



Evaluating Teacher/Leader Effectiveness

Laura Goe, Ph.D.

Webinar for Alaska SIG Districts

Presented by

***Alaska Comprehensive Center and
SERRC - Alaska's Educational Resource Center***

March 25, 2011

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



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



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

➤ American Institutes for Research

- Technical assistance, research

➤ Educational Testing Service

- Technical assistance, research, dissemination

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
- **Focus on models and measures that may help districts/schools/teachers improve performance**
 - *The ultimate goal of evaluation should be to improve teaching and learning*

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).

Teacher evaluation



When all you have is a hammer,
everything looks like a nail.

**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, abbreviate 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 of 1 2 3 4 5 (Circle One) Length of Evaluation: _____ minutes
 Meeting set with teacher to discuss evaluation: Date _____ Time _____
 Was pre-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 calm, comfortable learning environment				
Establishes and communicates realistic expectations and well-defined behavioral rules				
Maintains eye contact				
Displays self-confidence and self-control				
Maintains attendance				
Uses rewards with discretion				
Monitors student work				
Monitors student understanding and comprehension				
Communicates clearly				
Calls students by name				
Involves students in learning tasks				

1

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

Multiple measures of student learning

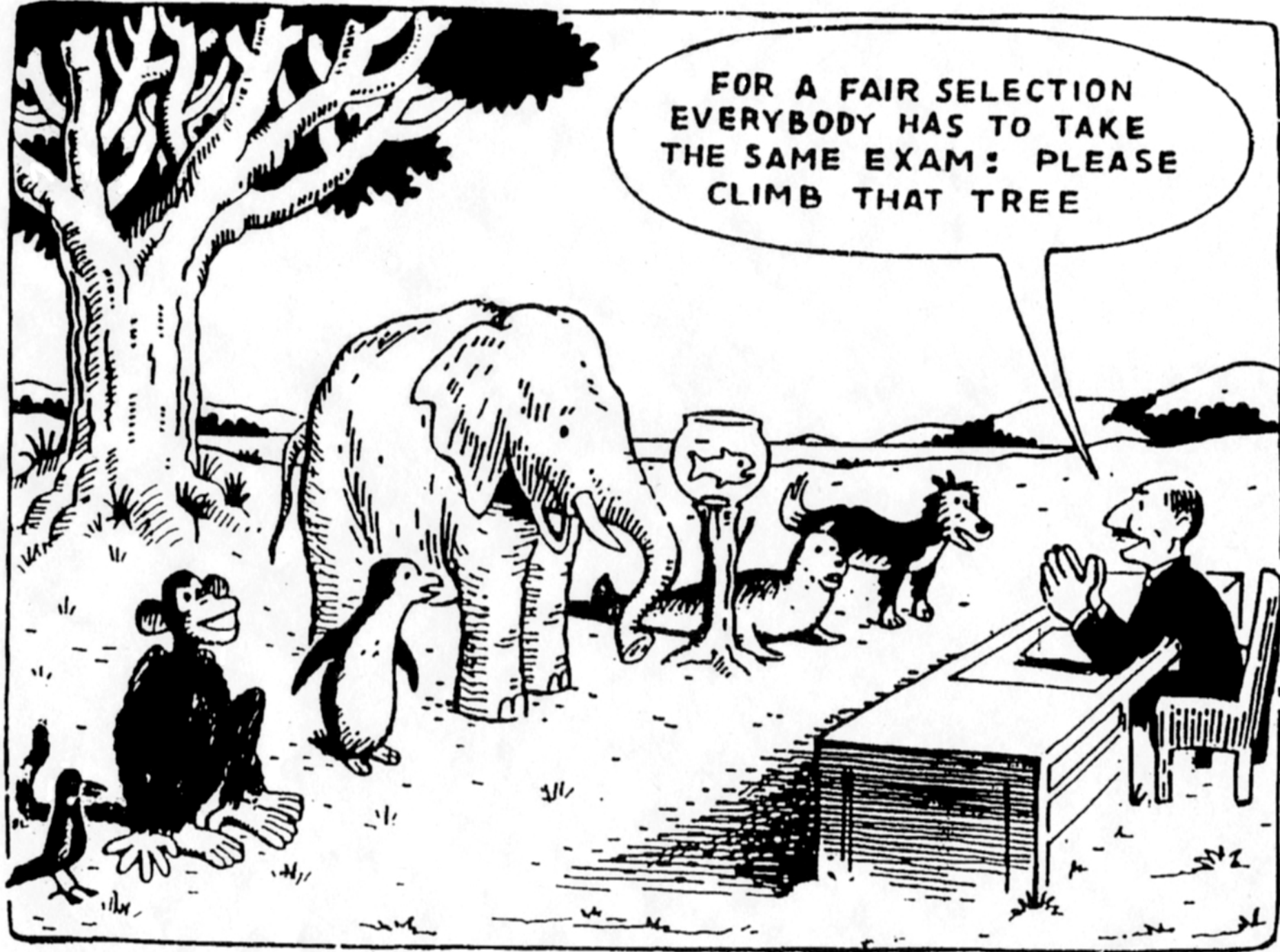
➤ Evidence of growth in student learning and competency

- Standardized assessments (state/district tests)
- Classroom-based assessments such as DRA, DIBELS, curriculum-based tests, unit tests

➤ Evidence of growth in skills and knowledge for specific purposes

- The 4 Ps: portfolios, projects, products, and performances
- Essays, written responses to complex questions

FOR A FAIR SELECTION
EVERYBODY HAS TO TAKE
THE SAME EXAM: PLEASE
CLIMB THAT TREE



Teacher observations: strengths and weaknesses

➤ 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	Positive Climate Negative Climate	Behavior Management Productivity	Concept Development Quality of Feedback Language Modeling
Upper Elementary/ Secondary	Teacher Sensitivity Regard for Student (Adolescent) Perspectives	Instructional Learning Formats	Content Understanding Analysis and Problem Solving Quality of Feedback

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 or no correlation
 - Kane et al. (2010) describes research linking observation scores with value-added scores (found some small, significant correlations)
 - Sartain et al. (2010) describes challenges in implementation, differences researcher/principal ratings
 - Consortium on Chicago School Research has ongoing project studying implementation and results of replacing the “checklist” with the Danielson Framework

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."

Observation instruments

Charlotte Danielson's Framework for Teaching

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

CLASS

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

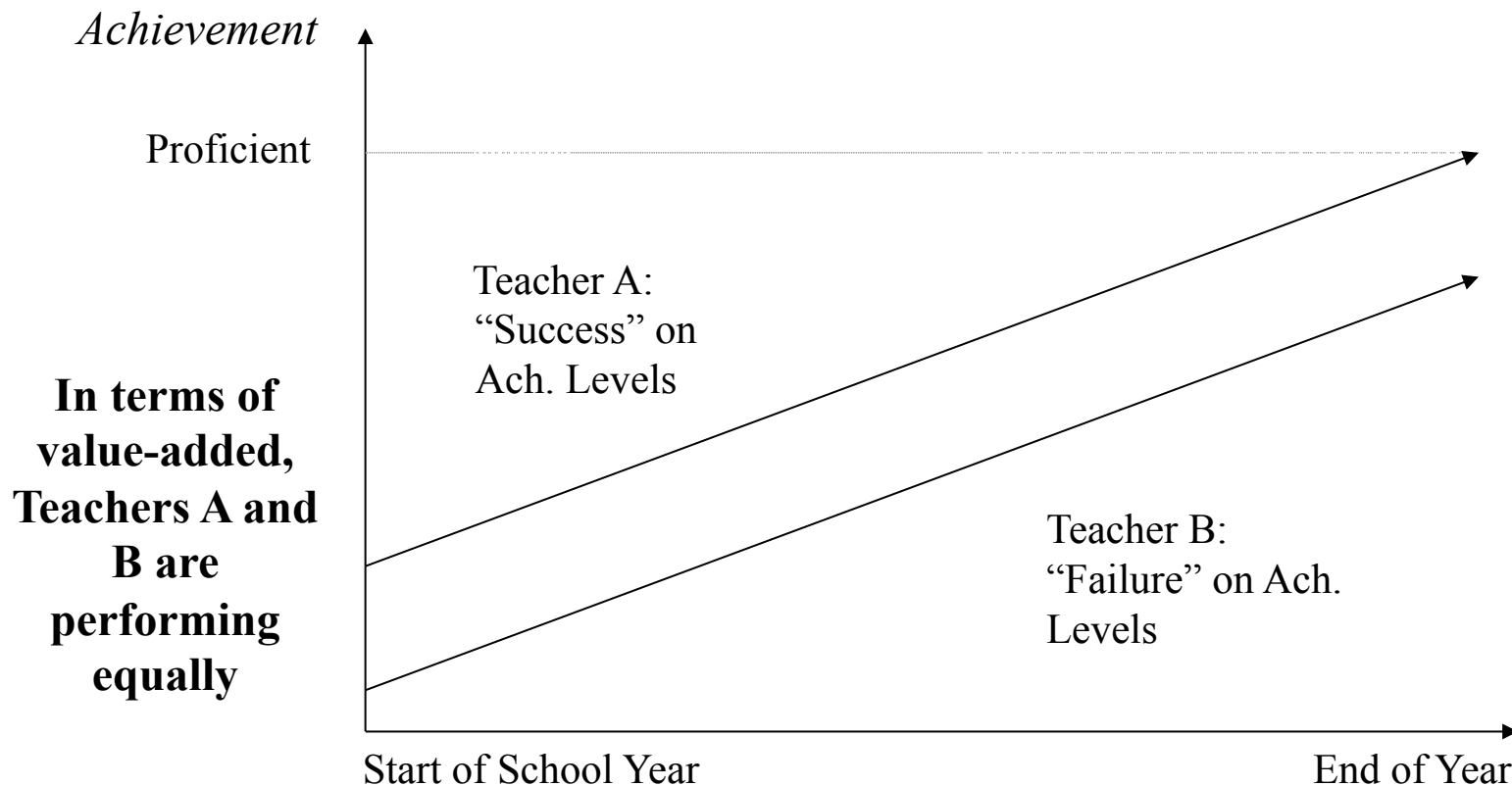
Kim Marshall Rubric

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

Most commonly used growth models

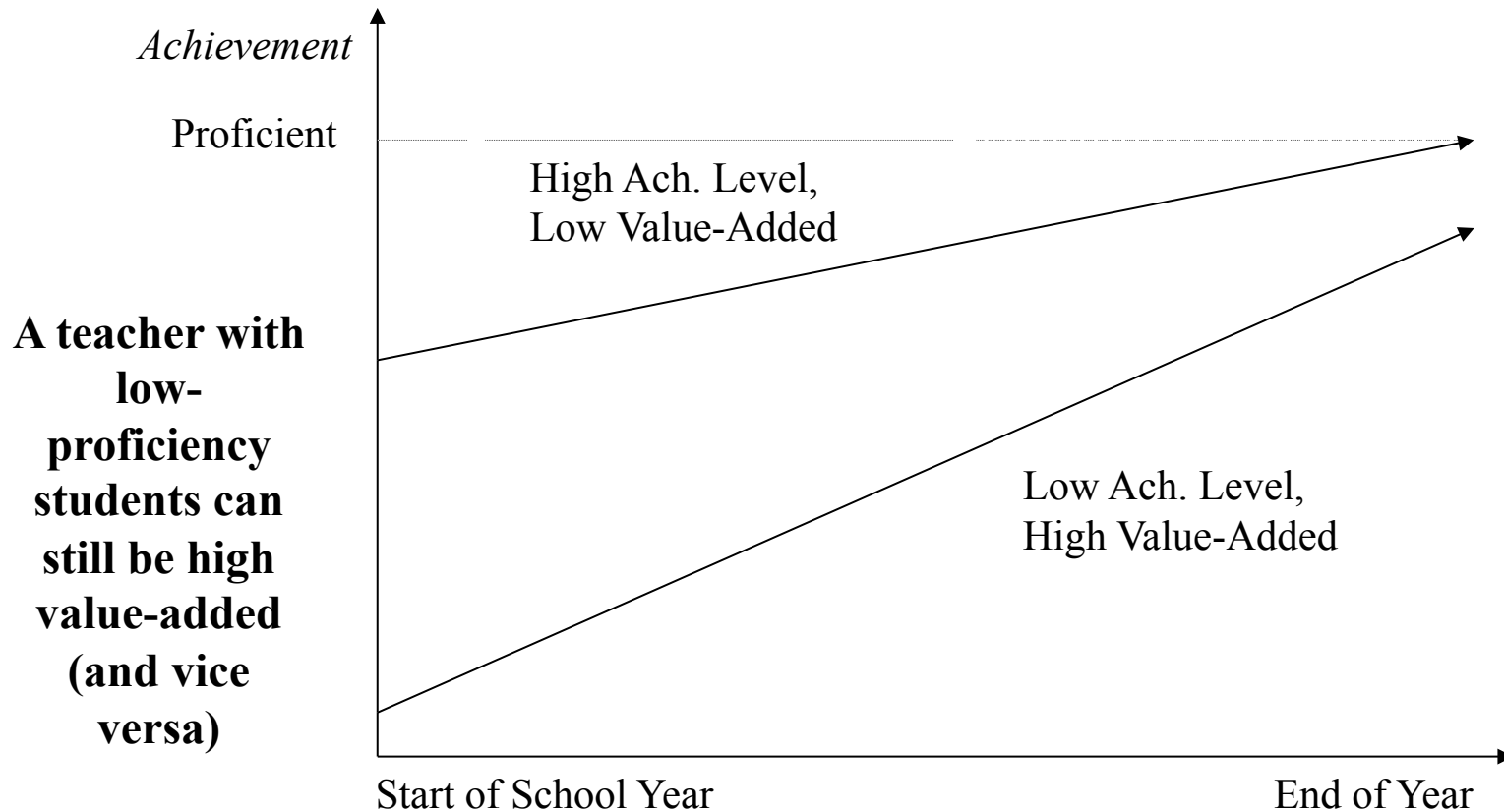
- Value-added models (requires prediction)
 - There are many versions of value-added models (VAMs), but results from the different models are quite similar
 - Most states and districts that use VAMs use the Sanders' model, also called TVAAS
 - Prior test scores (3+ years in the Sanders' model) are used to predict the next test score for a student
- Colorado Growth model (no prediction needed)
 - Focuses on “growth to proficiency”
 - Measures students against “academic peers”

Why growth models are better than status models (1)



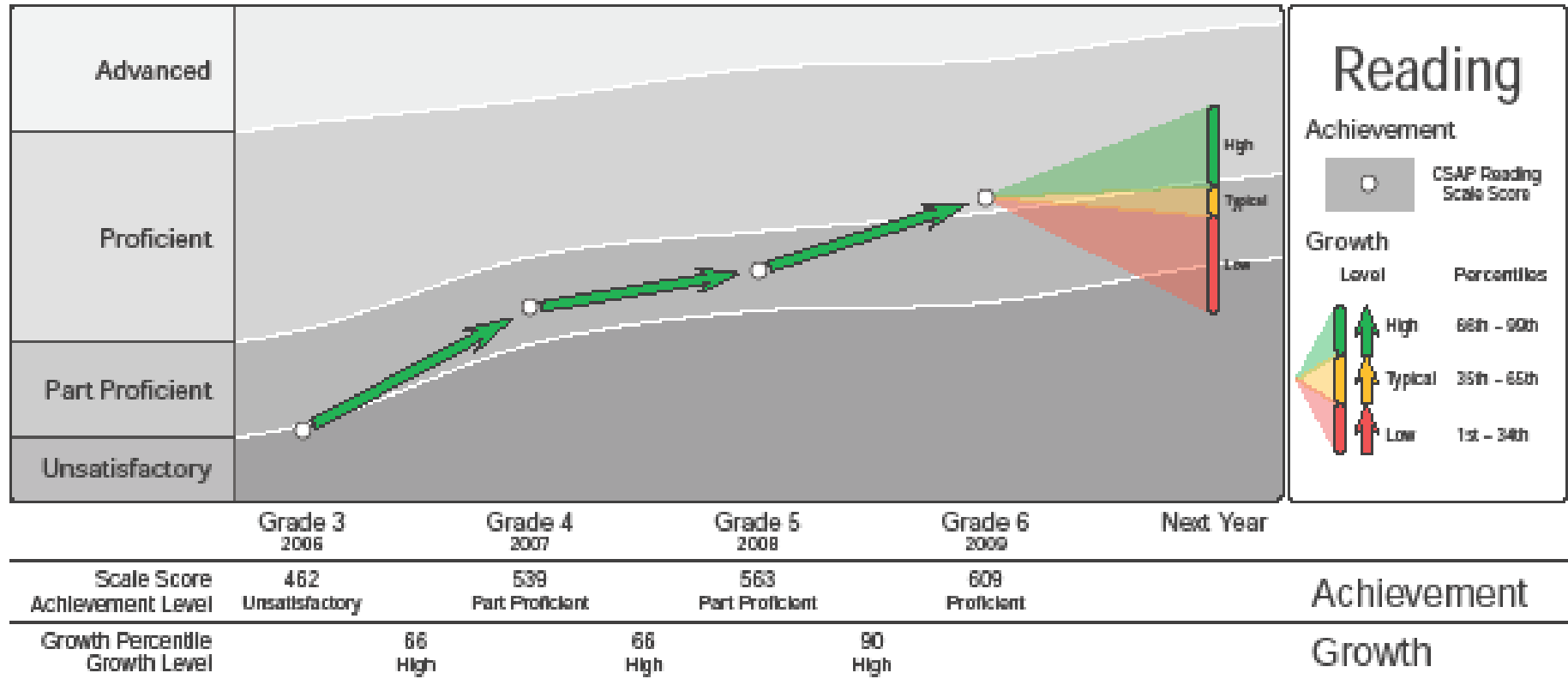
Slide courtesy of Doug Harris, Ph.D, University of Wisconsin-Madison

Why growth models are better than status models (2)



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

Sample student report: Colorado Growth Model



Slide courtesy of Damian Betebenner at www.nciea.org

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!

Cautions about using value-added for teacher evaluation

- Braun et al. (2010) provides some useful definitions and a good review of research; notes that most researchers are not comfortable with using VAMs as the sole measures of teacher effectiveness
- 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.”

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"

VAMs don't measure most teachers

- About 69% of teachers (Prince et al., 2006) can't be accurately assessed with VAMs
 - 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

Growth Models

SAS Education Value-Added Assessment System
(EVAAS)

[http://www.sas.com/govedu/edu/k12/evaas/
index.html](http://www.sas.com/govedu/edu/k12/evaas/index.html)

Colorado Growth Model

www.nciea.org

Questions to ask about measures and the models that incorporate them

- 1. Rigorous.** Are measures “rigorous,” focused on measuring students’ progress towards college and career readiness?
- 2. Comparable.** Are measures “comparable across classrooms,” ensuring that students are being measures with the same yardstick?
- 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?
- 5. 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?
- 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 to ensure success?

Evaluation Models

- Austin, TX
- Delaware
- Georgia
- Hillsborough, FL
- New Haven, CT
- Rhode Island
- TAP (Teacher Advancement Program)
- Washington, DC

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%202010-18-2010.pdf?>

[p=6CC6799F8C1371F6B59CF81E4ECD54E63F615CF1D9441A92E28BFA2A0AB27E3E&Type=D](http://www.gadoe.org/DMGetDocument.aspx/CK%20Standards%202010-18-2010.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/Assnt_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)

Austin Independent School District

Student Learning Objectives:

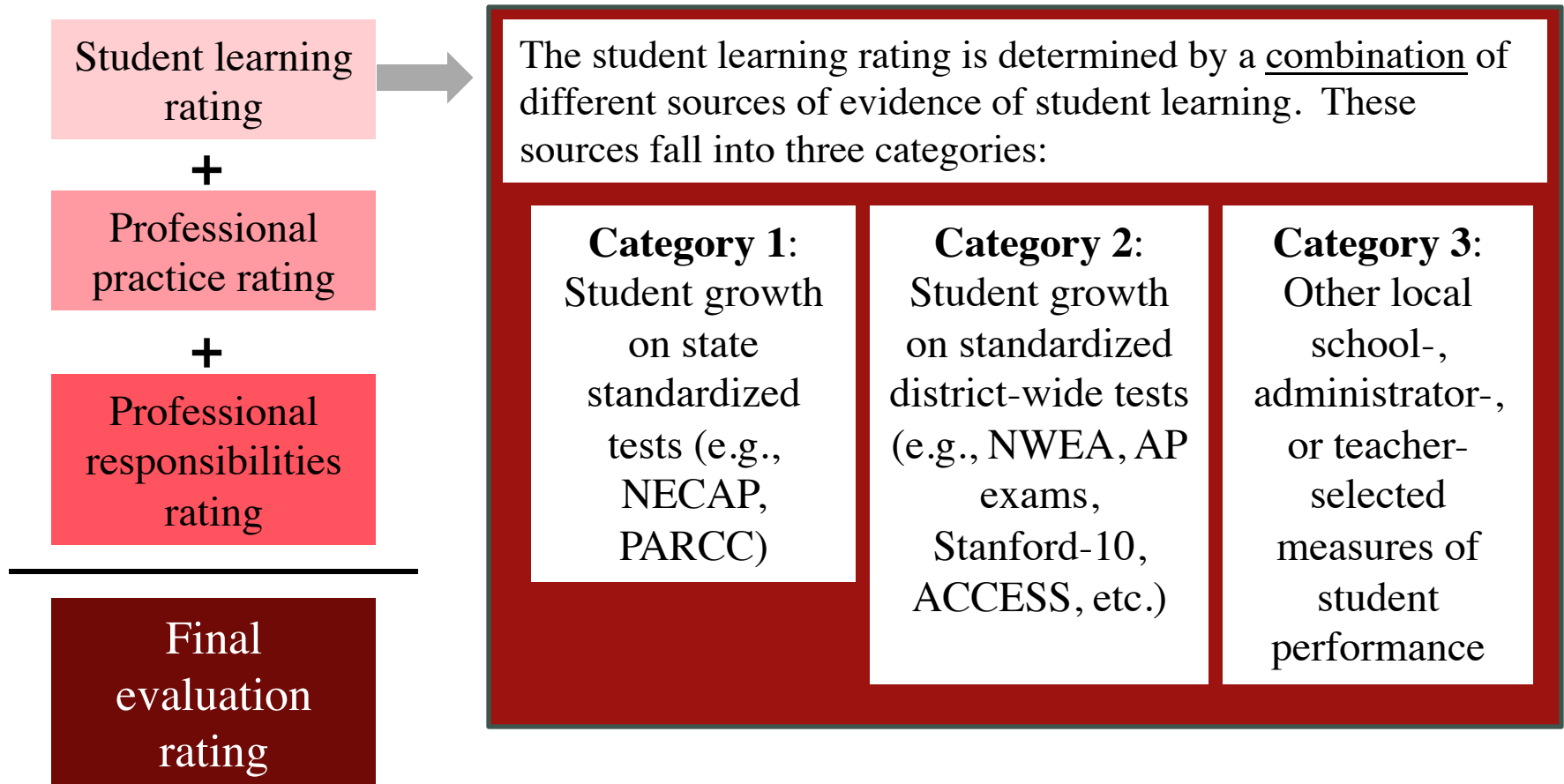
- Teachers determine two SLOs for the semester/year
- One SLO must address all students, other may be targeted
- Use broad array of assessments
- Assess student needs more directly
- Align classroom, campus, and district expectations
- Aligned to state standards/campus improvement plans
- Based on multiple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

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

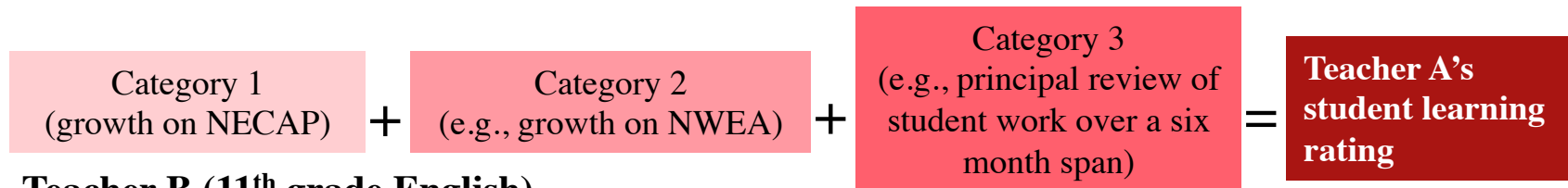
Rhode Island DOE Model: Framework for Applying Multiple Measures of Student Learning



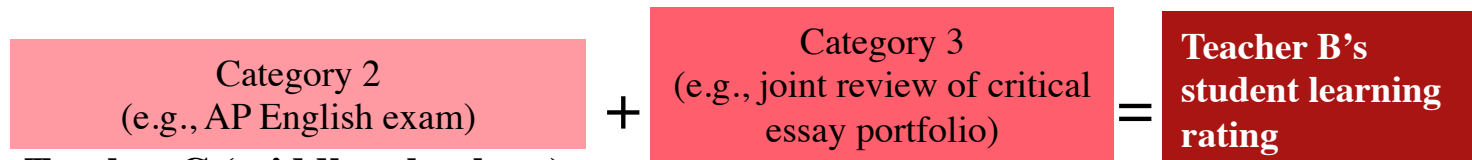
Rhode Island Model: Student Learning Group Guiding Principles

- “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.”

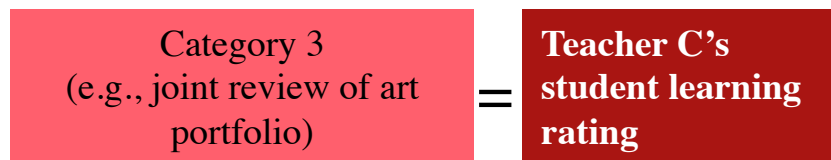
Teacher A (5th grade English)



Teacher B (11th grade English)



Teacher C (middle school art)



This teacher may use several category 3 assessments

New Haven goal-setting process

- Teachers administer formative/diagnostic assessments for each of his/her groups of students prior to the Goal-Setting Conference.
- During the Goal-Setting Conference, teachers set appropriate academic goals for students in collaboration with instructional manager.
- Secondary level: Goals for each of the teacher's individual classes, with academic goals focused solely on the knowledge and skills that are relevant to the content area.
- Elementary level: Where a teacher works primarily with one group of students (or a class) across multiple disciplines, the teacher will devise academic goals that cover the breadth of instruction with a focus on the priority learning areas.
- Teachers, in collaboration with their instructional manager, will determine the appropriate number of goals as well as whether or not the goals set are "acceptable" – i.e., aligned to standards, challenging but attainable, measureable, and based on assessment(s) that meet district criteria.
- If teacher and instructional manager are not able to agree on an appropriate set of goals, a third party individual (e.g., a district supervisor) will mediate and, if necessary, act as the final decision-maker.

New Haven Evaluators and support providers

- Instructional managers are responsible for giving final rating
- They may be principals, assistant principals, or “as necessary and appropriate, a designee”
- There are also coaches (instructional and content), lead teachers, and mentors
 - May have no teaching load or reduced load
 - May be itinerant or school-based

New Haven Measures by “group”

Group	Teachers by Subject and Grade	Growth Measures to Be Used in 2010 - 2011	Growth Measures to Be Used in the Long-term
1	General Ed (including Bilingual) (K-3)	<ul style="list-style-type: none"> Teacher and IM selected (2+) 	<ul style="list-style-type: none"> District-wide assessment aligned to guiding principles Portfolio-based assessment of 21st Century Competencies Teacher and IM selected (as needed)
2	General Ed (including Bilingual) (4-6)	<ul style="list-style-type: none"> CMT (Reading, Math, Writing) Teacher and IM selected (1+) 	<ul style="list-style-type: none"> CMT (Reading, Math, Writing) District-wide assessment aligned to guiding principles
3	English & Math (7-8)	<ul style="list-style-type: none"> CMT (Reading, Math, Writing) Teacher and IM selected (1+) 	<ul style="list-style-type: none"> Portfolio-based assessment of 21st Century Competencies Teacher and IM selected (as needed)
4	Social Studies, Science, & World Languages (7-8)	<ul style="list-style-type: none"> Teacher and IM selected (2+) 	<ul style="list-style-type: none"> District-wide assessment aligned to guiding principles Portfolio-based assessment of 21st Century Competencies Teacher and IM selected (as needed)
5	English, Math, Social Studies, Science, & World Languages (9-12)	<ul style="list-style-type: none"> Teacher and IM selected (2+) 	
6	Specials/Electives (e.g. Art, PE, Music, Tech Ed) (K-12)	<ul style="list-style-type: none"> Teacher and IM selected (2+) 	<ul style="list-style-type: none"> Portfolio-based assessment of 21st Century Competencies Teacher and IM selected (1+)
7	ESL (K-12)	<ul style="list-style-type: none"> CMT (Reading, Writing) where appropriate / applicable by grade Teacher and IM selected (1-2+) 	<ul style="list-style-type: none"> CMT (Reading, Writing) where applicable by grade District-wide LA assessment aligned to guiding principles, where appropriate Portfolio-based assessment of 21st Century Competencies Teacher and IM selected (as needed)
8	Special Education (K-12)	<ul style="list-style-type: none"> CMT or MAS (Reading, Math, Writing) where appropriate / applicable by grade and student inclusion Teacher and IM selected, based on IEP (1-2+) 	<ul style="list-style-type: none"> CMT or MAS (Reading, Math, Writing), where appropriate and applicable by grade District-wide assessment aligned to guiding principles, where appropriate Portfolio-based assessment of 21st Century Competencies Teacher and IM selected, based on IEP (as needed)
9	NHFT *not* primary instructors	<ul style="list-style-type: none"> Teacher and IM selected (2+) 	<ul style="list-style-type: none"> Teacher and IM selected (2+)

New Haven assessment examples

➤ Examples of Assessments/Measures

- Basic literacy assessments, DRA
- District benchmark assessments
- District Connecticut Mastery Test
- LAS Links (English language proficiency for ELLs)
- Unit tests from NHPS approved textbooks
- Off-the-shelf standardized assessments (aligned to standards)
- Teacher-created assessments (aligned to standards)
- Portfolios of student work (aligned to standards)
- AP and International Baccalaureate exams

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

“The ratings for the three evaluation components will be synthesized into a final summative rating at the end of each year. Student growth outcomes will play a preponderant role in the synthesis.”

Washington DC IMPACT: Educator Groups

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

DC Impact: Score comparison for Groups 1-3

	Group 1 (tested subjects)	Group 2 (non-tested subjects)	Group 3 (special education)
Teacher value-added (based on test scores)	50%	0%	0%
Teacher-assessed student achievement (based on non-VAM assessments)	0%	10%	10%
Teacher and Learning Framework (observations)	35%	75%	55%

Washington DC IMPACT: Instructions for teachers in non-tested subjects/ grades

“In the fall, you will meet with your administrator to decide which assessment(s) you will use to evaluate your students’ achievement. If you are using multiple assessments, you will decide how to weight them. Finally, you will also decide on your specific student learning targets for the year. Please note that your administrator must approve your choice of assessments, the weights you assign to them, and your achievement targets. Please also note that your administrator may choose to meet with groups of teachers from similar content areas rather than with each teacher individually.”

Washington DC IMPACT: Rubric for Determining Success (for teachers in non-tested subjects/grades)

TEACHER-ASSESSED STUDENT ACHIEVEMENT DATA (TAS) RUBRIC

LEVEL 4 (HIGHEST)

LEVEL 3

TAS 1: TEACHER-ASSESSED STUDENT ACHIEVEMENT DATA

TAS
1

Student scores on teacher assessments indicate, on average, **exceptional** learning, such as at least 1.5 years of growth or at least 90% mastery of content standards; assessments used are **approved** by the administration; and scores reported are **validated** by the administration.

Student scores on teacher assessments indicate, on average, **significant** learning, such as at least 1.25 years of growth or at least 80% mastery of content standards; assessments used are **approved** by the administration; and scores reported are **validated** by the administration.

Note: If a teacher uses more than one assessment, each will be rated individually and the scores will be averaged together.

Washington DC IMPACT: Rubric for Determining Success (for teachers in non-tested subjects/grades)

LEVEL 2

Student scores on teacher assessments indicate, on average, **some** learning, such as at least 1 year of growth or at least 70% mastery of content standards; assessments used are **approved** by the administration; and scores reported are **validated** by the administration.

LEVEL 1 (LOWEST)

Student scores on teacher assessments indicate, on average, **little** learning, such as less than 1 year of growth or less than 70% mastery of content standards; assessments used are **not approved** by the administration; or scores reported are **not validated** by the administration.

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

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
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 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.			

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"

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

Principal Evaluation: Interstate School Leaders Licensure Consortium (ISSLC) Standards

Standard 1: A school administrator is an educational leader who promotes the success of all students by facilitating the development, articulation, implementation, and stewardship of a vision of learning that is shared and supported by the school community.

Standards 2: A school administrator is an educational leader who promotes the success of all students by advocating, nurturing, and sustaining a school culture and instructional program conducive to student learning and staff professional growth.

Standard 3: A school administrator is an educational leader who promotes the success of all students by ensuring management of the organization, operations, and resources for a safe, efficient, and effective learning environment.

Principal Evaluation: Interstate School Leaders Licensure Consortium (ISSLC) Standards (cont'd)

Standard 4: A school administrator is an educational leader who promotes the success of all students by collaborating with families and community members, responding to diverse community interests and needs, and mobilizing community resources.

Standard 5: A school administrator is an educational leader who promotes the success of all students by acting with integrity, fairness, and in an ethical manner.

Standard 6: A school administrator is an educational leader who promotes the success of all students by understanding, responding to, and influencing the larger political, social, economic, legal, and cultural context.

Vanderbilt Assessment of Leadership in Education (VAL-Ed)

- “The instrument consists of 72 items defining six core component subscales and six key process subscales.
- Principal, Teachers, & Supervisor provide a 360-degree, evidenced-based assessment of leadership behaviors.
- Respondents rate effectiveness of 72 behaviors on scale 1=Ineffective to 5=Outstandingly effective.
- Each respondent rates the principal’s effectiveness after indicating the sources of evidence on which the effectiveness is rated.
- Two parallel forms of the assessment facilitate measuring growth over time.
- The instrument will be available in both paper and online versions.”

Vanderbilt Assessment of Leadership in Education (VAL-Ed) Definitions

- **"High Standards for Student Learning"** —There are individual, team, and school goals for rigorous student academic and social learning.
- **Rigorous Curriculum (content)** —There is ambitious academic content provided to all students in core academic subjects.
- **Quality Instruction (pedagogy)** —There are effective instructional practices that maximize student academic and social learning.
- **Culture of Learning & Professional Behavior** —There are integrated communities of professional practice in the service of student academic and social learning. There is a healthy school environment in which student learning is the central focus.
- **Connections to External Communities** —There are linkages to family and/or other people and institutions in the community that advance academic and social learning.
- **Performance Accountability** — Leadership holds itself and others responsible for realizing high standards of performance for student academic and social learning. There is individual and collective responsibility among the professional staff and students."

Vanderbilt Assessment of Leadership in Education (VAL-Ed)

- **"Planning**—Articulate shared direction and coherent policies, practices, and procedures for realizing high standards of student performance.
- **Implementing**—Engage people, ideas, and resources to put into practice the activities necessary to realize high standards for student performance.
- **Supporting**—Create enabling conditions; secure and use the financial, political, technological, and human resources necessary to promote academic and social learning.
- **Advocating**—Promotes the diverse needs of students within and beyond the school.
- **Communicating**—Develop, utilize, and maintain systems of exchange among members of the school and with its external communities.
- **Monitoring**—Systematically collect and analyze data to make judgments that guide decisions and actions for continuous improvement."

Vanderbilt Assessment of Leadership in Education (VAL-Ed)

This table represents the conceptual framework for VAL-Ed. Each cell represents the cross-section of one core component and one key process of principal leadership. Every item in the Principal, Supervisor, and Teacher Response form represents a cross-section of one core component and one key process.

Core Components	Key Processes					
	Planning	Implementing	Supporting	Advocating	Communicating	Monitoring
High Standards for Student Learning						
Rigorous Curriculum (content)						
Quality Instruction (pedagogy)						
Culture of Learning & Professional Behavior						
Connections to External Communities						
Performance Accountability						

Vanderbilt Assessment of Leadership in Education (VAL-Ed)

High Standards for Student Learning	Sources of Evidence Check Key Sources of Evidence						Effectiveness Rating Circle One Number to Indicate How Effective				
	Reports from Others	Personal Observations	School Documents	School Projects or Activities	Other Sources	No Evidence	Ineffective	Minimally Effective	Satisfactorily Effective	Highly Effective	Outstandingly Effective

How effective is the principal at ensuring the school ...

Planning	1. plans rigorous growth targets in learning for all students.							1	2	3	4	5
	2. plans targets of faculty performance that emphasize improvement in student learning.							1	2	3	4	5

North Carolina School Executive Evaluation Goals

- The principal/assistant principal performance evaluation process will:
- Serve as a guide for principals/assistant principals as they reflect upon and improve their effectiveness as school leaders;
 - Inform higher education programs in developing the content and requirements of degree programs that prepare future principals/assistant principals;
 - Focus the goals and objectives of districts as they support, monitor and evaluate their principals/assistant principals;
 - Guide professional development for principals/assistant principals; and
 - Serve as a tool in developing coaching and mentoring programs for principals/assistant principals.

North Carolina School Executive Evaluation Process

- Principals and Assistant Principals are evaluated annually
- Focus is “formative professional development” (non-threatening, collegial)
- Assesses performance in relation to NC Standards for School Executives
- All school executives and those who will evaluate them must complete approved state training on the rubric and evaluation process

Principal Evaluation Instruments

Vanderbilt Assessment of Leadership in Education

<http://www.valed.com/>

- Also see the VAL-Ed Powerpoint at http://peabody.vanderbilt.edu/Documents/pdf/LSI/VALED_AssessLCL.ppt

North Carolina School Executive Evaluation Rubric

<http://www.ncpublicschools.org/profdev/training/principal/>

- Also see the NC “process” document at <http://www.ncpublicschools.org/docs/profdev/training/principal/principal-evaluation.pdf>

Iowa’s Principal Leadership Performance Review

<http://www.sai-iowa.org/principaleval>

Ohio’s Leadership Development Framework

http://www.ohioleadership.org/pdf/OLAC_Framework.pdf

References

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>

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>

References (continued)

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>

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>

References (continued)

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>

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>

References (continued)

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>

Tymms, P., Jones, P., Albane, S., & Henderson, B. (2009). The first seven years at school. *Educational Assessment Evaluation and Accountability*, 21(1), 67-80.

<http://www.springerlink.com/content/wm06474757652100/>

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>

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?





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