NATIONAL COMPREHENSIVE CENTER FOR TEACHER QUALITY

Models for Evaluating Teacher Effectiveness

Laura Goe, Ph.D.

CCSSO National Summit on Educator Effectiveness

April 29, 2011 ♦ Washington, DC

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

ETS®

VANDERBILT®

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

2

The goal of teacher evaluation

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

TO IMPROVE TEACHING AND LEARNING

Some assumptions for this working session

1. States are interested in developing comprehensive teacher evaluation systems that include student learning growth and multiple measures
2. States would like to create systems that align with key priorities (rigor, comparability, two points in time)
3. States are interested not only in "compliance" but also improving 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).

Race to the Top definition of student achievement

Student achievement means—

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

(b) **For non-tested grades and subjects:** alternative measures of student learning and performance such as student scores on pre-tests and end-of-course tests; student performance on English language proficiency assessments; and other measures of student achievement that are rigorous and comparable across classrooms.

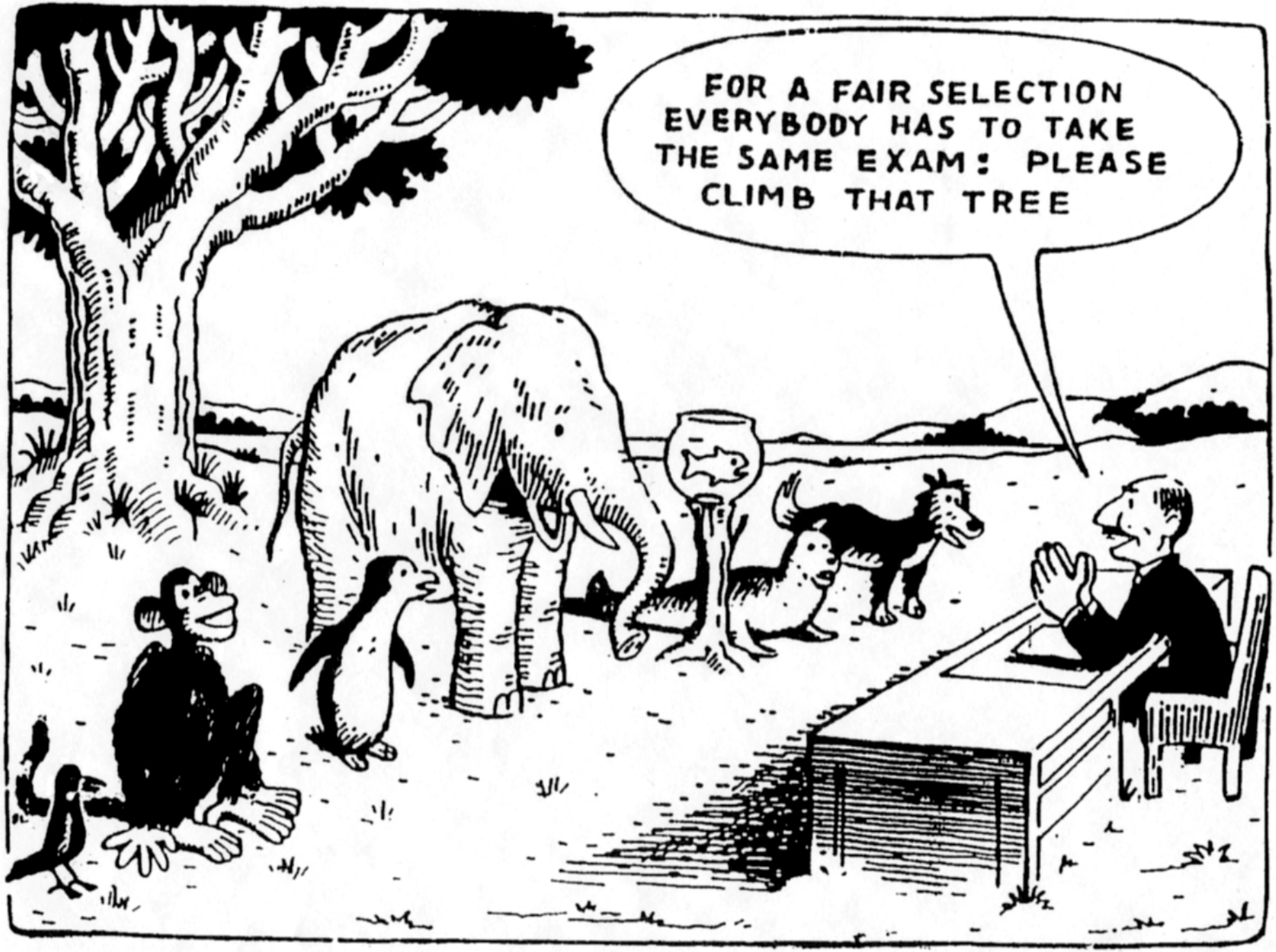
6

Multiple measures of student learning

- Standardized tests (state/district tests)
  - Typically use students' prior test scores from previous grades to show growth using growth models such as EVAAS (value-added) or Colorado Growth Model
  - With a pre- and post-test design, students may be tested in the same academic year (fall/spring)
- Classroom-based assessments such as DRA, DIBELS, curriculum-based tests, unit tests
  - Given in the classroom to individuals or groups of students
  - Measures growth in the academic year
  - Processes for using tests can be standardized

Multiple measures of student learning (continued)

- The 4 Ps: portfolios, projects, products, and performances
  - Examples: essays; written responses to complex questions in various subjects; research projects; capstone projects; art portfolios; live or videotaped music and theatrical performances; performance of specific physical activities for physical education; student-created videos in various subjects; products created by students in woodworking, welding, culinary arts, etc.


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

VAMs don't measure most teachers

- At least 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

Limitations of standardized tests

- **What students know:** Even in subjects and grades where we can measure student growth with standardized tests, such tests do not capture all important aspects of student learning growth
- **What students know *and can do*:** Curriculum-based tests and the 4 Ps may provide different information about student learning growth

Questions to ask about student growth measures

For evaluating teacher effectiveness

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 student growth 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 student growth measures

For improving teaching and learning

5. **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 student learning growth aspects of teacher 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?
2. **Professional growth.** Can results from the measures be aligned with professional growth opportunities?

*Models in this case 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.

Evaluation System Models

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

http://archive.austinisd.org/inside/initiatives/compensation/slos.phtml **Delaware** Model (Teacher participation in identifying grade/subject measures which then must be approved by state)

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

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

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

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

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

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

Evaluation System Models (cont'd)

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

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

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

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

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
Assessment • 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	Assessment • Variety of levels of questions (Beginning, Progressing, Proficient, Advanced) • Sufficient number of items • Grade level appropriate • Measures what is intended	Assessment • Addresses 2 or 3 levels of questions • Spread of questions is insufficient • Grade level appropriate • Mostly measures what is intended	Assessment • Addresses only 1 level of questions • Insufficient number of questions • Not grade level appropriate • Does not measure what is intended
Objective • Reflects a high need • Yearlong objective • Grade level appropriate • Deepens and extends knowledge for all students	Objective • Reflects a significant need • Yearlong objective • Grade level appropriate_	Objective • Addresses a need • Yearlong objective • Grade level appropriate	Objective • Does not address a need • Not a yearlong objective • Not grade level appropriate
Growth Target • Addresses more than 75% of students • Substantial growth expected (2 or more years) • Students and teachers exceeding expectations	Growth Target • 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	Growth Target • Addresses fewer than 75% of students • Moderate individual growth (less than one year) • Students and teachers barely meet expectations	Growth Target • Does not address 75% of students • Minor individual student growth (less than ½ year) • Students and teachers do not meet expectations

Draft 7-26-10

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

Student learning rating

+

Professional practice rating

+

Professional responsibilities rating

Final evaluation rating

The student learning rating is determined by a combination of different sources of evidence of student learning. These sources fall into three categories:

Category 1: Student growth on state standardized tests (e.g., NECAP, PARCC)

Category 2: Student growth on standardized district-wide tests (e.g., NWEA, AP exams, Stanford-10, ACCESS, etc.)

Category 3: Other local school-, administrator-, or teacher-selected measures of student performance

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)

Category 1 (growth on NECAP) + Category 2 (e.g., growth on NWEA) + Category 3 (e.g., principal review of student work over a six month span) = **Teacher A's student learning rating**

Teacher B (11th grade English)

Category 2 (e.g., AP English exam) + Category 3 (e.g., joint review of critical essay portfolio) = **Teacher B's student learning rating**

Teacher C (middle school art)

Category 3 (e.g., joint review of art portfolio) = **Teacher C's student learning rating**

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)	• Teacher and IM selected (2+)	• 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)	• CMT (Reading, Math, Writing) • Teacher and IM selected (1+)	• CMT (Reading, Math, Writing) • District-wide assessment aligned to guiding principles • Portfolio-based assessment of 21st Century Competencies • Teacher and IM selected (as needed)
3	English & Math (7-8)	• CMT (Reading, Math, Writing) • Teacher and IM selected (1+)	
4	Social Studies, Science, & World Languages (7-8)	• Teacher and IM selected (2+)	• 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)	• Teacher and IM selected (2+)	
6	Specials/Electives (e.g. Art, PE, Music, Tech Ed) (K-12)	• Teacher and IM selected (2+)	• Portfolio-based assessment of 21st Century Competencies • Teacher and IM selected (1+)
7	ESL (K-12)	• CMT (Reading, Writing) where appropriate / applicable by grade • Teacher and IM selected (1-2+)	• 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)	• CMT or MAS (Reading, Math, Writing) where appropriate / applicable by grade and student inclusion • Teacher and IM selected, based on IEP (1-2+)	• 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	• Teacher and IM selected (2+)	• 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"

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

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

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."

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

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.).

	☐ 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 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/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 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 including classroom observations

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

35

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