



Measuring Teacher Effectiveness in Untested Subjects and Grades

Laura Goe, Ph.D.

Educator Effectiveness Council

Annapolis, MD ♦ February 14, 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

The goal of teacher evaluation

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

**TO IMPROVE
TEACHING AND
LEARNING**

Teacher evaluation

**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, adapt 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 warm, 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 students				
Monitors student behavior				
Students display understanding and cooperation				
Classroom is orderly				
Calls students by name				
Involved students in learning tasks				



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

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*

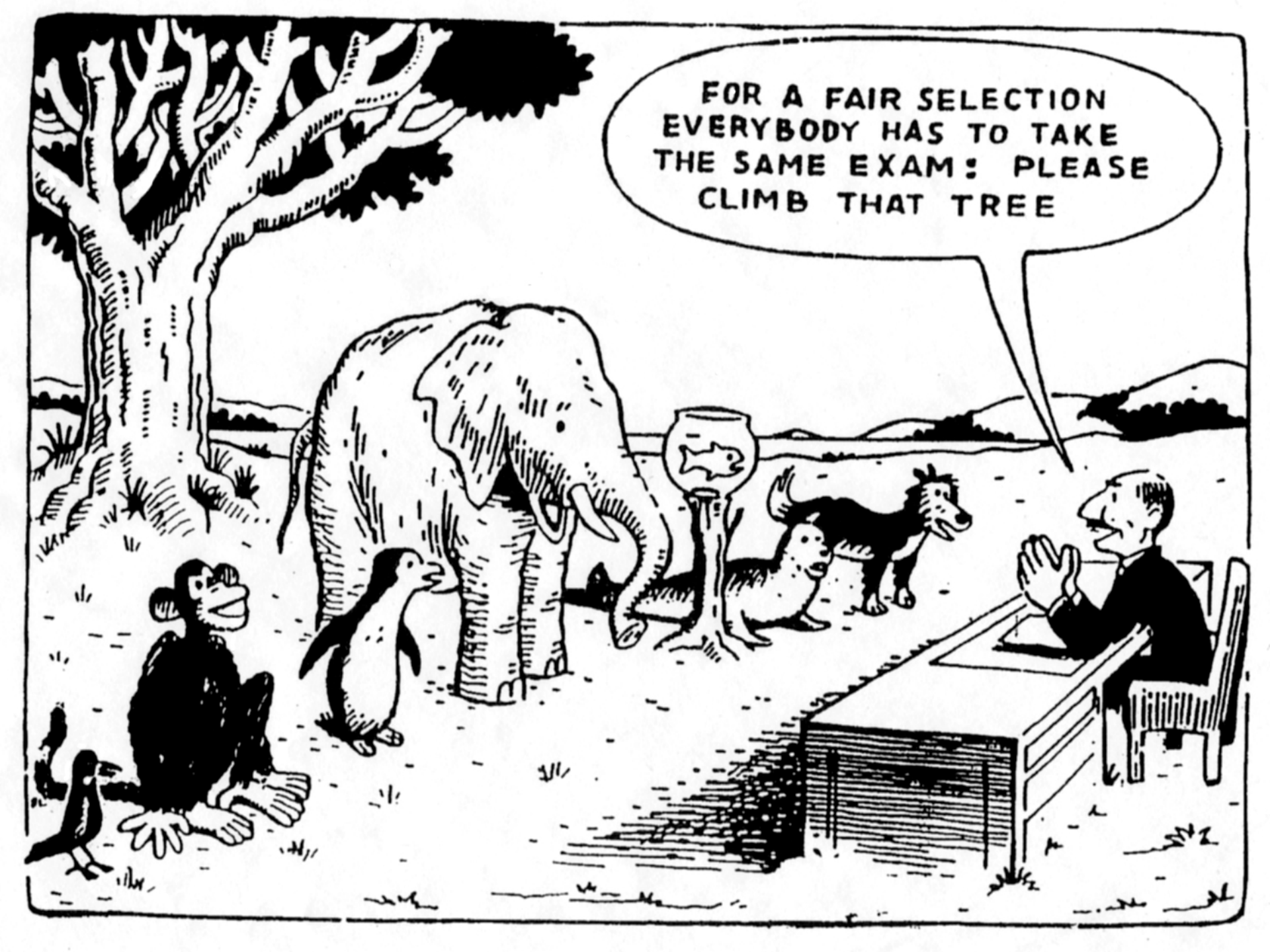
- Classroom observations
- Lesson plans, assignments, and student work

➤ Evidence of *professional responsibility*

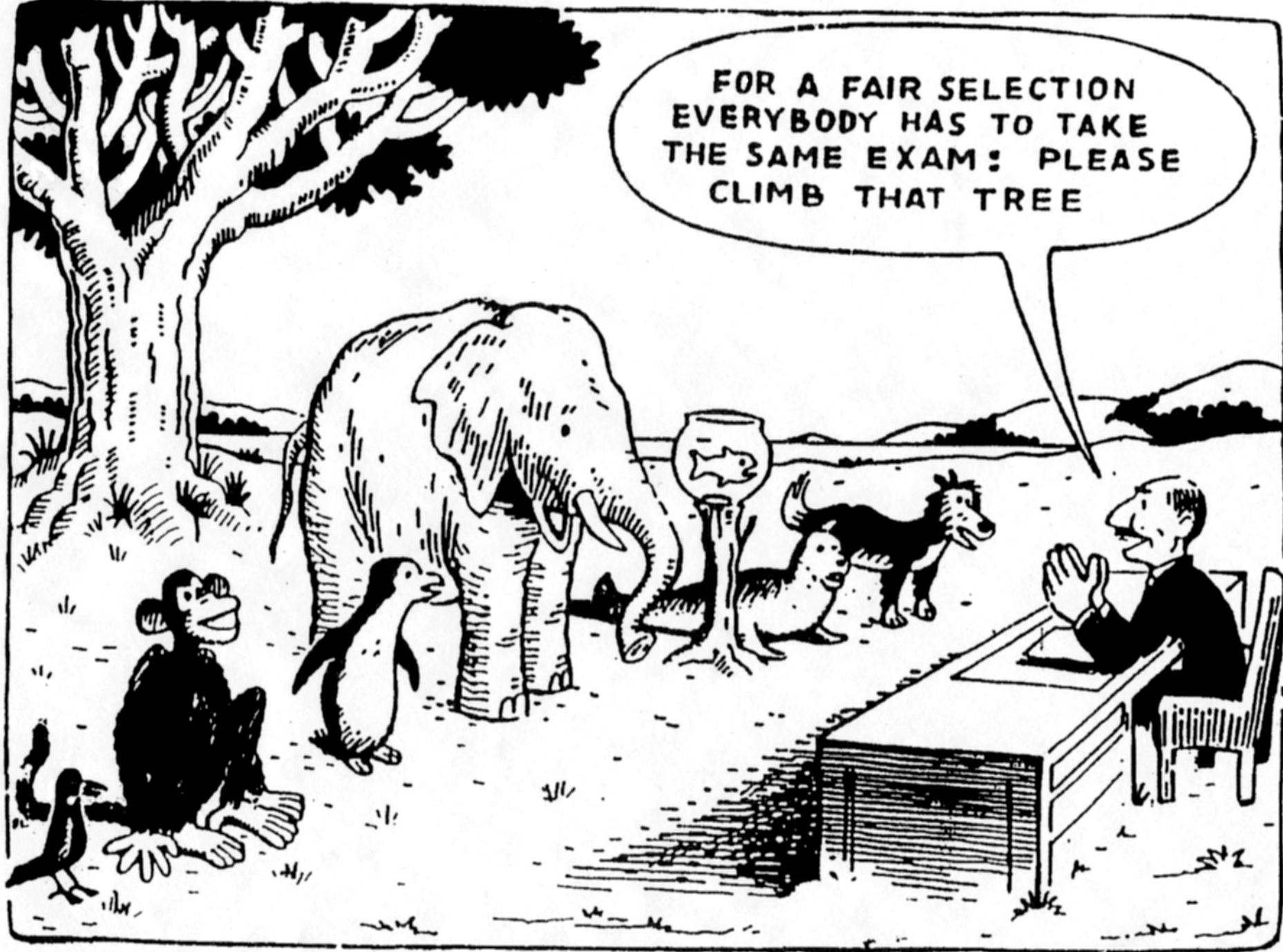
- Administrator/supervisor reports
- Surveys of students and/or parents
- An “evidence binder” created & presented by the teacher

Multiple measures of *student learning*

- **Evidence of growth in student learning and competency**
 - Standardized assessments (state/district tests)
- **Evidence collected by teachers and scored by groups of educators**
 - The 4 Ps: portfolios, projects, products, and performances
 - Essays, written responses to complex questions
- **Evidence collected and scored in classrooms**
 - Classroom-based assessments such as DRA, DIBELS, curriculum-based tests, unit tests



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



Federal priorities (August 2010)

- From “Race to the Top” and reiterated in the August 5, 2010 Federal Register (Vol. 75, No. 150) “Secretary’s Priorities for Discretionary Grant Programs”
 - Teachers should be evaluated using state standardized tests where possible
 - For non-tested subjects, other measures (including pre- and post-tests) can be used but must be **“rigorous and comparable across classrooms”** and must be **“between two points in time”**
 - Multiple measures should be used, such as multiple classroom evaluations

Evaluation System Models

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

<http://www.austinisd.org/inside/initiatives/compensation/slos.phtm>

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's 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?](http://www.gadoe.org/DMGetDocument.aspx/CK%20Standards)

[p=6CC6799F8C1371F6B59CF81E4ECD54E63F615CF1D9441A92E28BFA2](http://www.gadoe.org/DMGetDocument.aspx/CK%20Standards)

[A0AB27E3E&Type=D](http://www.gadoe.org/DMGetDocument.aspx/CK%20Standards)

Evaluation System Models (cont'd)

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

http://communication.sdhc.k12.fl.us/empoweringteachers/?page_id=317

Rhode Island 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 (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)

Questions to ask about measures (and one about the model)

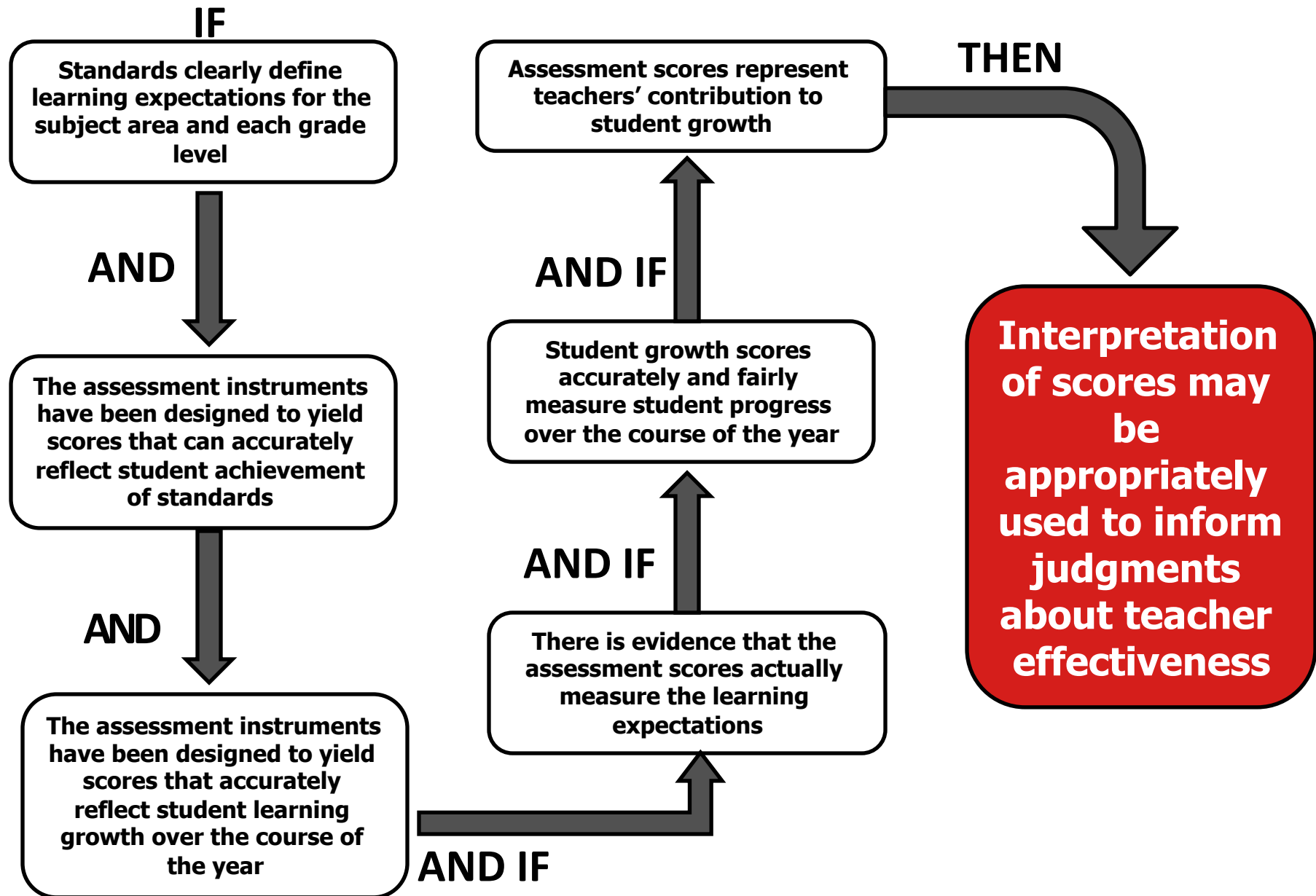
- 1. Rigorous.** Are measures “rigorous,” focused on measuring students’ progress towards college and career readiness?
- 2. Comparable.** Are measures “comparable across classrooms,” ensuring that students are being measures with the same yardstick?
- 3. Growth over time.** Do the measures enable student 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. 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?
- 6. 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?

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"

Validity (measures and models)

- Validity is ensured when the instrument/test accurately measures what it is intended to
- Validity cannot be fully established until you collect and analyze evidence; evidence on these models/measures is incomplete/non-existent
- Focus on models and measures that meet the 6 priorities (Rigorous, Comparable, Growth over time, Standards-based, Inclusive, Improve teaching)
- When data is collected, determine whether the measures are differentiating among stronger and weaker teachers (correlate results from multiple measures)
- Evaluate how the model and measures are working often



Propositions that justify the use of these measures for evaluating teacher effectiveness. (Adaptation based on Bailey & Heritage, 2010 and Perie & Forte (in press)) (Herman, Heritage & Goldschmidt, 2011). Slide used courtesy of Margaret Heritage.

Evidence of growth in student learning

- Evidence is strongest when it is
 - **Standardized**, meaning that all teachers used the assessment in exactly the same way
 - Gave the assessment on the same day
 - Gave students a specific amount of time to complete the test
 - Used the same preparation/instructions prior to the test
 - Recorded/reported results accurately
 - **Valid**, meaning that it measures what is intended
 - Items (questions) accurately capture students' understanding and knowledge
 - Progress towards proficiency in a subject is captured because there are sufficient items to measure students at all levels
 - **Recorded**, meaning that student progress can be compared across classrooms and schools

Challenges for Special Education Teachers and ELL Specialists

➤ Challenges for SWD and ELLs

- Small student numbers
- Not all special educators and ELLs instruct students in tested subjects
- Teacher contribution to social and behavioral growth would not be factored into results
- Teachers working with students on alternate standards
- Little research exploring whether growth rates are comparable
- Little research on the use of accommodations & their impact related to teacher effects

The “Caseload Teachers”

- For nurses, counselors, librarians, and other professionals who do not have their own classroom, what counts for you is your “caseload”
 - May be all the students in the school
 - May be a specific set of students
 - May be other teachers
 - May be all of the above!

Other teachers with “caseloads”

- For team teachers, special ed teachers, ELL teachers, other itinerant teachers
 - Caseload would be the students you provide instruction or assistance to
 - When students are shared between two teachers, those students belong to *both* teachers’ caseloads
 - This may be done as a percentage, or the shared student scores would be counted for each teacher

Measuring with the “4 Ps”

- The 4 Ps: Student products, presentations, portfolios, or projects
- Must be standards-based
- Students demonstrate their proficiency by
 - Giving a performance (dance, music, drama, art, speech, etc.)
 - Creating and displaying something (artwork, a bookshelf, etc.)
 - Completing a portfolio or project with multiple entries reflecting specific knowledge and skills

Measuring Teacher Effectiveness: System Design Suggestions

- *The goal of all teacher evaluation should be to improve teaching and learning*
 - Give all teachers opportunities to improve their practice and outcomes
 - Provide assistance in determining areas where growth is needed and planning strategies to address them
 - Provide resources to help them improve
 - Align professional development opportunities with teachers' growth plans

Questions?





Laura Goe, Ph.D.
E-Mail: lgoe@ets.org
www.tqsource.org
www.laurago.com