Understanding the Evaluator Calibration Report (ECR)

Below is a description of each section of the Evaluator Calibration Report is described in detail.

1. BIAS/OBJECTIVITY

Evaluators must consistently demonstrate objectivity and a lack of bias in evidence collection.

Objective evaluators provide descriptions of evidence that are completely free of bias, opinions, and subjective judgments. Descriptive statements do not summarize or offer opinions; evidence may include specific numbers and/or time references (e.g. "The teacher provided three discussion questions for each small group") or direct discourse: T- "How are you today?" S - "Good, great." Objective evaluators report what they see and hear without making judgments.

If you didn't score effective or higher on the bias/objectivity section of the calibration examination, it is likely that your descriptions of evidence demonstrated bias or subjective judgment. When a skilled evaluator offers evidence, it is important to understand that accurate, unbiased descriptions answer such questions as:

- **What did the teacher do?** (Not, How did the teacher behave?)
  Biased: The teacher was friendly to everyone.
  Unbiased: The teacher called on every student by name.

- **What did the teacher (or student) say?** (Not, How did the teacher express herself?)
  Biased: The teacher impatiently and anxiously told the student to get their things.
  Unbiased: The teacher said "Hurry up, we're running out of time. Get your things and come sit on the carpet."

- **How many students performed what action for how long?** (Not, How would you characterize the behavior of students?)
  Biased: Students were uncontrolled, rude, and acting like hooligans.
  Unbiased: Students frequently interrupted one another, and six students were out of their seats for more than 10 minutes.

- **What was the appearance of the environment?** (Not, How would you judge the condition of the classroom?
  Biased: The teacher's classroom was really messy.
  Unbiased: Desks were arranged in groups of four, and the aisles were free from obstructions.

How can skills in the bias/objectivity category be improved?

Remediating a tendency toward bias in descriptions of evidence is both a linguistic and a judgment issue. One way to quickly check one's descriptions is to scan for the use of adverbs ending in "-ly"— these are almost always an indication that some measure of "judgment" is in play (e.g. harshly, angrily, hardly, apparently, clearly).

It's important to avoid trying to *interpret* what the teacher is doing
Biased: The teacher tried to engage the students with the short story.
Unbiased: The teacher read a short story aloud; students listened.

Be cautious to avoid ascribing meaning to a teacher's actions.
Biased: The teacher's compassion was apparent.
Unbiased: The teacher and students discussed two poems about death and bereavement. The teacher said "the poem made her feel sad."

Discussions of bias are important ways of building objectivity in this category. Practice evidence collection with colleagues and compare notes.

2. ALIGNMENT

Evaluators must be able to consistently align evidence to appropriate performance indicators and standards.

Evidence must be both aligned to the teaching standards and accurately reflect the context of the evidence. Evidence must be matched with the correct standard and performance indicator.

If you didn't score effective or higher on the "alignment" section of the calibration examination, you were unable to "match" the evidence you collected to the most appropriate elements of the standard or performance indicator. When a successful observer aligns evidence, he/she is able to recognize the standard or performance indicators that the evidence supports.

For example, evidence such as "In the discussion of Title IX, the teacher called on every student by name, and alternately selected males and females to answer questions," might be aligned with Standard 4's performance indicator, 4.a. "Interactions with students," but considering the context, the evidence is most appropriately aligned with the next performance indicator, 4.b. "Supports student diversity."

A consideration of context is particularly important when evidence may be attributed to more than one performance indicator or standard. Generally the same evidence can be used for up to two performance indicators.

How can you improve your skills in the "alignment" category?

Observers will improve their ability to align evidence accurately by developing a strong familiarity with the Teacher Practice rubric, and apply this rubric in many different contexts.

3. ACCURACY

There are two criteria in the accuracy category; (1) exact agreement and (2) score differential.

Exact Agreement
Exact agreement indicates the percentage of the observer's responses that exactly match the master scores. Over time, increasing the percentage of responses that exactly match master scores works to ensure inter-rater reliability. Agreement at or above 60% indicates a high level of accuracy.

Score Differential
Score differential averages the overall distance between the observer's scores and the master scores.

When a successful evaluator scores evidence, he/she consistently evaluates it as "ineffective" "developing" "effective" or "highly effective". Scores that fall between +/-.4 indicate a high level of accuracy.

Low agreement scores/high differential scores indicate you were unable to accurately score evidence according to the rubric.

If your scores vary in a (+) positive direction, the values you have assigned to evidence tend to exceed master score values indicating inflated scores; negative (-) score differentials indicate that the values you have assigned to evidence tend to be lower than the master score values. This often happens when evaluators "chunk" evidence without discriminating for performance indicators.

How can skills in the "accuracy" category be improved?

Over time, observers improve accuracy with practice. In-depth discussions of the meanings and expectations attached to rating categories (such as "developing" or "highly effective") can clarify areas of disagreement. Paired observation practice can significantly improve the performance of each observer in this category and lead to greater inter-rater agreement (consistency across all scorers across the district). Providing rationale for why a score is not higher (what is missing) or lower (what evidence was present) can ensure accuracy and provide more productive feedback.