

Teacher Observation Calibration Protocols

Purpose: Calibration protocols are procedures used to increase calibration between appraisers and between campuses throughout the year. When used strategically, they can help increase scoring accuracy by providing appraisers opportunities to practice collecting defensible evidence for ratings. They also help appraisers develop a deeper understanding of what effective instruction looks like across a variety of contexts and ensure that each appraiser in the district is aligned in how they are evaluating teachers.

Appraiser Development vs. Calibration: At times, these protocols may be used to coach and develop appraisers through practice and feedback. At other times, the protocols may be used to evaluate whether appraisers are appropriately calibrated to the district's standards.

Best Practices to Implement Protocols:

- Create a schedule of calibration activities that span the year
- Designate times for appraisers at different campuses to calibrate together
- Ensure that district leaders are certified and calibrating alongside campus appraisers
- Decide what it means for two appraisers to be calibrated to each other on a given calibration activity. (e.g. Do they need to match on every rating? Do they need to be within one on each dimension? Do they need to match on a certain percentage of ratings?)

High Quality Evidence: During calibration debriefs, discuss evidence before sharing ratings so that everyone is aligned about not only what the rating should be but *why*. Quality evidence is objective and states exactly what the teacher said or did, or what students said or did.

- Low quality evidence: *The teacher checked for understanding after modeling one problem.*
- High quality evidence: *After modeling one problem, the teacher asked each student to attempt a second problem on individual white boards (15 x 24). The teacher circulated as students worked and wrote down common errors she was seeing in kids' work. Then, all students raised their boards so she could see how many students were able to complete the problem successfully on their own. 15 out of 22 students were successful.*

Should ratings collected during calibration protocols be used as formal ratings? While the decision is entirely up to the school, the ratings collected during a calibration activity are likely to be accurate since each rating was agreed upon by more than one person. For that reason, we recommend using these as formal ratings if that makes sense for your district or school. As you decide, be sure to solicit teacher input.

If two appraisers disagree about a rating, how do we decide who is "right"? Using evidence collected during the observation, appraisers should discuss which rating makes the most sense based on the teacher observation rubric and then come to a consensus. The practice of debating and grounding discussion in evidence is perhaps the most important part of calibration activities because it promotes a deeper understanding of how to appraise instruction using the rubric. When in doubt, rely on scripted evidence.

What should we do if appraisers don't calibrate to each other during a calibration activity? During a single calibration activity, districts should not be concerned if appraisers aren't calibrated. Continue engaging in calibration activities to become increasingly aligned over time. If a trend emerges in which appraisers or campuses are consistently not calibrated, the district and/or campus should create a plan to increase appraiser validity and reliability. Next steps could include the following:

- Re-train appraiser(s) on the district's teacher observation rubric.
- Norm on what constitutes each performance level on the rubric for a specific subject or grade level.



- Until calibration is established or re-established, have two appraisers conduct each scored observation.
- Assign each teacher two appraisers and use the average scores of both appraisers.
- Increase individualized coaching of appraisers who are not highly calibrated.

**All protocols are a suggestion. Districts are encouraged to adapt these protocols to meet their needs.*

Calibration Protocols

	Time Estimate	Calibration Protocol Description	When would this be useful?
Co-Observation	30-45 minutes	Two or more appraisers observe the same live lesson at the same time, score 2-3 predetermined rubric dimensions and then use the evidence collected to norm on ratings.	<ul style="list-style-type: none"> ○ This can be used by appraiser managers as a coaching tool to develop appraisers' accuracy in rating and their ability to collect high-quality evidence. ○ It can be used by peers to increase their calibration to each other. ○ It can also be used to assess how calibrated to the rubric an appraiser is.
Single Dimension Walkthrough	60-90 minutes	Two or more appraisers conduct short co-observations of multiple teachers (districts select time for short observations such as 5 minutes, 10 minutes, etc.). Appraisers rate each teacher on only one rubric dimension.	<ul style="list-style-type: none"> ○ Two or more appraisers conduct short co-observations of multiple teachers (districts select time for short observations such as 5 minutes, 10 minutes, etc.). ○ Appraisers rate each teacher on only one rubric dimension.
Campus Walkthrough	3-6 hours	Campus leadership team conducts short (10-15 minute) observations across many or all classrooms on a campus.	<ul style="list-style-type: none"> ○ Full campus walkthroughs can provide leadership teams a view of strengths and areas of weakness in instructional practices across their entire campus, especially if appraisers score teachers they don't normally observe. ○ This protocol can help increase alignment across a campus' leadership team.
Student Actions vs. Teacher Actions Co-Observation	30-45 minutes	Two or more appraisers observe the same lesson (either live or videoed). One person scripts only what students say and do. The other person scripts only what the teacher says and does.	<ul style="list-style-type: none"> ○ This protocol is useful for developing appraisers' ability to collect quality evidence using not only teacher actions but also student actions. ○ The debrief conversation will help appraisers develop a deeper understanding of the teaching rubric.
Virtual Synchronous Lesson Co-Observation	30-45 minutes	Two or more appraisers observe the same live, synchronous lesson at the same time, score certain rubric dimensions, and then discuss ratings together.	<ul style="list-style-type: none"> ○ These protocols are especially useful to train appraisers to evaluate instruction in a new context (virtual) and using an adapted virtual instruction rubric. ○ If your district needs to conduct scored observations virtually, we recommend implementing calibrated co-observations of virtual instruction. ○ This can be used by appraiser managers to develop appraisers' accuracy and ability to use high-quality evidence to rate teachers using the observation rubric.
Virtual Asynchronous Co-observation	Varies	Two or more appraisers collect evidence on a few predetermined rubric dimensions using asynchronous instruction, and then discuss ratings together.	

Co-Observation	
Description	<ul style="list-style-type: none"> Two or more appraisers observe the same live lesson at the same time, score 2-3 predetermined rubric dimensions, and then use the evidence collected to norm on ratings.
When is this useful?	<ul style="list-style-type: none"> This can be used by appraiser managers as a coaching tool to develop appraisers' accuracy in rating and their ability to collect high-quality evidence. It can be used for peers to increase their calibration to each other. It can also be used to assess how calibrated an appraiser is to the rubric.
Before the observation	<ul style="list-style-type: none"> Decide which appraisers should participate in the co-observation and which teacher to observe Schedule the co-observation and inform the teacher. Decide which part of the rubric to score. Avoid scoring the entire rubric. Rather, pick a few strategic dimensions to focus on.
Time (minutes)	Process
0-5	<ul style="list-style-type: none"> Immediately before co-observation, the teacher's main appraiser shares necessary context with the other appraiser(s).
5-20	<ul style="list-style-type: none"> Appraisers go into the classroom together and collect observable evidence on the pre-determined dimensions and takes notes.
20-25	<ul style="list-style-type: none"> If needed, each appraiser takes a few minutes to finish organizing evidence. Assign one person to be the facilitator
25-45	<ul style="list-style-type: none"> The facilitator names a dimension, and each appraiser shares the evidence they collected. Then, appraisers share their rating and rationale. Each appraiser notes their partner's rating to refer to later. If appraisers have scored a dimension differently, come to an agreement on which rating is best supported by the evidence. When in doubt, rely on scripted evidence.

Single Dimension Walkthrough	
Description	<ul style="list-style-type: none"> ○ Two or more appraisers conduct short co-observations of multiple teachers (districts select time for short observations such as 5 minutes, 10 minutes, etc.). ○ Appraisers rate each teacher on only one rubric dimension.
When is this useful?	<ul style="list-style-type: none"> ○ This protocol is useful if there is a specific dimension that appraisers are not normed on how to rate. ○ This protocol can also give school leaders a “snapshot” of how the entire school is performing a certain skill and can be used to collect data on a specific area of focus for the school.
Before the observation	<ul style="list-style-type: none"> ○ Decide which appraisers should participate in the co-observation and which teachers to observe. The more teachers the better! ○ Schedule the co-observation and inform the teachers. ○ Decide which dimension of the rubric to score ○ Create a walkthrough schedule to ensure that appraisers are able to visit as many teachers as possible during the allotted time, avoid going into classrooms when a teacher is not teaching, and prevent more than 2 appraisers being in a classroom at the same time. ○ If there are more than 3 appraisers participating in the walkthrough, divide the appraisers into pairs. Assign one appraiser in each group to be the timekeeper so that the group can stay on schedule.
Time (minutes)	Process
0-5	<ul style="list-style-type: none"> ○ Principal shares context for why the dimension was chosen ○ Team reviews the rubric dimension descriptors
5-55 (10 teachers at 5 minutes each)	<ul style="list-style-type: none"> ○ Appraiser team(s) follow the observation schedule spending allocated time in each classroom ○ In each classroom, collect evidence and assign a rating. ○ If the appraisers don’t see evidence of the selected dimension, they should take notes on what they do see without assigning a rating.
55-90	<ul style="list-style-type: none"> ○ If needed, each appraiser takes a few minutes to finish assigning a rating to each teacher on the designated dimension. ○ If there are multiple appraiser teams participating, each team should share their ratings and align on any mismatched ratings before rejoining the teams together into one group. ○ Assign one person to be the facilitator. ○ The facilitator names a teacher and each appraiser (or each team) shares their evidence as well as the rating they assigned. ○ Discuss the teachers which appraisers scored differently. ○ Come to an agreement on which rating is best supported by the evidence. ○ After the group reaches a consensus and rates each teacher on the selected dimension, discuss the trends that emerged and implications for instructional leadership. What next steps will the instructional leadership team take to support areas of weakness that emerged?

Campus Walkthrough	
Description	<ul style="list-style-type: none"> ○ Campus leadership team conducts short (10-15 minute) observations across many or all classrooms on a campus.
When is this useful?	<ul style="list-style-type: none"> ○ Full campus walkthroughs can provide leadership teams a view of strengths and areas of weakness in instructional practices across their entire campus, especially if appraisers score teachers they don't normally observe. ○ This protocol can help increase alignment across a campus' leadership team.
Before the observation	<ul style="list-style-type: none"> ○ Determine the purpose of the walkthrough, which teachers you will observe (if not all), and which rubric dimensions you will observe for (if not all). ○ Schedule the walkthrough. ○ Inform teachers of the walkthrough including what the leadership team intends to do with the observation data they collect. Alternatively, ask teachers to volunteer to be part of the walkthrough and give them the choice of whether ratings would be considered formal ratings. ○ Arrange for at least one district leader to join as an outside observer. ○ Clear each leader's schedule to allow for at least a half-day of uninterrupted time (barring emergencies that may come up during that time). ○ Create a walkthrough schedule to ensure that appraiser teams are able to visit as many teachers as possible during the allotted time and avoid going into classrooms when a teacher is not teaching. It is recommended to divide the classrooms between the teams in order to observe as many teachers as possible. ○ If there are more than three appraisers participating in the walkthrough, divide the appraisers into pairs. Assign one appraiser in each group to be the timekeeper so that the group can stay on schedule.
Time (minutes)	Process
0-15	<ul style="list-style-type: none"> ○ Meet as a group to review the procedure, schedule, and groupings. ○ Review the purpose of this walkthrough. ○ Encourage appraisers to leave notes of praise in each teacher's classroom.
15-135	<ul style="list-style-type: none"> ○ Appraiser team(s) follow the observation schedule, spending 10-15 minutes in each classroom. ○ In each classroom, collect evidence and assign evidence to each dimension. ○ After 10-15 minutes inside the classroom, appraiser teams step into the hallway and finish jotting down and organizing evidence.
135-165	<ul style="list-style-type: none"> ○ Appraisers use the evidence collected to assign ratings to each teacher. ○ Appraiser team(s) reconvene and compile ratings into one data spreadsheet. ○ Group teachers in a way that will support the leadership team's analysis of the data. For example, decide whether to group the teachers by subject, grade level, etc.
165-180	<ul style="list-style-type: none"> ○ Once the data has been compiled, take 10-15 minutes to independently review the data for trends.
180-240	<p>As a team, discuss trends that you see in the data:</p> <ul style="list-style-type: none"> ○ Where are the strengths across the school? ○ Where are there areas of weakness?
After the walkthrough	<ul style="list-style-type: none"> ○ Each teacher's appraiser shares brief feedback with the teacher. For example, the appraiser may share one area of strength and an area of growth. This can look like an abbreviated post-conference, or can come in the form of written feedback. ○ District leadership and principal determine strategic next steps to leverage areas of strength and develop areas of growth across the school. ○ Principal shares strengths with school staff.

Student Actions vs. Teacher Actions Co-Observation	
Description	<ul style="list-style-type: none"> Two or more people observe the same lesson (either live or recorded). One person scripts only what students say and do. The other person scripts only what the teacher says and does.
When is this useful?	<ul style="list-style-type: none"> This protocol is useful for developing appraisers' ability to collect quality evidence using not only teacher actions but also student actions. The debrief conversation will help appraisers develop a deeper understanding of the teaching rubric.
Before the observation	<ul style="list-style-type: none"> Decide which appraisers should participate in the co-observation and which teacher to observe. Schedule the co-observation and inform the teacher. Decide which dimensions of the rubric to score and review the descriptors. Preview the lesson plan and objectives.
Time (minutes)	Process
0-5	<ul style="list-style-type: none"> Immediately before co-observation, the teacher's main appraiser shares necessary context with the other appraiser(s). Decide who will script student actions and who will script teacher actions.
5-20	<ul style="list-style-type: none"> Appraiser #1 scripts everything students say and do. Appraiser #2 scripts everything teachers say and do.
20-25	<ul style="list-style-type: none"> Each appraiser sorts their evidence by rubric dimension. Assign one person to be the facilitator.
25-45	<ul style="list-style-type: none"> The facilitator names a dimension and each appraiser shares the evidence they collected (teacher or student) for that dimension. Both appraisers come to a consensus about how to score each dimension based on both teacher and student evidence.

Virtual Synchronous Lesson Co-Observation	
Description	<ul style="list-style-type: none"> ○ Two or more appraisers observe the same live, synchronous lesson at the same time, score certain pre-determined rubric dimensions and then discuss ratings together.
When is this useful?	<ul style="list-style-type: none"> ○ This protocol is especially useful to train appraisers to evaluate instruction in a new context (virtual) and using an adapted virtual instruction rubric. ○ If your district needs to conduct scored observations virtually, we recommend implementing calibrated co-observations of virtual instruction. ○ This can be used by appraiser managers to develop appraisers' accuracy and ability to use high-quality evidence to rate teachers using the observation rubric. ○ It can be used for peers to increase their calibration to each other.
Before the observation	<ul style="list-style-type: none"> ○ Decide which appraisers should participate in the co-observation and which teacher to observe. ○ Schedule the co-observation and inform the teacher. ○ Decide which dimensions of the rubric to score and review the descriptors. ○ If possible, preview the lesson plan and objectives.
Time (minutes)	Process
0-5	<ul style="list-style-type: none"> ○ Immediately before co-observation, the teacher's main appraiser shares necessary context with the other appraiser(s). ○ Review what each dimension should look like in a virtual setting.
5-20	<ul style="list-style-type: none"> ○ Appraisers script the lesson and collect observable evidence on the pre-determined dimensions.
20-25	<ul style="list-style-type: none"> ○ Each appraiser sorts their evidence by rubric dimension. ○ Appraisers meet in a separate virtual space (video call, for example). ○ Assign one person to be the facilitator,
25-45	<ul style="list-style-type: none"> ○ The facilitator names a dimension, and each appraiser shares the evidence they collected. Then, appraisers share their rating and rationale. Each appraiser notes their partner's rating to refer to later. ○ If appraisers have scored a dimension differently, come to an agreement on which rating is best supported by the evidence. When in doubt, rely on scripted evidence.

Virtual Asynchronous Lesson Co-Observation	
Description	<ul style="list-style-type: none"> ○ Two or more appraisers collect evidence on a few predetermined rubric dimensions using asynchronous instruction and then discuss ratings together.
When is this useful?	<ul style="list-style-type: none"> ○ This protocol is especially useful to train appraisers to evaluate instruction in a new context (virtual) and using an adapted virtual instruction rubric. ○ If your district needs to conduct scored observations virtually, we recommend implementing calibrated co-observations of virtual instruction. ○ This can be used by appraiser managers to develop appraisers' accuracy and ability to use high-quality evidence to rate teachers using the observation rubric. ○ It can be used for peers to increase their calibration to each other.
Before the observation	<ul style="list-style-type: none"> ○ Decide which appraisers should participate in the co-observation and which teacher to evaluate. ○ Determine which dimensions on the rubric to rate ○ Determine what you will use to collect evidence; for example, posted assignments, recorded lessons, student work, etc. ○ Determine a timeline for reviewing sources of evidence and rating the teacher. For example, the appraisers may have one week to review the pre-determined sources of evidence and rate.
Time (minutes)	Process
Varies	<ul style="list-style-type: none"> ○ Appraiser uses pre-determined sources of evidence to create a "case" for why the teacher should receive a certain rating. ○ Rate each pre-determined rubric dimension.
15-30 minutes	<ul style="list-style-type: none"> ○ Appraisers meet in a virtual space (video call, for example). ○ Assign one person to be the facilitator.
	<ul style="list-style-type: none"> ○ The facilitator names a dimension, and each appraiser shares the evidence they collected. Then, appraisers share their rating and evidence. Each appraiser notes their partner's rating to refer to later. ○ If appraisers have scored a dimension differently, come to an agreement on which rating is best supported by the evidence. When in doubt, rely on scripted evidence.