



## Establishing Inter-Rater Reliability

### **What would you suggest as the frequency of appraisal calibrations?**

After initial appraiser certification (which must include a calibration component) re-calibration is required at least annually for all appraisers. In addition, appraisers are required to be re-certified at minimum every three years.

### **How does inter-rater reliability work if we move to a hybrid or virtual model for next year?**

TEA will be publishing guidance on teacher observation for virtual settings in late August. Districts can use this guidance to norm on both scoring as well as coaching feedback for virtual lessons, however, if possible districts are encouraged to wait until instruction returns to the in-person format before conducting formal appraisals. (Note: this may not be possible in all situations, but is something to consider.) Additional resources for observing and scoring virtual lessons, likely to include sample scored video observations of virtual lessons, will be published later this fall.

### **How do you recommend using outside data such as student growth scores to look for evaluation score inflation?**

As part of a district's local teacher designation system, districts are required to review the congruence of teacher observation data with student growth data. The expectation is that observation and growth will be positively correlated, such that higher growth scores would be found for teachers who also receive higher observation scores. When a district reviews this data and finds instances of skew, such as a teacher with high observation scores has low student growth scores, the district is required to develop a plan to address areas of skew. This will need to include the cause of the skew. Is it explicable or inexplicable skew? Is the cause related to teacher practices, the appraiser not being properly calibrated, the student growth measure not being valid and reliable, etc.?

### **If a district has good inter-rater reliability, but administrators are collectively inflating scores, will the approach to calibrate them differ?**

If appraisers consistently inflate observation scores, then they likely are calibrated with each other, but not calibrated to the rubric. In other words, their scores are not valid. The approach would be to ensure all appraisers are calibrated to the rubric itself and know what kind of teacher behaviors and student behaviors align to each rating category of the rubric. For example, if there is inflation because appraisers

are assigning scores due to familiarity with the teacher (e.g., what I have seen the teacher do in the past), then norming around assigning scores based *only* on what was observed would be appropriate.

**There are a lot of observation instruments. How would you characterize each of them regarding the level of inference required by an observer?**

The goal of requiring a robust teacher appraiser certification process that includes a calibration component, as well as requiring at minimum a re-calibration process for all appraisers at least annually, is to minimize the level of inference or subjectivity on the part of the appraiser. For the purposes of TIA, districts must use a teacher appraisal system that aligns to §21.3521. A best practice to reduce the level of inference required by an observer is to increase calibration procedures and protocols, such that over time, a district would be using the teacher observation rubric with increasing fidelity, validity and reliability, thereby reducing the effect of inference/subjectivity.

**When administrators script or take notes by dimension during an observation, does that mitigate the concern about cognitive load?**

Yes, taking notes can mitigate excessive cognitive load during observations. Taking notes is a good idea because after the observation, the appraiser can refer back to these notes if they are unable to remember certain details. Also, when appraisers confer with one another for purposes of calibration and norming as an appraiser team, having notes provides a reference point for justifying scores.

**Do you recommend creating “look-fors” to accompany the use of the observation instrument?**

Yes. An excellent professional development activity is to have both teachers and appraisers generate the list of “look fors” using the observation rubric. Given a specific descriptor in the rubric, what exactly does this mean the teacher is doing during the lesson? What exactly does this mean the students are doing during the lesson. Districts that use the rubric descriptors to specify exactly what this looks like in the classroom greatly increase understanding of the rubric (and associated scores) for both teachers and appraisers.

**Could you characterize what items to look for that would accompany an instrument?**

The “look fors” are examples of observed teacher and student behaviors that correspond to the descriptors in the rubric. Often rubric language is intentionally worded broadly so that it can apply across grade levels and content. A best practice is for teachers and appraisers to develop the list of “look fors” collaboratively. An example from T-TESS Dimension 2.4:

| T-TESS Descriptor 2.4                                                                             | Look Fors (to be created collaboratively by teachers and appraisers)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|---------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| The teacher differentiates instruction, aligning methods and techniques to diverse student needs. | <ul style="list-style-type: none"> <li>• Teacher adjusts the process of instruction via groupings by learning styles, by academic content proficiency levels, by language proficiency levels, etc.</li> <li>• Teacher adjusts the required student output by providing multiple ways for students to demonstrate proficiency/mastery, e.g. using visuals, performance-based options, including options for hand-written and computer/digitally produced work, etc.</li> <li>• Students have multiple opportunities and formats to answer questions</li> </ul> |