



## Semi-Structured Calibration Activity Protocol

### Goals of the Activity

- Read, score, and discuss two papers from a single task in common
- Build common understandings of the meaning of the scoring rubric and calibrate to a common scoring standard
- Engage in evidence-based discussions and evaluation of student work
- Reflect on information from scoring and analysis of student work to support instructional decisions, or revisions in task prompts or instruction.

**Purpose:** This is a semi-structured, teacher-led learning activity in which teachers organize themselves to share the results of their performance assessments - student work samples - with other teachers in their school; to gain confidence in their ability to score accurately using common scoring rubrics by constructing their own understanding of what the score levels mean; and to reflect on the student work to improve the design of their tasks and instruction. Teachers score and calibrate on two work samples that have been brought to the table by one of the group members. There are no external facilitators instructing teachers how to score - the teachers construct this knowledge amongst themselves in a professional community. This protocol provides a structured set of guidelines for teachers who voluntarily engage in this process together. It suggests ways to structure constructive and productive professional conversations that are beneficial for teachers' learning and for their instructional practice.

**Context:** While it is best if teachers are able to convene within subject-specific departments (and even with teachers who teach the same content and/or grade levels), it may be beneficial to teachers within a school to meet in heterogeneous groups of teachers across content areas and grade levels. This would allow all teachers within a school to begin with a common framework and understanding of "college-ready" work. However, ultimately, it is peers with a similar content-specific focus and expertise who can best provide the kind of feedback on subject-specific tasks and instruction that will support improvements in teaching and learning.





The approximate time needed for this formative calibrating experience is 2.5 hours in a single professional development session. (The approximate time allocations are only suggestions based on experience. The agenda and schedule can be adjusted based on individual school/department needs. As teachers gain more and more experience with this activity and familiarity with each other, certain steps may be skipped, e.g., Steps 1, 2, and 4). The session can be broken down into two sessions with the first session taking 1 hour, 10 minutes, and the second session taking 1 hour, 25 minutes. The preference is to score and calibrate on at least two work samples because experience indicates that agreement on one sample does not effectively result in calibration or confidence in scoring.

The "steps" below are suggestions only - to set up the group norms and roles and to prepare teachers who may be new to scoring using the common rubrics. These may be modified at the discretion of the team based on the experience level of the team members with the common rubrics as well as familiarity and comfort with each other.

***PREPARATION: Prior to the session, a volunteer teacher who has implemented a performance assessment should select two samples for scoring – one that represents a typical performance in the class and one that represents a “low” performance. All student identifiers should be removed from the samples and enough photocopies made for each participant. The performance task associated with the work samples should also be photocopied and provided to each participant.***



	Suggested time (approx)	Calibration Activity	Resources Needed
<b>PART I: Calibration Activity 1 (About 1 hour, 10 min)</b>			
1.	5 min.	<b>Assign Roles:</b> Elect a team captain to keep the group on task; a time-keeper to remind the group of time and keep the group moving forward in the agenda; a resource organizer; and a recorder (to take notes on questions, issues, agreements)	
2.	5 min.	<b>Orientation to the scoring rubric</b> <ul style="list-style-type: none"> <li>Q&amp;A - Teachers raise questions about the meaning of the rubric dimensions or wordings. Peers share their own understandings.</li> </ul>	Copies of the appropriate or relevant rubric
3.	5 min.	<b>Orientation to the task</b> from which student sample is drawn <ul style="list-style-type: none"> <li>Q&amp;A - Peers ask clarifying questions of the teacher who generated the teaching task to better understand the task and its requirements (including content standards addressed, required or optional texts). Remind teachers that the purpose is not to critique the task but to better understand the task and what the teacher expected students to do.</li> <li>Discussion - Group comes to understand the expectations of the task, and the required or provided texts or other materials</li> </ul>	Copies of the performance task used to generate the student work samples
4.	5 min.	<b>Agree on / remind teachers of norms</b> for discussing teacher and student work Suggestions: <ul style="list-style-type: none"> <li>Focus of critique should be the work, not the student</li> <li>Keep the names of students confidential</li> <li>Be honest and precise in evaluating the student work</li> </ul>	Chart paper to record agreed upon norms. The norms should be posted at every calibration session thereafter.





		<ul style="list-style-type: none"> <li>• Cite evidence in your evaluation of student work; Evidence should be tied to the relevant rubric</li> <li>• Disagree respectfully but work toward agreement</li> </ul>	
5.	Up to 45 min.	<p><b>Calibration Activity 1</b></p> <ol style="list-style-type: none"> <li>1. Independently read the first work sample. (5-7 min)</li> <li>2. Resource organizer provides the scoring rubric.</li> <li>3. Using the relevant rubric and scoring form, each individual marks up the rubric descriptors that describe the work sample and on the scoring form, notes evidence from the work sample that supports the scores for each dimension. (10-13 min)</li> <li>4. Each participant charts their scores as in <b>Figure 1</b> below. (5 min)</li> <li>5. Team captain leads a discussion of the scores, noting areas of agreement and discussing areas of disagreement, with the goal of coming to consensus. (20 min.)</li> </ol>	<ul style="list-style-type: none"> <li>• Scoring rubric for selected work sample</li> <li>• Pens, post-it notes, highlighter markers</li> <li>• Chart paper, Markers</li> </ul>
6 .	15 min.	<p><b>Debrief Calibration Activity 1</b></p> <ul style="list-style-type: none"> <li>• What seems to be difficult for this student? What kind of feedback would you give to this student to improve their work?</li> <li>• What instructional supports are needed to help this student improve on the next writing task?</li> </ul>	<ul style="list-style-type: none"> <li>• Chart paper, Markers</li> </ul>
	10 min.	<b>Take A Break OR Complete the rest at the next session</b>	
<b>PART 2: Calibration Activity 2 (About 1 hour, 25 min)</b>			
7.	Up to 45 min.	<p><b>Calibration Activity 2</b></p> <ol style="list-style-type: none"> <li>1. Independently read the second work sample. (5-7 min)</li> <li>2. Resource organizer provides the scoring rubric.</li> <li>3. Using the relevant rubric and scoring form, each individual marks up the rubric descriptors that describe the work sample and on the scoring form,</li> </ol>	<ul style="list-style-type: none"> <li>• Scoring rubric for selected work sample</li> <li>• Pens, post-it notes, highlighter markers</li> <li>• Chart paper, Markers</li> </ul>

© 2014 by The Board of Trustees of The Leland Stanford University. This work is licensed under a [Creative Commons Attribution 4.0 International Public License](https://creativecommons.org/licenses/by/4.0/) and should be attributed as follows: "Semi-Structured Calibration Activity Protocol was authored by Stanford Center for Assessment, Learning, & Equity (SCALE)."



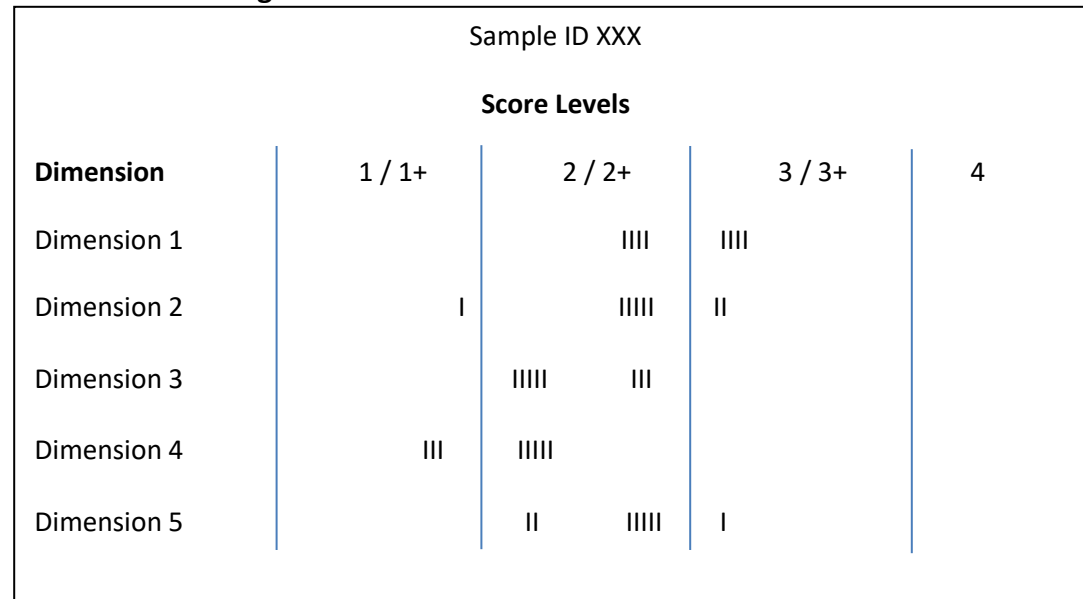


		notes evidence from the work sample that supports the scores for each dimension. (10-13 min) 4. Each participant charts their scores as in <b>Figure 1</b> below. (5 min) 5. Team captain leads a discussion of the scores, noting areas of agreement and discussing areas of disagreement, with the goal of coming to consensus. (20 min.)	
8.	15 min.	<b>Debrief Calibration Activity 2</b> <ul style="list-style-type: none"> <li>• What seems to be difficult for this student? What kind of feedback would you give to this student to improve their work?</li> <li>• What instructional supports are needed to help this student improve their work on the next writing task?</li> </ul>	• Chart paper, Markers
9.	15 min.	<b>Reflection Discussion</b> <ul style="list-style-type: none"> <li>• <b>Question to the teacher who generated the task:</b> Are there any changes you would make to the task based on scoring these samples?</li> <li>• <b>Question for group discussion:</b> Are there any changes you would suggest to the task to support improvements in student performance on the task?</li> </ul>	
10.		<b>Decide whose task and samples you will score and discuss next time</b>	





**FIGURE 1. Charting Scores for Calibration Session**



Note: Tic marks represent one teacher’s scores for each dimension of scoring. In this example, there are eight teachers’ scores represented in the chart.

