



**Student Learning and
Achievement in Measuring
Teacher Effectiveness**

Laura Goe, Ph.D.
Policy Challenge Conference
*Multiple Measures of Teacher Effectiveness:
Bridging Research and Practice*

REL-NEI, NECC, TQ Center,
& Nellie Mae Education Foundation
Marlborough, MA ♦ 6/25/2010

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 Learning & Teaching Research Center at ETS

NATIONAL COMPREHENSIVE CENTER FOR TEACHER QUALITY 2 www.tqsource.org

**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
- Learning Point Associates
 - Technical assistance, research, fiscal agent
- ETS
 - Research & dissemination, technical assistance

NATIONAL COMPREHENSIVE CENTER FOR TEACHER QUALITY 3 www.tqsource.org

To be discussed...

- Measurement challenges
- Using multiples measures
- Implications for policy and practice
 - Delaware
 - Tennessee
 - AFT Innovation Grant states (NY & RI)
 - Other states/districts
- What's next

NATIONAL COMPREHENSIVE CENTER
for TEACHER QUALITY 4 www.tqsource.org

The goal of teacher evaluation

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



**TO IMPROVE
TEACHING AND
LEARNING**

NATIONAL COMPREHENSIVE CENTER
for TEACHER QUALITY 5 www.tqsource.org

The quandary

- Tension between federal (and sometimes state) pressures and teacher (and sometimes union) beliefs about what should constitute measures and evidence
- Teachers want a greater focus on the job they are doing (instructional quality)
 - Assumes shared responsibility for student learning
- Federal pressures are towards a greater focus on outcomes (student learning)
 - Assumes that teachers are primarily responsible for student learning

NATIONAL COMPREHENSIVE CENTER
for TEACHER QUALITY 6 www.tqsource.org

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).

Validity in measurement

➤ **Validity:** The extent to which evidence and theory support an interpretation of scores for a particular use of the measure

- We want to know: Is our measure telling us how teachers perform in the areas we care about? What are appropriate uses of these results?
- Instruments (such as tests) do not “have” validity
 - Validity lies in how well the instrument measures the domain we care about and *how the results are used*

Interpretation of evidence

• Is the instrument being used for the purposes for which it was designed?

• Does the instrument capture what it is intended to, or is it impacted by factors unrelated to teaching?

• Do the interpretations being drawn from the scores go beyond what the valid uses of the instrument?

Measures of teacher effectiveness

- **Evidence of growth in student learning and competency**
 - Standardized tests, pre/post tests in untested subjects
 - Student performance (art, music, phys ed, etc.)
 - Curriculum-based tests given in a standardized manner
 - Classroom-based tests such as DIBELS & DRA
- **Evidence of instructional quality**
 - Classroom observations
 - Quality of assignments and resulting student work
- **Other evidence (varies, based on local values)**
 - Administrator/supervisor reports
 - Surveys of students and/or parents
 - An "evidence binder" created & presented by the teacher

Measures that help teachers grow

- 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 "evidence binders")
- Measures that give teachers opportunities to discuss the results with evaluators, administrators, colleagues, teacher learning communities, mentors, coaches, etc.
- Measures that are directly and explicitly aligned with teaching standards (local, state, or national)
- Measures that are aligned with professional development offerings
- Measures which include protocols and processes that teachers can examine and comprehend

Growth opportunities for *all* teachers

The point is *not* that accountability systems lack value. They serve an important purpose. But alone they touch too few teachers. We need evaluation systems that promote the development of *all* teachers, not just those having difficulty. We need teacher evaluations that help and encourage the tenured teacher to perform to maximum capabilities. In addition, we need evaluations that help the outstanding teacher—the virtuoso performer—to (a) use his or her strengths to maximum efficiency and (b) share these strengths with other teachers.

Duke, DL; Stiggins, RJ. (1986.) *Teacher Evaluation: Five Keys to Growth*. West Haven, CT: National Education Association. ERIC # ED275069 (full text, pg 15)

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.

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

One type of measurement: student achievement gains

- Value-added models (VAMs) are the most sophisticated way to measure student growth
- VAMs are a version of growth models
 - There are many versions, but results from the different models are generally similar
 - Prior test scores (at least 3 years in the Sanders' model) are used to predict the **next** test score for a student
 - If average student performance in a classroom is better than predicted, they can be said to have an effective teacher

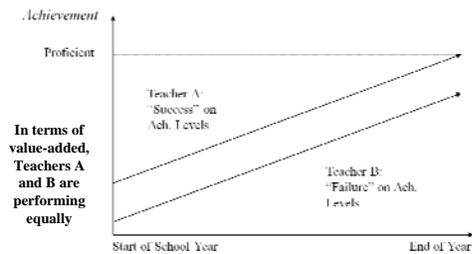
Two Ways VAMs May Be Used for Accountability

- **School Focus (state and district level)**
 - Identification for "needs improvement"
 - Rewards or sanctions
 - Could be used for NCLB purposes
- **Teacher Focus (district level)**
 - One measure in a teacher evaluation system
 - Pay for Performance
 - Promotion/dismissal

Data requirements for VAMs

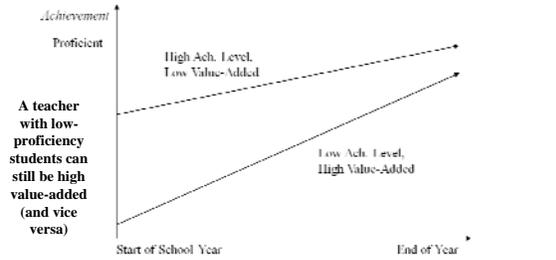
- To use VAMs, you must have the following
 - Unique identifiers for each *teacher*
 - Unique identifiers for each *student*
 - A link between the student and each of his or her teachers in the data system
 - Accurate, complete data going as far back as possible, for both students and teachers
 - Student achievement scores for several years (to be used to predict next year's score)

Why VAMs are better than status models (1)



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

Why VAMs are better than status models (2)



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

What the Research Says About Value-Added

- Researchers found that scores cannot be *solely* attributed to teachers' influence; VAMs provide a summary score of the "contribution of various factors toward growth in student achievement" (Goldhaber & Anthony, 2003, p. 38).

What the Research Says About Value-Added (Cont'd)

- Very little teacher effectiveness is explained by observable characteristics, and while teachers vary in their contribution to students' achievement score gains, researchers did not know what caused the variation (Rivkin, Hanushek, & Kain, 2005).

What VAMs can't tell you

- VAMs can't tell you **why** students in a particular classroom are scoring higher than expected
 - Maybe there is a narrow instructional focus on test content
 - Or maybe the classroom offers a rich, engaging curriculum that fosters deep student learning
- **How teachers are getting results matters, not just the results themselves**

NATIONAL COMPREHENSIVE CENTER
FOR TEACHER QUALITY 22 www.tqsource.org

Is it possible to accurately measure teacher effectiveness with VAMs?

- We could get an estimate that would be close to the teacher's true contribution to student learning growth if
 - Teachers were randomly assigned to schools
 - Students were randomly assigned to teachers
- However, given differences in resources, school environment, etc., we would only know how the teacher did with **that** group of students in **that** setting

NATIONAL COMPREHENSIVE CENTER
FOR TEACHER QUALITY 23 www.tqsource.org

VAMs don't measure most teachers

- About 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

NATIONAL COMPREHENSIVE CENTER
FOR TEACHER QUALITY 24 www.tqsource.org

Measurement challenges for Special Educators and ELL Specialists

- Challenges for SWD and ELLs
 - Small student numbers per teacher
 - Various co-teaching models make it difficult to evaluate teachers unique contributions
 - 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 and trajectories
 - Little research on the use of accommodations & their impact related to teacher effects

Non-VAM tests (accepted under Washington, DC's IMPACT evaluation system)

- DC Benchmark Assessment System (DC BAS)
- Dynamic Indicators of Basic Early Literacy Skills (DIBELS)
- Developmental Reading Assessment (DRA)
- Curriculum-based assessments (e.g., Everyday Mathematics)
- Unit tests from DCPS-approved textbooks
- Off-the-shelf standardized assessments that are aligned to the DCPS Content Standards
- Rigorous teacher-created assessments that are aligned to the DCPS Content Standards
- Rigorous portfolios of student work that are aligned to the DCPS Content Standards

Standardization is key for using other tests for teacher evaluation

- Standardizing how curriculum- or classroom-based tests are given is key to ensuring that tests are "rigorous and comparable across classrooms"
- Ensure that tests meet district approval
- For subject-matter tests, ensure that
 - Tests are given on the same day, at the same time, for the same length of time, with supervision
 - Teachers agree to appropriate "test prep" rules

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 set amount of time to complete test/task
 - Provided the same preparation/instructions prior to the test
- **Valid**, meaning that it measures what is intended
 - Items (questions) or performances/products accurately capture/reflect students' understanding and knowledge
 - Progress towards proficiency in a subject is accurately captured because there are sufficient items to measure students at all levels, both high- and low-performing
- **Accurately Recorded**, meaning that student progress can be compared across classrooms and schools


28
www.tqsource.org

Delaware

➤ DE already has a good evaluation system

- Using Danielson's Framework for Teaching
- Added a 5th Domain: "Student Improvement"

➤ Student Improvement Domain

- Showing Student Improvement
- Aligning Assessments to Teacher Data-Driven Goals
- Measuring Student Improvement
- Reflecting on Student Improvement


29
www.tqsource.org

Current work in Delaware

➤ Just completed a statewide workshop

- How to incorporate multiple measures
- How to measure student learning growth, particularly in non-tested subjects/grades

➤ Meeting again in July, bringing together working groups of teachers

- Teams divided by subject taught
- With guidance from facilitators, teams will determine what measures to use for each subject and grade (ex: K-12th grade music)


30
www.tqsource.org

Delaware's challenges

- How to show contributions to student achievement for various professionals like counselors
- How to coordinate pre- and post-tests in a standardized way, particularly when schools may start at different times—as much as 2 weeks later for some school districts
- What policies and procedures to develop to ensure fidelity and integrity of the system
- How to measure state test score growth when the state tests used in the coming year will be new

Tennessee

- TN has the country's greatest value-added system with longitudinal teacher-student linked data going back for 20 years or so
- But there is no statewide teacher evaluation system—varies widely from district to district
- Like everyone else, they are struggling to figure out how to use student learning growth in their teacher evaluation system

Tennessee's challenges

- Whether to try to put a statewide system into place that will mandate how districts should measure teachers in non-value-added subjects/grades or whether to offer guidance and let districts determine the measures
- Who to include on teams to determine appropriate measures for performance & product subjects
- Who should evaluate performance/products

AFT's Innovation Grant work

- New York and Rhode Island are both developing comprehensive teacher evaluation systems that include measures of student learning growth
- District teams (including union representatives, superintendents, and administrators) working with experts to learn about measuring teacher performance
 - Rhode Island requires that 51% of teacher evaluation be based on student achievement growth; RIDE determines measures
 - New York is just developing statewide teaching standards; recently decided that 40% of teacher evaluation must be based on student learning growth, including 20% on standardized test scores
 - Incorporating student learning growth is biggest challenge

Other states and districts

- Colorado: new legislation (affects tenure)
- Louisiana: "soup to nuts" value-added
- Ohio: includes principal evaluation
- Texas: longitudinal data is a plus
- Austin: Student Learning Objectives
- Los Angeles: currently using a "checklist" but moving quickly to develop new system
- Washington, DC: IMPACT system

What's next?

- The lack of research to guide the use of student learning growth in teacher evaluation is hampering efforts to move forward
- But federal priorities and state legislation will keep the pressure on, so districts and states will have to experiment and see what works
- **Best to adopt a "pilot" approach which includes continual review & revision**

References

Goldhaber, D., & Anthony, E. (2003). *Teacher quality and student achievement* (No. UDS-115). New York, NY: ERIC Clearinghouse on Urban Education.

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>

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). *Teachers, schools, and academic achievement*. *Econometrica*, 73(2), 417 - 458.

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
