**Reviewing Professional Practice Calibration Framework**

**Utilizing an Observation Protocol can have many purposes**

* ensure consistent and uniform scoring of teacher practice during observations within and across schools and districts
* developing common language and shared expectations
* supporting educators through high quality feedback
* peer observations as a form of professional development, and building a culture of professional learning community within our schools

**Why it’s important to continually calibrate**

Personnel evaluating teachers in all models participated in training and calibration of observations leading up to and throughout the first year of full implementation. Continual calibration is critical as evaluators conduct more and more observations, as personnel evaluating teachers change within schools and districts, and as drift naturally occurs over time. To calibrate observations of Professional Practice a variety of sessions could be utilized. We suggest calibrating multiple times a year at the school level and at least once a year at an LEA level.

**Two levels of Calibration**

**LEA Level:**

1. All personnel evaluating teachers’ practice watch a video of classroom instruction and utilize Protocol 1.
2. All personnel evaluating teachers’ practice observe a teacher’s practice live and utilize Protocol 2.

\*Note: these could be completed in grade-span groups (Elementary, Secondary or Elementary, Middle, High School) but it is also beneficial to see multiple grade spans

**School Level:**

1. All personnel evaluating teachers’ practice watch a video of classroom instruction and utilize Protocol 1.
2. All personnel evaluating teacher’s practice (or a team of 2-3) observe a teachers’ practice live at their school and utilize Protocol 2.
3. All personnel evaluating teachers’ practice (or a team of 2-3) observe a teachers’ practice live at a different school and utilize Protocol 2.

****

**Additional Opportunities for using Calibration for Professional Development of Teachers**

**Calibrating with Teachers:**

Teachers can also be included in calibration of observations to support their understanding of high-quality instruction and use of the rubric in their evaluation. This can be powerful professional development and can be structured in a variety of ways including, but not limited to the following:

1. *Teachers in a school within a grade-level team or content area:* Grade level teams or departments within a school can utilize either of the two protocols. All RI Model districts have an FFTPS account they can utilize for PD with teachers and many other websites have video libraries of teaching.
2. *Teachers in a content area across a district:* Teachers could see what teaching looks like at other schools and grade levels. Additionally, if a school has only one or two art teachers, the arts teachers in the district could come together for a calibration session.
3. *Teachers observing different grade-levels and/or content areas:* We encourage teachers to calibrate their observation of subjects and grades outside their own, as this can oftentimes be highly beneficial.
4. *Teachers only focusing on one component in observation:* Each month teachers could focus on a different component of the rubric and observe one another in a focused way to identify a variety of successful strategies in diverse contexts. If utilizing this approach, Protocol 3 would be most helpful.

**Calibrating Professional Practice of Support Professionals and Building Administrators using holistic rubrics:**

Calibration can also take place even with the use of holistic rubrics which are not scored at the end of individual school site visits or observation of practice. If personnel evaluating Building Administrators or personnel evaluating Support Professionals wanted to calibrate their observations of Professional Practice we encourage the use of Protocol 4.

**Protocol 1: Video Observation Calibration Protocol**

1. **Identify a video** that you’d like to use.
2. **Observe the video of the teaching episode** as a group and individually record evidence (free of bias and interpretation). Each member of the group is responsible for taking notes in their preferred format (EPSS, hand-written notes, iPad, etc). The group can watch the video together or separately.
3. After the observation, individual evaluators should **independently sort evidence and score** for each component based on the evidence they collected. In addition to the rating, evaluators should be prepared to provide rationale to support their score.
4. Once each evaluator has had a chance to score independently and identify evidence, the group shoul**d share and discuss component level ratings and rationale** together.

**SCORING DEBRIEF NORMS**

* One member serves as facilitator
* Establish conversation time limits (e.g. plan to complete Domain 2 by x time)
* Hold one another accountable to bias and interpretation
* Every member shares their component score
* If the scores are the same, name why the score is that level
* If the scores are different, have a conversation regarding rationale in order to reach consensus by grounding the conversation in two questions:
  + *What does the rubric level descriptor say exactly?*
  + *Where does the preponderance of evidence fall?*
* Repeat this process for each component

1. After completing the scoring, consider the **reflection questions.**

**CONNECT, REFLECT, AND PLAN**

* What components were most challenging in reaching consensus? What caused this challenge?
* Were there instructional practices that were interpreted differently?
* How was this protocol helpful in aligning our instruction language and expectations?
* How can others experience this learning?
* Are there any significant next steps that have come from this conversation that need action steps?

**Protocol 2: In-Person Observation Calibration Protocol**

1. **Identify a teacher** willing to volunteer for an unofficial classroom visit that will include an observation scored by multiple evaluators (but will not count as an official observation for their evaluation).
2. **Observe a teaching episode** as a group and individually record evidence (free of bias and interpretation). Each member of the group is responsible for taking notes in their preferred format (EPSS, hand-written notes, iPad, etc).
3. After the observation, individual evaluators should **independently sort evidence and score** for each component based on the evidence they collected. In addition to the rating, evaluators should be prepared to provide rationale to support their score.
4. Once each evaluator has had a chance to score independently and identify evidence, the group should **reconvene to share and discuss component level ratings and rationale**.

**SCORING DEBRIEF NORMS**

* One member serves as facilitator
* Establish conversation time limits (eg plan to complete Domain 2 by x time)
* Hold one another accountable to bias and interpretation
* Every member shares their component score
* If the scores are the same, name why the score is that level
* If the scores are different, have a conversation regarding rationale in order to reach consensus by grounding the conversation in two questions:
  + *What does the rubric level descriptor say exactly?*
  + *Where does the preponderance of evidence fall?*
* Repeat this process for each component

1. After completing the scoring, consider the **reflection questions.**

**CONNECT, REFLECT, AND PLAN**

* What components were most challenging in reaching consensus? What caused this challenge?
* Were there instructional practices that were interpreted differently?
* How was this protocol helpful in aligning our instruction language and expectations?
* How can others experience this learning?
* Are there any significant next steps that have come from this conversation that need action steps?

**Protocol 3: Single Component Observation Calibration Protocol**

1. **Select the single component** to be the focus of the observation.
2. **Identify a video of teaching or a teacher** willing to volunteer for an unofficial classroom visit by other teachers (that will not count as an official observation for their evaluation).
3. **Observe a teaching episode** as a group and individually record evidence (free of bias and interpretation) that reflects the selected component. Each member of the group is responsible for taking notes in their preferred format (EPSS, hand-written notes, iPad, etc).
4. After the observation, individual observers should **independently review evidence and score** the component based on the evidence they collected. In addition to the rating, observers should be prepared to provide rationale to support their score.
5. Once each observer has had a chance to score independently and identify evidence, the group should **reconvene to share and discuss component level rating and rationale.**

\*NOTE: If desired observers can provide structured feedback to the teacher, rather than providing a component score.

**SCORING DEBRIEF NORMS**

* One member serves as facilitator
* Establish conversation time limits (eg plan to complete Domain 2 by x time)
* Hold one another accountable to bias and interpretation
* Every member shares their component score (if applicable)
* If the scores are the same, name why the score is that level
* If the scores are different, have a conversation regarding rationale in order to reach consensus by grounding the conversation in two questions:
  + *What does the rubric level descriptor say exactly?*
  + *Where does the preponderance of evidence fall?*
* Identify practices that worked well in the lesson and provide suggestions or questions for those that were less successful.

1. After completing the observation, consider the **reflection questions.**

**CONNECT, REFLECT, AND PLAN**

* What was most challenging in reaching consensus about this component? What caused this challenge?
* Were there instructional practices that were interpreted differently?
* How was this protocol helpful in aligning our instruction language and expectations?
* How can we share our experience and learning with others?

**Protocol 4: Professional Practice Calibration Protocol for Personnel Evaluating Support Professionals and/or Building Administrators**

1. **Identify a Support Professional or Building Administrator** willing to volunteer for an unofficial site visit/assessment of practice that will include an observation scored by multiple evaluators (but will not count as an official site visit/assessment of practice for their evaluation).
2. **Conduct the site visit/assessment of practice** as a group and individually record evidence (free of bias and interpretation). Each member of the group is responsible for taking notes in their preferred format (EPSS, hand-written notes, iPad, etc).
3. Afterwards, individuals should **independently sort evidence** for each component based on the evidence they collected. Since this model is holistic the focus of this exercise should be around the collection and sorting of evidence. Individuals should be prepared to provide rationale to support their distribution of evidence.
4. Once each evaluator has had a chance to sort their evidence, the group should **reconvene to share and discuss evidence alignment and rationale**. Evaluators should try to reach consensus on where evidence fits into the rubric. Evaluators can generate a formative score based on the evidence collected together and any other possible evidence that may have been collected to this point. The focus of this activity is about aligning evidence to each component and being able to discuss and differentiate between the different levels of scoring as a group. A formative score can be assigned at the end of the calibration session on the components that the group feels they have collected enough evidence to support but it is not necessary to score all components.

**SCORING DEBRIEF NORMS**

* One member serves as facilitator
* Establish conversation time limits (eg plan to complete Domain 2 by x time)
* Hold one another accountable to bias and interpretation
* Every member shares their evidence alignment and rationale
* If the formative scores are the same, name why the score is that level
* If the formative scores are different, have a conversation regarding rationale in order to reach consensus by grounding the conversation in two questions:
  + *What does the rubric level descriptor say exactly?*
  + *Where does the preponderance of evidence fall?*
* Repeat this process for each component

1. After completing the scoring, consider the **reflection questions** on the following page.

**CONNECT, REFLECT, AND PLAN**

* What components were most challenging in reaching consensus? What caused this challenge?
* Were there professional practices that were interpreted differently?
* How was this protocol helpful in aligning our language and expectations?
* How can others experience this learning?
* Are there any significant next steps that have come from this conversation that need action steps?