



Evaluating Teacher and Principal Effectiveness

Laura Goe, Ph.D.

Legislative Conference

Council of Chief State School Officers

March 29, 2011

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



Laura Goe, Ph.D.

- Former teacher in rural & urban schools
 - Special education (7th & 8th grade, Tunica, MS)
 - Language arts (7th grade, Memphis, TN)
- Graduate of UC Berkeley's Policy, Organizations, Measurement & Evaluation doctoral program
- Principal Investigator for the National Comprehensive Center for Teacher Quality
- Research Scientist in the 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, dissemination

➤ 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**
 - Focus on models and measures that are closely aligned with teaching standards and student subject/content standards

Research Behind the Push for New Evaluation Measures and Systems

- Value-added research shows that teachers vary greatly in their contributions to student achievement; qualifications are poor predictors (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)

The role of teaching standards

- A set of practices teachers should aspire to
- A teaching tool in teacher preparation programs
- A guiding document with which to align:
 - Measurement tools and processes for teacher evaluation, such as classroom observations, surveys, portfolios/evidence binders, student outcomes, etc.
 - Teacher professional growth opportunities, based on evaluation of performance on standards
- A tool for coaching and mentoring teachers:
 - Teachers analyze and reflect on their strengths and challenges and discuss with consulting teachers

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
- Portfolio/Evidence Binder

➤ Evidence of *professional responsibility*

- Administrator/supervisor reports
- Parent surveys

Teacher evaluation

**TEACHER CLASSROOM EVALUATION CHECKLIST
(LONG FORM)**

Notes 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 # 1 2 3 4 5 (Circle One) Length of Evaluation: _____ minutes
 Meeting set with teacher to discuss evaluation: Date _____ Time _____
 Was an 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 expectations				
Maintains eye contact				
Displays self-confidence and self-control				
Maintains attendance				
Uses rewards with discretion				
Moves around the room				
Students (reports) understanding and comprehension				
Communicates clearly				
Calls (reports) by name				
Involves (reports) in learning tasks				



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

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"
- Resource-intensive 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

Most commonly used growth models

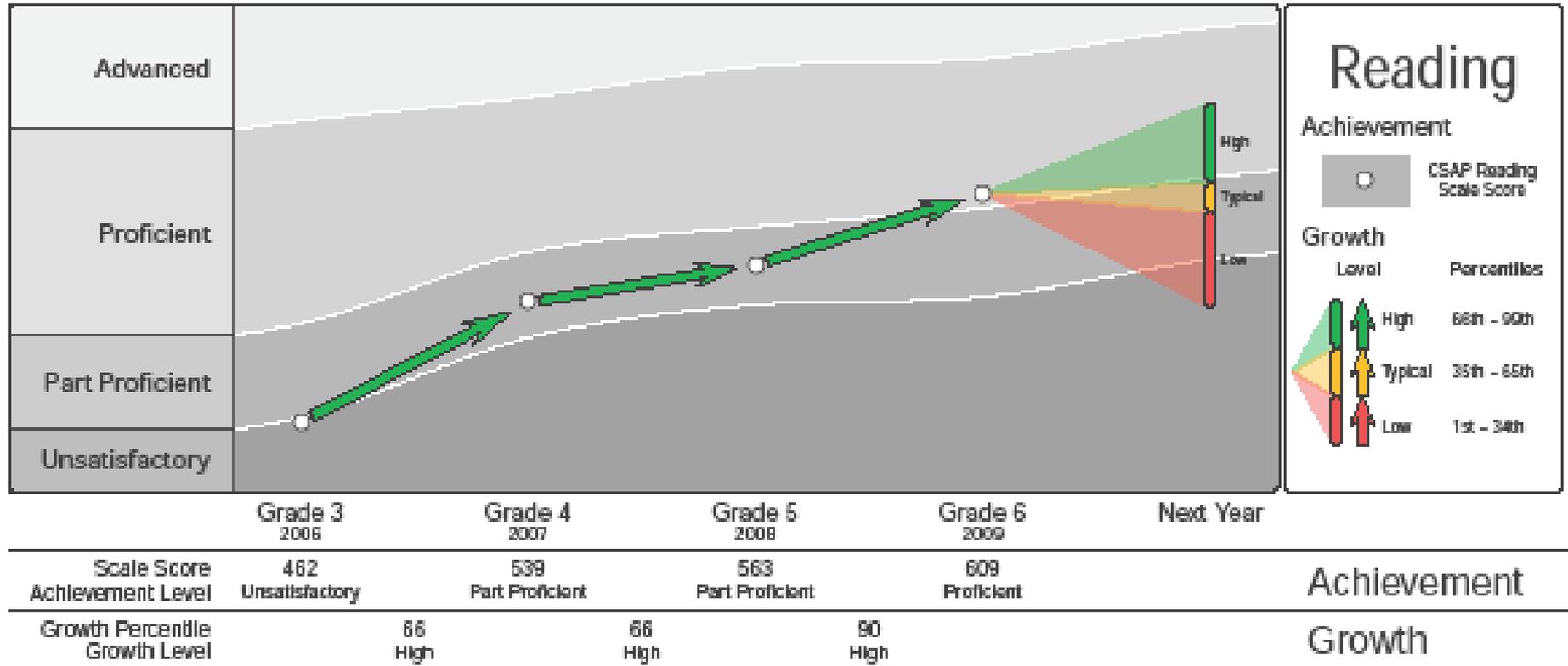
➤ Value-added models

- There are many versions of value-added models (VAMs), and results from the different models may vary depending on what is included in the model
- Most states and districts that use VAMs use the Sanders' model, also called EVAAS
- Prior achievement test scores are used to predict the next test score for students

➤ Colorado Growth model

- Focuses on “growth to proficiency”
- Measures students against “academic peers”

Sample student report: Colorado Growth Model



Slide courtesy of Damian Betebenner at www.nciea.org

What Growth Models Cannot Tell You

- Growth models are really measuring *classroom* effects, not teacher effects
- Growth 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!

Growth models don't measure most teachers

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

Measuring teacher effectiveness with student learning growth

- Not everything students know and can do shows up in a standardized test score
- Include multiple measures of student achievement, accomplishment, and progress
- Valid measures of student learning growth between two points may be useful in determining which students are making the most progress in a given teacher's classroom
- However, caveats apply!

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
- **Collect evidence in a standardized manner!**

Questions to ask about measures of 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 yardstick?
- 3. Growth over time.** Do the measures enable student learning growth to be assessed “between two points in time”?

Questions to ask about measures of teacher effectiveness (cont'd)

- 4. Standards-based.** Are the measures focused on assessing growth on important high-quality grade level and subject standards for students?
- 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 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.

How districts/states are using student achievement in teacher evaluation models

- Four basic types have emerged, with most systems being variations of these
- If you evaluate them against the “Questions to ask about measures” and “Questions to ask about models” you find substantial differences
- The goal is to create/adopt measures and models that allow you to answer “yes” to all or most of the questions

Student Learning Objectives (SLOs)

Teacher selects assessments, tests students, sets objects, then tests again at end of year

Examples:

- Rhode Island, New Haven CT, Washington DC
- Other districts have used SLOs in pay-for-performance plans (Denver CO, Austin TX, Charlotte-Mecklenburg NC)

Variations include:

- Rubric to determine “rigor” of SLO
- Group/team SLOs in addition to individual SLOs

Challenges include:

- Administrator burden (approving assessments, supervising, making determinations of success)

The Delaware/NYSUT Model

Teachers meet work in subject/grade alike teams to identify appropriate assessments; agree to use in a standardized manner

Examples

- Delaware, 6 districts in New York participating in the AFT Innovation Grant Project

Variations include:

- Statewide model (all teachers in state use same instruments & processes) vs. within-district model (all teachers in district use same instruments & processes)

Challenges include:

- Identifying & approving measures, scoring, standardization

The Hillsborough Model

Creating (or identifying) pre- and post-assessments for every grade and subject, essentially standardizing assessments for all

Examples

- Hillsborough, FL

Variations:

- None so far

Challenges include:

- Like other models, much easier to do at the district level; local control over teacher evaluation and differences in schedules, curriculum, resources, etc. make it difficult to do statewide
- Very expensive to create valid tests and assessments

The Teacher Advancement Program (TAP) Model

Teachers in tested grades/subjects receive their own value-added scores; teachers in non-tested grades/subjects receive the school-wide average for tested teachers

Examples:

- Many schools, some districts throughout the US

Variations:

- Some variation in observation component, not in testing

Challenges include:

- Students of teachers in non-tested subjects/grades are not assessed against the appropriate standards; their efforts and progress are not included in the evaluation

Trends in scoring teacher effectiveness

- Several trends are emerging in how to use measures (including student achievement) to arrive at a teachers' effectiveness score
- States/districts struggling with questions such as
 - Should we have a minimum score for each component of the model, or should a high score on observations make up for a low score on student achievement?
 - Should we evaluate teachers differently depending on grade, subject, specialty?
 - How much of the score should be based on student achievement growth and how much on other categories such as classroom practice and professional responsibilities?
 - Should other measures of student learning growth be included in evaluating teachers for whom standardized tests are used?

New Haven “matrix”

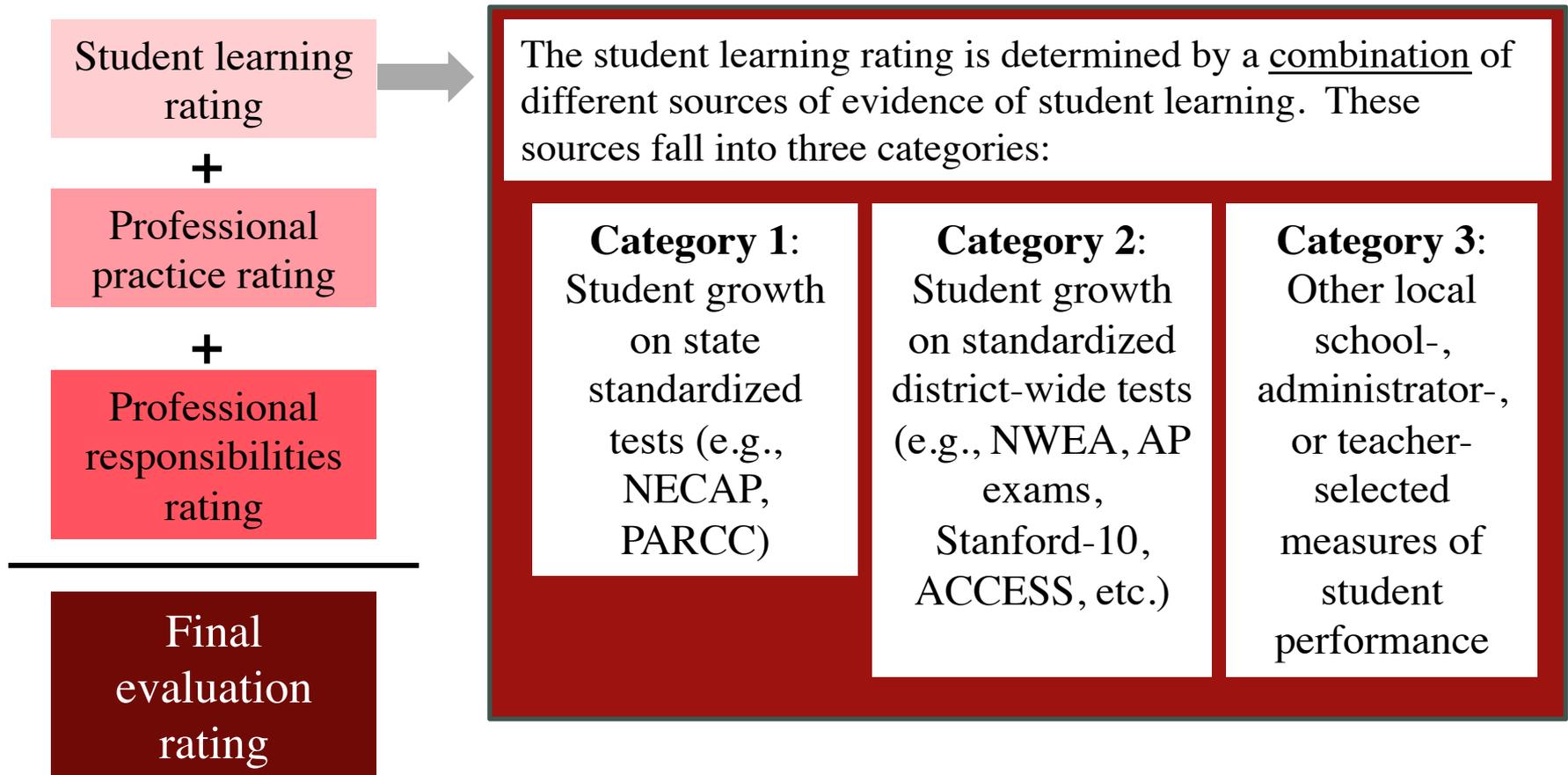
		Student Learning Growth				
		1	2	3	4	5
Instructional Practice and Professional Values	1	1	1	2	3*	3*
	2	1	2	2	3	4*
	3	1	2	3	4	5
	4	2*	3	4	4	5
	5	3*	3*	4	5	5

“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
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

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



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%

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

The role of principal standards

- A set of practices principals should aspire to
- A teaching tool for administrator preparation programs
- A guiding document with which to align measurement tools and processes for principal evaluation
 - Teacher, parent, and student surveys/interviews
 - Analysis of schooling outcomes (student achievement growth, promotion, school completion), resource use, recruitment & retention of effective teachers
 - Principal behavior, practice, activities (observations, records)

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

A school administrator is an educational leader who promotes the success of all students by

- 1.** Facilitating the development, articulation, implementation, and stewardship of a vision of learning that is shared and supported by the school community.
- 2.** Advocating, nurturing, and sustaining a school culture and instructional program conducive to student learning and staff professional growth.
- 3.** 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)

A school administrator is an educational leader who promotes the success of all students by

4. Collaborating with families and community members, responding to diverse community interests and needs, and mobilizing community resources.
5. Acting with integrity, fairness, and in an ethical manner.
6. 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)

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

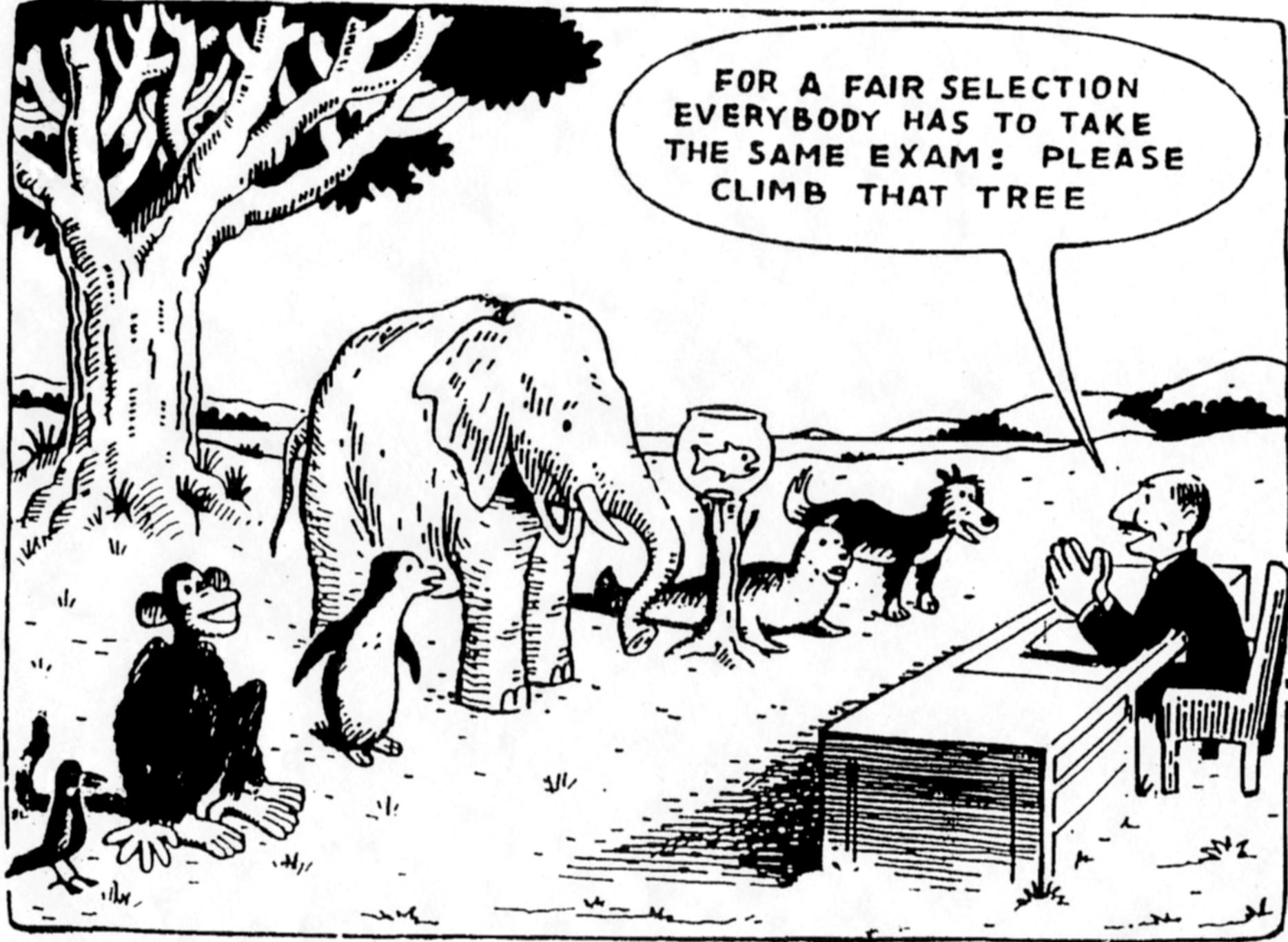
North Carolina School Executive Evaluation Goals

- 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.

Final thoughts

- We must continue to ask ourselves: *How is this instrument, model, system, measure, or process going to improve teaching and learning?*
- If our standards are clear indicators of the knowledge and skills we expect from teachers, principals, and students, then we should include them in our measures and models (not just those that are “easy” to measure)
- Changes in evaluation are revolutionary but must also be evolutionary: revisit, adjust, and evaluate measures, models, and processes; avoid rigidity

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



Popular 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

Some evaluation models to explore

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 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/Asnt_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)

Some principal evaluation instruments & models to explore

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

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>

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

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>

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>

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>

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