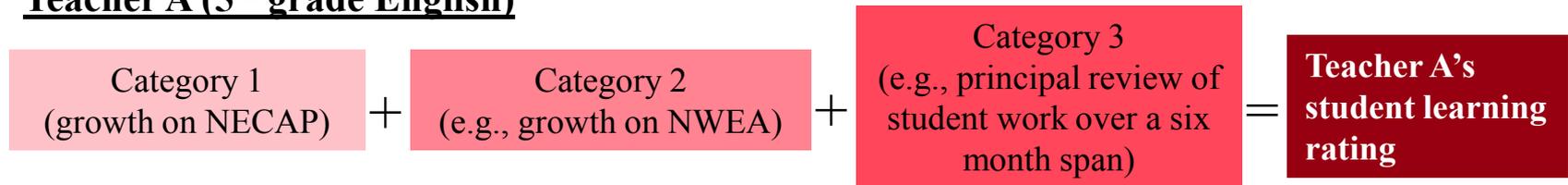
# Rhode Island DOE Model: Framework for Applying Multiple Measures of Student Learning



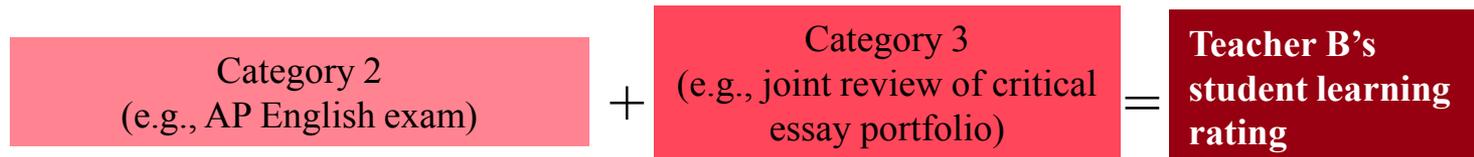
# “Rhode Island Model”: Student Learning Group Guiding Principles

- “Not all teachers’ impact on student learning will be measured by the same mix of assessments, and the mix of assessments used for any given teacher group may vary from year to year.”

## Teacher A (5<sup>th</sup> grade English)



## Teacher B (11<sup>th</sup> grade English)



## Teacher C (middle school art)



*This teacher may use several category 3 assessments*

# “Rhode Island Model” Strengths and Weaknesses

## ➤ Strengths

- Includes teachers in evaluation of student learning (outside of standardized tests)
- Teachers will benefit from conducting assessment of student learning at the classroom level

## ➤ Weaknesses

- Heavily administrator/evaluator driven process
- Not always using **two points in time**
- Not clear how or if **rigor** will be determined
- Not **comparable across classrooms** since teachers decisions about objectives, assessments, and measures may vary widely

# 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) at least three times per year by different observers (usually one administrator and two teachers who have been appointed to the role)
- 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

# TAP Strengths

## ➤ Strengths

- Value-added becomes everyone's responsibility, which should encourage efforts from teachers in non-tested subjects to support teachers in tested subjects
- Multiple yearly observations should be more informative and produce more reliable information about practice
- Professional development aligned with results is required
- Some evidence that schools participating in TAP may show improvement in achievement scores *in tested subjects*

# Tap Weaknesses

## ➤ Weaknesses

- Concerns about “fairness” when only a few teachers’ student achievement and progress toward learning goals “counts”
- Tells you nothing about how teachers in other subjects are performing in terms of student learning growth (grades are not always good indicators)
- Not using **two points in time** *in non-tested subjects*
- **Rigor** not determined *in non-tested subjects*
- Not **comparable across classrooms** because *in non-tested subjects*, teachers are receiving scores based on math and English/language arts, so their scores are not related to their efforts but rather to the efforts of colleagues in tested subjects

# IMPACT sorts teachers into groups that are evaluated differently

- Group 1: general ed teachers for whom value-added data can be generated
- Group 2: general ed teachers for whom value-added data *cannot* be generated
- Group 3: special education teachers
- Group 4: non-itinerant English Language Learner (ELL) teachers and bilingual teachers
- Group 5: itinerant ELL teachers
- Etc...

# Score calculation for Groups 1 & 2

	Group 1 (tested subjects)	Group 2 (non-tested subjects)
Teacher value-added (based on test scores)	50%	0%
Teacher effectiveness (based on non-VAM assessments)	0%	10%
Teacher practice (based on observations)	40%	80%
Other (school-wide value added & commitment to the school community)	10%	10%

# Non-VAM tests (accepted under Washington, DC's IMPACT evaluation system)

- DC Benchmark Assessment System (DC BAS)
- Dynamic Indicators of Basic Early Literacy Skills (DIBELS)
- Developmental Reading Assessment (DRA)
- Curriculum-based assessments (e.g., Everyday Mathematics)
- Unit tests from DCPS-approved textbooks
- Off-the-shelf standardized assessments that are aligned to the DCPS Content Standards
- Rigorous teacher-created assessments that are aligned to the DCPS Content Standards
- Rigorous portfolios of student work that are aligned to the DCPS Content Standards

# DC IMPACT Strengths

## ➤ Strengths

- Uses multiples measures to assess effectiveness (observations, assessment of student learning)
- Permits the use of many types of (test-based) assessments for students in non-tested subjects and grades
- Includes what is important in the system (in order to encourage specific teacher behaviors)
- Using **two points in time** in tested subjects (less clear if this is done in non-tested subjects)

# DC IMPACT Weaknesses

## ➤ Weaknesses

- No multiple measures of student learning growth for teachers in *tested* subjects and grades
- Huge differences in how teachers are measured
- Not clear how **rigor** is determined *in non-tested subjects*
- Not **comparable across classrooms** *in non-tested subjects*, because teachers are choosing from among many different assessments rather than agreeing to use the same ones in specific subjects and grades

# Georgia KEYS

**STUDENT ACHIEVEMENT** - “Annual teacher evaluations shall as a minimum take into consideration the following: (1) the role of the teacher in meeting the school’s student achievement goals, including the academic gains of students assigned to the teacher.” Georgia Code 20-2-210 (b) (1) and (a)

*“In making a determination of the academic gains of the students assigned to a teacher, evaluators should make every effort to have available and to utilize the results of a wide range of student achievement assessments, including those utilized by the teacher, set by the local board of education, or required under this article.” Georgia Code 20-2-210 (b) (1) and (c)*

**Student Achievement Teacher Standard 1: The teacher has a positive impact on student learning and academic achievement.**

**SA 1.1 Students taught by the teacher demonstrate the Georgia Performance Standard (GPS) related academic achievement progress on measures of student learning including state-mandated achievement tests or other measures as determined by the school district (e.g., teacher-developed assessments, department or district common assessments, benchmark tests, student work samples, portfolios, etc.).**

	<input type="checkbox"/> Not Evident	<input type="checkbox"/> Emerging	<input type="checkbox"/> Proficient	<input type="checkbox"/> Exemplary
Continuum of Improvement	No quantifiable evidence exists that student achievement has increased, based on pre- and post-assessments using measures identified by the school district.	Quantifiable evidence exists that student achievement has increased, but has not met the established benchmark identified by the school district.	Quantifiable evidence exists that student achievement has met the benchmark based on pre- and post-assessments using measures identified by the school district.	Quantifiable evidence exists that student achievement has exceeded the benchmarks based on multiple measures of student learning including pre- and post-measures identified by the school district and also includes data from multiple measures of student learning.

# Georgia KEYS for Non-tested subjects

**SA 1.2 Students taught by the teacher of content areas not addressed by the Georgia Performance Standards (GPS) demonstrate academic achievement progress on measures of student learning as determined by the school district (e.g., teacher-developed assessments, department or district common assessments, benchmark tests, student work samples, portfolios, etc.).**

	Not Evident	Emerging	Proficient	Exemplary
<b>Continuum of Improvement</b>	No quantifiable evidence exists that student achievement has increased, based on pre- and post-assessments using measures identified by the school district.	Quantifiable evidence exists that student achievement has increased, but has not met the benchmarks based on pre- and post-assessments using measures identified by the school district.	Quantifiable evidence exists that student achievement has met the benchmarks based on pre- and post-assessments using measures identified by the school district.	Quantifiable evidence exists that student achievement has exceeded the benchmarks based on multiple measures of student learning including pre- and post-assessments identified by the school district.

**Example 1**  
**Teacher Generated Performance Standards**  
 This option is similar to Example 3 in SA 1.1. A district-wide group of teachers could collaborate to determine proficiency or progress standards for a given subject. The type of assessment would depend on the skills and knowledge that students are expected to master. Art and music classes, for example, may require students to demonstrate skills through performance. Art students might be required to amass a portfolio that exhibits progress and eventual mastery of certain skills. Band students may be required to make recordings or give live performances.

**Example 2**  
**Certification Based Assessment**  
 Students in some fields, such as career and technical education, can seek certification that they have mastered certain skills. These certification tests may have been developed by national associations, state boards, or private companies. Districts may choose to adopt some of these tests as assessments of proficiency for their own coursework. This strategy has the advantage of holding students to a recognized standard and allowing for comparisons to students outside the district. Drawbacks may include the monetary cost of testing and the challenge of finding tests that are representative of course content.

**Example 3**  
**National Standards**  
 Some subjects may be covered by standards set by a national organization. For example, physical education students may be assessed using the President's Physical Fitness Test. Students that achieve passing scores may be considered proficient, and progress can be measured across multiple testing periods. In addition, information from sporting associations may be used to assess students' knowledge of the rules and strategies of various sports.

# Georgia KEYS Strengths & Weaknesses

## ➤ Strengths

- Rubric for measuring teacher contribution to student learning is easy to understand
- Includes examples of multiple measures of student learning for all teachers, including those in tested grades and subjects
- “Teacher Generated Performance Standards” should make scores **comparable across classrooms** in non-tested subjects since they are done district-wide

## ➤ Weaknesses

- Not clear that student growth is measured between **two points in time** in all cases
- Not clear how **rigor** is determined

# Hillsborough, FL

- Stated goal is to evaluate **every** teacher's effectiveness with student achievement growth, even teachers in non-tested subjects and grades
- Undertaking to create pre- and post-assessments for all subjects and grades
- Expanding state standardized tests and using value-added to evaluate more teachers
- Part of a multiple measures system

# Hillsborough Strengths & Weaknesses

## ➤ Strengths (based on completion of system)

- Teacher and union involved in evaluation system decisions
- Teachers may be able to recommend tests they are already using
- All teachers included, not just tested subjects
- Should be **comparable across classrooms** in district
- Student growth should be measured between **two points in time** in all cases (when all subjects have appropriate tests)
- **Rigor** should be determined in test validation process

## ➤ Weaknesses

- Very expensive to create tests for all grades and subjects
- Takes teachers out of the assessing/scoring/improving instruction loop

# Delaware Model

- Standardized test will be used as part of teachers' scores in some 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 state for approval
- Teachers/groups of teachers take primary responsibility for determining student growth
- State will monitor how assessments are "working"

# Delaware Model: Strengths

## ➤ Strengths

- Teacher-driven process (assumes teachers are the experts in assessing their students' learning growth)
- Great professional growth opportunity as teachers work together across schools to determine assessments, score student work, etc.
- Should be **comparable across classrooms** in state
- Student growth should be measured between **two points in time** for all subjects and grades
- **Rigor** should be determined in assessment approval process

# Delaware Model: Weaknesses

## ➤ Weaknesses

- Validity issues
  - Whether delivery of assessments is standardized
  - Teacher training to score student work & assessments
- Time must be built in for teachers to work together on scoring (particularly for rubric-based assessments)

# Questions to ask about models

- Are they “rigorous and comparable across classrooms”?
- Do they show student learning growth “between two points in time”?
- Are they based on grade level and subject standards?
- Do they allow teachers from all subjects to be evaluated with evidence of student learning growth?

# Keys to Measuring teacher Effectiveness

- Measure what is *required* (i.e., federal/state legislation and incentives)
- Measure what is *valued* (i.e., all the things we expect teachers to do)
- Familiarize teachers with tools and processes of the evaluation
- Develop and make available to teachers and evaluators the standards by which teachers will be evaluated
- Measure performance against the standards

# Questions?





**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](http://www.tqsource.org)