

The Wyoming Department of Education would like to thank all the districts that willingly shared their current Body of Evidence plans for this project. The intent of posting sample BOE plans is to provide clear examples of different approaches that satisfied the criterion when reviewed during the 2009 BOE Peer Review. Since the BOE plans are the property of the identified districts, please contact the individual district directly should you want to use any part of their plan.

The following example is what a **course-based, common assessment approach** section for consistency might look like:

Section 3: Consistency

In order to meet the consistency criterion on the BOE Peer Review Scoring Guide, the submitted plan must include evidence of the following:

- For **open-ended assessments**, the district plan describes clear procedures to be used to ensure inter-rater reliability and defines a desired, acceptable rate. Data are presented that support implementation of the stated procedures.
- For **closed-ended assessments**, the district plan describes clear procedures to be used to ensure reliability and defines a desired, acceptable rate. Data are presented that support implementation of the stated procedures.
- If **teacher judgment** is part of the plan, the plan describes procedures to ensure reliability of judgment across assessments within a course & across teachers. There is clear documentation that judgment is anchored to the performance standards. Data are presented that support implementation of the stated procedures.

Evidence in plan to support required criteria for consistency:

- The procedures used to ensure inter-rater reliability on open-ended assessments are described.
 - Inter-rater reliability data that meets acceptable rates (inter-rater reliability to meet or exceed 80% exact agreement and 98% exact + adjacent agreement) is included.
 - The procedures used to ensure reliability on closed-ended assessments are described.
 - Desired, acceptable rates of reliability on closed-ended assessments are stated.
 - Reliability data on closed-ended assessments (to meet or exceed average reliability coefficients greater than 0.85) is included.
 - Procedures used to ensure reliability of teacher judgment across assessments within a course and across multiple teachers are described.
 - Reliability data of teacher judgment is included.
-

Section 3: Consistency

Ensuring Consistency in Decision Making

The Sublette County School District #1 Body of Evidence system ensures a consistent decision making process for each student through:

1. Identifying the particular assessment evidence/scores for making Body of Evidence decisions and collecting the same data on each student;
2. Requiring multiple opportunities and multiple formats for students to demonstrate proficiency on standards;
3. Interpreting the Body of Evidence data the same way for each student using a consistent system and aggregation rules;
4. Providing clear requirements for awarding transcript endorsements;
5. Providing an appeals processes throughout the system.

Training for Administration and Guidance Staff

Additionally, during start up of the system, the administration and guidance counselor(s) have received training and practice on applying sets of Body of Evidence data to ensure compliance with the system. These practice activities are scored and an agreement rating of 100% is required for data sets on students that are four year Pinedale High School students with an agreement rate of 75 % required for data sets on students transferring into the system. (Please see Section 1, Overview of Course Based Common Assessment Approach for additional explanation on transfer students.)

Reliability on Open Ended Assessments

The Sublette County School District #1 system utilizes open ended assessments are part of the Body of Evidence system. These assessments are primarily hands-on demonstrations of skills, abilities and application of knowledge and are scored using a rubric. These assessments are focused on sampling the **cognitive depth** of knowledge implied in the standards.

Performance assessments are designed with clear directions, requirements, scoring procedures and rubrics to minimize variability and increase the reliability of the assessment tasks. Teachers are encouraged to provide scoring rubrics and examples of assessment products to assist students in gaining additional understanding of what is possible and to what level of quality their work should be in order to earn the various performance ratings (Appendix B.)

For reliability purposes results are examined from student cohort to cohort on these open ended assessments. If the percentage of students scoring in the four performance levels changes by

more than 10% then the results are examined to identify the causes for this variance. A set of questions is used to help determine the cause of variance in student performance.

1. Is instruction causing the variance?
2. Is this group of students, although randomly assigned, different from other groups of students in some way?
3. Is there an issue with assessment conditions?
4. Has the assessment been altered?
5. Have there been events that could cause the variance?

The answers to these questions are used to search for inconsistency and to eliminate it as much as is possible from the system.

Validity on open ended assessments is achieved through blueprinting each assessment and matching all tasks to the appropriate standards, benchmarks and cognitive complexity in their content area.

Inter-rater Reliability on Open Ended Assessments (Appendix B)

- Department Chairs randomly select 10% of the assessments administered during the semester for rescoring to ensure ongoing inter rater reliability.
- All teachers that have administered the assessment participate in the re scoring.
- Teachers score the projects independently.
- Scores are recorded, compared and discussed; detailing the reasons for the ratings.
- Agreement rates must be at 98% plus or minus one point.
- If this level of agreement is not reached a new set of papers are scored until the desired rate of inter rater reliability is achieved.

Scoring Procedures for Open Ended Assessments

In an effort to minimize the variability and increase the reliability of open ended assessments the district has developed these procedures for teachers:

1. Provide clear, written scoring procedures and rubric descriptions;
2. Provide anchor papers or anchor projects as guides after the piloting phase;
3. Provide scorers with training and practice on scoring specific performance assessments;
4. Provide multiple scorers with training for scoring open ended assessments to increase inter rater reliability.
 - Content Chairs conduct interdepartmental training with all content area members. Each term the Content Chairs randomly selects assessments to be re-scored by all staff that administered the assessment. Inter rater agreement must be 98% between scorers with a 1 point adjacent agreement (Appendix B, Math - Inter Rater Agreement Data).
5. Providing ongoing training to teachers on scoring open ended assessments.

6. For courses with a single teacher the district has established procedures that require the teacher to calibrate their scoring from cohort of students to cohort. This is accomplished by having the teacher:
 - Review the performance descriptors for the content area,
 - Review the rubrics for the specific task,
 - Review the scoring of past cohort samples and
 - Seeding the current set of tasks with samples from previous administrations to ensure consistent evaluation of the tasks (Appendix B, Health)

Reliability and Validity Data on Closed Ended Assessments

The Sublette County School District #1 system utilizes some closed ended assessments. These assessments are primarily multiple choice, short answer and matching; having generally one right answer per question. These assessments are designed to cover the **breadth** of the standards, benchmarks and content.

The Sublette County School District #1 Body of Evidence system uses the Chronbach's alpha for establishing reliability for closed ended assessments (Appendix E, Reliability Data). Results are examined from student cohort to cohort on these closed ended assessments. An acceptable reliability coefficient for teacher developed assessments is determined to be .50 or higher (Eden and Frisbie, 1986). Should the Chronbach's alpha measure show a reliability coefficient of less than .50 the district analyzes the reasons for the variance and for determining what steps if any need to be taken to improve reliability. The examination of the assessment and reliability data start with this set of questions:

1. Is instruction causing the variance?
2. Is this group of students, although randomly assigned, different from other groups of students in some way?
3. Is there an issue with assessment conditions?
4. Has the assessment been altered?
5. Have there been events that could cause the variance?

Validity on closed ended assessments is achieved through blueprinting each assessment and matching all items to the appropriate standards, benchmarks and cognitive complexity in their content area.

Teacher Judgment

Sublette County School District **does not use teacher judgment** as a Body of Evidence measurement construct.