NATIONAL COMPREHENSIVE CENTER FOR TEACHER QUALITY

State Approaches to Measuring Student Growth for the Purpose of Teacher Evaluation

Introduction: Lynn Holdheide, Vanderbilt University

Presentation: Laura Goe, Ph.D., ETS

May 2012

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

ETS®

VANDERBILT®

Series of Webinars

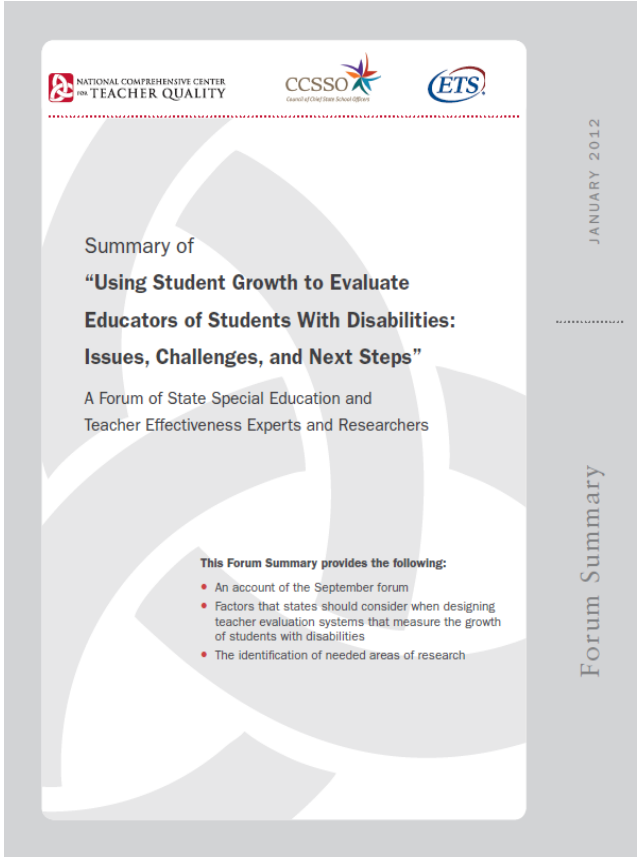
- **Webinar 1 (recorded):** State Approaches to Measuring Student Growth for the Purpose of Teacher Evaluation
- **Webinar 2:** Challenges and Considerations in Measuring the Growth of Students With disabilities

  Date: Thursday, June 7, 2012 **Time: 1 to 2:30 p.m. Eastern Time**
- **Webinar 3:** Measuring the Growth of Students Participating in the Alternate Assessment

  Date: Thursday, August 9, 2012 **Time: 1 to 2:30 p.m. Eastern Time**

A Forum of State Special Education and Teacher Effectiveness Experts and Researchers

- Identify challenges in using the growth of students with disabilities to evaluate educators.
- Develop considerations for states when designing systems that include the academic growth of students with disabilities.
- Identify needed areas of research.
- Develop initial recommendations to inform policy and practice, where feasible.



Summary Report available online:
http://www.tqsource.org/pdfs/TQ_Forum_SummaryUsing_Student_Growth.pdf

Webinar 1 Learning Targets

Seeks to build the capacity of participants to:

- Recognize and understand various approaches states and districts are using or contemplating for measuring student growth within educator performance evaluations.
- Actively participate in the subsequent webinars specific to measuring growth of students with disabilities.

Laura Goe, Ph.D.

- Principal Investigator for the National Comprehensive Center for Teacher Quality
- Research Scientist in the Performance Research Group at ETS

The Goal of Teacher Evaluation

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

TO IMPROVE TEACHING AND LEARNING

Measures and Models: Definitions

- Measures are the instruments, assessments, protocols, rubrics, and tools that are used in determining teacher effectiveness.
- Models are the state or district systems of teacher evaluation, including all of the inputs and decision points (e.g., measures, instruments, processes, training, and scoring) that result in determinations about the effectiveness of individual teachers.

Race to the Top
Definition of *Student Growth*

- ***Student growth*** means the change in student achievement (as defined in this notice) for an individual student between two or more points in time.

  A state also may include other measures that are rigorous and comparable across classrooms (p.11).

8

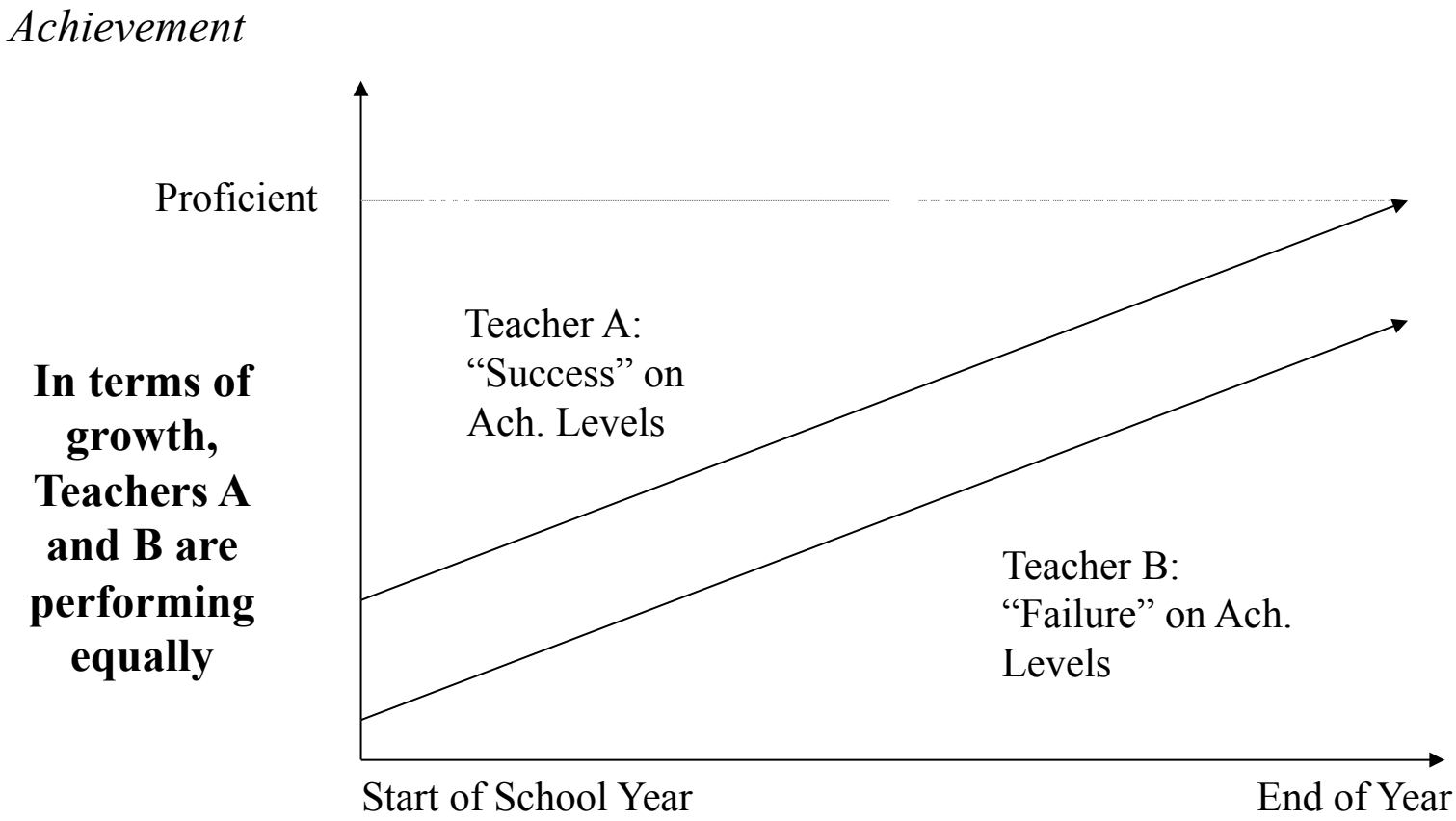
Multiple Measures of Teacher Effectiveness

- **Evidence of *growth in student learning and competency***
  - Standardized tests, pre/post tests in untested subjects
  - Student performance (e.g., art, music)
  - Curriculum-based tests given in a standardized manner
  - Classroom-based tests such as DIBELS
- **Evidence of *instructional quality***
  - Classroom observations
  - Lesson plans, assignments, and student work
  - Student surveys, such as Harvard's Tripod
  - Evidence binder (next generation of portfolio)
- **Evidence of *professional responsibility***
  - Administrator/supervisor reports, parent surveys
  - Teacher reflection and self-reports, records of contributions

What Nearly All State and District Models Have in Common

- Value-added or Colorado Growth Model will be used for those teachers in tested grades and subjects (Grades 4-8 ELA and Math in most states).
- States want to increase the number of tested subjects and grades so that more teachers can be evaluated with growth models.
- States are trying various approaches to measuring teachers' contribution to student growth in nontested subjects and grades.

Growth vs. Proficiency Models (1)

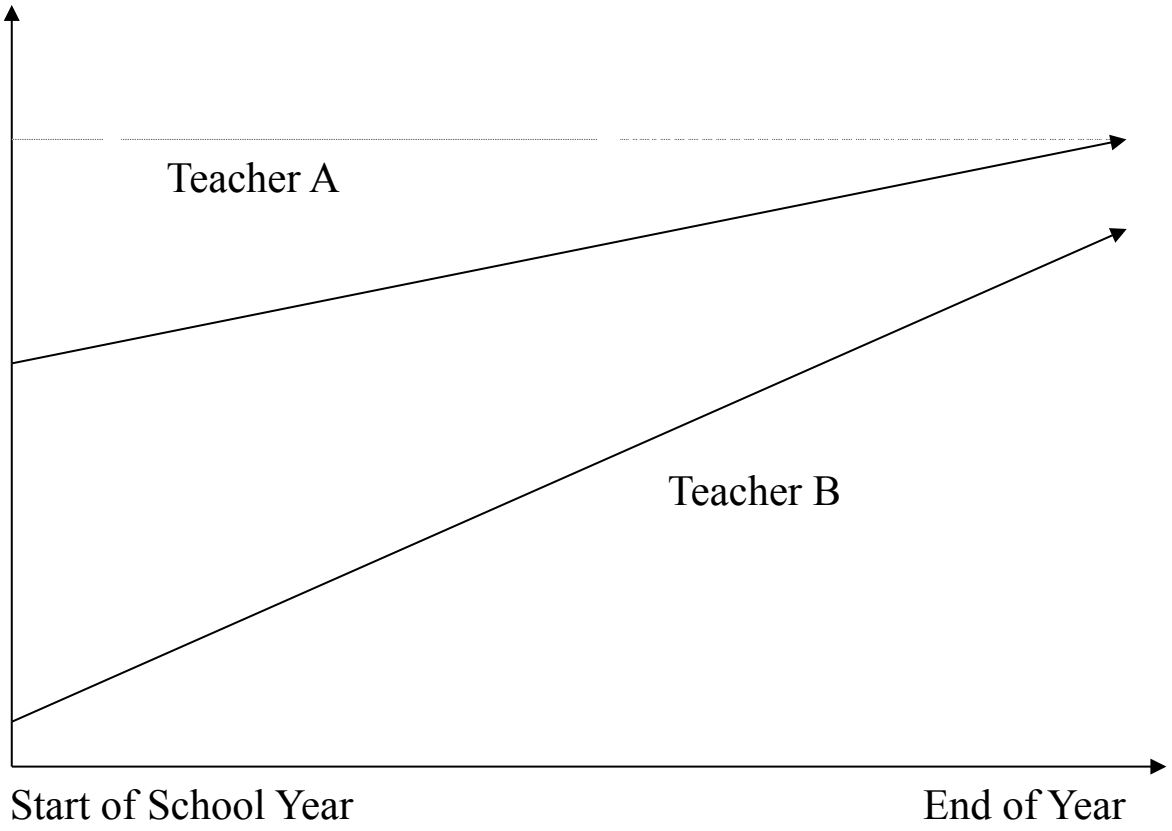


In terms of growth, Teachers A and B are performing equally

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

Growth vs. Proficiency Models (2)

A teacher with low-proficiency students can still be high in terms of GROWTH (and vice versa)



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

Value-Added Models

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

13

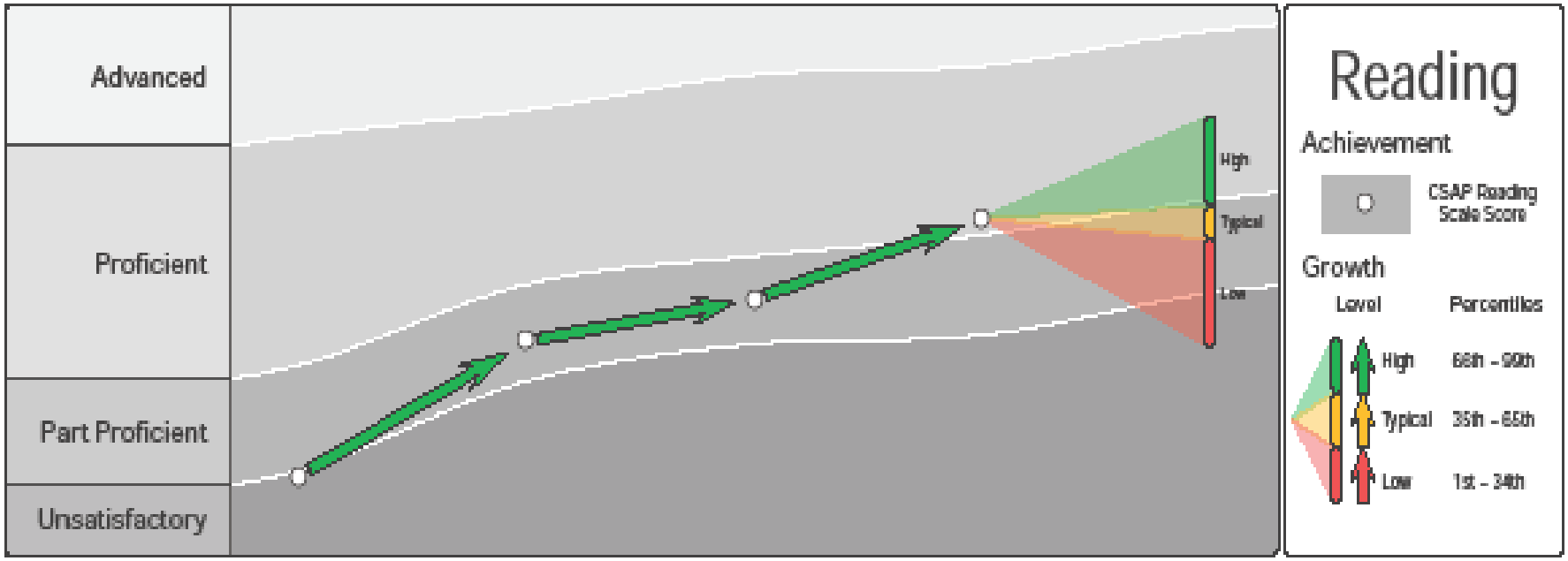
Evidence of Teachers' Contribution to Student Learning Growth

- Value-added can provide useful evidence of teacher's contribution to student growth
- "It is not a perfect system of measurement, but it can complement observational measures, parent feedback, and personal reflections on teaching far better than any available alternative." (Glazerman et al., 2010, p. 4).

Most popular growth models: Colorado Growth Model

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

Colorado Growth Model



	Grade 3 2006	Grade 4 2007	Grade 5 2008	Grade 6 2009	Next Year	
Scale Score Achievement Level	462 Unsatisfactory	539 Part Proficient	563 Part Proficient	609 Proficient		Achievement
Growth Percentile Growth Level		66 High	66 High	90 High		Growth

Slide courtesy of Damian Betebenner at www.nciea.org

What Value-added and Growth Models Cannot Tell You

- Value-added and growth models are really measuring *classroom,* 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!

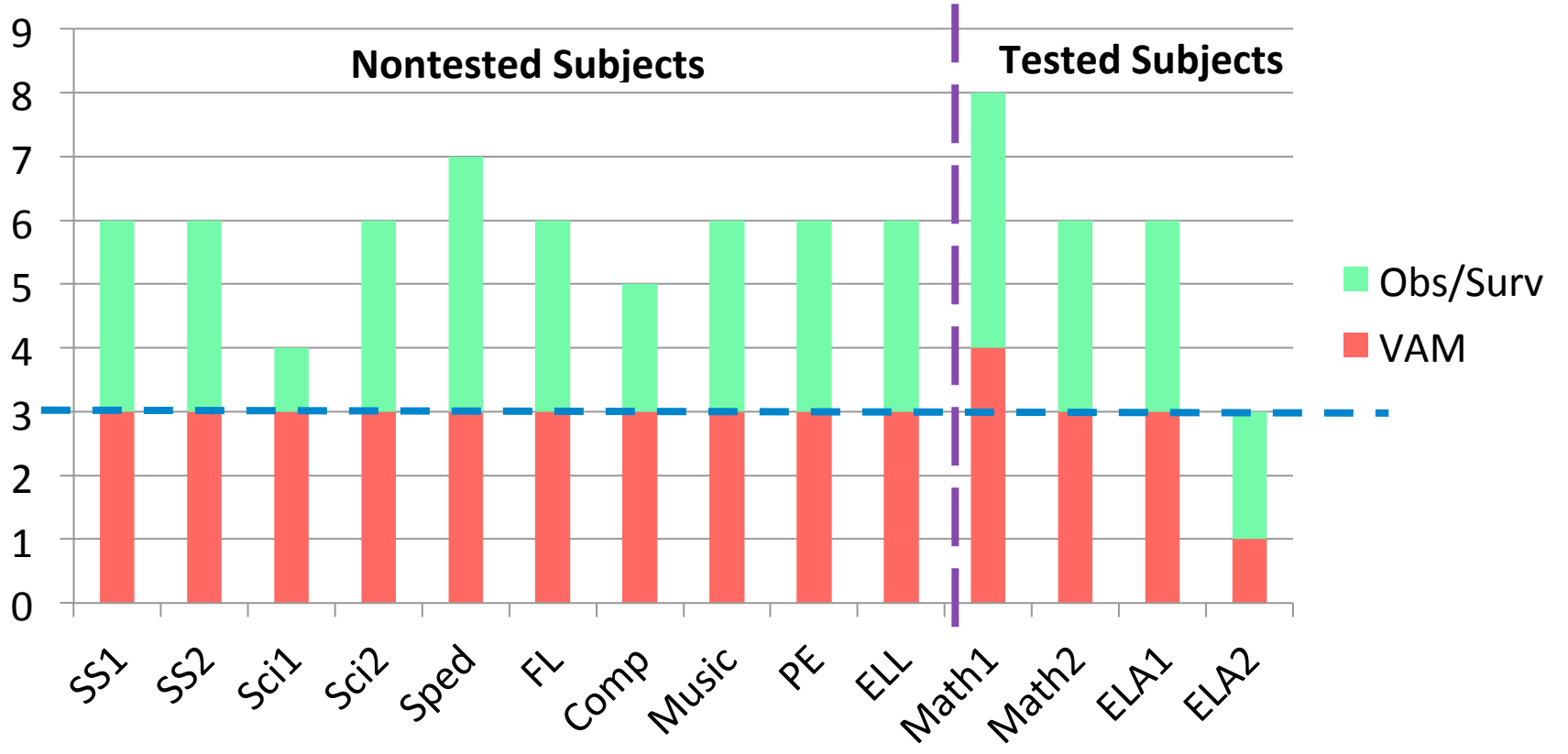
Recommendation from NBPTS Task Force (Linn et al., 2011)

Recommendation 2: Employ measures of student learning explicitly aligned with the elements of curriculum for which the teachers are responsible. This recommendation emphasizes the importance of ensuring that teachers are evaluated for what they are teaching.

Measuring Teachers' Contributions to Student Learning Growth: A Summary of Current Models

Model	Description
Student learning objectives	Teachers assess students at beginning of year and set objectives, then assess again at end of year; principal or designee works with teacher, determines success.
Subject- and grade-alike team models ("Ask a Teacher")	Teachers meet in grade-specific and/or subject-specific teams to consider and agree on appropriate measures that they will all use to determine their individual contributions to student learning growth.
Content Collaboratives	Content experts (external) identify measures and groups of content teachers consider the measures from the perspective of classroom use; may not include pretest and posttest measures.
Pretest and posttest model	Identify or create pretests and posttests for every grade and subject.
Schoolwide value-added	Teachers in tested subjects and grades receive their own value-added score; *all other teachers get the schoolwide average.*

Schoolwide VAM Illustration for Middle School

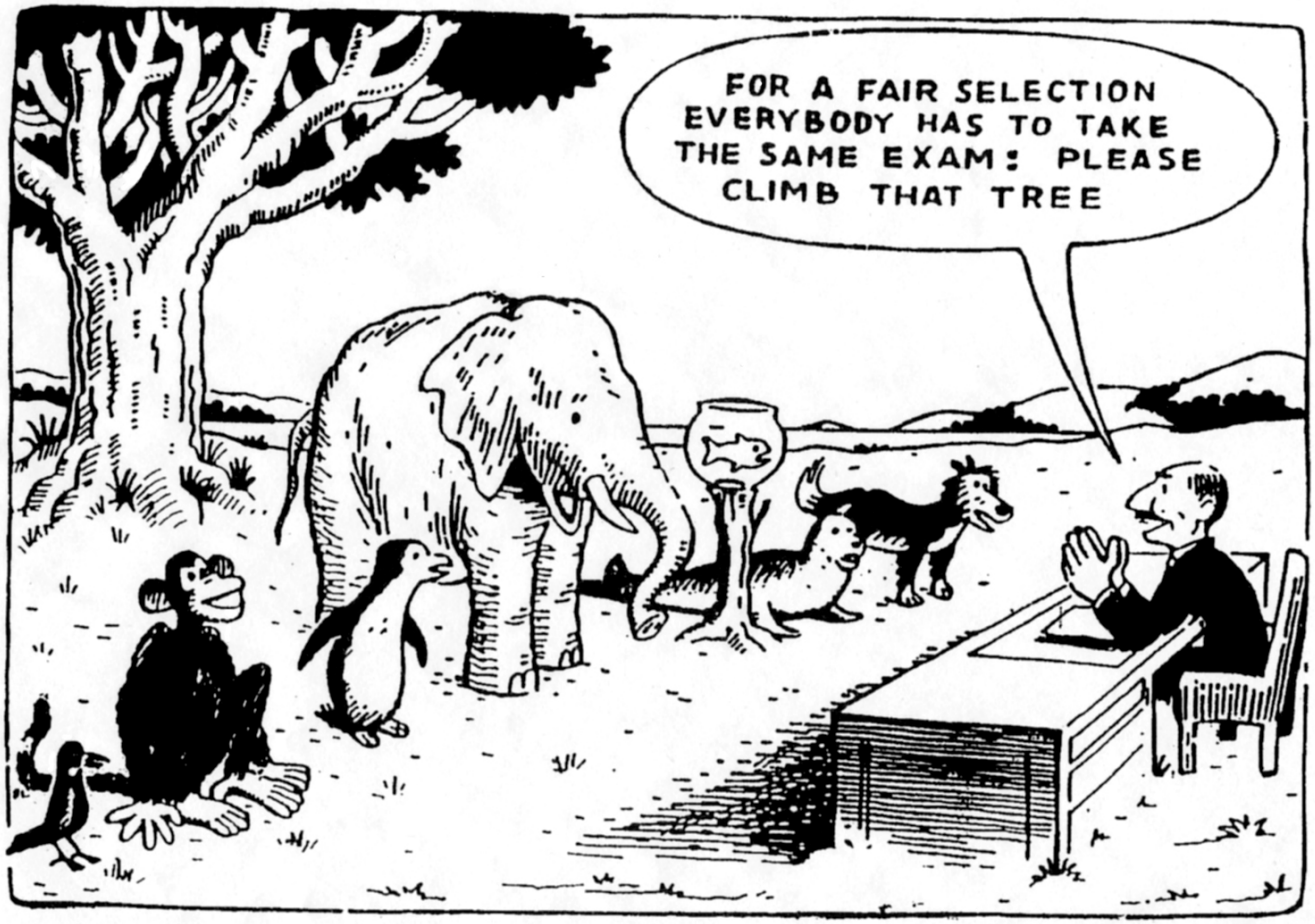


Tested subject teachers receive their own value-added score while nontested subject teachers receive a schoolwide average for their value-added score.

Differentiating Among Teachers

- “It is nearly impossible to discover and act on performance differences among teachers when documented records show them all to be the same.” (Glazerman et al., 2011, p.1)

Multiple Assessments Needed, Not One-Size-Fits All



Reprinted from The Educational System Comic *at weknowmemes.com*

Rhode Island's SLO Language

- "Student Learning Objectives are not set by educators in isolation; rather, they are developed by teams of administrators, grade-level teams or groups of content-alike teachers and, are aligned to district and school priorities, wherever possible." (p. 12)

  From Rhode Island's "*Guide to Measures of Student Learning for Administrators and Teachers 2011-2012*"
  http://www.ride.ri.gov/educatorquality/educatorevaluation/Docs/GuideSLO.pdf

The 4 Ps (Projects, Performances, Products, Portfolios)

- Yes, they can be used to demonstrate teachers' contributions to student learning growth
- Here's the basic approach
  - Use a high-quality rubric to judge initial knowledge and skills required for mastery of the standard(s).
  - Use the same rubric to judge knowledge and skills at the end of a specific time period (e.g., grading period, semester, year).

Assessing Musical Behaviors: The Type of Assessment Must Match The Knowledge or Skill

4 types of musical behaviors:	Types of assessment
1. Responding 2. Creating 3. Performing 4. Listening	1. Rubrics 2. Playing tests 3. Written tests 4. Practice sheets 5. Teacher Observation 6. Portfolios 7. Peer and Self-Assessment

Slide used with permission of authors Carla Maltas, Ph.D. and Steve Williams, M.Ed. (See reference list for details.)

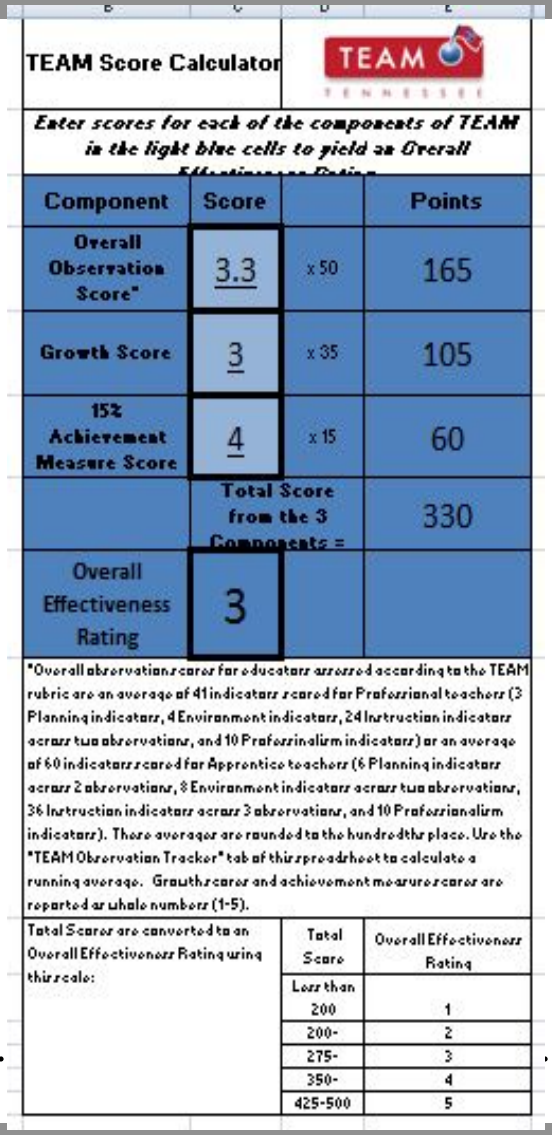
Collect Evidence in a Standardized Way (to the Extent Possible)

- Evidence of student learning growth
  - Locate or develop rubrics with explicit instructions and clear indicators of proficiency for each level of the rubric.
  - Establish time for teachers to collectively examine student work and come to a consensus on performance at each level.
    - Identify “anchor” papers or examples.
  - Provide training for teachers to determine how and when assessments should be given, and how to record results in specific formats.

Three Approaches to Combining Measures

- AIR's Sheri Frost Leo and Lisa Lachlan-Haché (2012) have written a really useful paper on combining and weighting measures
  - **Numerical approach:** Measures of teacher performance are quantified and added or averaged into a teacher effectiveness "score."
  - **Profile approach:** Performance data are gathered and maintained separately, without adding or averaging the results across metrics; then placed into rating categories for each of the measures.
  - **Holistic approach:** Principal may exercise considerable judgment in combining information to arrive at a final score.

An Example of the Numerical Approach: Tennessee Overall Score Calculator



TEAM Score Calculator

TEAM TENNESSEE

Enter scores for each of the components of TEAM in the light blue cells to yield an Overall Effectiveness Rating

Component	Score		Points
Overall Observation Score*	3.3	x 50	165
Growth Score	3	x 35	105
15% Achievement Measure Score	4	x 15	60
	Total Score from the 3 Components =		330
Overall Effectiveness Rating	3		

*Overall observation scores for educators assessed according to the TEAM rubric are an average of 41 indicators scored for Professional teachers (3 Planning indicators, 4 Environment indicators, 24 Instruction indicators across two observations, and 10 Professionalism indicators) or an average of 60 indicators scored for Apprentice teachers (6 Planning indicators across 2 observations, 8 Environment indicators across two observations, 36 Instruction indicators across 3 observations, and 10 Professionalism indicators). These averages are rounded to the hundredths place. Use the "TEAM Observation Tracker" tab of this spreadsheet to calculate a running average. Growth scores and achievement measures scores are reported as whole numbers (1-5).

Total Scores are converted to an Overall Effectiveness Rating using this scale:

Total Score	Overall Effectiveness Rating
Less than 200	1
200-	2
275-	3
350-	4
425-500	5

- Overall Observation Score x 50
- Growth Score x 35
- Achievement Measure Score x 15

- Overall Effectiveness Rating
  - 1 = Less than 200
  - 2 = 200+
  - 3 = 275+
  - 4 = 350+
  - 5 = 425-500

An Example of the Profile Approach: 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—teacher is very high on one area (practice or growth) and very low on the other area.

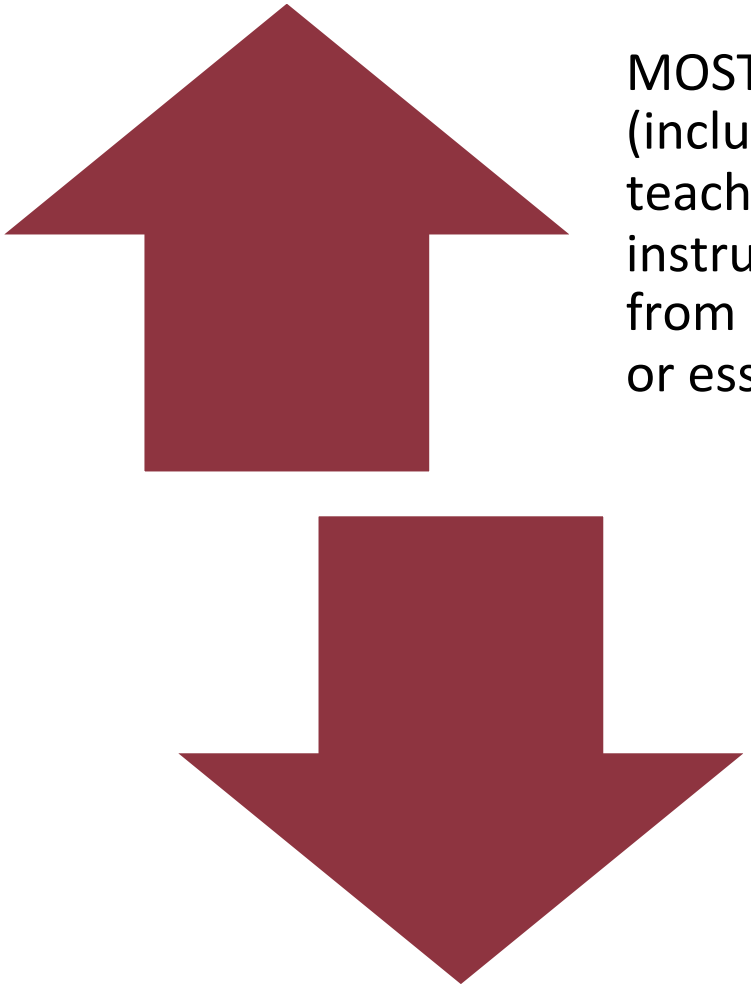
Teacher Evaluation in Isolated and/or Low-Capacity Districts

- External evaluators may need to be brought in for very small, isolated districts.
  - For example, a district where the superintendent is also principal, history teacher, and bus driver.
  - May also be needed when evaluators' objectivity is impacted by factors such as fear of losing teachers or damaging long-term relationships in the community.
- Evaluators could be "exchanged" across districts within a specific region ("you evaluate mine, and I'll evaluate yours") or regional evaluators could serve a set of districts.

Measures That Help Teachers Grow

- Measures that include protocols and processes that teachers can examine and comprehend
- Measures that are directly and explicitly aligned with teaching standards
- 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 portfolios)
- Measures that give teachers opportunities to discuss the results for formative purposes with evaluators, administrators, teacher learning communities, mentors, coaches, etc.
- Measures that are aligned with and used to inform professional growth and development offerings

Using Student Learning Outcomes to Inform Teacher Professional Growth



MOST helpful: Student assessments (including 4Ps) that provide information teachers can use immediately to adjust instructional strategies, such as results from benchmark or interim assessment or essays scores with rubrics

LEAST helpful: Student assessments that provide a snapshot of students' skills at a single point in time after most instruction is complete, such as last year's state standardized test results

Considerations for Implementing Measurement Systems

- Consider whether human resources and capacity are sufficient to ensure fidelity of implementation.
  - Poor implementation threatens validity of results.
- Establish a plan to evaluate measures to determine if they can effectively differentiate among teacher performance.
  - Need to identify potential "widget effects" in measures.
  - If measure is not differentiating among teachers, may be faulty training or poor implementation, *not the measure itself.*
- Examine correlations among results from measures.
- Evaluate processes and data each year and make needed adjustments.

Before You Implement Teacher Evaluation Systems, Ask Yourself…

- How will this component of the teacher evaluation system impact teaching and learning in classrooms and schools?
- How will this component look different in low-capacity vs. high-capacity schools?
- How will reporting on this component be done (to provide actionable information to teachers, principals, schools, districts, teacher preparation programs, and the state)?
- How will we know if this component is working as we intended?

References and Resources

Betebenner, D. W. (2008). *A primer on student growth percentiles.* Dover, NH: National Center for the Improvement of Educational Assessment (NCIEA). Retrieved from http://www.cde.state.co.us/cdedocs/Research/PDF/Aprimeronstudentgrowthpercentiles.pdf

Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D. O., & Whitehurst, G. J. (2011). *Passing muster: Evaluating evaluation systems.* Washington, DC: Brown Center on Education Policy. Retrieved from http://www.brookings.edu/reports/2011/0426_evaluating_teachers.aspx#

Leo, S. F., & Lachlan-Haché, L. (2012). *Creating summative educator effectiveness scores: Approaches to combining measures*. Washington, DC: American Institutes for Research.

Linn, R., Bond, L., Darling-Hammond, L., Harris, D., Hess, F., & Shulman, L. (2011). *Student learning, student achievement: How do teachers measure up?* Arlington, VA: National Board for Professional Teaching Standards. Retrieved from http://www.nbpts.org/index.cfm?t=downloader.cfm&id=1305

Malta, C., & Williams, S. (2010, January 27). *Meaningful assessment in the music classroom*. Presented at Missouri Music Educators Association Conference, Jefferson City, MO. Retrieved from http://dese.mo.gov/divimprove/curriculum/fa/AssessmentintheMusicClassroom.pptx

New Haven, CT Teacher Evaluation & Development System: http://www.nhps.net/scc/index

Race to the Top Application: http://www2.ed.gov/programs/racetothetop/resources.html

Rhode Island Educator Evaluation System: http://www.ride.ri.gov/educatorquality/educatorevaluation/

Tennessee Teacher Evaluation System: http://team-tn.org/

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. Retrieved from http://widgeteffect.org/downloads/TheWidgetEffect.pdf

Laura Goe, Ph.D.
609-619-1648
lgoe@ets.org
www.lauragoe.com
https://twitter.com/GoeLaura

National Comprehensive Center for Teacher Quality
1000 Thomas Jefferson Street NW
Washington, DC 20007
www.tqsource.org