



Preparing Educators for Evaluation and Feedback

Planning for Professional Learning

Lynn Holdheide

Senior Technical Assistance Consultant

June 2014

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





Welcome and Introductions

Module Overview

- **Module Goals:** at the end of the day, you should...
 - Be able to identify a high-quality evaluation training plan and understand how training is integral to a system of instructional improvement.
 - Identify evaluation training approaches for different state contexts and for all educators impacted by the evaluation system.
 - Recognize the critical role of assessing and monitoring evaluators' skills to ensure validity of evaluation results.
 - Explain what makes an *evaluator* training process high-quality and helps administrators develop strong skills in providing feedback.
 - Consider next steps for communicating about an evaluation training approach that is appropriate for your state or district context.

Agenda

- Welcome and Introductions
- The Big Picture: Developing a Comprehensive *Evaluation* Training Plan
- Characteristics of High-Quality Evaluator Training
- Completing Your Comprehensive Evaluation Training Plan

Activity: Confidence Statements (Handout 1)

How confident are you that...

1. Educators in our state have a solid understanding of the state and district requirements and processes (e.g., measures, timelines, documentation) for educator evaluation?
2. Educators in our state have access to strong professional learning opportunities about the new evaluation system and can implement their role successfully?
3. Evaluation data collected in the new system is reliable, accurate, and useful for providing high quality feedback?

After discussing, place a sticky note representing your level of confidence on the 10-point scale for each question on the chart paper.

Debrief

- For each question, please share out: What specifically gives you confidence? Or what undermines your confidence?
- On a sticky note, write the one question you have when you hear the term “evaluation training.”



The Big Picture: Developing a Comprehensive *Evaluation* Training Plan

The Big Picture:

“It’s About More Than Evaluators”

There is a tendency to overlook:

- Educators (teachers, principals, assistant principals) *being* evaluated
- Staff (central office, HR managers, administrative staff, information technology staff) supporting evaluators

Fully preparing educators requires considering:

- Who is involved in evaluations and in what role?
- What knowledge, supports, and opportunities will people in each role need?

Comprehensive Evaluation Preparation

Training Plan Design Decisions

1

**Roles and
Responsibilities**

2

**Audiences, Format,
and Content**

3

Timelines

Communication

Assessing Effectiveness

6

Sustainability

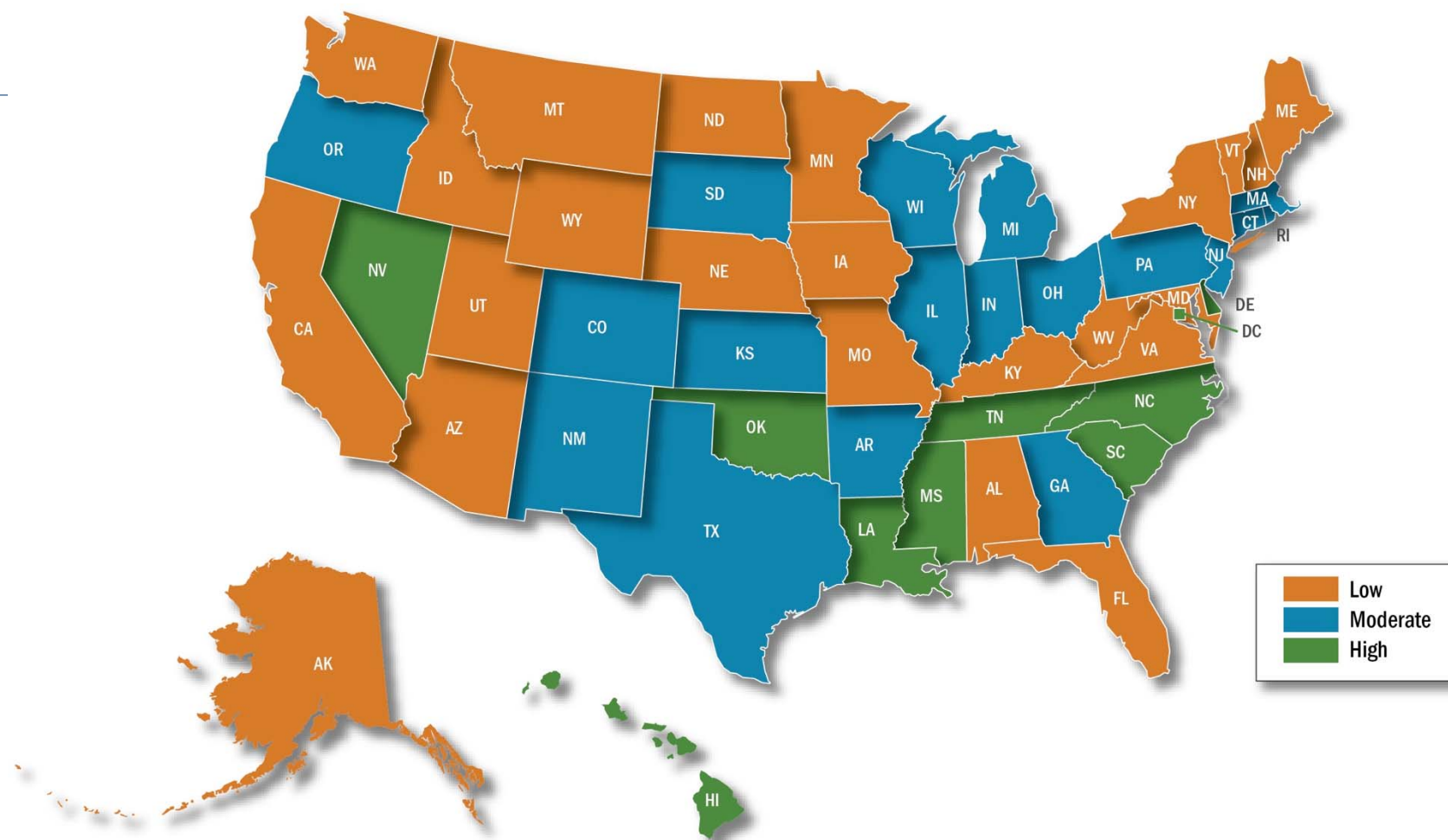
Roles and Responsibilities

- **Regulatory Framework:** What do your state's laws and regulatory guidance on evaluation training require from different actors:
 - State education agency (SEA)?
 - Regional service areas?
 - Districts?
 - Schools?

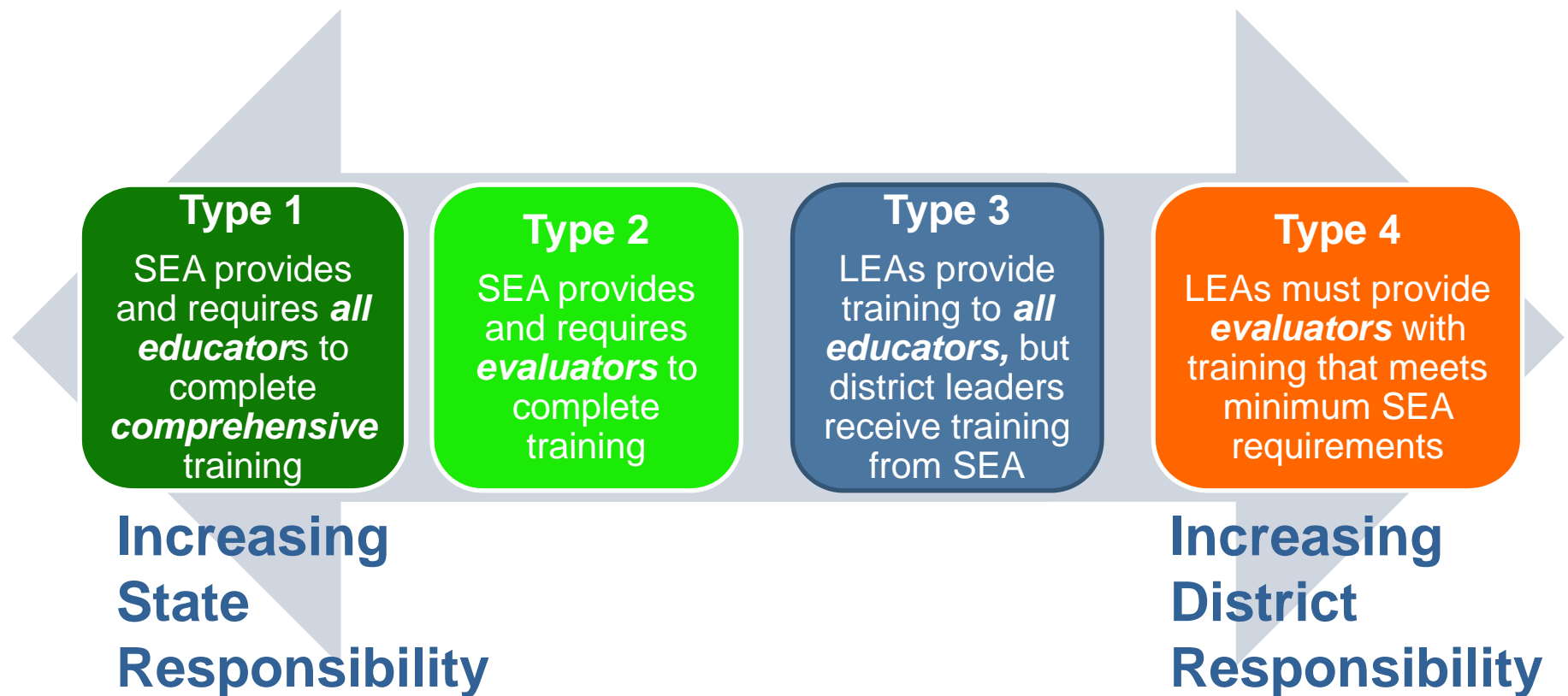
Context: Level of State Control

State-Level Evaluation System (High)	Elective State-Level Evaluation System (Medium)	District Evaluation System With Required Parameters (Low)
<ul style="list-style-type: none"> Determines the components, measures, frequency, and types of evaluators. All districts must implement the state model with little flexibility. 	<ul style="list-style-type: none"> Mandates student growth measures, models, and weights; but leaves observation measures and other protocols up to local education agencies (LEAs). Offers state model but allows districts to choose alternatives if they meet state criteria. 	<ul style="list-style-type: none"> Provides general guidance, requires certain components (observations), and may use an approval process; but allows LEAs wide latitude in selecting components and creating the system.

State Control Over Evaluation Systems



State Versus District Roles: What's Your Context?



Audiences: Types of Training

SEA-Provided Training

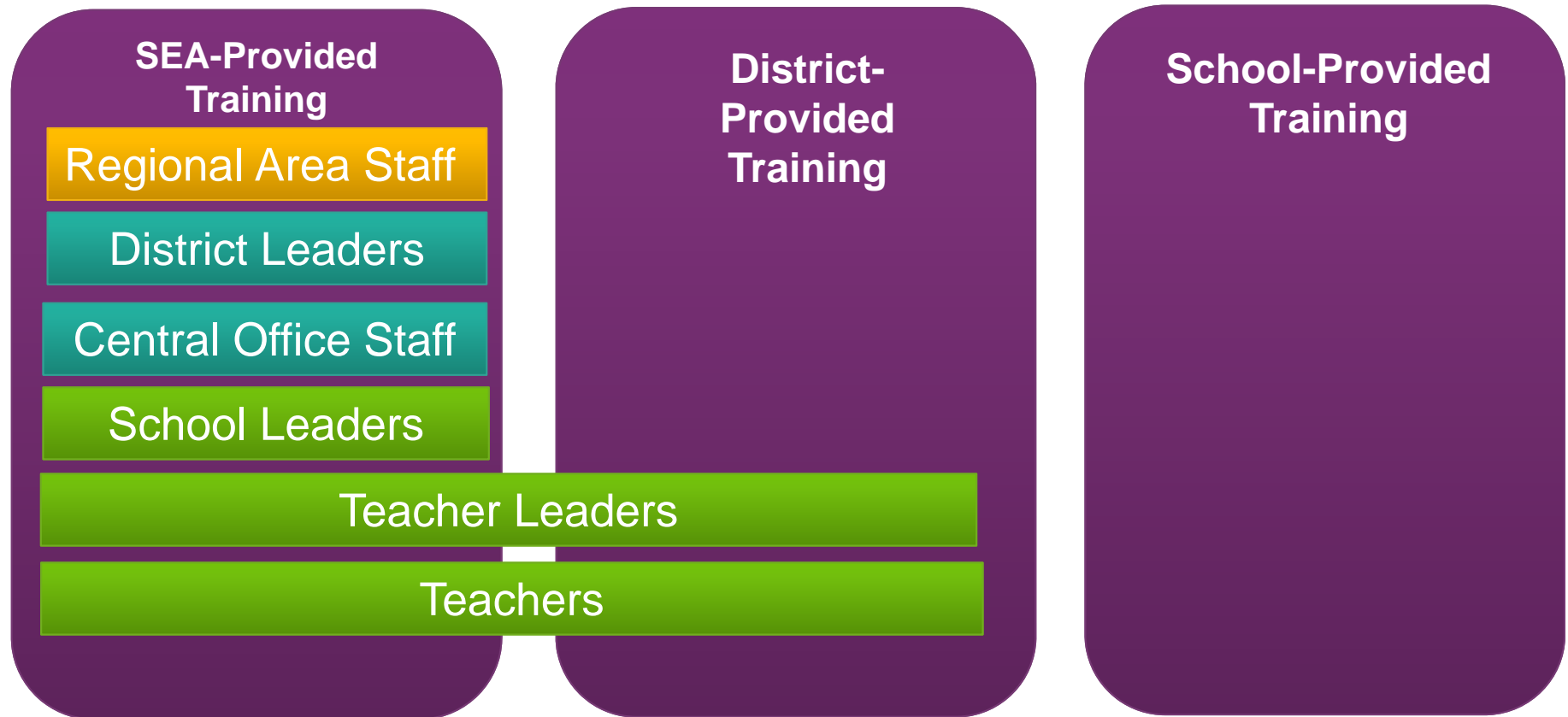
- District leadership training
- Educator orientation (principals and teachers)
- Evaluator training (superintendents, principals, vice principals, peer evaluators)

District-Provided Training

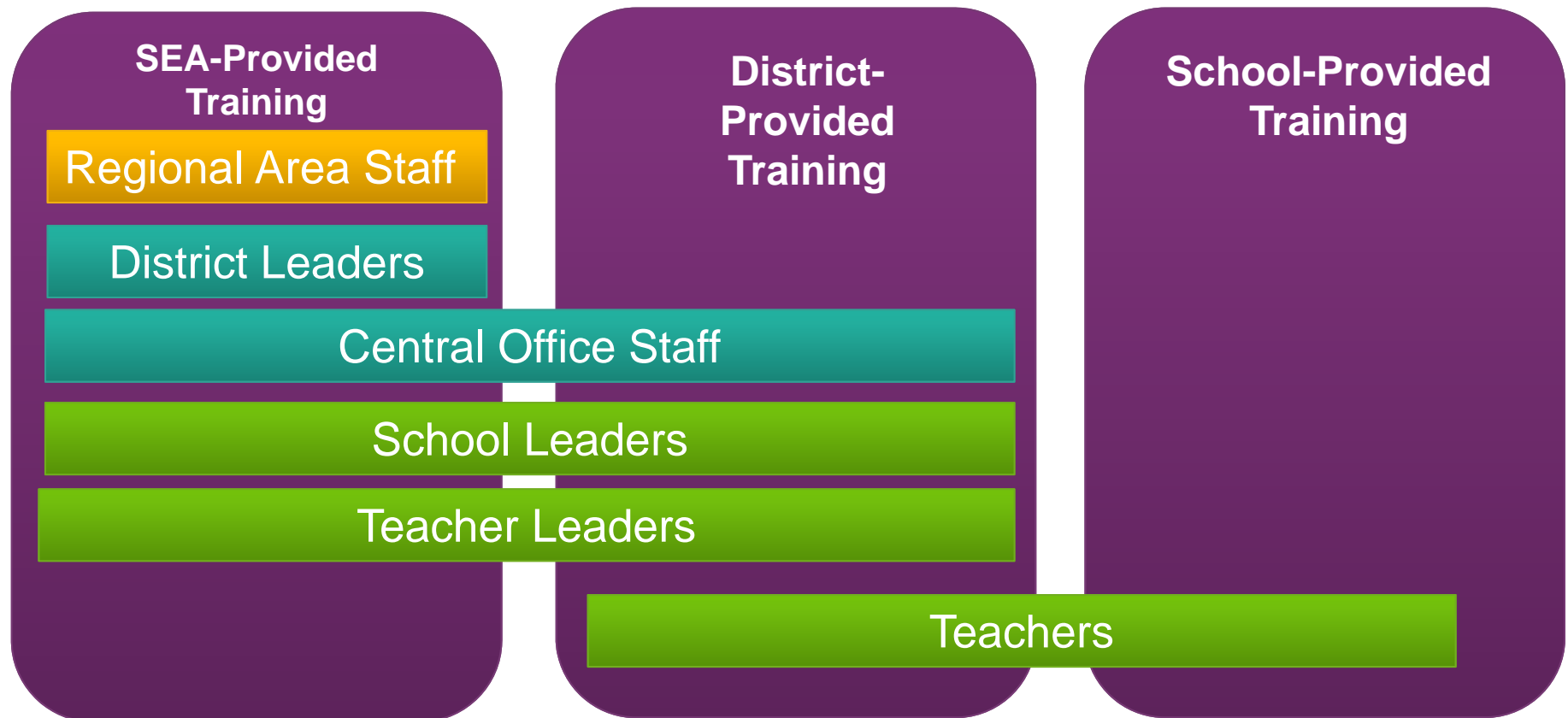
- School leadership team training
- Central office training
- Educator orientation
- Evaluator training



High-Control State

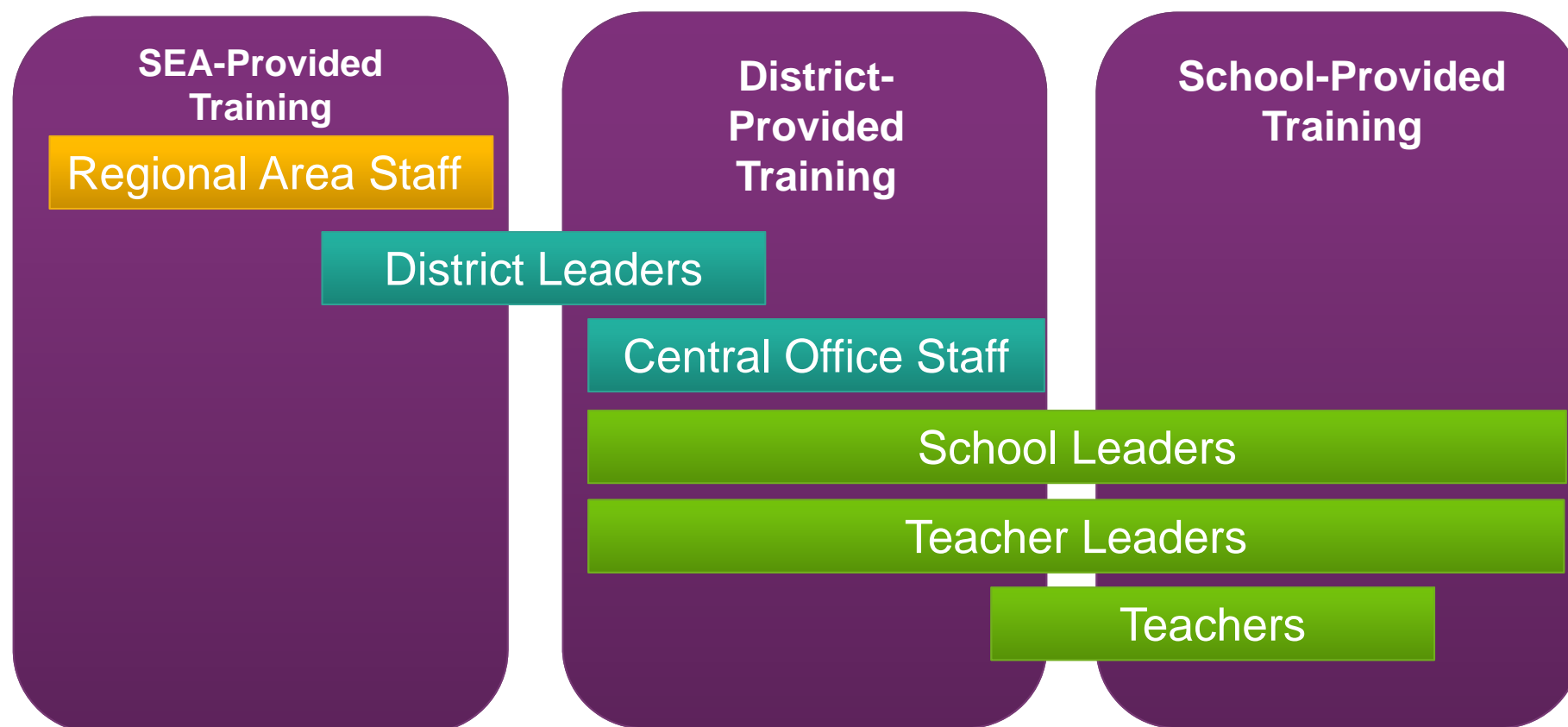


Medium-Control State





Low-Control State





Formats

State Level

- In-person, train-the-trainer
- Online, self-directed
- Online, webinar or video
- Symposia or periodic conferences

District or School Level

- In-person, facilitated
- Subject or grade-level teams or professional learning communities
- Hybrid: partially online, partially in-person



Activity: Identifying Your Roles

Activity:

Identifying Your State's Roles

Use **Handout 2: Roles, Responsibilities, and Resources** and complete Steps 1–3 at your table.

- **Step 1.** In the **green columns**, place an “S” in any box the state will take responsibility for and a “D” in any box that districts will take responsibility for.
- **Step 2.** In the **purple columns**, list existing resources from the state that can support both SEAs and districts.
- **Step 3.** Prioritize the list of SEA-identified roles for your state by considering:
 - Which roles will be the greatest challenge for your SEA? Why?
 - In which roles will districts need the most support? Why?

Debrief

Option 1

Each table “conferences” with another table. Together, they compare the state and district roles they selected as well as and their prioritization. The two tables must produce a single, consolidated table that represents the group’s consensus.

Option 2

Each state team presents its state roles and prioritization to the whole group, in turn.



Characteristics of High-Quality Evaluator Training

Activity: Quick Jot

- Within the next two minutes, work with a partner to generate a list of the characteristics of high-quality evaluator training.

High-Quality Evaluator Training is...

- Comprehensive
- In-depth
- Concrete
- Hands-on
- Assessed
- Continuous

Comprehensive Training Covers...

**Observing
educators'
practice**

**Analyzing
non-
observation
evidence**

**Facilitating
observation
conferences**

**Coaching
educators
and
providing
feedback**

**Understanding
and analyzing
student
growth data**

**Combining
measures
for
summative
scoring**

**Guiding
creation of
professional
development
plans**

**Managing
time and
technology
to be
efficient**

Critical Training Elements

High-Value Resources

- Examples of practice (artifacts, videos of classroom instruction, sample data)
- Master scored videos, artifacts, and exemplars

Crucial Processes

- Master scoring process
- Opportunities to practice data collection and scoring, and to receive immediate feedback
- Assessing and certifying observers
- Calibration monitoring and support

What's Our Goal? Interrater Reliability Versus Rater 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 consistency in judgments over time, in different contexts and for different educators.

What Is Interrater Reliability?

Do Raters A and B demonstrate interrater reliability?

Teacher	Component Score	
	<i>Rater A</i>	<i>Rater B</i>
Teacher A	1	2
Teacher B	2	3
Teacher C	3	4
Teacher D	4	5

The diagram illustrates interrater reliability by showing the scores assigned by two raters, Rater A and Rater B, to four teachers. Rater A's scores are 1, 2, 3, and 4, while Rater B's scores are 2, 3, 4, and 5. The scores for each rater are circled, and a bracket indicates a +1 difference between the first and last scores for each rater, suggesting a consistent rating pattern.

Illustrating Rater Agreement

Component	Component Score			Type of Agreement
	<i>Rater A</i>	<i>Rater B</i>	<i>Master Scorer</i>	
1	4	4	4	Exact Agreement
2	3	2	3	Adjacent Agreement
3	1	4	4	?
4	3	3	1	?

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.

Rater Agreement: Key to Impacting Practice

“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.”

“For teacher evaluation policy to be successful, it will have to be implemented in such a way that a common language and understanding of teaching is fostered.... Observers will be more likely to score reliably, and teachers will have views of their own instruction that are more consistent with those of external observers.”

(Gitomer et al., in press)

Calculating Agreement

Component	Component Score			More Than One Point Off
	<i>Rater A</i>	<i>Rater B</i>	<i>Master Scorer</i>	
Subcomponent 1	4	3	4	No
Subcomponent 2	2	1	3	Yes
Subcomponent 3	1	3	4	Yes
Subcomponent 4	4	3	1	Yes
Subcomponent 5	3	4	2	Yes
Component Score (Average)	2.8	2.8	2.8	

Evaluator Skills That Promote Rater Reliability and Agreement

- **Objectivity:** Records evidence that is free of bias, opinion, and subjectivity.
- **Alignment:** Correctly aligns evidence to framework criteria that reflect the context of the evidence.
- **Representation:** Records a preponderance of evidence for scoring criteria; accurately reports the classroom and artifact data.
- **Accuracy:** Assigns numerical scores similar to the scores that master observers assign.

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.

The higher the stakes, the higher the need for strong interrater agreement and reliability.

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



Evaluator Skills that Promote Rater Reliability and Agreement

- **Objectivity:** records evidence that is free of ‘bias, opinion, and subjectivity’
- **Alignment:** correctly aligns evidence to framework criteria that reflect the context of the evidence
- **Representation:** records a preponderance of evidence for scoring criteria and accurately report the classroom and artifact data.
- **Accuracy:** assigns numerical scores similar to the scores master observers assign.

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?

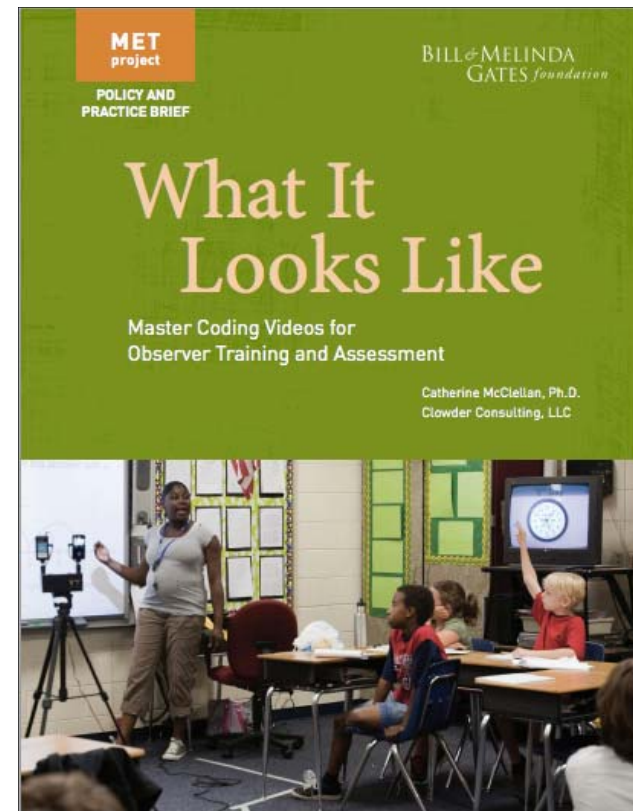
- **Supplemental training** on hard to score sections, for example:
 - Learning to focus on student responses
 - Weighing competing evidence
 - Understanding what a specific element looks like in classrooms

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

Master Scoring Process

- 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, p. 2)



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 (McClellan, 2013).

Concrete Examples: Types of Master-Coded Videos to Support Observations

Video Type	Purpose in Training	What Video Shows	Length
Benchmark	Clarifies each performance level	Clear examples	Two to seven minutes
Rangefinder	Clarifies boundaries between adjacent performance levels	High and low examples within levels (“a high 3 and a low 4”)	Two to seven minutes
Practice	Provides opportunity to observe, score, and receive feedback	Fairly clear-cut instances of most or all aspects of practice	20 to 30 minutes
Assessment	Helps determine whether observers have attained sufficient accuracy	Fairly clear-cut instances of most or all teaching components	20 to 30 minutes

American Federation of Teachers (AFT) i3 Master Scoring Process

- Part of the AFTs' evaluator training for the Investing in Innovation (i3) grant.
- Held two three-day master coding “boot camps” for about 80 observers from Rhode Island and New York.
- Trained master coders to work on an ongoing basis to code master videos (McClellan, 2013).

Master Coding Explained



<https://vimeo.com/96864796>

Master Coding in Action



<https://vimeo.com/96869509>

Discussion

- What seemed valuable to you about using a master coding process?
- What seemed challenging or raised a concern for you?
- What questions do you have about developing a master coding process in your own state or district context?

Comprehensive Evaluation Preparation

Training Plan Design Questions

1

**Roles and
Responsibilities**

2

**Audiences, Format,
and Content**

3

Timelines

Communication

Assessing Effectiveness

6

Sustainability



Setting Timelines

Considerations:

- **Requirements:** Does your state have internal or federally mandated timelines? Can you back-map your training timelines to ensure that districts can meet the requirements?
- **Cumulative:** Are your timelines created to build educator capacity at an appropriate pace and over time?
- **Staggered:** Are you focusing first on building district leadership team capacity and then moving to educator and evaluator training?

Activity: Examining Example Training Timelines

3

- **See Handout 3: Examples of Training Timelines and Plans.**
 - Wisconsin
 - Arkansas
 - Colorado
- **At your table, examine the timelines:**
 - Are they cumulative? If so, how?
 - Are they staggered? If so, how?
 - How would you strengthen or improve the timelines?
 - What elements of these timeline examples can inform your own planning?

Communication: Transparency and Feedback Loops



1. Be proactive and transparent: Communicate your training plan to teachers, principals, parents, and community stakeholders.



2. Use training as an opportunity to communicate with educators about the overarching goals and purposes of the system.



3. Use training as an opportunity to gather feedback about the evaluation system, materials, and processes.



References

Gitomer, D. H., Bell, C. A., Qi, Y., McCaffrey, D. F., Hamre, B. K., & Pianta, R. C. (in press). The instructional challenge in improving teaching quality: Lessons from a classroom observation protocol. *Teachers College Record*.

McClellan, C. (2013). *What it looks like: Master coding videos for observer training and assessment* (MET Project Policy and Practice Brief). Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from http://www.metproject.org/downloads/MET_Master_Coding_Brief.pdf

Lynn Holdheide
615-308-2525
lholdheide@air.org

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

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

Center on
GREAT TEACHERS & LEADERS

at American Institutes for Research ■

