

Engaging Teachers in Evaluation Reform

Presentation to the Asia-Pacific Assessment
Education Conference

Sabrina Laine, Ph.D.

Vice President, Education Program

September 2013

PERFORMANCE MANAGEMENT Advantage
Evaluation & Professional Growth

at American Institutes for Research ■



American Institutes for Research

Established in 1946, with headquarters in Washington, D.C., American Institutes for Research (AIR) is an independent, nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally.

As one of the largest behavioral and social science research organizations in the world, AIR is committed to empowering communities and institutions with innovative solutions to the most critical challenges in education, health, workforce, and international development.

Session Goals

- Measuring teacher effectiveness—what is the state of the art?
 - Multiple measures
 - Research and best practice
 - Design and implementation challenges
- Engaging stakeholders in decision making
- Strategies for authentic engagement in evaluation reform

Adaptive or Technical Challenge?



“Indeed, the single most common source of leadership failure we’ve been able to identify...is that people, especially those in positions of authority, treat adaptive challenges like technical problems” (Heifetz & Linsky, 2002, p. 14).

Technical Versus Adaptive Challenges

■ Technical Challenges

- Can be fixed by experts and by implementation of best practices
- Easy to identify and have solutions that can be implemented quickly

■ Adaptive Challenges

- Require people to change their values, behaviors, and attitudes
- Necessitate learning new ways of doing business
- Often difficult to identify
- Must be solved by the people with the problem
- Often require experiments, innovations, and new learning
- Can take longer to implement

Why Measure Teacher Effectiveness?

- There are many valid reasons, both formative and summative, to measure teacher effectiveness
- *The **ultimate** goal of all measurement of teacher effectiveness should be...*

**TO IMPROVE TEACHING AND
LEARNING**

Two Important Goals for Evaluation

Accountability: We are interested in ensuring that measures are comparable and rigorous and that they correctly identify students' learning growth compared with other students in the same grade/subject.

Instructional improvement: We are interested in ensuring that teachers actively and regularly collect data on students' performance toward standards and adjust and differentiate instruction accordingly.

Teacher Evaluation: How Far Have We Come?

- Teachers recognize the importance of establishing a common language and vision about instruction and professional practice.
- Many states and districts in the United States have prioritized efforts to engage stakeholders in teacher evaluation design processes to support successful implementation.
- There is rapid development and validation of measures.
- There is progress in developing and implementing evaluator training.

Teacher Evaluation: How Far Have We Come?

- Enhancement of data systems to include workforce performance information
- Broad recognition in the education and policy community of the need for more valid and reliable data to inform improvements in instruction and high-stakes employment decisions

Essential Components of the Design Process

- **Component 1:** Specifying Evaluation System Goals
- **Component 2:** Securing and Sustaining Stakeholder Investment and Cultivating a Strategic Communication Plan
- **Component 3:** Selecting Measures
- **Component 4:** Determining the Structure of the Evaluation System
- **Component 5:** Selecting and Training Evaluators
- **Component 6:** Ensuring Data Integrity and Transparency
- **Component 7:** Using Teacher Evaluation Results
- **Component 8:** Evaluating the System

Practical Guide to Designing Comprehensive Evaluation Systems

<http://www.gtlcenter.org/tools-publications/online-tools/teacher-evaluation>

The screenshot shows the website header for the Center on GREAT TEACHERS & LEADERS at American Institutes for Research. It includes social media icons for Twitter, Facebook, RSS, and YouTube. A search bar contains the text 'e.g. teacher retention' and a green 'SEARCH' button. A navigation menu has six items: ABOUT US, TECHNICAL ASSISTANCE, TOOLS & PUBLICATIONS, CONVENINGS, LEARNING HUB, and CONTACT US, each with a downward arrow. Below the menu is a 'Home' link. The main content area features a section titled 'Teacher Evaluation Practical Guide' with a horizontal line underneath. The text describes the nationwide effort to build better teacher evaluation systems. A key resources section lists a downloadable guide: *A Practical Guide to Designing Comprehensive Teacher Evaluation Systems*. On the right side, there is a large image of three blue flower-like icons over a background of people. Below the image is an orange button labeled 'Navigate the guide'. At the bottom right, a 'Home' section lists two items: '1. Evaluation System Goals' and '2. Stakeholder Investment & Communication Plan'.

Center on
GREAT TEACHERS & LEADERS
at American Institutes for Research ■

Twitter Facebook RSS YouTube

e.g. teacher retention **SEARCH**

ABOUT US TECHNICAL ASSISTANCE TOOLS & PUBLICATIONS CONVENINGS LEARNING HUB CONTACT US

Home

Teacher Evaluation Practical Guide

Across the nation, states and districts are in the process of building better teacher evaluation systems that not only identify highly effective teachers but also systematically provide data and feedback that can be used to improve teacher practice.

This site compiles key resources from the GTL Center website as well as other national websites to support these efforts. Key resources include the following:

- Our downloadable guide: [A Practical Guide to Designing Comprehensive Teacher Evaluation Systems](#)

Navigate the guide

Home

1. Evaluation System Goals
2. Stakeholder Investment & Communication Plan

Evaluation System's Purpose/Goals

- Purpose should drive all decisions regarding:
 - Measurement selection and weight
 - Evaluation format (e.g. frequency of observations, pre-post observation conferences)
 - Data collection needs
- Higher stakes point to measures that are technically defensible (e.g., valid and reliable)
- Improved teacher capacity points to measures that identify effective teaching practices

Multiple Measures of Teacher Effectiveness

- **Evidence of *growth in student learning***
 - 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

Why Multiple Measures?

- Creates a more accurate and complete representation of teacher performance
- Takes into account full range of what teachers do
- Increases confidence in teacher evaluation results
- Increases ability to make informed human capacity decisions (e.g., recruitment, retention, tenure, compensation)

Why Multiple Measures?

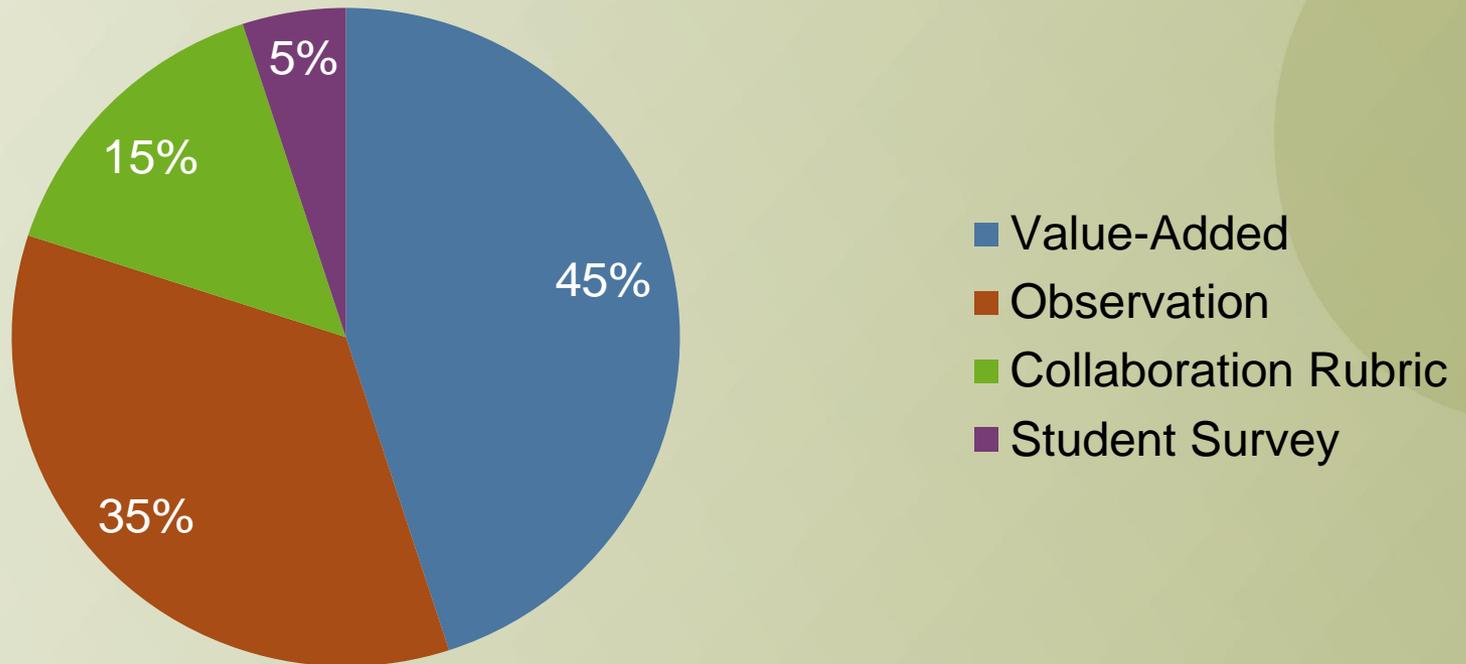
- Helps identify the following:
 - **Why** one teacher is better than another
 - Effective instructional practices
 - Ways in which the teacher could improve
- Absence of quality feedback inhibits teacher growth (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2006)

Challenges With Implementing Multiple Measures

- Establishing coherence
 - Within standards and between instruments
- Maintaining fidelity
 - Increased human and resource capacity
 - Enhanced training and support
 - Continual monitoring
- Allocating time
 - To train, implement, and support

Determining the Structure of the Evaluation System

Measures



Impact and Reliability

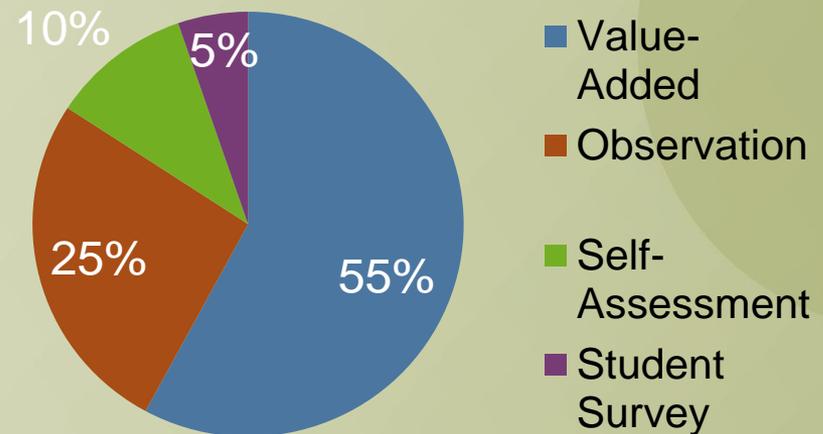
- Not all measures are equally useful.
- Not all measures are equally reliable.
- Measures' weights should vary according to:
 - Their ability to accurately measure student progress
 - Their demonstrated impact on student achievement
 - Their demonstrated impact on teacher practice
- Weights assigned to each measure dictate how each component will factor into final evaluation ratings.

Example Weights

■ Evaluation System Goals

- Teacher Accountability in Student Growth
- Compensation Decisions
- Student & Family Engagement

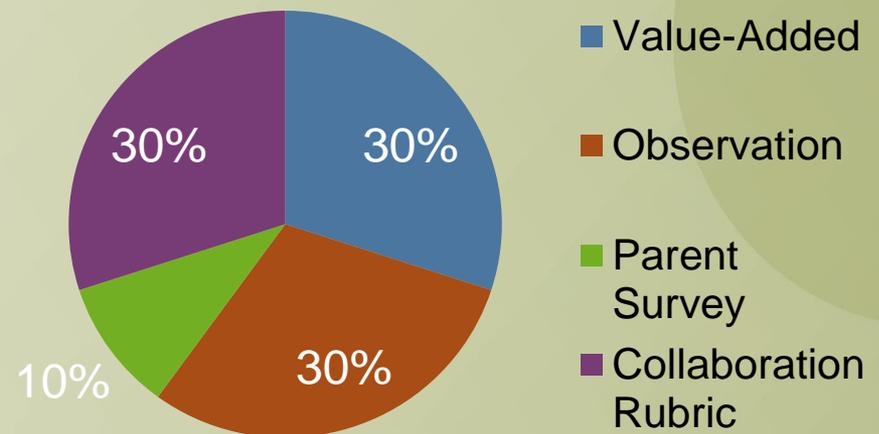
Measures



Example Weights

- Evaluation System Goals
 - Improve Teacher Capacity
 - Teacher Accountability in Student Growth
 - Collaboration

Measures



To Measure Growth, You Need Baseline Data

- Baseline data can be historic (found) or current (collected).
 - Historic (found) data includes all prior history on students' proficiency on specific standards.
 - Current (collected) data includes all efforts made by the state, district, school, or teacher to establish students' current levels of proficiency on specific standards.
- Each one has advantages and drawbacks.

Evidence of Teachers' Contribution to Student Learning Growth

- Value-added can provide useful evidence of teacher's contribution to student growth.
- “It is not a perfect system of measurement, but it can complement observational measures, parent feedback, and personal reflections on teaching far better than any available alternative” (Glazerman et al., 2010, p. 4).

What Value-Added and Other Growth Models Cannot Tell You

- Value-added and growth models are really measuring *classroom*, not teacher, effects.
- Value-added models can't tell you why a particular teacher's students are scoring higher than expected.
 - The teacher might be focusing instruction narrowly on test content.
 - The teacher might be offering a rich, engaging curriculum that fosters deep student learning.
- **How** the teacher is achieving results matters!

Schoolwide Growth

- Who can be evaluated with schoolwide growth measures?
 - Individual teachers in untested grades and subjects
 - Principals and other administrators
 - Other school personnel

Schoolwide Growth

- What do we mean by *schoolwide growth measures*?
 - Often, this means taking existing value-added or growth measures in reading and mathematics and applying them to individual teachers in other subjects or to administrators. (Teachers or administrators may select measures.)
 - We need not focus only on existing measures in reading or mathematics, however (e.g., team-based student learning organizations [SLOs] or subjects in which a small number of teachers represent the work of the school). Measures could be extended to nonacademic areas as well.

What Is Known About Schoolwide Growth: Are the Measures Good?

- Newer information on educator perceptions of schoolwide measures, e.g., Tennessee Year 1 Implementation report:¹
 - Administrators consistently noted that having schoolwide value-added scores has led to an increase in collaboration among teachers and a higher emphasis on academic standards in all subjects.
 - Teachers in subjects and grades that do not yield an individual value-added score do not believe it is fair to have 35 percent of their evaluations determined by schoolwide scores.

¹ Source: http://www.tn.gov/education/doc/yr_1_tchr_eval_rpt.pdf

The Four Ps (Portfolios, Projects, Performances, Products)

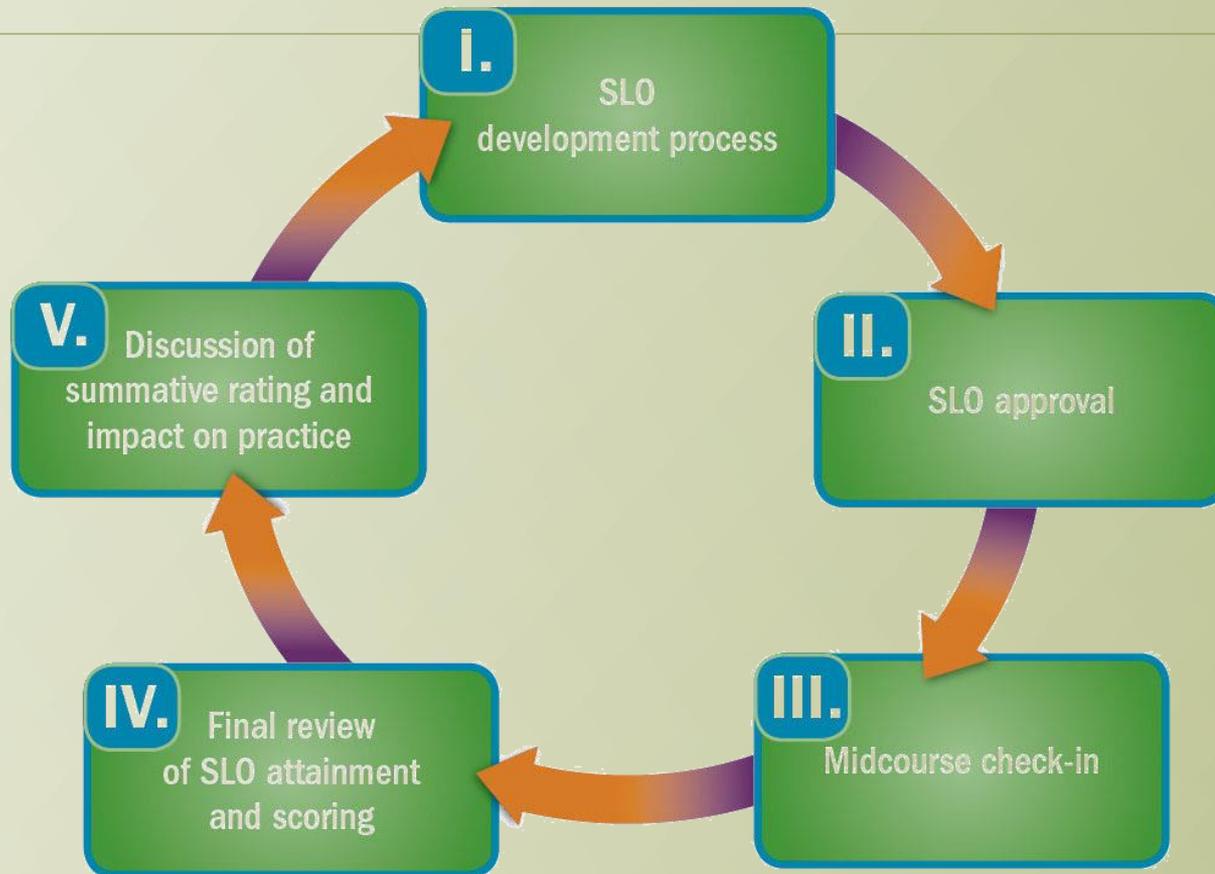
- Almost any measure, including student portfolios, projects, performances, and products (the Four Ps), can be used to demonstrate teachers' contributions to student learning growth if teachers share objectives and assessments and if they work together on scoring.
 - Use a high-quality rubric to judge initial knowledge and skills required for mastery of the content standard(s); then, use the same rubric to judge knowledge and skills at the end of a specific period (unit, grading period, semester, year, etc.)
 - Or use other assessments or student history to establish baseline on standards

Approach: SLOs

- Who can be evaluated with SLOs?
 - Individual teachers in all grades and subjects
 - Individual teachers in nontested grades and subjects
 - Teams of teachers
 - Principals or other administrators
 - Other school personnel
- What do we mean by “student learning objectives”?
 - An SLO is a measurable, long-term academic goal informed by available data that a teacher or teacher team sets at the beginning of the year for all students or for subgroups of students.



The SLO Evaluation Cycle



SLO Approaches

LEVEL 1

Set by teacher or teacher team using available assessments

Increasing Teacher Agency

LEVEL 2

Set by teacher or teacher team using assessment list or ranking

LEVEL 3

Set by teacher or teacher team using common assessments

LEVEL 4

Set by local education agency using common assessments and common growth targets

Increasing Comparability

What Is Known About SLOs?

- Teachers report increased focus on student achievement and data use as a result of the SLO goal-setting process, practices that align with evidence-based practice (Donaldson, 2012; What Works Clearinghouse, 2009).
- Teachers using SLOs value the opportunity to analyze data and plan instruction as part of the SLO process, and they report feeling “empowered” and took a more active role in their evaluation after SLOs were implemented (Community Training and Assistance Center [CTAC], 2013; TNTP, 2012).
- Some positive correlations have been found between the quality of SLOs and student achievement and between the number of objectives met by teachers and student achievement, but mixed results point to a need for more research (Austin Independent Schools, 2010; CTAC, 2013).

Teacher Collaboration

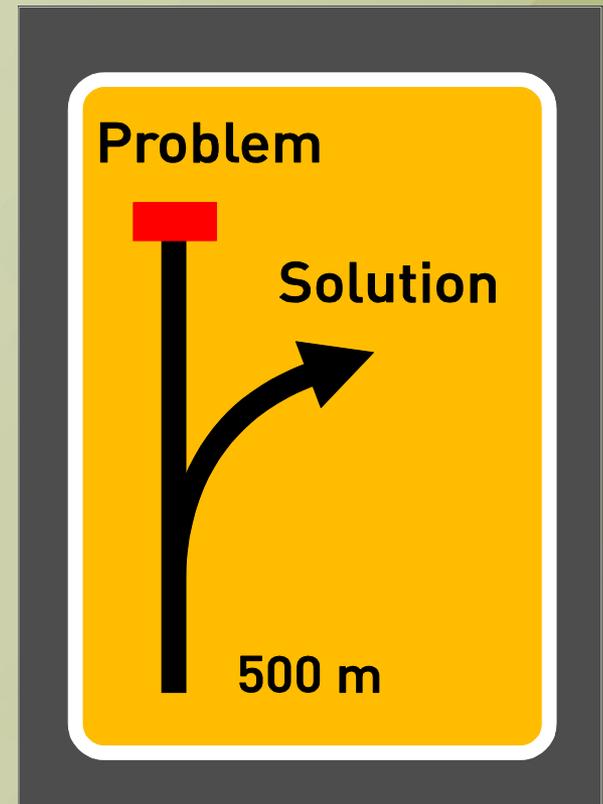
- Teachers don't need to assess in isolation.
 - Collaborate/share great lesson plans, materials, assessments, and so on across classrooms, schools, and districts (by content area, grades taught).
 - Work together to grade projects, essays, and so on by using technology when meeting in person is not feasible.
 - Working together encourages consistency in scoring, increasing validity and comparability of results.

Teacher Evaluation: Common Design and Implementation Challenges



10 Areas Where Challenges Occur

1. Stakeholder Engagement and Buy-In
2. Communication
3. Knowledge Management
4. Training
5. Creating Professional Growth Systems
6. Using Evaluation Data
7. Alignment and Integration
8. Adapting to Changes in Context
9. Scaling Up
10. Sustainability



1—Stakeholder Engagement and Buy-In

- Foundation of a successful effort to implement and sustain a robust educator evaluation system
- Along with communication issues, lack of engagement and buy-in is frequently cited reason for reform failure
- Beginning of creating a culture of trust and well-being

2—Communication

- Develop strategic communication plans that outline *multiple ways* to reach educators, parents, the community, and other important stakeholders about the effort.
- The communication plan must be designed to specifically inform educators throughout the change process.

3—Knowledge Management

- Ensure that all stakeholders understand the evaluation system, its components—including teaching standards—the evaluation process, and how the results will be used and affect their work.

4—Training

- At the heart of an evaluation system are personnel who can accurately assess teacher performance, communicate the results of that assessment to teachers, and help them plan for their professional growth.
- High-quality training is essential to ensure validity so that decisions that result from evaluation processes will be both fair and defensible.
- Training serves to ensure that evaluators agree on (1) the evidence and
- (2) what the evidence means in terms of a score.
- Calibration among evaluators requires ongoing training and should be done regularly to ensure that evaluators agree on the evidence and what it means.
- High-quality training builds trust!

5—Creating Professional Growth Systems

- Evaluation systems should be based on a theory of how teachers grow throughout their careers.
- The theory of development should guide professional learning opportunities and supports necessary to improve teacher practice.
- Creating the systems relies on the ability to accurately assess *all* teachers across a continuum of performance and to provide high-quality feedback.
- Evaluation instruments and processes should be aligned with state teaching standards, which outline the performance expectations for teachers.
- Professional growth opportunities should be job-embedded, collaborative, and sustained. The expectation is that growth happens over time. Opportunities for sharing and reflection around professional growth also should be built into the system.

6—Using Evaluation Data

- Evaluation as a process to improve teaching and leadership practice relies on valid and reliable data that are accessible to educators.
- Some data are more useful than others for helping teachers to adjust instruction. Data that are closer to the classroom and to specific instruction will be most useful for changing teacher practice.
- Observations followed quickly by specific feedback about instructional practice may be most helpful to teachers.
- Evaluation data should be part of a larger talent management system or approach.

7—Alignment and Integration

- Evaluation systems should not be layered on top of other initiatives but rather integrated and embedded into a larger system of educator talent development.
- Evaluation systems should align with and connect to statewide and districtwide priorities (e.g., college- and career-ready initiatives, teacher preparation).

8—Adapting to Changes in Context

- Creating dynamic systems is key to adapting to changes in leadership or policy direction.
- Focus on codifying essential practices and elements and building broad support to weather changes.

9—Scaling Up

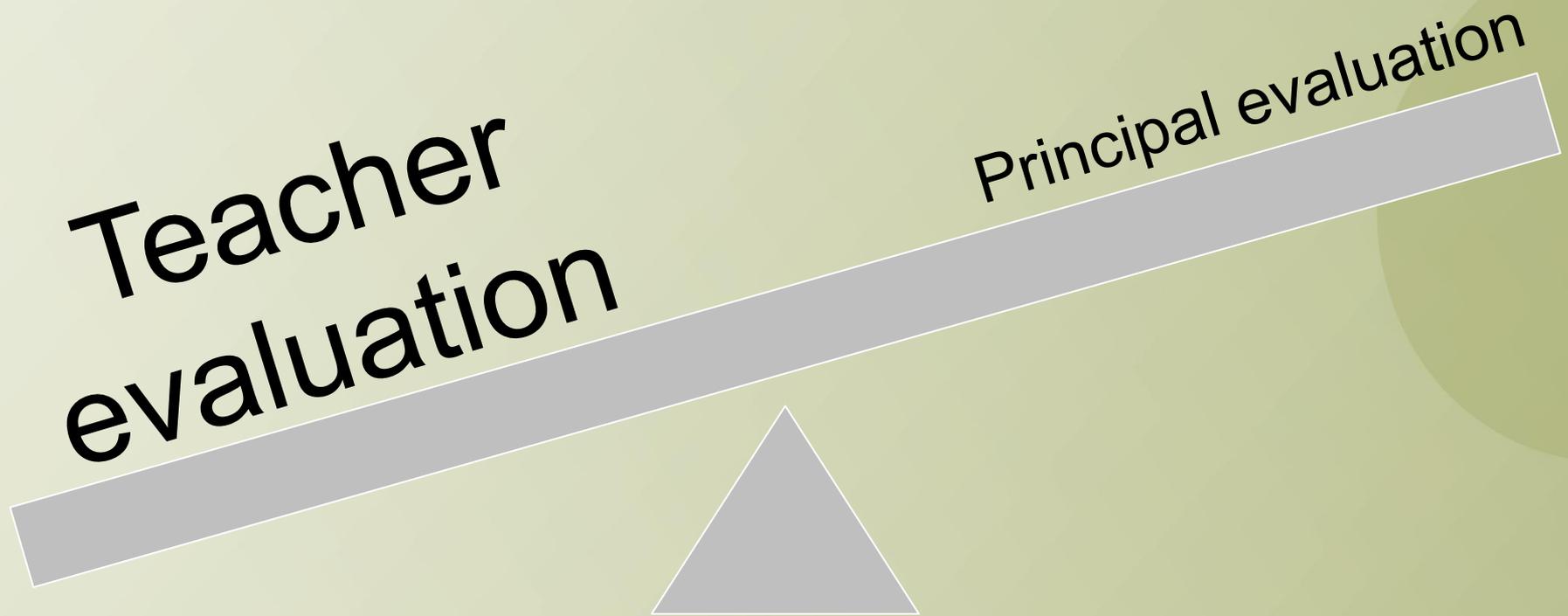
- Scaling up the evaluation system should be thoughtful and strategic and should be developed early in the design phase.
- Implement a continuous improvement plan (e.g., pilot, test, revise) as you scale.
- Collecting and analyze pilot data. Develop a set of questions you want answered by the pilot data before you implement the pilot.
- Consider who should participate in the pilot.

10—Sustainability

- Sustaining high-quality, robust educator evaluation systems is challenging in three ways:
 - Costs
 - Time
 - Teacher and principal beliefs about the value of the system and process

The Special Case of Leader Evaluation

The dialogue about evaluation virtually ignores leaders.



Embedding Professional Learning in Evaluation: Essential Conditions

- A culture of trust, continuous learning, and collaborative inquiry within each school
- Effective school leaders charged with the authority to support and monitor teacher learning and provide evidence-based feedback on practice
- Resources to support the formation of teams of teachers for job-embedded professional learning
- Availability of knowledgeable facilitators to ensure team time is purposeful and productive

Essential Conditions cont.

- Ample common collaborative learning time for teachers to analyze and reflect on their teaching
- Prioritization and allocation of resources to support sustained job-embedded learning
- Careful alignment among professional learning, school and district goals and instructional resources (Coggshall, et. al, 2012)

1 GENERATING GREATER TEACHING EFFECTIVENESS

Three Ways Evaluation Systems Can Propel Continuous Improvements of Teaching and Learning

SHARED UNDERSTANDING OF EFFECTIVE PRACTICE

Helping develop shared understandings of the contours of effective teaching in every school community

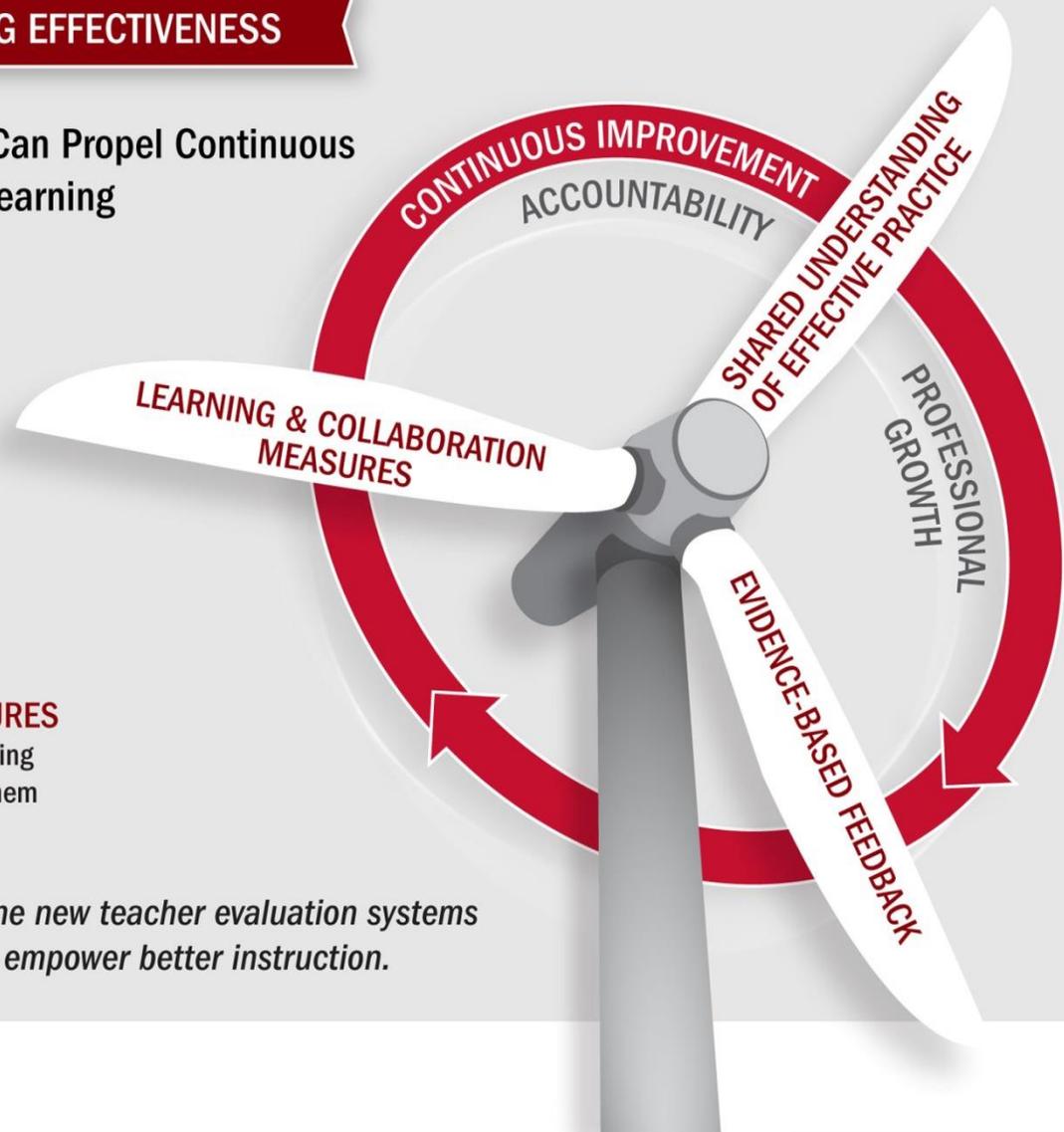
EVIDENCE-BASED FEEDBACK

Providing timely, relevant, and sufficient evidence-based feedback to teachers on the effectiveness of their practice

LEARNING & COLLABORATION MEASURES

Accounting for teachers' engagement in learning and collaboration by including measures of them as part of the evaluation

With these three elements in place, the new teacher evaluation systems taking shape around the country can empower better instruction.



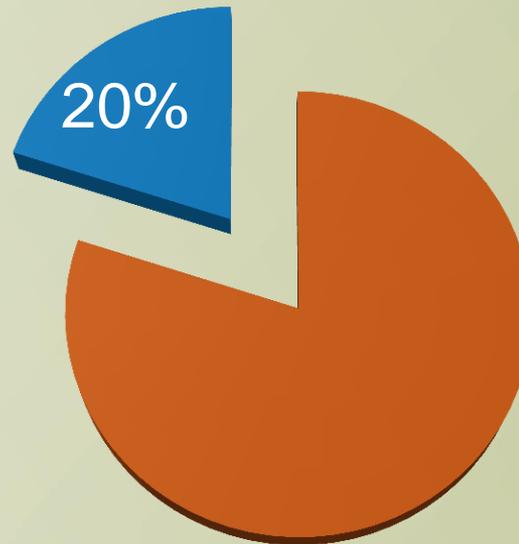
Teacher Voice in Evaluation Reform

- Why is teacher voice often not included in evaluation reforms?
 - Time
 - Concerns about setting false hopes
 - Other reasons?



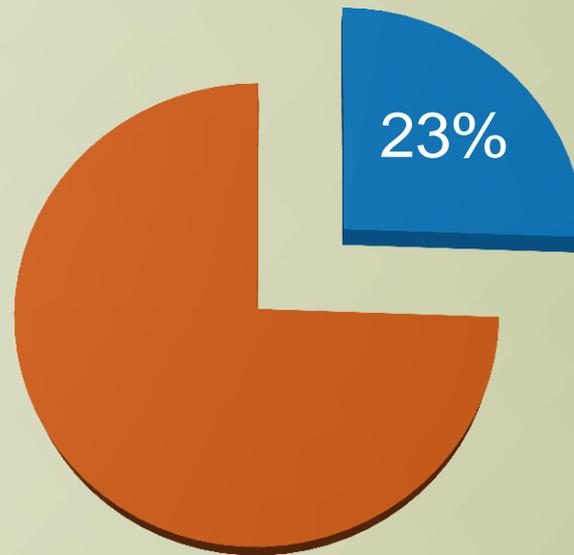
What Teachers Have to Say

Only 20 percent of teachers believe that they are regularly consulted about what happens in their schools.



What Teachers Have to Say

Only 23 percent of teachers believe that district leaders speak to them to gain a stronger sense of teachers' concerns.



Teacher Voice in Evaluation Reform

- Why should teacher voice be included in evaluation reforms?
 - Sound design
 - Effective implementation
 - Sustainability
 - Teacher leadership



What Is *Everyone at the Table*?

- A **free**, online resource dedicated to bringing teachers to the policy table: <http://www.everyoneatthetable.org>



What Is *Everyone at the Table*?

- A communications and engagement package that promotes **structured, constructive, even-handed** teacher dialogue on evaluation reform



Why Engaging Teachers in Evaluation Reform Is Important: Leaders' Perspectives

- View the [Discussion Starter Video](#) and Leaders' Involvement Guide



Principal-led Teacher Engagement

- Understand and engage teachers
- Establish a task force that includes authentic opportunities for teachers to lend insight and input
- Establish engagement teams
- Assess feedback and proceed based on feedback

The Benefits of Teacher Engagement

- Allowing more perspectives, which leads to better decisions
- Encouraging broader ownership of implementation
- Improving communication
- Building trust
- Shaping a collaborative culture

What Makes Teacher Engagement Effective?

- Transparent outcomes and expectations
- Teachers' interests, concerns, and natural language
- Broad, inclusive participation
 - Ground rules
 - Dialogue
 - Impartial discussion materials
 - Trained moderators



Impartial Discussion Materials

- Public Agenda's *Choicework* framework
- Breaks “either-or” patterns of thinking
- Three or four choices
- Focuses on trade-offs and choices

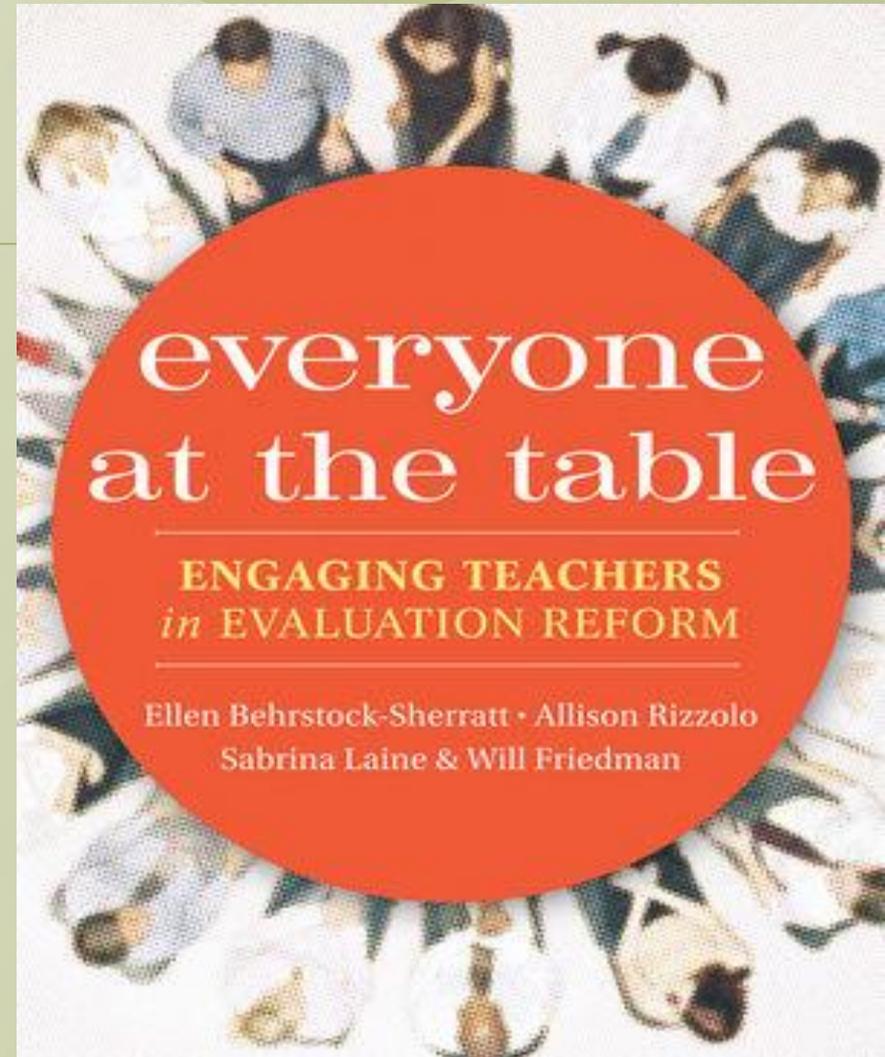


Trained Moderators

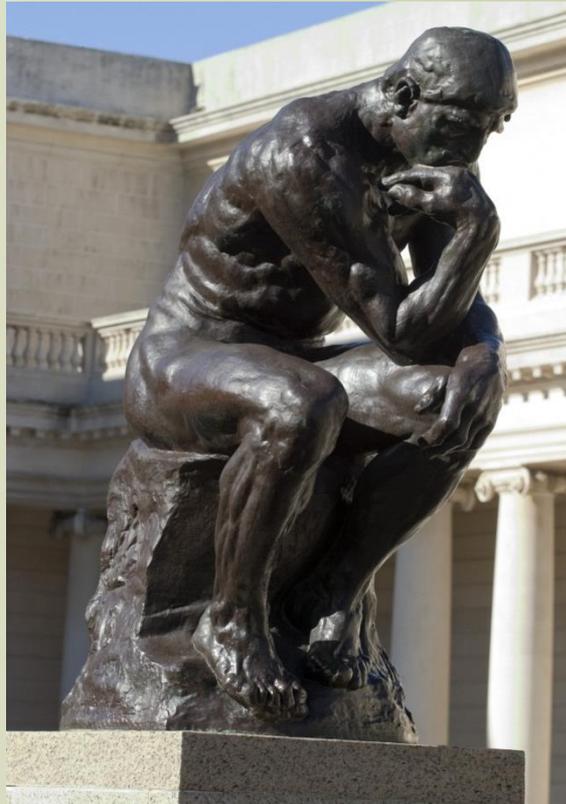
- Unbiased
- Manage time
- Facilitate group interaction
- Keep the discussion focused but not controlled
- Foster critical thinking and full consideration of pros and cons
- Model collaborative attitudes



Everyone at the Table: The Book



Questions?



GTL Center Resources

Teacher and Principal Evaluation Practical Guides

Need guidance to help design your teacher and leader evaluation systems? Access key resources, guiding questions, and practical examples through these interactive sites.

<http://www.gtlcenter.org/tools-publications/online-tools/teacher-evaluation>

<http://www.gtlcenter.org/tools-publications/online-tools/principal-evaluation>

References

Austin Independent Schools. (2010). *AISD REACH program update*. Retrieved from http://archive.austinisd.org/inside/docs/ope_09-83_RB_Reach_TAKS_and_SLOs.pdf

Bill & Melinda Gates Foundation. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study*. Seattle, WA: Author. Retrieved from http://www.metproject.org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy*, 1(2), 176–216. <http://www.teacherpolicyresearch.org/portals/1/pdfs/Reducing%20Entry%20Requirements%20EPF%202006.pdf>

References (continued)

- Community Training and Assistance Center. (2013). *It's more than money: Teacher Incentive Fund—Leadership for educators' advanced performance Charlotte-Mecklenburg Schools*. Boston, MA: Author.
- Donaldson, M. L. (2012). *Teachers' perspectives on evaluation reform*. Washington, DC: Center for American Progress.
- Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D. O., & Whitehurst, G. J. (2010). *Evaluating teachers: The important role of value-added*. Washington, DC: Brown Center on Education Policy at Brookings. Retrieved from http://www.brookings.edu/reports/2010/1117_evaluating_teachers.aspx
- Grissom, J. A., Kalogrides, D., & Loeb, S. (2012). *Using student test scores to measure principal performance* (NBER Working Paper No. 18568). Retrieved from <http://www.nber.org/papers/w18568>

References (continued)

- Heifetz, R. A., & Linsky, M. (2002). *Leadership on the line: Staying alive through the dangers of leading*. Cambridge, MA: Harvard Business School Press.
- Polikoff, M. S., Porter, A. C., & Smithson, J. (2011). How well aligned are state assessments of student achievement with state content standards? *American Educational Research Journal*, 48(4), 965–995.
- TNTP. (2012). *Summer report: Creating a culture of excellence in Indiana schools*. Indianapolis: Indiana Department of Education.
- What Works Clearinghouse. (2009). *Using student achievement data to support instructional decision making*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/practice_guides/dddm_pg_092909.pdf

Sabrina W.M. Laine
slaine@air.org
973-727-1446 (cell)

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
877-322-8700
gtlcenter@air.org
www.gtlcenter.org | www.air.org

PERFORMANCE MANAGEMENT Advantage
Evaluation & Professional Growth

at American Institutes for Research ■

