



NATIONAL COMPREHENSIVE CENTER
FOR **TEACHER QUALITY**

Evaluating Teacher Effectiveness: An Overview

Laura Goe, Ph.D.

Ohio Task Force

Columbus, Ohio ♦ January 26, 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**

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

**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 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 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 rules clearly				
Calls students by name				
Involved students in learning tasks				



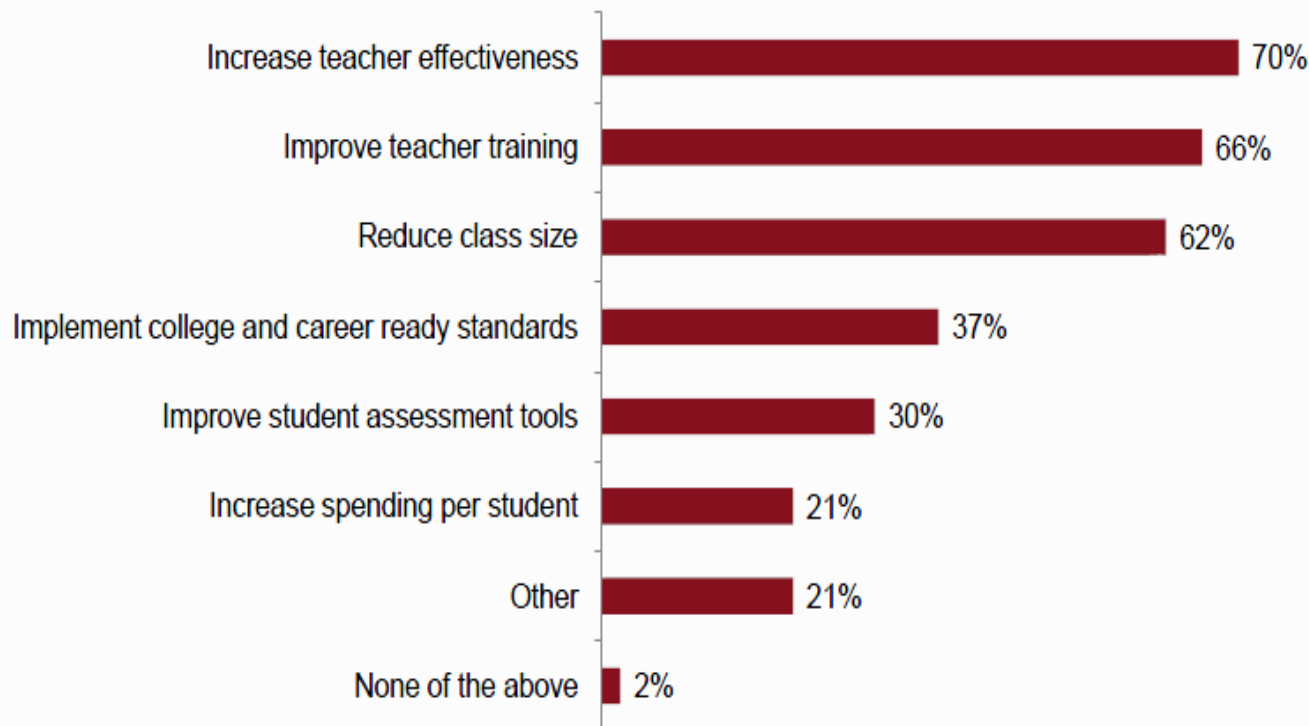
“When all you have is a hammer, everything looks like a nail.”

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)

- Many Washington Insiders focus on teachers; more than two-thirds suggest increasing teacher effectiveness (70 percent) or improving teacher training (66 percent) will improve K-12 student outcome.
- A number of respondents (62 percent) indicate that reducing the number of students per class would translate to better results for K-12 students in the United States.

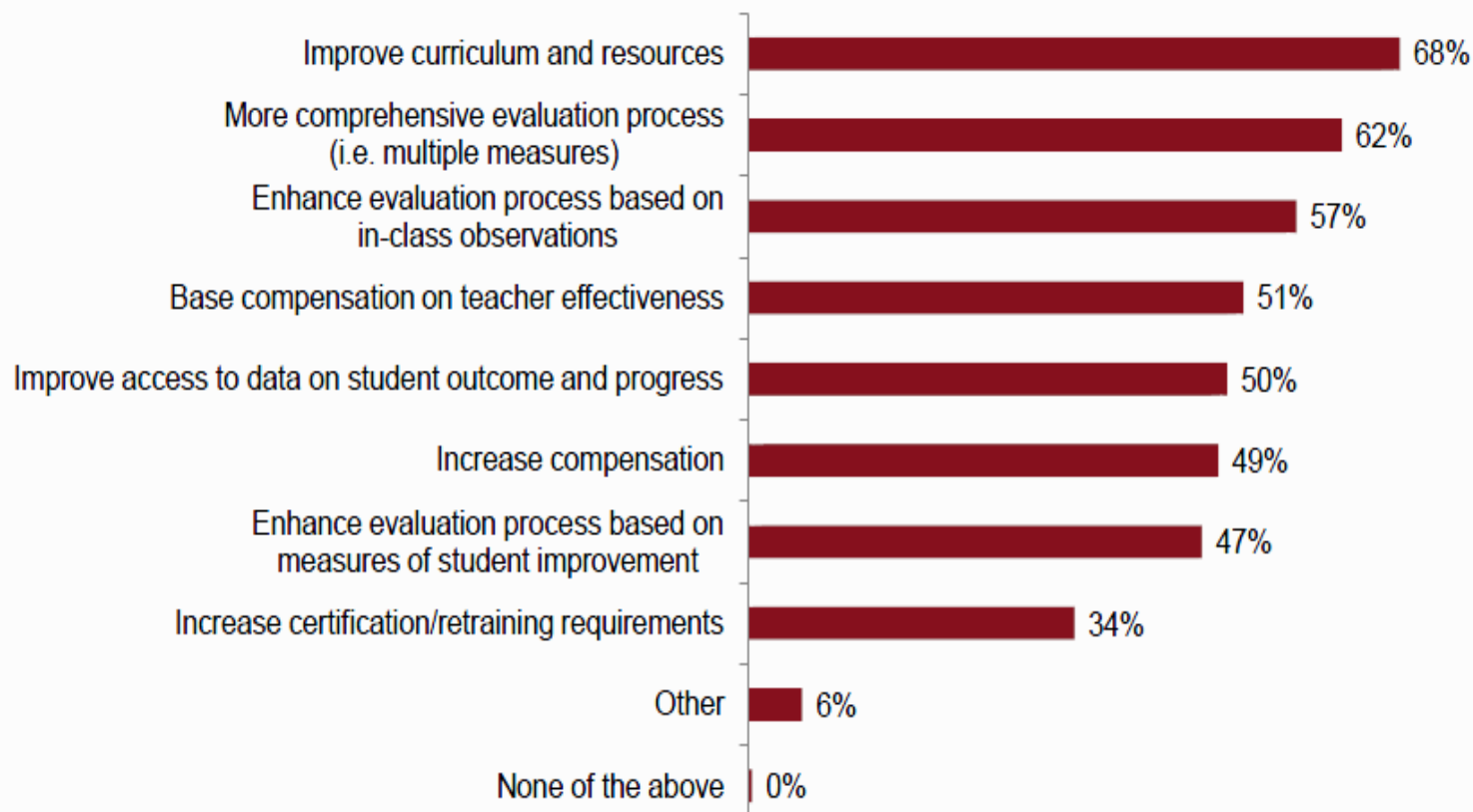
Ways to Improve K-12 Education Outcomes*



*In your opinion, which of the following will significantly improve student outcome as it pertains to K-12 education in the United States. Please select all that apply.

- In general, Washington Insiders believe that improved resources and more comprehensive evaluations are the best ways to improve teacher effectiveness. More than half of respondents identified “improve curriculum and resources” (68 percent), “more comprehensive evaluation process” (62 percent), or “enhance evaluation process based on in-class observations” (57 percent) as ways to improve teacher effectiveness.
- Only 34 percent of respondents believe that increasing certification or training requirements is likely to improve teacher effectiveness.

Ways to Improve Teacher Effectiveness*



* In your opinion, which of the following is likely to improve teacher effectiveness? Please select all that apply.

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*

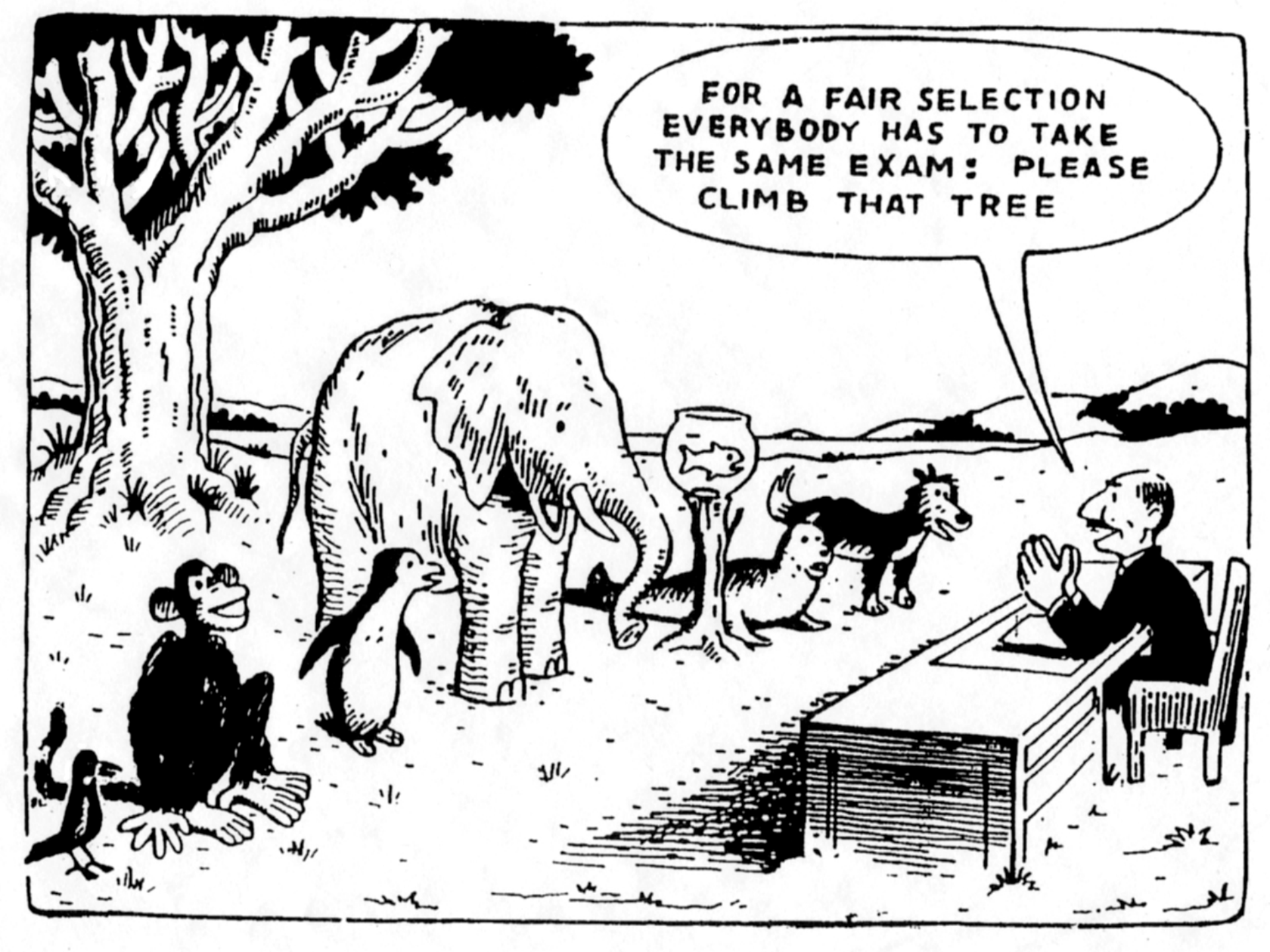
- Classroom observations
- Lesson plans, assignments, and student work

➤ Evidence of *professional responsibility*

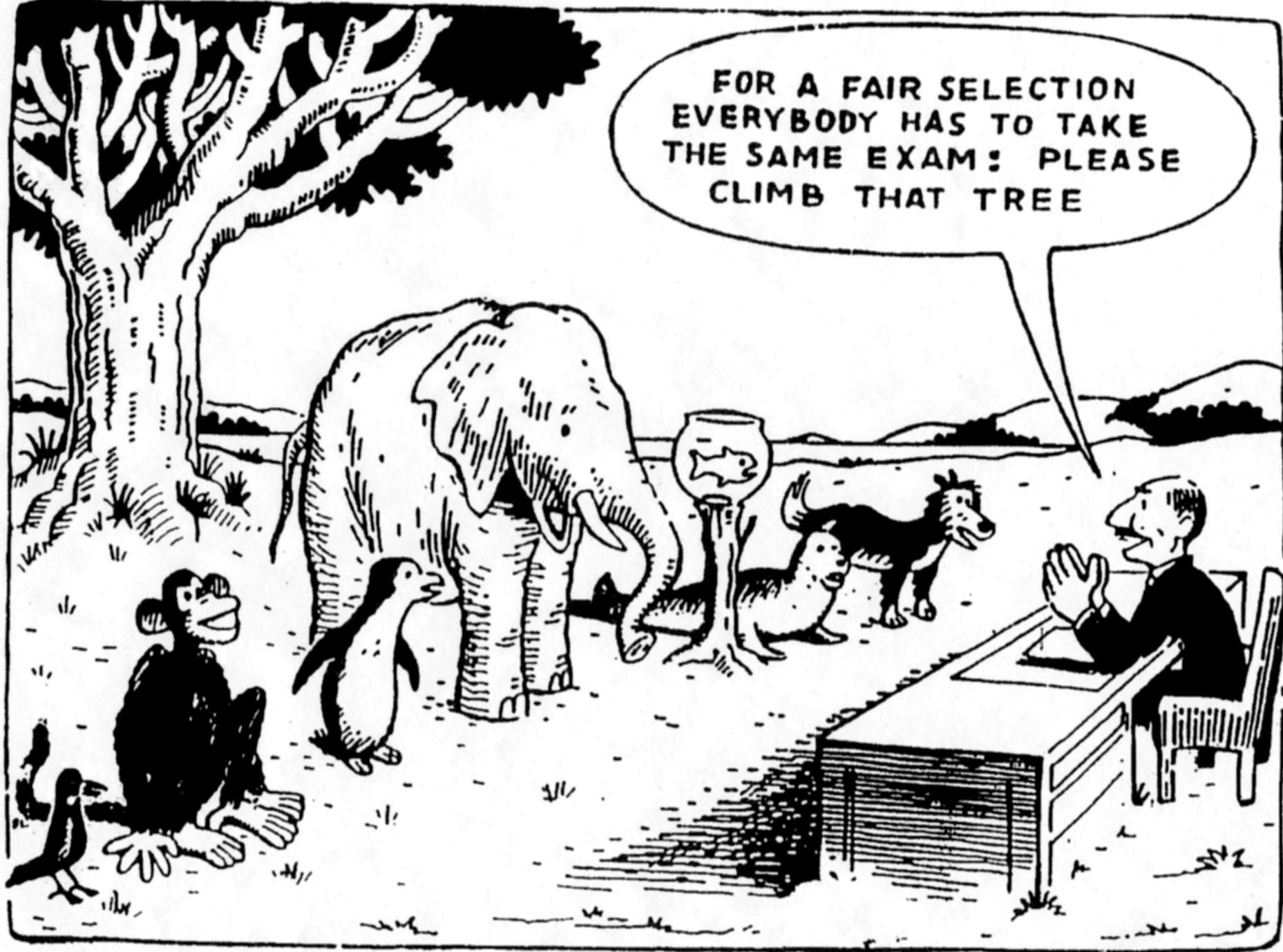
- Administrator/supervisor reports
- Surveys of students and/or parents
- An “evidence binder” created & presented by the teacher

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

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 points in time**
- A state may also include other measures that are
 - **rigorous**
 - **comparable across classrooms** (pg 11)

Potential Definition

➤ **Rigorous**

- Rigorous measures may mean that *high expectations for student progress towards college- and career-readiness* are exhibited
 - Measure designed to measure students' mastery of grade-level standards for that subject.

Potential Definition

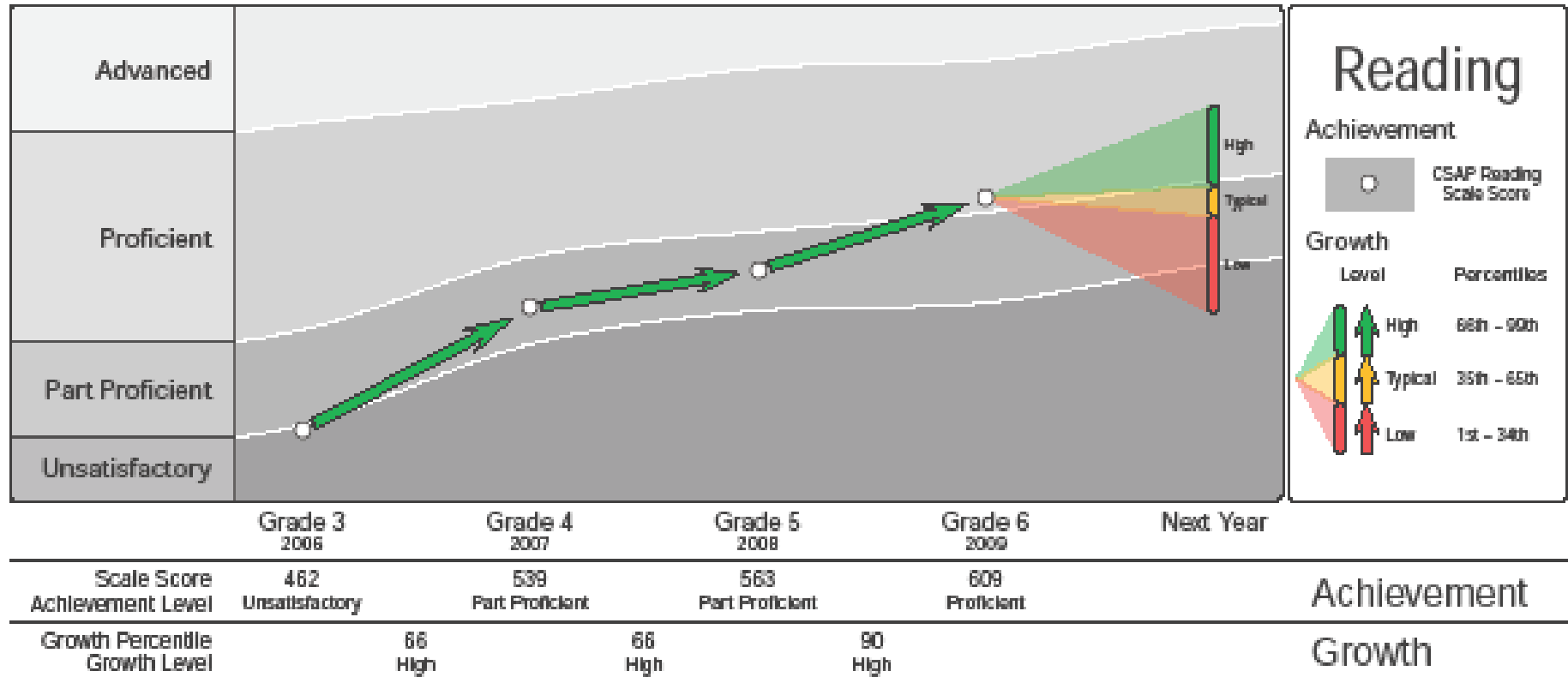
➤ **Between two points in time**

- May mean assessments that occur as close as possible to the beginning and end of a course, so that the maximum growth towards subject/grade standards can be shown.
 - Examples:
 - Pre and post test given in a particular course
 - Student portfolio demonstrating mastery of standards-based knowledge and skills (measured over time)

Frequently used growth models

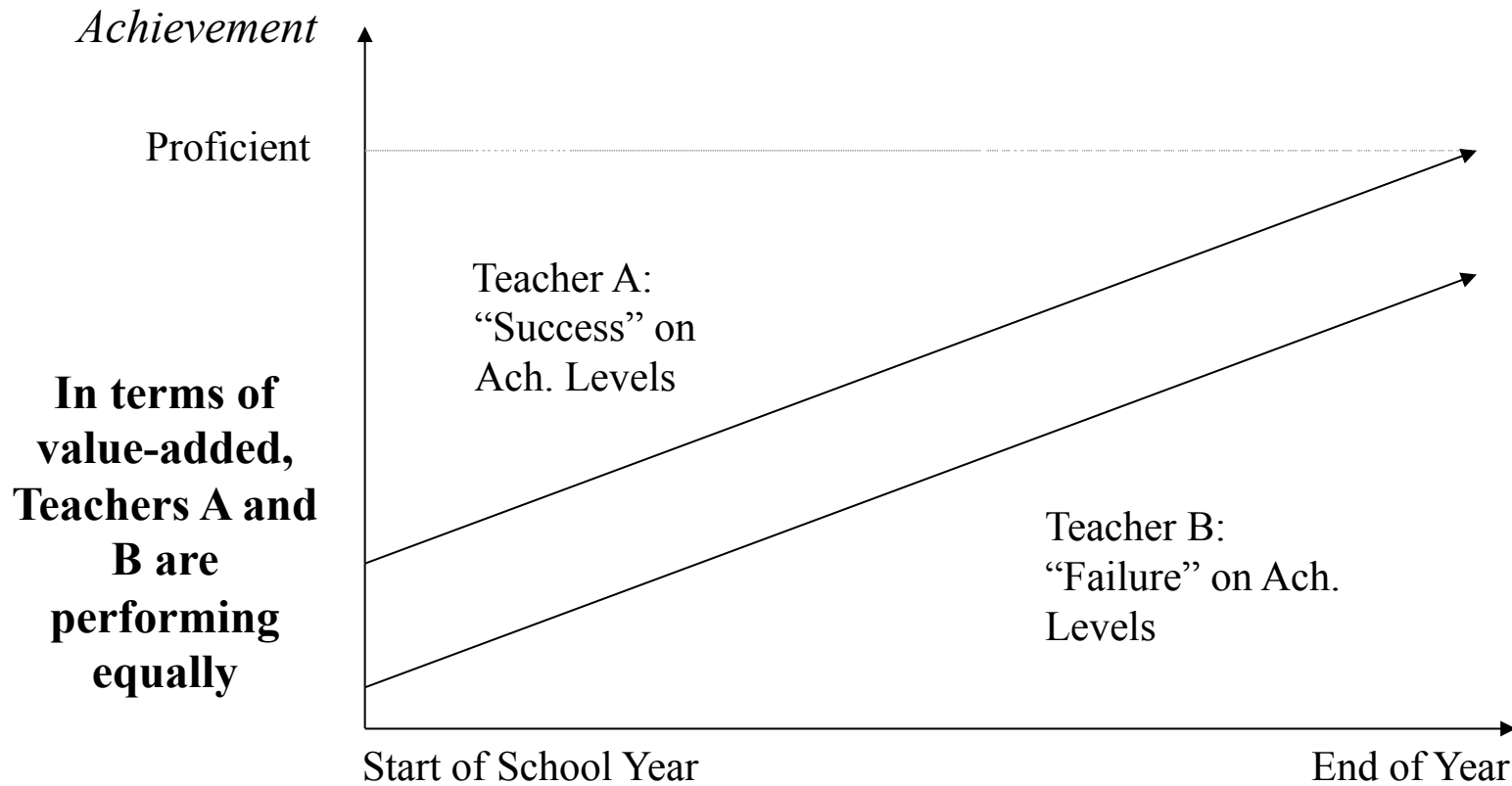
- Value-added models (requires prediction)
 - There are many versions of value-added models (VAMs), but results from the different models are quite similar
 - Most states and districts that use VAMs use the Sanders' model, also called TVAAS
 - Prior test scores (3+ years in the Sanders' model) are used 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”

Sample student report: Colorado Growth Model



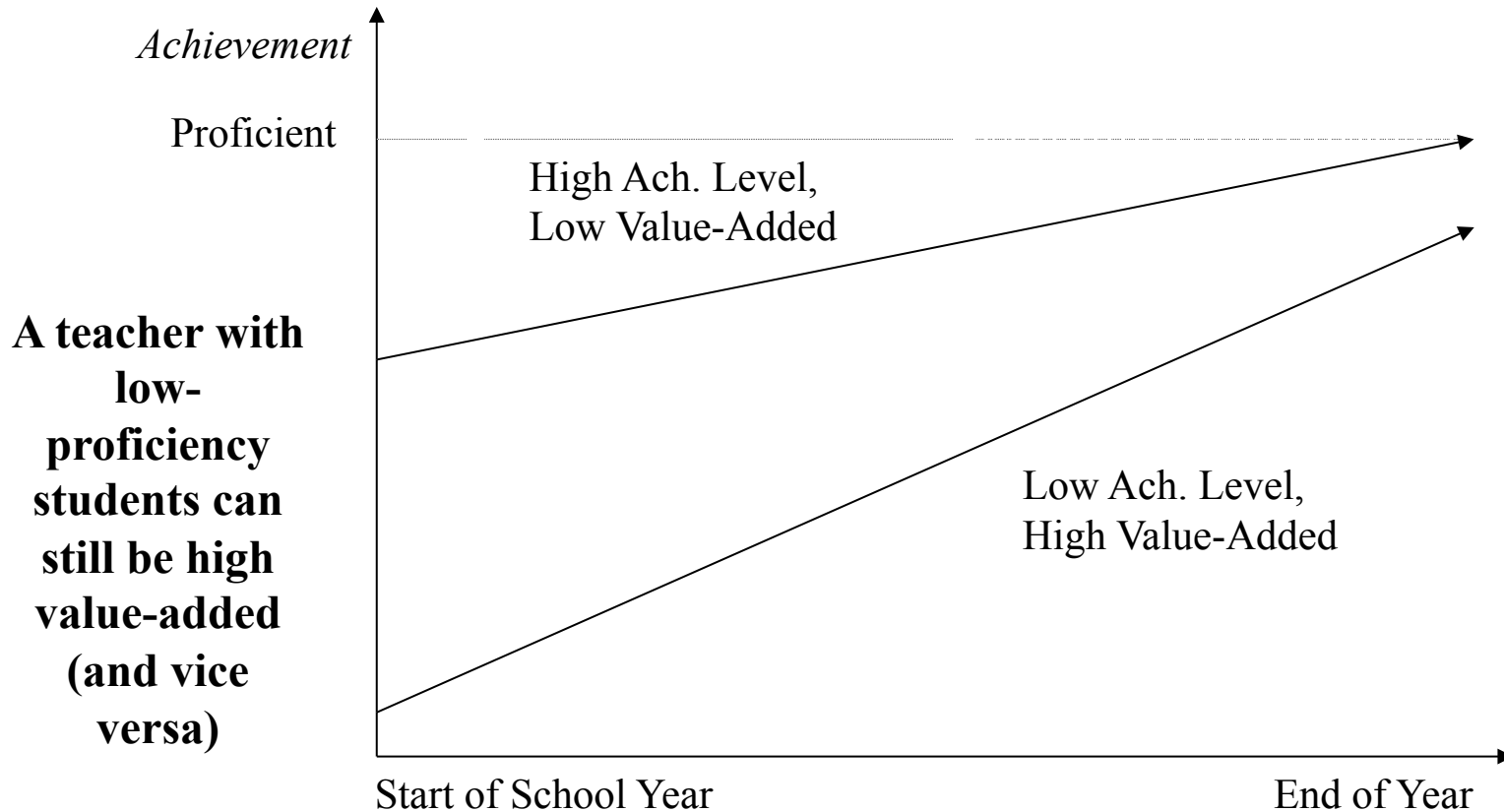
Slide courtesy of Damian Betebenner at www.nciea.org

Why growth models are better than status models (1)



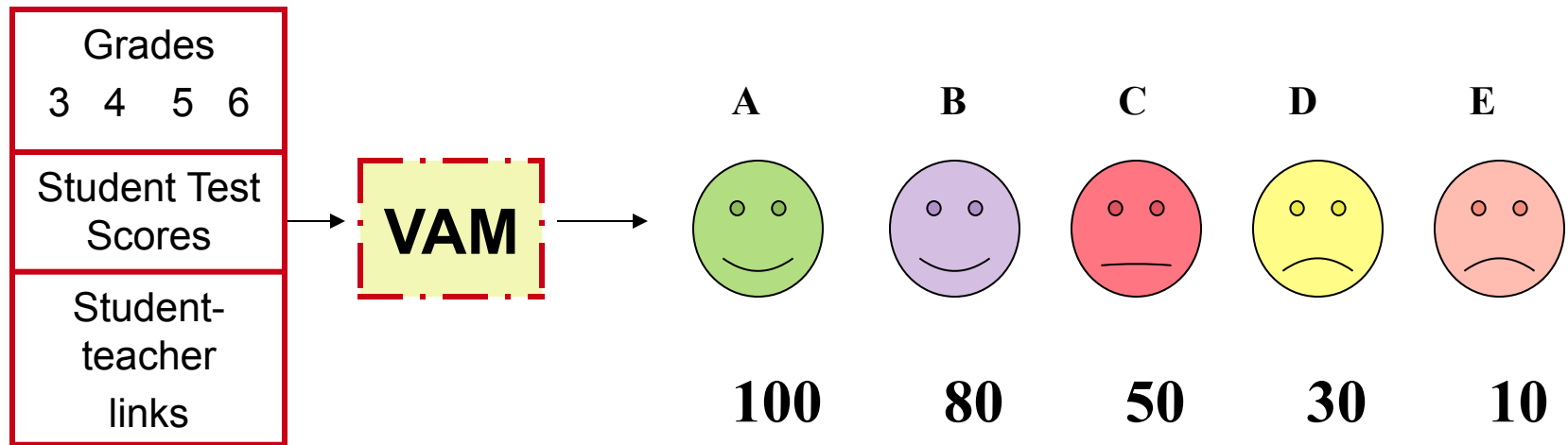
Slide courtesy of Doug Harris, Ph.D, University of Wisconsin-Madison

Why growth models are better than status models (2)

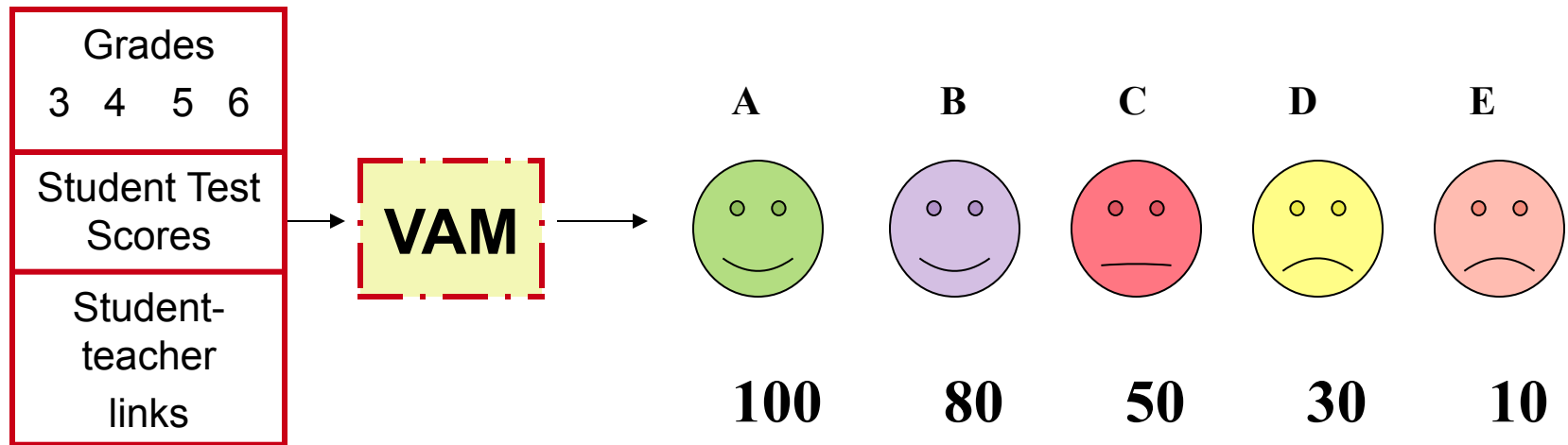


Slide courtesy of Doug Harris, Ph.D, University of Wisconsin-Madison

Teacher effects



Classroom effects



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"

The “other 69%”

- In one state, 69% of teachers (Prince et al., 2006) could not 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
- May be higher in many states—75-80%

Models

- Specify the elements of the evaluation system such as
 - How often teachers are to be evaluated
 - Who will collect and evaluate evidence on teachers' performance
 - What training will be required in order to conduct observations and evaluate other types of evidence
 - Whether student learning results (such as standardized tests) will be used as a component of teachers' scores

(continued, next slide)

Models (continued)

- What percentage of a teachers' total scores will be based on student achievement vs. other measures
- State teaching standards to be used to guide measurement of performance
- Several levels for teacher performance (such as "highly effective," "effective," or "ineffective")
- Consequences for failure to meet acceptable performance levels, such as referral to a Peer Assistance and Review (PAR) program
- Which categories of measures or which specific measures are to be used, i.e., observations (category) or Charlotte Danielson's Framework (specific)

Questions to ask about models

- Do they include measures that are “rigorous and comparable across classrooms”?
- Do they include measures that show learning growth “between two points in time”?
- Do they include measures that are aligned with and focused on grade level and subject standards?
- Do they allow teachers from all subjects to be evaluated with evidence of student learning growth?
- Will using this model help improve teaching and learning?

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

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>

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>

References (continued)

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>

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>

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