Evaluating Teacher Effectiveness: Selecting Measures

Laura Goe, Ph.D.

SIG Schools Webinar

August 12, 2011

The Role of the TQ Center

A federally-funded partnership created to help states carry out the teacher quality mandates of ESEA

- *American Institutes for Research*
  - Technical assistance, research, dissemination
- *Educational Testing Service*
  - Technical assistance, research, dissemination
- *Vanderbilt University*
  - Students with special needs, at-risk students

TQ Center Resources to Support Educator Effectiveness

Online Tools

- **Practical Guide to Designing Teacher Evaluation Systems**

- **Teacher Evaluation Models in Practice**

- **Guide to Evaluation Models**

- **Innovation Configurations**

- **Other Research Briefs and Online Resources**

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
- **The implementation challenges are enormous**
  - Few models exist for states and districts to adopt or adapt
  - Many districts have limited capacity to implement comprehensive systems, and states have limited resources to help them

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 (measures, instruments, processes, training, and scoring, etc.) that result in determinations about individual teachers' effectiveness

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
  - Evidence binder (next generation of portfolio)
- **Evidence** of *professional responsibility*
  - Administrator/supervisor reports, parent surveys
  - Teacher reflection and self-reports, records of contributions

Measures that help teachers grow

- Measures that motivate teachers to examine their own practice against specific teaching 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 aligned with professional development offerings
- Measures which include protocols and processes that teachers can examine and comprehend
- Measures that provide information teachers can use to make immediate adjustments in instruction

Considerations for choosing and implementing measures

- Consider whether human resources and capacity are sufficient to ensure fidelity of implementation
    - Having the "right" measures is only the beginning
    - Poor implementation will call validity into question
- Conserve resources by encouraging districts to join forces with other districts or regional groups
    - Don't "reinvent the wheel;" instead, share your documents, processes, instruments, etc.
    - Share data analysis staff and results across districts
        - This makes it possible to identify schools/districts where better implementation and processes may need to be considered

Validity is a <u>process</u>

- Herman et al. (2011) state, "Validity is a matter of degree (based on the extent to which an evidence-based argument justifies the use of an assessment for a specific purpose)." (pg. 1)
- Starts with defining the criteria and standards you want to measure, *then* choosing measures
- Requires judgment about whether the instruments and processes are giving accurate, helpful information about performance
- Verify validity by
  - Comparing results on multiple measures
  - Multiple time points, multiple raters

Validity of classroom observations is highly dependent on training

- Even with a terrific observation instrument, the results are meaningless if observers are not trained to agree on evidence and scoring
- *A teacher should get the same score no matter who observes him*
  - This requires that all observers be trained on the instruments and processes
  - Occasional "calibrating" should be done; more often if there are discrepancies or new observers
  - Who the evaluators are matters less than that they are adequate trained and calibrated
  - Teachers should also be trained on the observation forms and processes to improve validity of results

Value-added and Colorado Growth Model

- EVAAS uses prior test scores to predict the next score for a student
  - Teachers' value-added is the difference between actual and predicted scores for a set of students
- Colorado Growth model
  - Betebenner 2008: Focus on "growth to proficiency"
  - Measures students against "academic peers"
- Ongoing concerns about using growth models for teacher evaluation
  - Researchers have raised numerous cautions (see my July 28 Texas and Southeast Comp Center presentation for studies and findings)

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 (4-8 ELA & 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 generally at a loss when it comes to measuring teachers' contribution to student growth in non-tested subjects and grades

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 assesses again at end of year; principal or designee works with teacher, determines success
Subject & 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
Pre-and post-tests model	Identify or create pre- and post-tests for every grade and subject
School-wide value-added	Teachers in tested subjects & grades receive their own value-added score; *all other teachers get the school-wide average*

SLOs + "Ask a Teacher" (Hybrid model)

- Concerns about SLOs are 1) rigor, 2) comparability, and 3) administrator burden
- A "rigor rubric" helps with first concern
- Combining SLOs with aspects of the "Ask A Teacher" model will help with all 3 concerns
  - Teachers discuss and agree to use particular assessments and measures of student learning growth, ensuring great rigor and comparability
  - Teachers work together on aspects of scoring which improves validity and comparability and lightens the administrator burden

Considerations

- 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 different measures
- Evaluate **processes and data** each year and make needed adjustments
- Publish findings of evaluations of both overall system and specific measure

Final thoughts

- The limitations:
  - There are no perfect measures
  - There are no perfect models
  - Changing the culture of evaluation is hard work
- The opportunities:
  - Evidence can be used to trigger support for struggling teachers and acknowledge effective ones
  - Multiple sources of evidence can provide powerful information to improve teaching and learning
  - Evidence is more valid than "judgment" and provides better information for teachers to improve practice

Evaluation System Models

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%2010-18-2010.pdf?p=6CC6799F8C1371F6B59CF81E4ECD54E63F615CF1D9441A92E28BFA2A0AB27E3E&Type=D

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

http://communication.sdhc.k12.fl.us/empoweringteachers/

Evaluation System 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

References (continued)

Betebenner, D. W. (2008). *A primer on student growth percentiles. Dover, NH: National Center for the Improvement of Educational Assessment (NCIEA).*

http://www.cde.state.co.us/cdedocs/Research/PDF/Aprimeronstudentgrowthpercentiles.pdf

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

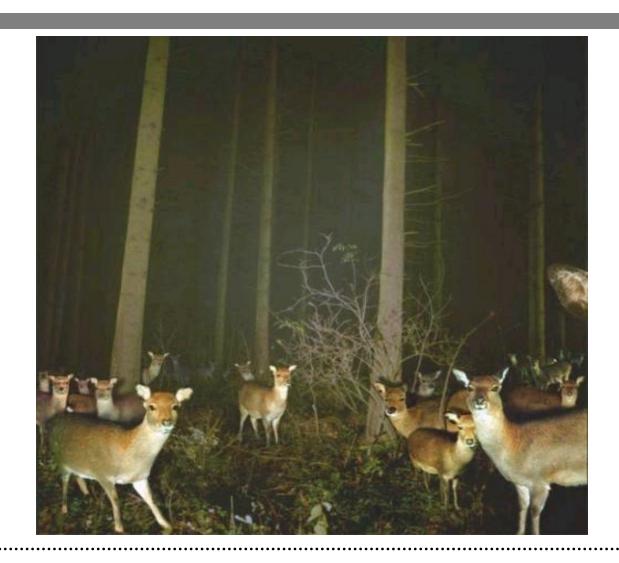http://www.econ.ucsb.edu/~jon/Econ230C/HanushekRivkin.pdf

Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, 12(3), 247-256.*

http://www.sas.com/govedu/edu/ed_eval.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?

Some Options for Next Steps

- Follow-up Q & A webinar
  - Friday, August 19th
  - 11:00 a.m. PDT / 2:00 p.m. EDT
- Further reading (www.tqsource.org)
  - *Evaluating All Teachers: Measuring Student Growth in Non-tested Subjects and for Teachers of At-Risk Students*
  - *A Practical Guide to Designing Teacher Evaluation Systems*

Some Options for Next Steps

- Collaborative dialogue with school, district and state leaders
    - As a team, what did you learn from the webinar that's most helpful to your work on selecting evaluation measures?
    - What are your goals for your teacher evaluation system and how will they guide you in selecting measures?
    - How will you ensure that the measures you choose support the established standards of teacher effectiveness?
    - Considering your local context, what are your top considerations when selecting evaluation measures?
    - What will be your data system requirements to ensure data integrity and usability?
    - What are your immediate next steps to do next week?
    - What are your next steps over the next 2-3 months with regard to selecting evaluation measures?
    - What information do you still need to accomplish these tasks?

Laura Goe, Ph.D.
609-734-1076
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