

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 **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.

CONSISTENCY

Overview

Five year cycle Big Horn County School District Number One has developed a systematic plan to address the process of monitoring, evaluating, revising, aligning, and implementing curriculum. This Curriculum Plan involves a five-year rotation cycle that describes the responsibilities of each of the content area curriculum teams. The elementary and secondary subcommittees and grade/subject level teams that report to the district curriculum teams review standards, assessments and evaluation data. This provides for consistency throughout the curriculum process. Program changes as well as staff development needs are identified through the work of these teams.

Common assessment procedures Consistency within the district assessments involves commonality of assessment procedures and scoring across the district. Common rubrics are used with performance assessments. Criteria for proficiency have been established during development of the assessments. These criteria appear on the cover sheet of each assessment. Each proficiency level has been carefully identified for each component of the rubric. Clearly delineated student behavior at each score point has been developed in rubrics. Training has been provided for all staff in areas such as the scoring of writing. Within the curriculum cycle test items are examined to determine whether “student memorability” could occur. Replacement items are selected at a comparable cognitive level.

Common rubrics

Data review Data from assessments (See samples at the end of the Fairness section.) is examined annually in order to ensure that the decisions made concerning students’ performance relating to standards are consistent and valid. The building level principals and the curriculum office monitor assessment results annually to provide another level of consistency.

Inter-rater Reliability

Our district conducts a district-wide writing assessment each year. Papers are read by two readers. If there is more than one point

Double scoring process for open-ended assessments

difference, a third reader is utilized. Some papers with these discrepancies are then reviewed and discussed to determine the reasons for discrepancy. This year the district has moved this inter rater reliability effort to all content areas and open-ended district assessments. To make this process manageable for teachers, we double score two common assessments at each grade level where standards are embedded. Assessments that need a third read are scored by another content teacher. Teachers then meet to review their scores and to discuss the process. (See pages 3-14.)

Inter rater reliability

The district is tracking the inter-rater consistency overtime to see if it remains similar. (See sample data sheets pages 3-14 of this section.) Our goal is to have inter-rater reliability to meet or exceed 80% exact agreement and 98% exact and adjacent agreement. We are also collecting anchor papers, which will allow us to have examples for new teachers or for future scoring trainings. (See samples at the end of this section on pages 17-26.)

In-service time has been used for content teams to score projects and compare results. Cutscores have been determined and grading and scoring practices have been established. This has allowed consistent decisions to be made regarding student performance across schools and over time.

Close-Ended Assessment Reliability

Split-half reliability

Currently, we are attempting to use a simple split-half reliability check on closed-ended assessments. The process of obtaining split-half reliability is begun by “splitting in half” all items of a test that are intended to probe the same area of knowledge in order to form two “sets” of items. If the scores are consistent it leads one to believe that it is most likely measuring the same thing. Simply put: If items on a test can be divided into two halves and give the same results, your test is reliable. We are aiming for reliability coefficients greater than .85 on close ended assessments. (See sample at end of this section on page 16 which is in math because we currently don’t have assessments in language arts or health that are close-ended.)

Teacher judgment

*Note: Teacher judgment is not a part of our present assessment system.