

# Writing Calibration Protocol

## Rhode Island Department of Education

---

The Writing Calibration Protocol presented here provides a process for bringing educators together to discuss student work in order to reach consensus about how to score it based on a rubric or other scoring criteria. The process is especially important now, as schools and LEAs increasingly use common assessments to evaluate student expectations contained in the Common Core State Standards (CCSS). Calibration makes scoring of student work more consistent and more aligned to the standards upon which rubrics are based. This protocol is intended to help all content area educators realize these benefits as they implement assessments based on the CCSS in English Language Arts & Literacy in History/Social Studies, Science and Technical Subjects.

Protocol Module:

- I: Introduction and Background Information** explains the purpose of calibration and provides an overview of the process.
- II: Calibration Protocol** explains each step in calibrating with either a holistic or analytic rubric.
- III: Calibration with Anchor Papers** provides an overview of calibrating with anchor papers that have already been identified.
- IV: Extending the Calibration Process** discusses other benefits and uses of calibration.
- V: Appendix** includes resources.

### **I: Introduction and Background Information**

#### **Purpose of Calibration**

Calibration enables a group of teachers and other educators to reach agreement on how to score student responses to a given assessment task. Through discussion of student work, educators work through differences in the way they understand the rubric or evaluate qualities in the student work that would cause them to differ in their scoring. For example, they may look at wording in the rubric that requires students to “support ideas with sufficient and relevant evidence.” Initially, they may be unclear about what the rubric means by “sufficient” or disagree about whether the evidence in a particular sample of student work is relevant enough. By the end of the calibration session, they will have reached agreement about such issues and be interpreting and applying the rubric in the same way. When they go on to score the remaining papers written to that task, their results will be more reliable because it is likely that any given paper would have been scored the same by another evaluator in the group. Their results will also be more accurate, in that the group has collectively and deeply thought through the rubric before applying it.

## Calibration Process

During a calibration session, educators read and score several student responses to a given task. Led by a facilitator, they note how their scores differ and attempt to understand and resolve the reasons for those differences. Discussion is focused on two areas: the meaning of the rubric and the “weight and worth” of the evidence from the student work (Rhode Island Skills Commission, 2005). When the scorers have resolved their issues and reached consensus on a score for one paper, they move on to the next paper. The process continues until differences are resolved and the group is consistently scoring student work.

Because calibration is part of a larger process of scoring the set of student work, materials are developed during the session to be used later in scoring. A note taker records the score of each paper and any decisions the group made to resolve difficult or critical issues. Papers that the group scored with consensus become Anchor Papers, and both the notes and the Anchor Papers are used by educators during the scoring process to ensure that their scores remain consistent.

## Scoring Process

Following calibration, the actual scoring of papers may happen in different ways. The group may divide up the student responses and score them independently, either scoring their own classes or scoring a cross-section of classes. Or the group may stay together and engage in some form of collaborative scoring enabling them to continue to compare scores and discuss student work (Ainsworth & Viegut, 2006).

The scoring process may include steps for periodically checking on the consistency of scoring. For example, papers may be selected randomly to be scored independently by two scorers with the scores compared to see if they are indeed still consistent. In situations where the reliability of results is very important, as with high stakes testing, all papers may be evaluated by two scorers as an additional validation of the students’ scores.

## Terminology

Some of the terms used in this document require clarification.

**Task** refers to any assignment that requires an extended written response from students, such as an essay question or constructed response.

**Student work, student response,** and **paper** are used interchangeably to refer to the student’s response to the task.

**Scorer, evaluator** and, more generally, **educator** are used interchangeably to refer to individuals assigning a grade, score or performance level to student work during a calibration or scoring session. This group will include classroom/content area teachers, but may also include administrators, special education teachers, specialists (reading, media, speech pathologists, school

psychologists, for example) and other educators who can contribute to and learn from the calibration process.

**Scoring criteria** refers to any set of expectations that are the basis for evaluating student work, such as a rubric.

**Analytic rubrics** identify several different **criteria** or traits (such as organization, development of ideas, and knowledge of language and conventions), which evaluators score separately before assigning an overall score to the paper. In most rubrics, descriptions define what is expected of students at each score point/level of performance. The rubric **scale** refers to the number of possible score points or performance levels identified. For example, a three-point scale might rate papers as 1-2-3 or as below proficient-proficient-above proficient.

**Holistic rubrics** integrate all criteria or traits into one description at each score point/performance level, which scorers use to make one judgment about the overall effectiveness of the paper.

## II: Calibration Protocol

### Prior to the Calibration Session

- **Identify participants and roles**

Everyone who will score student papers is included in the session. Teachers who administered the task to their students typically are the core group. Administrators, teachers in other courses and grades, and educators in other roles are often included. A facilitator who is familiar with the process leads the session. One of the participants should be identified as a note taker to record important decisions made by the group.

- **Schedule the session**

In deciding the amount of time needed, consideration should be given to the complexity of the task, type of rubric, and familiarity of the group members with the task, rubric, and process. By one estimate, it takes 35 minutes to introduce a session and score one student paper with an analytic rubric (Center for Collaborative Education, 2012, p. T9). Typically it takes less time for each additional paper as the process continues.

- **Select student work to use in the process**

Prior to the session, the facilitator reads through most or all of the set of student work to select a range of low, middle and high papers to use in the process. Others may assist with this. The number of papers needed varies depending upon the task and rubric. For example, with a four-point scale rubric, educators need to reach consensus about what student work looks like at four different performance levels. A rule of thumb may be to have at least two pieces of student work times the number of points in the rubric scale. Fewer papers are needed at the extremes of high and low and more are needed in the middle where educators must distinguish between adequate/proficient performance and performance just above and below (Daiker & Grogan, 1991).

- **Prepare materials**

A copy of the following is needed for each educator:

- Task and any texts or graphics that accompany it
- Rubric
- Copies of student work that will be used in the process along with sufficient copies of the rubric for scoring each paper

Ensure that papers do not identify students in any way. Label papers with letters (A, B, C) rather than numbers to avoid confusion later with scores. Do not staple the set together.

## Calibration Session

### Calibration Session with a Holistic Rubric

- **Introduce the process**

The facilitator explains the purpose of calibration (reaching consensus about how student papers will be scored by everyone) and briefly describes the steps in the process. Norms for group discussion may be presented.

- **Describe the context of the task**

For example, the leader might say: “Today we are looking at student work from a common task that was administered to all 9<sup>th</sup> grade English students in the school as part of the mid-year exam. The school-wide rubric for writing literary analysis was used.”

- **Read the task, any associated texts or graphics, and the rubric**

Educators review the task and rubric and may ask clarifying questions or discuss vague language. The note taker records any agreed-upon changes for the future.

- **Read several pieces of student work**

Educators read several papers (4-5) that will be used in the process and sort them as strong, medium or weak responses based on their overall impression. The facilitator asks educators to report their results with a show of hands. This step allows educators to see a range of student work before they begin to score individual pieces, and the facilitator begins get a sense of how close or divergent the group is in its thinking. Differences are not discussed at this point.

- **Score the first paper**

A paper that is in the high range may be presented first. Individually, educators read and score the student work. Each highlights or underlines words and phrases in the rubric’s descriptions that best match qualities of the work. He or she also notes evidence in the student work that supports the scoring of the rubric. Each scorer then decides on an overall score for the paper.

The facilitator asks educators to report their scores with a show of hands or with a score sheet. The facilitator charts the scores and then leads a calibration discussion that is grounded in the rubric. A scorer who is high or low compared to the group explains his or her scoring first. A scorer with a different score or view responds. The facilitator requires that all comments refer to specific language in the rubric and be supported with evidence in the student work. As scorers respond to each other’s comments, discussion stays focused on an area of disagreement until it is resolved. The facilitator checks for changes in the group’s scores. If consensus has been reached, the group moves to the next paper. If not, scorers who still differ from the group identify remaining issues for discussion. The note taker records the score and any important notes from the discussion that explain why the paper was given this score (Rhode Island Skills Commission, 2005).

- **Continue to score papers until consistency at each performance level is reached**

The process continues with papers at different levels. Papers in the higher and then lower range may be presented first, so that scorers gain confidence placing papers at different performance levels. Typically, the group takes most time deciding between papers that meet expectations for satisfactory performance and papers that do not, such as deciding between “proficient” and “below proficient,” and scoring papers that are borderline between levels. Papers with special issues may be introduced later in the process. When the group is consistently scoring at each performance level, scoring is calibrated.

- **Keep the process moving**

The facilitator should not allow repeated debate to slow down the process. For example, if a scorer insists that proficient papers should have no mistakes in conventions even though the rubric does not require this, the facilitator will need to be forceful in stating that scoring is based on the language of the rubric not on individual opinion. If a scorer continues to score higher or lower than the rest of the group even after discussion, the facilitator reminds scorers that the aim of calibration is to reach agreement collaboratively, which may mean recognizing and accepting the judgment of the group rather than holding to a position that others do not accept.

- **Explain next steps for scoring the full set of student work.**

The facilitator explains when and how scoring will take place and ensures that scorers have access to the anchor papers and any important notes related to scoring decisions.

## Calibration Session with an Analytic Rubric

When scoring an analytic rubric, each of the criteria is scored separately. The scores are averaged or totaled in the way indicated by the rubric. Certain criteria may be weighted more heavily than others. These characteristics of analytic rubrics factor in the calibration process. The first few steps in the process are the same as with a holistic rubric but the process of scoring individual papers differs significantly.

- **Introduce the process**

The facilitator explains the purpose of calibration (reaching consensus about how student papers will be scored by everyone) and briefly describes the steps in the process. Norms for group discussion may be presented.

- **Describe the context of the task**

For example, the leader might say: “Today we are looking at student work from a common task that was administered to all 9<sup>th</sup> grade English students in the school as part of the mid-year exam. The school-wide rubric for writing literary analysis was used.”

- **Read the task, any associated texts or graphics, and the rubric**

Educators review the task and rubric and may ask clarifying questions or discuss vague language. The note taker records any agreed-upon changes for the future.

- **Read several pieces of student work**

Educators read several papers (4-5) that will be used in the process and sort them as strong, medium or weak responses based on their overall impression. The facilitator asks educators to report their results with a show of hands. This step allows educators to see a range of student work before they begin to score individual pieces, and the facilitator begins get a sense of how close or divergent the group is in its thinking. Differences are not discussed at this point.

- **Score the first paper**

A paper that is in the high range may be presented first. Individually, educators read and score the student work. Reading across each of the rubric descriptions for the first criterion, the scorer highlights or underlines words and phrases that best describe qualities of the student work. He or she also notes evidence in the student work that supports the scoring of the rubric. The scorer decides on one performance level score for the criterion and checkmarks or circles it. Each criterion is scored in this way. As the final step, the scorer uses the criterion scores to determine one performance level score for the paper as a whole.

The facilitator asks educators to report their scores with a show of hands or with a score sheet. The facilitator charts the scores for the paper as a whole. This gives the group an overall sense of how the paper scored. The facilitator moves directly to the first criterion, charts the scores, and begins the discussion by asking a scorer who is high or low compared to the group to explain his or her scoring. A scorer with a different score or view responds. The facilitator requires that all comments refer to

specific language in the rubric and be supported with evidence in the student work. As scorers respond to each other's comments, discussion stays focused on an area of disagreement until it is resolved. The facilitator checks to see if the group has reached consensus on a score for this criterion, and if so moves on to the next. If not, scorers who still differ from the group identify remaining issues for discussion. When scores are agreed upon for all criteria, they are averaged or totaled according to what is stated in the rubric to arrive at the performance level score for the paper. Because the group had reached agreement on the criteria, they will likely agree with this score. If they do not, they continue discussion to resolve issues and agree on an overall score (Rhode Island Skills Commission, 2005).

- **Continue to score papers until consistency at each performance level is reached**

The process continues with papers at different levels. Papers in the higher and then lower range may be presented first, so that scorers gain confidence placing papers at different performance levels. Typically, the group takes most time deciding between papers that meet expectations for satisfactory performance and papers that do not, such as deciding between "proficient" and "below proficient," and scoring papers that are borderline between levels. Papers with special issues may be introduced later in the process. When the group is consistently scoring at each performance level, scoring is calibrated.

In some cases, it is decided that adjacent scores (one point apart) on individual criteria are allowed if scorers have reached consensus on the performance level score for a paper as a whole based on the point range indicated in the rubric. While the group is working toward consensus, however, they still discuss and resolve adjacent scores (Ainsworth & Viegut, 2006). The decision regarding adjacent scores may be based on how the criteria scores will be used.

- **Keep the process moving**

The facilitator should not allow repeated debate to slow down the process. For example, if a scorer insists that proficient papers should have no mistakes in conventions even though the rubric does not require this, the facilitator will need to be forceful in stating that scoring is based on the language of the rubric not on individual opinion. If a scorer continues to score higher or lower than the rest of the group even after discussion, the facilitator reminds scorers that the aim of calibration is to reach agreement collaboratively, which may mean recognizing and accepting the judgment of the group rather than holding to a position that others do not accept.

- **Explain next steps for scoring the full set of student work.**

The facilitator explains when and how scoring will take place and ensures that scorers have access to the anchor papers and any important notes related to scoring decisions.



### III: Calibration with Anchor Papers

Often calibration is done with anchor papers identified prior to the calibration session. Rather than scorers deciding on the student work that will “anchor” each of the performance levels, the process begins with anchor papers already decided. Educators learn to align their scoring to these papers and the rubric. This process is sometimes referred to as training of scorers.

Anchor papers are often used when a task will be administered more than once or scored by more than one group. For example, a common assessment may be given to students across an LEA or across multiple LEAs in a state. The initial identification of anchor papers is done by smaller group of educators during a calibration session or similar process. They may be a cross-section of the larger group or they may be chosen as individuals who bring special expertise to the process. The anchor papers and scoring notes that they develop are used to train other educators to score the same task.

When anchor papers and scoring notes are being compiled for further use, care is taken to select anchor papers that are clear examples and to annotate them with explanations of the scoring decisions. An example of a packet for training scorers with anchor papers can be found at Rhode Island Assessment ELA Scoring Module: Item Training for Booker Washington.

#### Scoring Training with Anchor Papers

The process of training a group of scorers using anchor papers is more directive because the purpose is to enable scorers to understand and agree to scoring decisions about student work that have already been made. The following provides an overview of the process.

- The materials needed for each participant include:
  - Copy of task and rubric
  - Anchor paper with accompanying annotations for each performance level of the rubric; the papers include scores
  - 1-2 additional anchor papers for each performance level, not revealing the scores, for use as practice papers
  - Additional copies of the rubric for use in scoring the practice papers
- Facilitator introduces purpose, steps in the process, and context of task. Scorers read the task and rubric.
- Scorers individually read and analyze each anchor paper beginning with the anchor paper for the highest performance level/score point of the rubric. Scorers note where qualities in the papers relate to language in the rubric using the annotations as a guide. The facilitator leads a discussion of the anchor papers in relation to the performance level/score point descriptions to allow scorers to internalize their understanding of how the anchor papers were scored (MA Department of Education, 2007).

- Scorers individually read and score the first practice paper using the rubric and referring to the anchor papers for guidance. The facilitator charts the group’s individual scores. The score that was pre-assigned to the practice paper is revealed. The facilitator leads a group discussion that helps scorers who differed from the practice paper score to reconsider the rubric, student work and anchor papers in order to reach understanding and agreement with the score. The purpose is not to debate the pre-assigned scores (Kansas State Department of Education, 2008).
- The process continues with additional practice papers until the scorers are able to align their scores to the anchor scores. An alternative approach is for scorers to score several papers first before discussing them one by one.

#### **IV: Extending the Calibration Process**

The calibration process has other uses and benefits.

- Anchor papers from tasks that all use the same general rubric (for example, an argument writing rubric) can be “attached” to the rubric as samples of performance. Papers that represent the proficient or above performance levels are sometimes referred to as exemplars or benchmarks at the grade level.
- Annotated anchor papers can be used in professional development sessions to model how to look at and score student work using rubrics and anchor standards (Center for Collaborative Education, 2012).
- Calibration provides an opportunity to revise a rubric to reflect more accurately a school or department’s expectations. If a new rubric based on the CCSS is developed, it may be piloted with a task and revised during a scoring calibration session.
- The calibration process can be a professional learning opportunity for educators, enabling them to deepen and solidify their common understanding of expectations for student work and of the standards upon which they are based. Ainsworth and Viegut refer to collaborative scoring in general as “powerful professional practice” in which educators share insights and expertise while deeply examining student papers (2006, p. 87).
- Often when time and effort has been invested in a calibration process, the scores of all students are collected and used for a purpose beyond individual grades, such as to assess overall student achievement of academic expectations or to guide instruction.

## Appendix

PARCC Generic Rubrics for Prose Constructed Response Items (DRAFT)

[Grade 3](#)

[Grades 4 -5](#)

[Grades 6-11](#)

[Rhode Island Interim Assessment Constructed Response Rubric](#)

## References

- Ainsworth, L., & Viegut, D. (2006). *Common Formative Assessments: How to Connect Standards-Based Instruction and Assessment*. Thousand Oaks, CA: Corwin Press.
- Center for Collaborative Education. (2012). *Quality Performance Assessment: A Guide for Schools and Districts*. Boston, MA: Center for Collaborative Education.
- Daiker, D., & Grogan, N. (1991). Selecting and Using Sample Papers in Holistic Evaluation. *Journal of Advanced Composition*, 11, No. 1, 159-171.
- Kansas State Department of Education. (2008). *Introduction to Analytic Scoring for the Kansas Writing Assessment*. Topeka, KA: KA State Department of Education.
- MA Department of Education. (2007). *2007 MCAS Technical Report*. Malden, MA: MA Department of Education.
- Rhode Island Skills Commission. (2005). *Protocol for Calibrating and Scoring ELA Tasks*. Providence, RI: Rhode Island Department of Education and the Education Alliance at Brown University.