Teacher Evaluation for Professional Growth

Design and Implementation Considerations

Angela Minnici, Director
June 2014
The mission of the Center on Great Teachers and Leaders (GTL Center) is to foster the capacity of vibrant networks of practitioners, researchers, innovators, and experts to build and sustain a seamless system of support for great teachers and leaders for every school in every state in the nation.
Workshop Overview

- Purpose of Teacher Evaluation: What Are Your Goals?
- Benefits and Challenges of Using Multiple Measures of Teacher Performance
  - Observation Measures
  - Student Growth Measures
  - Surveys
- Weighting and Combining Multiple Measures
- Teacher Evaluation to Support Professional Learning
- Questions/Comments
Purposes of Teacher Evaluation
What Is the Purpose?

- As a team, determine which five key purposes best describe why you are evaluating teachers.
- You may select only five from the list.
Benefits and Challenges of Multiple Measures of Teacher Performance
Choosing the Right Measures

Ask yourself:

- What do we want to measure?
- How will the pieces of evidence support our goals/purpose?
Multiple Measures of Teacher Effectiveness

Evidence of Instructional Quality

- Classroom observations
- Lesson plans, assignments, and student work
- Student surveys such as Harvard’s Tripod
- Evidence binder/portfolio
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 Dynamic Indicators of Basic Early Literacy Skills (DIBELS)
Multiple Measures of Teacher Effectiveness

Evidence of Professional Responsibility

- Administrator/supervisor reports, parent surveys
- Teacher reflection and self-reports, records of contributions
Measures That Help Teachers Grow

- Include protocols and processes that teachers can examine and comprehend.
- Directly and explicitly align with teaching standards.
- Motivate teachers to examine their own practice against specific standards.
Measures That Help Teachers Grow

- Allow teachers to participate in or co-construct the evaluation (such as portfolios).
- Give teachers opportunities to discuss the results for formative purposes with evaluators, administrators, teacher learning communities, mentors, coaches, etc.
- Align with and inform professional growth and development offerings.
Multiple Measures: Observation
Teacher Observations:

Strengths

- Great for teacher formative evaluation (if observation is followed by opportunity to discuss)
- Helps evaluator (principals or others) understand teachers’ needs across school or across district
Teacher Observations:

Weaknesses

- Only as good as the instruments and the observers
- Considered “less objective”
- Expensive to conduct (personnel time, training, calibrating)
- Variance in validity of observation results depending on who is conducting observations and the level of training and calibration of observers
Observations are the traditional measure of teacher performance.

Teachers feel they have some control over the process and outcomes.

They report that having a conversation with the observation and receiving constructive feedback after the observation is greatly beneficial.

Evidence-centered discussions can help teachers improve instruction.

Peer evaluators often report that they learn new teaching techniques.
When Teachers Don’t Value Observations, It’s Because…

- They do not receive feedback at all.
- The feedback they receive is not specific and actionable.
- The observer suggests actions but is unable to offer the means and resources to carry out those actions.
  - Mentors/coaches, other support personnel
  - Time for individual growth planning/activities
  - Protected time for collaboration with others
Implementation Considerations for Teacher Observations
Interrater Reliability and Agreement

- Interrater *reliability* is the relative similarity between two or more sets of ratings.

- Interrater *agreement* is the degree to which two raters, using the same scale, give the same rating in identical situations.

- Rater *reliability* refers to consistency in judgments over time in different contexts and for different educators.

“The degree of observer agreement is one indicator of the extent to which there is a common understanding of teaching within the community of practice” (Gitomer et al., in press).
Why Does It Matter?

- Reliability and agreement are important for evaluators conducting observations, assessing artifact reviews, and approving and scoring student learning objectives.

- Reliability and agreement are essential to:
  - Bridge the credibility gap.
  - Train and certify raters.
  - Monitor system performance.
  - Make human resource decisions.
  - Link professional development to evaluation results.
What Is Interrater Reliability?

Do Raters A and B demonstrate interrater reliability?

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Component Score</th>
<th>Rater A</th>
<th>Rater B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher A</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Teacher B</td>
<td>2</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Teacher C</td>
<td>3</td>
<td>+1</td>
<td>4</td>
</tr>
<tr>
<td>Teacher D</td>
<td>4</td>
<td></td>
<td>5</td>
</tr>
</tbody>
</table>
### Illustrating Rater Agreement

<table>
<thead>
<tr>
<th>Component</th>
<th>Component Score</th>
<th>Type of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Rater A</strong></td>
<td><strong>Rater B</strong></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
## Calculating Agreement

<table>
<thead>
<tr>
<th>Component</th>
<th>Component Score</th>
<th>More Than One Point Off</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rater A</td>
<td>Rater B</td>
</tr>
<tr>
<td>Subcomponent 1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Subcomponent 2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Subcomponent 3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Subcomponent 4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Subcomponent 5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Component Score (Average)</td>
<td>2.8</td>
<td>2.8</td>
</tr>
</tbody>
</table>
Interrater Reliability and Agreement: How Much Is Enough?

There is currently no standard for the level of agreement or reliability for the use of measures in high-stakes performance evaluation. Experts tend to agree, however, that at minimum:

- Absolute agreement should be 75 percent.
- Kappa rating should be .75.
- Intra-class correlations should be .70.
What Affects Reliability and Agreement in Observation?

Quality of the Observation Instrument

- Number and complexity of components and indicators
- Clarity and consistency of language
- Meaningful, realistic distinctions across levels of performance
- Likelihood of seeing the described practice in the classroom
What Affects Reliability and Agreement in Observation?

Observer Bias: What are some of the various “lenses” that might bias:

- A teacher evaluator?
- A principal evaluator?
Example Activity: Strategizing for Reducing Bias

Handout 2: Common Sources of Bias

- **Step 1.** At your table, work as a group to match each common rater error with a possible strategy evaluators can use to avoid the error.

- **Step 2.** After each match, discuss other possible strategies you have seen used or that you think might be effective.
Example Activity: Strategizing for Reducing Bias

Answer Key

- 1 = D
- 2 = G
- 3 = A
- 4 = F
- 5 = B
- 6 = E
- 7 = H
- 8 = C
What Affects Reliability and Agreement in Observation?

**Context**
- Relationship to observed teacher
- Other demands on observer’s time
- Level of students
- Particular challenges of students
- How results will be used (e.g., high stakes versus formative)
A Corrective Lens: The Observation Instrument

Quality of the Observation Instrument

- Number and complexity of components and indicators
- Clarity and consistency of language
- Meaningful, realistic distinctions across levels of performance
- Likelihood of seeing the described practice in the classroom
A Corrective Lens:
Observation Format

Frequency and Number of Observations/Observers

- More frequent, shorter observations
- Observations by more than one observer
In findings from the MET Project’s three-year study, the Bill & Melinda Gates Foundation (2013) reports that “there are many roads to reliability” (p. 18).
Improving Reliability and Agreement

Disciplining Evaluator Judgment

- No matter how much you train or how high-quality your instrument is, total objectivity in any type of measurement is impossible.

- **Training Goal:** “Disciplining” evaluators’ professional judgment and developing common understanding of effective instruction/leadership practice.
What Does Research Say?

Improving Observational Score Quality: Challenges in Observer Thinking (Bell et al., 2013)

- Based on data from two studies funded by the Bill & Melinda Gates Foundation, Measures of Effective Teaching (MET) and Understanding Teaching Quality (UTQ)

- Examined four protocols and analyzed calibration and certification scores

- For a subsample of UTQ observers, captured “think aloud” as they made scoring decisions and engaged them in stimulated recall session
Key Research Finding: Some Dimensions Are Harder to Score

**Harder to Score Reliably:** High Inference, focused on student/teacher interactions = more uncertainty

- Instructional Techniques
- Emotional Supports

**Easier to Score Reliably:** Low Inference = less uncertainty

- Classroom Organization
- Classroom Environment

*Source: Bell et al., 2013*
Why? Getting in Observers’ Heads

Observers say some dimensions are more challenging to score because…

- They feel the scoring criteria for the dimension were applied inconsistently by the master scorer.
- They feel the scoring criteria for the dimension were applied in ways that they did not agree with or understand.

Source: Bell et al., 2013
Training Take-Aways

Ensure that observers have opportunities to learn to:

- Use the rubric language in explaining their scoring decisions.
- Consistently take notes that gathers useful evidence.
- Avoid making scoring decisions during note-taking.
- Resort back to the scoring criteria when uncertain.
- Score using videos and in live classroom observations.

Source: Bell et al., 2013
Observation Mode and Rater Drift

Effect of Observation Mode on Measures of Secondary Mathematics Teaching (Casabianca et al., 2013)

- Observation mode (video versus live) has minimal effect on rater reliability.
What Improves Evaluators’ Reliability and Agreement?

- **Concrete examples** of the practices described in the rubric at each performance level for both evaluators and educators

- **Opportunities to practice** scoring, receive immediate feedback on scoring, and **regularly calibrate** their scores against master scorers on an ongoing basis
What Improves Evaluators’ Reliability and Agreement?

- **Supplemental training** on hard to score sections, for example:
  - Learning to focus on student responses
  - Weighing competing evidence
  - Understanding what a specific element looks like in classrooms
What Improves Evaluators’ Reliability and Agreement?

- An **assessment and/or certification test** to ensure that evaluators can meet a minimum level of reliability and agreement before evaluating educators

- **Ongoing recalibration**, opportunities to collaborate with fellow observers to strengthen skill in difficult-to-score components

- Annual refresher and recertification test
What Does Observation Training Usually Look Like?

Learning the Observation Framework

- The **educational philosophy** and **research base** used to develop the instructional or leadership framework and observation rubrics

- The purpose and logic for each **performance level** and **scale** in the framework or rubric

- The framework or rubric structure and the core performance behaviors included in each dimension or component
What Does Observation Training Usually Look Like? Phase 2

Learning to Apply the Observation Framework

1. Explore each core practice via short one- to two-minute video clips illustrating the practice.

2. Explore what each practice looks like at each level of performance and discuss why the practice fits.

3. Practice with 10- to 15-minute classroom videos to identify rubric elements in the observed practice and initial practice with scoring.

4. Practice scoring full-length classroom videos, discussing scoring decisions, and calibrating scores against master scores.
What Does Observation Training Usually Look Like? Phase 3

- Assessment tests to demonstrate evaluator’s mastery of necessary skills and reliability/agreement
- Recalibration and reassessment as needed
- Ongoing recalibration to retain accuracy and reliability
- Annual recertification
Master scored videos are “videos of teachers engaged in classroom instruction that have been assigned correct scores by people with expertise in both the rubric and teaching practice” (McClellan, 2013).
Master Scoring Process

- Creates a library of videos that can be used for:
  - Rater assessment and ongoing calibration
  - Orienting teachers to the framework
  - Teacher professional development

- Creates a cohort of master observers who can assist in training and coaching other evaluators

- Provides formative feedback to improve and refine the observation rubric

Source: McClellan, 2013
Multiple Measures: Student Growth
How Do We Measure Contributions to Learning Growth in the Following Cases?

- Teachers of nontested subjects (e.g., social studies, K–2, art, drama, band)

- Teachers of certain student populations and situations in which standardized test scores are not available or utilized
  - Teachers of students assessed on alternate assessments
  - Smaller teacher caseloads for some student groups (e.g., students with disabilities, English language learners)
Range of State and District Approaches

- Existing measures
- Rigorous *new* measures
- Portfolios/products/performance/projects
- Student learning objectives
### Existing Assessments

<table>
<thead>
<tr>
<th>Strengths of This Measure</th>
<th>Challenges for This Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Already exist</td>
<td>Validity (whenever a measure is used in a way that was not intended)</td>
</tr>
<tr>
<td>Teacher familiarity and use</td>
<td>Concern over content validity</td>
</tr>
<tr>
<td>Not creating additional assessments/work</td>
<td>Fidelity and standardization</td>
</tr>
<tr>
<td>Possibly formative in nature</td>
<td></td>
</tr>
</tbody>
</table>

- **Delaware, Tennessee, Rhode Island**
  - Assembled group of practitioners
  - Tightly facilitated meetings
  - Group-recommended measures
  - Expert panel approves measures

- **Center on Response to Intervention**
  - Progress monitoring tools
  - Tiers I, II, and III
## New Assessments

### Hillsborough County, Florida
- Race to the Top Grantee
- Pre- and post-assessment for each course
- Scores averaged over three years to determine teacher effectiveness

<table>
<thead>
<tr>
<th>Strengths of This Measure</th>
<th>Challenges for This Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Tests can be made to match specific grade or subject standards.</td>
<td>- More tests!</td>
</tr>
<tr>
<td>- Assessments can be created to meet standards of validity and reliability.</td>
<td>- Time and cost-intensive approach</td>
</tr>
<tr>
<td>- Same assessment can be given across district/teachers.</td>
<td>- Paper-and-pencil tests that may not be appropriate as the sole measure, particularly in subjects requiring students to demonstrate knowledge and skills (art, music, etc.)</td>
</tr>
<tr>
<td></td>
<td>- Capacity to build valid and reliable assessments</td>
</tr>
</tbody>
</table>
**Strengths of This Measure**
- Evidence of growth can be documented over time using performance rubrics.
- Portfolios and projects can reflect skills and knowledge that are not readily measured by paper-and-pencil tests.

**Challenges for This Measure**
- Training for interrater reliability
- Logistical challenge for group raters
- Ensuring rigor

---

**New York and Rhode Island Districts Participating in the AFT Innovation (i3) Project**
- As in Delaware, teachers identify existing measures already used in classrooms.
- Develop pretests to establish knowledge and skills students need prior to project.
- Panel of experts and practitioners evaluate and approve measures.
Multiple Measures: Student Learning Objectives
Student Learning Objectives as a Measure of Student Growth

A student learning objective (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.
Where Are SLOs Being Used?

- A review of publically available documents shows that 23 states require and two states allow individual teacher SLOs (Lacireno-Paquet, Morgan, & Mello, 2014).
- 12 Race to the Top states require or recommend SLOs for at least some teachers.
- Many Teacher Incentive Fund grantees use SLOs.
## Use in Districts and States

<table>
<thead>
<tr>
<th>District/State</th>
<th>Required to Use SLOs</th>
<th>Use of SLO Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austin, TX</td>
<td>Teachers and administrators</td>
<td>Compensation &amp; Evaluation</td>
</tr>
<tr>
<td>Denver, CO</td>
<td>All teachers</td>
<td>Compensation</td>
</tr>
<tr>
<td>Georgia</td>
<td>Teachers in NTGS</td>
<td>Evaluation</td>
</tr>
<tr>
<td>Hazelwood, MO</td>
<td>All teachers in SIG schools</td>
<td>Evaluation</td>
</tr>
<tr>
<td>Indiana</td>
<td>All teachers in the default model</td>
<td>Evaluation</td>
</tr>
<tr>
<td>Kentucky</td>
<td>All teachers</td>
<td>Evaluation</td>
</tr>
<tr>
<td>Louisiana</td>
<td>Teachers in NTGS</td>
<td>Evaluation</td>
</tr>
<tr>
<td>Maryland</td>
<td>Teachers in NTGS</td>
<td>Evaluation</td>
</tr>
<tr>
<td>McMinnville, OR</td>
<td>All teachers</td>
<td>Compensation</td>
</tr>
<tr>
<td>New Haven, CT</td>
<td>Teachers in NTGS</td>
<td>Evaluation</td>
</tr>
<tr>
<td>New York</td>
<td>Teachers in NTGS</td>
<td>Evaluation</td>
</tr>
<tr>
<td>Ohio</td>
<td>Offered for all teachers and principals</td>
<td>Evaluation</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>All teachers</td>
<td>Evaluation</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>All teachers</td>
<td>Evaluation</td>
</tr>
</tbody>
</table>

*Source: Lachlan-Haché, Matlach, Cushing, Mean, and Reese (2013)*
Why Are SLOs Being Widely Used?

- Reinforce evidence-based teaching practices.
- Can be used with all teachers.
- Are adaptable.
- Encourage collaboration.
- Acknowledge the value of educator knowledge and skill.
- Connect teacher practice to student learning.
What Are SLOs?

- A goal that demonstrates a teacher’s impact on student learning within a given interval of instruction.

- A measurable, long-term academic target written by an individual teacher or a teacher team.

- A process that allows teachers to demonstrate their impact on student learning within a given interval of instruction.
  - Student baseline data are collected.
  - Appropriate objectives are set for students.
  - Students are assessed at the end of the interval.
Most SLOs include or address criteria like the following:

- Baseline and trend data
- Student population (general and special needs)
- Interval of instruction
- Standards and content
- Assessment(s)
- Growth target(s)
- Rationale for growth target(s)
# How Are SLOs Developed?

## SLO Checklist

<table>
<thead>
<tr>
<th>Baseline and Trend Data</th>
<th>Student Population</th>
<th>Interval of Instruction</th>
<th>Standards and Content</th>
<th>Assessments</th>
<th>Growth Target(s)</th>
<th>Rationale for Growth Target(s)</th>
<th>Instructional Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>❑ Identifies sources of information about students</td>
<td>❑ Covers all students in the class (or in the case of a targeted SLO, covers all students in the subgroup)</td>
<td>❑ Matches the length of the course (e.g., quarter, semester, trimester, year)</td>
<td>❑ Specifies how the SLO will address applicable standards from the highest ranking of the following: (1) Common Core State Standards (2) Academic Content Standards (3) National standards put forth by education organization</td>
<td>❑ Identifies assessments that have been reviewed by content experts to effectively measure course content and reliably measure student learning as intended</td>
<td>❑ Ensures all students in the course have a growth target</td>
<td>❑ Demonstrates teacher knowledge of students and content</td>
<td>❑ Highlights the instructional methods that will best support the student achievement goals set forth in the SLO</td>
</tr>
<tr>
<td>❑ Draws upon trend data, if available</td>
<td>❑ Describes the student population and considers any contextual factors that may impact student growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>❑ Explains why target is appropriate for the population</td>
<td>❑ Discusses how the teacher will differentiate instruction in support of this SLO</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>❑ Addresses observed student needs</td>
<td>❑ Uses data to identify student needs and determine appropriate growth targets</td>
</tr>
</tbody>
</table>

---

Center on GREAT TEACHERS & LEADERS at American Institutes for Research
SLO Approaches

Increasing Teacher Agency

Type 1
Set by teacher or teacher team using available assessments

Type 2
Set by teacher or teacher team using assessment list or ranking

Type 3
Set by teacher or teacher team using common assessments

Type 4
Set by local education agency using common assessments and common growth targets

Increasing SLO Comparability

Student Learning Objectives

Strengths of This Measure

- Provide the opportunity to discuss teacher expectations and goals and reinforce teacher practices.
- Feedback from SLOs can provide detailed instructional goals for educator professional development plans.
- Schoolwide SLO attainment can be easily translated to a results measure for principal evaluation, and growth priorities can be integrated into the school growth/improvement plan.
- They are flexible and can be tailored to specific grade levels, subjects, students, and individual teachers.
- SLOs encourage collaboration among educators to set and achieve goals and provide educators with ownership over their evaluations.

The Illinois Performance Evaluation Advisory Council (PEAC) decided to use SLOs in its model evaluation system for Type III assessments and recommends that districts use SLOs for Type III assessments.
## Challenges for This Measure

- SLOs can be time-intensive to develop and evaluate while meeting requirements for rigor and comparability.
- SLOs require high-quality assessments, which may be difficult to identify or challenging to develop.
- SLOs require specific guidance to help educators define appropriate differentiated targets for students.
- SLO use in principal evaluation should be balanced with a measure of SLO quality, through district-level review processes.
- SLOs can require a shift in school culture to support continuous improvement of educators.
SLOs: Setting Growth Targets
Establishing Growth Targets

Growth targets:
- Are informed by baseline trend data:
  - Previous end-of-year assessment(s)
  - Preassessment scores
  - Individual student trend data
  - Schoolwide trend data
  - Scores on predictive assessments
  - Previous end-of-year grades
  - Beginning of year school work
- Include specific indicators of growth.
- Are set so that all students included in the SLO can demonstrate developmentally appropriate growth.
Examples of SLO Growth Targets

Basic Growth Target
- All students have the same growth target.
- *Example:* All of my students will grow by 20 points by the end of the semester.
Examples of SLO Growth Targets

**Simple Average Growth Calculation**

- Growth targets are determined by a common formula, but each student has a different growth target based on his or her preassessment score.

- *Example:* Based on the preassessment score, students will score halfway between their baseline score and 100.

- If a student scored 50 on the preassessment, his or her growth target is 75. Or, if a student scored 40 on the preassessment, his or her growth target is 70.
Examples of SLO Growth Targets

Tiered Growth Target

- Group students based on their preassessment scores.
- Divide students into three or more categories (e.g. low, middle, advanced).
- Example:

<table>
<thead>
<tr>
<th>Preassessment Score</th>
<th>Growth Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–45 points</td>
<td>65</td>
</tr>
<tr>
<td>46–70 points</td>
<td>75</td>
</tr>
<tr>
<td>70+ points</td>
<td>85</td>
</tr>
</tbody>
</table>
Examples of SLO Growth Targets

Advanced Tiered Growth Target

- Students have a tiered target based on their preassessment.
- Divide students into three or more categories.
- Students have to reach the greater of the two targets.

*Example:*

<table>
<thead>
<tr>
<th>Preassessment Score</th>
<th>Growth Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–45 points</td>
<td>65 or +35 points, whichever is greater</td>
</tr>
<tr>
<td>46–70 points</td>
<td>75 or +15 points, whichever is greater</td>
</tr>
<tr>
<td>70+ points</td>
<td>85 or +14 points, whichever is greater</td>
</tr>
</tbody>
</table>
Examples of SLO Growth Targets

Individual Student Growth Target

- Each student has their own growth target.
- Can use tiered targets as a guide for individual targets and adjust up or down as appropriate.
- Example:

<table>
<thead>
<tr>
<th>Student Name</th>
<th>Preassessment</th>
<th>Growth Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>John G.</td>
<td>43</td>
<td>71</td>
</tr>
<tr>
<td>Sally M.</td>
<td>68</td>
<td>77</td>
</tr>
<tr>
<td>Mary S.</td>
<td>55</td>
<td>74</td>
</tr>
<tr>
<td>Chris N.</td>
<td>34</td>
<td>65</td>
</tr>
</tbody>
</table>
SLO Growth Target Writing Activity
SLO Activity

- Review the SLO baseline data provided.
- Analyze the baseline data to determine:
  - What the information is telling you about your students
  - What kind of growth target would you develop for these students?
  - If using a tiered target, what are some natural groupings of students?
- Discuss your analysis and approach with a colleague.
Questions to Consider

• What type of baseline and trend data are available to teachers?
• How are teachers supported in analyzing baseline data?
• How are evaluators supported in knowing “how much growth is enough”?
<table>
<thead>
<tr>
<th>State</th>
<th>SLO Policy</th>
<th>Resources</th>
</tr>
</thead>
</table>
| Georgia     | • SLOs are set by the district.  
• The state reviews and approves the SLOs and provides supports for district implementation, including the development, review, and approval of SLO assessments.  
• The district creates course-level SLOs for all nontested grades and subjects, which teachers modify and use with specific targets for their individual classrooms. | Georgia SLO Operations Manual                  |
| Indiana     | • Teachers develop SLOs, and administrators or groups of district leaders approve them.  
• The state provides trainings, webinars, and resources to assist with SLO development.                                                                                                                   | Indiana RISE Handbook  
Indiana SLO Manual                                |
### State SLO Policies and Resources

<table>
<thead>
<tr>
<th>State</th>
<th>SLO Policy</th>
<th>Resources</th>
</tr>
</thead>
</table>
| New York     | • The state specifies the number and types of SLOs teachers must create based on teaching assignment and specifies which assessments are acceptable for the SLO.  
• Districts in New York can establish requirements or recommendations for assessments and rating scales to be used with SLOs to ensure consistency of expectations across schools. | [New York Locally Selected Measures](#)  
[New York SLO Resources](#) |
| Ohio         | • Teachers develop SLOs, and administrators or groups of district leaders approve them.  
• The state plans to randomly audit SLOs at the local level. | [Ohio SLO Process](#) |
| Rhode Island | • Teachers develop SLOs, and administrators or groups of district leaders approve them.  
• **Principal SLO measures, with district review process.** | [Rhode Island SLO Materials](#) |
Additional SLO Examples and Resources

- Austin, Texas: Austin SLO Manual
- Denver, Colorado: Denver Schools Student Growth Objective and Monitoring Process
- Connecticut: Connecticut’s System for Educator Evaluation and Development
- District of Columbia: DCPS IMPACT Guidebooks
- Maryland: Maryland Teacher and Principal Evaluation Guidebook
Multiple Measures: Surveys
Student surveys are a useful tool to assess teacher effectiveness:

- Students can accurately assess real differences between teachers.
- Student perceptions predict student outcomes, including student achievement and social and emotional skills.
- Student perceptions are a reliable measure of teacher performance.
- Students can provide valuable feedback to teachers.

More states and districts are using them as part of educator evaluations (e.g., Colorado, Georgia, and Massachusetts).
Selecting a Student Survey

- What is the intended use of the measures?
  - Explore individual teacher behavior or practice.
  - Track student perceptions or responses.

- What aspects of teacher effectiveness should be measured?
  - Determine which teacher behaviors (constructs) are important to measure.
  - Ensure constructs can be used to support teacher growth.

- What is the validity and reliability of the measures?

- What is the cost and feasibility of a survey?
How will the data be analyzed?

- How will the data be disaggregated? (e.g., Do you want individual teachers to have their own reports?)
- What level of analysis do you want (i.e., school, teacher, or student level)?
- Will survey results be linked to other data (e.g., teacher observation scores, disciplinary data, student grades)?
What Next?

- Provide results in a timely manner.
- Present results in a way that is meaningful to teachers.
  - What do the scores on each dimension actually mean?
  - Anchor results in a meaningful way.
- Use results to determine areas of strength and areas to improve.
- Connect results to professional development experiences.
- Connect results to other measures of teacher effectiveness.
- Integrate results into coaching conversations.
Tripod

- The Tripod survey developed by Ronald Ferguson at Harvard University is the most widely used survey within current teacher evaluation systems.
- The survey is composed of seven dimensions that connect to three overarching “tripods” of instruction: content knowledge, pedagogical skills, and relationships.

Source: Bill & Melinda Gates Foundation (2012)
<table>
<thead>
<tr>
<th>Construct</th>
<th>Definition (as defined by test developer)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caring</td>
<td>Caring about students (encouragement and support)</td>
</tr>
<tr>
<td>Captivating</td>
<td>Captivating students (learning seems interesting and relevant)</td>
</tr>
<tr>
<td>Conferring</td>
<td>Conferring with students (students sense their ideas are respected)</td>
</tr>
<tr>
<td>Controlling</td>
<td>Controlling behavior (culture of cooperation and peer support)</td>
</tr>
<tr>
<td>Clarifying</td>
<td>Clarifying lessons (success seems feasible)</td>
</tr>
<tr>
<td>Challenging</td>
<td>Challenging students (press for effort, perseverance, and rigor)</td>
</tr>
<tr>
<td>Consolidating</td>
<td>Consolidating knowledge (ideas get connected and integrated)</td>
</tr>
</tbody>
</table>

*Source: Bill & Melinda Gates Foundation (2012)*
My Student Survey

- Focuses on six roles teachers take during instruction.
- Validated as part of Georgia’s Race to the Top initiative

<table>
<thead>
<tr>
<th>Construct</th>
<th>Definition (as defined by test developer)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presenter</td>
<td>Ability to present information and structure lessons</td>
</tr>
<tr>
<td>Manager</td>
<td>Ability to manage a classroom and foster productivity</td>
</tr>
<tr>
<td>Counselor</td>
<td>Awareness of student needs and teacher-student relations</td>
</tr>
<tr>
<td>Coach</td>
<td>Providing feedback and challenging students</td>
</tr>
<tr>
<td>Motivational Speaker</td>
<td>Engaging and investing students in learning</td>
</tr>
<tr>
<td>Content Expert</td>
<td>Knowledge of subject and encouraging student thinking</td>
</tr>
</tbody>
</table>

Source: Balch (2012)
Weighting and Combining Measures
Keep in Mind . . .

- Any rating decision rests on human judgment at some point in the process.
- The amount of evidence required to justify a rating is related to the import of the decision.
  - Decisions about professional development (less evidence)
  - Granting tenure (more evidence)
  - Dismissal (most evidence)
Principles for Rating Performance

- Consider trends and patterns of evidence, not just point-in-time data.
- Use evidence to inform decisions.
- Think in terms of “consistent and credible” evidence rather than “valid and reliable” data.
Common Rating Approaches

- Numerical Approach
- Profile Approach
- Holistic Rating Approach
Numerical Approach: Teacher Evaluation Example

- Identify weight associated with each measure.
- Assign points to each measure and add or average together.
- Create and apply score ranges for each summative rating.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Indiv. Score</th>
<th>Weight</th>
<th>Final Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom observations</td>
<td>88%</td>
<td>25%</td>
<td>0.22</td>
</tr>
<tr>
<td>Professional goal setting</td>
<td>90%</td>
<td>10%</td>
<td>0.09</td>
</tr>
<tr>
<td>Professionalism</td>
<td>76%</td>
<td>15%</td>
<td>0.11</td>
</tr>
<tr>
<td>Student growth</td>
<td>84%</td>
<td>50%</td>
<td>0.42</td>
</tr>
</tbody>
</table>

**Summative teacher effectiveness score** 0.84

<table>
<thead>
<tr>
<th>Does Not Meet Standards</th>
<th>Partially Meets Standards</th>
<th>Meets Standards</th>
<th>Exceeds Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0–0.19</td>
<td>0.20–0.54</td>
<td>0.55–0.89</td>
<td>0.90–1.0</td>
</tr>
</tbody>
</table>
Profile Approach

- Gather and maintain evidence for multiple measures and rate educators separately on each measure.
- Combine results from disparate measures using a matrix, lookup table, or series of decision rules.

<table>
<thead>
<tr>
<th>Summative Student Growth Rating</th>
<th>Summative Professional Practice and Responsibility Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Distinguished</td>
</tr>
<tr>
<td>4</td>
<td>Highly effective</td>
</tr>
<tr>
<td>3</td>
<td>Highly effective</td>
</tr>
<tr>
<td>2</td>
<td>Effective</td>
</tr>
<tr>
<td>1</td>
<td>Minimally effective</td>
</tr>
</tbody>
</table>
Holistic Rating Approach

Review the body of collected evidence and interpret it using the performance rubric to issue a single holistic rating for the educator.

- Evidence and Other Factors
  - Teacher’s or Principal’s background and experience
  - Evaluation evidence
  - Local context, district priorities

- Evaluator Judgment

- Effectiveness Score or Rating

- Discussion With Teacher/Principal
Most Systems Use a Hybrid Approach

- Balances strengths and weaknesses of each pure approach.
- Incorporates stakeholder input and local context.
- Acknowledges the multiple levels of decision-making in rating performance.
- Breaks down the system into more easily communicated components.
Optional Implementation Rules

Minimum Competence Thresholds

- Create decision rules related to minimum standards for some or all performance criteria that supersede other rules.
- Apply these rules to all or some educators (e.g., veteran, those nearing tenure).

Proficiency Progression

- Choose the performance criteria that are most critical for proficiency in the first year or phase.
- Increase minimum requirements year by year until desired proficiency standards are met.
Designing a Rating System and Setting Cut Scores

Considerations

- Where you set the bar will affect an educator’s final rating and the distribution of scores.
- Using model performance data can help predict outcomes.
- Ensure that technical and policy needs and priorities are taken into account.
- Ensure that the components and the overall system are valid.
Teacher Evaluation to Support Professional Learning
Shifting Perspectives: Integrating Evaluation and Professional Learning

Integrating Performance Evaluation into Professional Learning for Educators can mean:

**Professional Learning Opportunities**
- Practice Frameworks
- Instructional Coaching
- Using Data

**Evaluation Process**
- Collecting and Analyzing Data
- Reflecting
- Giving and Receiving Feedback
- Goal Setting

**Professional Learning & Growth**
Shifting Perspectives: Integrating Evaluation and Professional Learning

- Evaluation “training” as a one-shot, one-time event is insufficient, unsustainable, and a poor investment.
- Consider: What do you gain by investing in preparing educators for evaluation as part of your broader state and/or district professional learning system?
More Trust, More Buy-in

Relevant, hands-on professional learning opportunities can increase teachers’ and leaders’ trust in and support for the evaluation system.
Better Data

Comprehensive, high-quality training for evaluators helps ensure the data collected are:

- Fair
- Defensible
- Accurate
- Useful
Better Feedback, Better Outcomes

Relevant, hands-on learning opportunities:

- Improve the **usefulness** and **accuracy** of feedback.
- Ensure that **coaching** and **supports** are offered.
- Prepare and support teachers and leaders to **take the lead** in their own professional growth.
Professional learning opportunities related to performance evaluation are integral to the long-term improvement and sustainability of the evaluation system itself.
Most important!

When educator preparation for evaluation is integrated with **professional learning systems**, it becomes a critical, reinforcing step in building educators’ capacity to deliver high-quality leadership and instruction.
High-Quality Professional Learning
What Is Effective Professional Learning?

Focused  Active  Collaborative

Ongoing, Embedded, and Differentiated

Source: Archibald, Coggshall, Croft, and Goe (2011)
Opportunities for Professional Learning

Discuss with your team:

What percentage of time each day are your teachers engaged in high-quality, effective professional learning?
Opportunities for Professional Learning in Teacher Evaluation
Evaluation Cycle

Step 1: Self-Assessment and Reflection

Step 2: Goal Setting

Step 3: Preobservation Conference*

Step 4: Observation

Step 5: Postobservation Conference

Step 6: Summative Conference
### Potential Evaluation Data Sources

<table>
<thead>
<tr>
<th>Data Sources</th>
<th>Evidence/Data Element Examples</th>
<th>Summary Data</th>
</tr>
</thead>
</table>
| Rubric-based observations of practice | ▪ Percentage of students on task  
▪ Number of higher-order questions  
▪ Narrative descriptions, running records | ▪ Practice ratings               |
| Artifacts                             | ▪ Unit plans  
▪ Classroom newsletter  
▪ Student behavior plan  
▪ Team action-planning protocol | ▪ Professionalism/practice ratings |
| Assessments of student learning       | ▪ Student work portfolio  
▪ Standardized tests  
▪ Student performance assessments | ▪ Student learning objectives  
▪ Value-added measure scores         |
| Student/parent perception surveys     | ▪ Mean standard scores                                                | ▪ Perception scores                |
Using Evaluation Data for Self-Reflection and Goal Setting

Center on GREAT TEACHERS & LEADERS at American Institutes for Research
Teacher Self-Reflection and Goal Setting

In this step of the evaluation data use cycle, teachers take charge of their own growth in the following ways:

• Analyzing the impact of their practice on student learning
• Engaging in reflection on their practice
• Setting focused professional and student learning goals, with concrete steps to get there
• Actively collaborating with colleagues to problem solve
• Adjusting their plans as a result of this reflection
Using Evaluation Data for Formative Feedback
Formative Evaluation: Pre- and Post-Observation Conferences

Goals
- Provide authentic, specific, and evidence-based feedback on practice.
- Encourage teacher self-reflection on practice.
- Identify professional learning opportunities.

Activities
- Collaborative conversation(s) between the teacher and evaluator to share evidence, ask questions, and provide feedback
- Adjustment to goals and plans
- Connection to resources and supports
Translating Data Into Feedback

- Evaluators can translate evidence about teacher practice into actionable feedback to teachers.
- Such feedback, when delivered effectively, will result in productive conversations and greater likelihood that professional learning will occur.
Characteristics of High-Quality Feedback

Teachers learn best when feedback is:

- Tied to specific teaching standards.
- Specific, detailed, and evidence based.
- Timely and frequent.
- Constructive, with effective use of questioning.
Using Evaluation Data: Example of a Feedback Conversation

Observing Instruction to Build Capacity
Waterford High School, California

Carolyn Viss, chair of the mathematics department at Waterford High School, conducts a coaching session based on classroom observation of a core algebra class. She acknowledges good instructional practices and offers solutions for specific dilemmas. (5:18 minutes)

Discussion: To what extent did this feedback session promote learning? What data or evidence did Ms. Viss refer to in her questions?

Video courtesy of the U.S. Department of Education’s former Doing What Works initiative.

Download video at: https://vimeo.com/84709443. The password is DWWVideo.
Using Summative Evaluation Data for Individual and Schoolwide Planning
Summative Evaluation

Goals

- Summarize evaluation data for individual teachers.
- Identify patterns in teacher and student performance across the school to inform the allocation of resources.
- Identify professional learning and career opportunities.

Activities

- Analysis of individual teacher evaluation data to determine final rating(s)
- Conversations between the teacher and evaluator about performance trends and patterns
- Connection of resources, supports, and opportunities for the next evaluation cycle
Activity: Using Data to Plan for Differentiated Professional Learning

- Read **Handout 4**, focusing on the scenario with Ms. Blue and her summative evaluation plan.
- Answer the guiding questions in Handout 4.
- Discuss your recommendations for a professional learning plan for Ms. Blue.
- Include the following elements in your plan:
  - Professional learning goals
  - Professional learning activities
  - Success metrics
Activity: Using Data to Plan for Schoolwide Professional Learning

- Read **Handout 5**.
- As a table group, read the summary ratings data for teachers at Orange School.
- Discuss and share:
  - What patterns do you see in the data?
  - What are Orange School’s greatest areas of need?
  - Based on these data, what professional learning activities would you include in a schoolwide professional learning plan?
Principles of Effective Planning

Using Evaluation Data

- Teacher evaluation data help ensure better allocation of resources, including teacher time.
- Human judgment is an inevitable—and critical—piece.
- To fill in the gaps around the data, those closest to the work of teaching and learning should be included in the planning.
- Necessary conditions for teacher learning must exist.
- Professional learning is high stakes!
Group Activity
Group Activity

In your table groups:

- List the **main challenges** associated with using evaluation data to inform professional learning and providing high-quality professional learning opportunities.

- Think about specific challenges at the **district and school** levels.

- Write a different challenge on each self-adhesive note.

- We will collect the self-adhesive notes and facilitate a large-group discussion.
Essential Conditions for Effective Professional Learning

Structural?
Cultural?
Technical?
Fiscal?
Essential Conditions Definitions

- **Structural:** The policies, programs, practices, and structures that need to be in place
- **Culture:** The necessary climate or attitude, including changes in culture, to be successful
- **Technical:** The technical aspects—skill, expertise, content knowledge—that are necessary to get this work done
- **Fiscal:** The financial supports necessary to do this work or the costs associated with doing this work
### Additional Resources

<table>
<thead>
<tr>
<th>What</th>
<th>Resource Topic or Title</th>
<th>Where It’s Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summative Evaluation</td>
<td>Observation Data Collection Tool</td>
<td><a href="http://bloomboard.com/schools">http://bloomboard.com/schools</a></td>
</tr>
<tr>
<td>What</td>
<td>Resource Topic or Title</td>
<td>Where It’s Available</td>
</tr>
<tr>
<td>------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>---------------------------------------------</td>
</tr>
</tbody>
</table>
References


Advancing state efforts to grow, respect, and retain great teachers and leaders for all students