**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.

## Section 3: Consistency

The district describes procedures to ensure reliability of teacher judgment across assessments, within a course and across multiple teachers.

### A. Open Ended Assessments

All of the Passport assessments in the representative content areas submitted (Science Appendix 3A; Fine & Performing Arts Appendix 3B) are open-ended assessments. Common scoring rubrics are provided for each Passport assessment and are used by all indicator course teachers. Indicator Course Manuals are provided as a part of training to each teacher, outlining common implementation procedures and guidelines.

The use of student anchor papers, including exemplars at each performance level, allows teachers to score student work using common rubrics. The district participated in the Wyoming Assessment Activities Consortium examination of piloted student work to learn and practice a process for creating exemplar or anchor papers. The district applied these processes to generate exemplars for the Passport assessments. These processes calibrate the consistency of indicator course teachers in the scoring of Passport assessments. Implementation and troubleshooting sessions are held for all teachers to standardize the implementation of the administration and timing of the Passport assessments. Training is provided annually to new staff since the original implementation (Appendix 3C).

Multiple double-scoring sessions have been held to ensure inter-rater reliability and to prevent drift. In each session approximately 10% of student papers were double scored (Appendix 3D). All teachers from all nine departments who teach indicator courses, from East Junior High, Farson/Eden High, Independence High, Roosevelt Learning Center, and Rock Springs High individually scored packets of student work. The student work was representative of all teachers and all levels of student work in the district. After each 5 samples of student work were scored, group members shared their scores and dialogued about the work. This process of reviewing 5 samples at a time continued until approximately 10% of the student papers were double-scored. Scores that differed by more than 1 ranking were discussed until consensus was achieved.

In order to meet the consistency requirements, inter-rater reliability estimates must be computed for open-ended assessments. While the most popular methods would be the simple percent-agreement figure, in which percent agreement is calculated by adding up the number of cases that received the same rating by both judges and dividing that number by the total number of cases rated by the two judges; or the Pearson correlation coefficient which is a common measure of the linear dependence between two variables *X* and *Y*. For consistency purposes in SSD#1, these procedures are limited in that results can be calculated only for one pair of judges at a time and for one item at a time.

The district states an acceptable rate for inter-rater reliability. Data are present.

In situations where multiple judges are used, such as is the case with the SSD#1 Passport System, an approach to computing a consistency estimate of inter-rater reliability is to compute Cronbach's alpha coefficient (Crocker & Algina, 1986). Cronbach's alpha coefficient is a measure of internal consistency and is useful for understanding the extent to which the ratings from a group of judges hold together to measure a common dimension. The major advantage of using Cronbach's alpha comes from its capacity to yield a single consistency estimate of inter-rater reliability across multiple judges (Stemler, 2004, p. 6). Within the social sciences, a Cronbach's alpha coefficient of.70 or higher is considered acceptable (See Appendix 3E). Below, for conceptual purposes, is the formula for the standardized Cronbach's alpha:

$$\alpha = \frac{N \cdot \bar{c}}{\bar{v} + (N-1) \cdot \bar{c}}$$

| Discipline | Passport | Cronbach's Alpha Coefficient |
|---|---|---|
| Science | Endangered Species-Morphology | 0.900 |
| Science | Endangered Species-Classification | 0.954 |
| Science | Endangered Species-Interdependence | 0.920 |
| Science | Endangered Species-Endangerment | 0.998 |
| Science | Endangered Species-Communication | 0.670 |
| Science | Genetic Engineering-Concepts | 0.918 |
| Science | Genetic Engineering-Communication | 0.950 |
| Science | Chemistry-Conducts | 0.950 |
| Science | Chemistry-Representation | 0.981 |

| | | |
|---|---|---|
| Science | Chemistry-Conclusion | 0.984 |
| Science | Chemistry-Communication | 0.956 |
| Science | Chemistry-Concepts | 0.989 |
| Science | Physics-Design | 0.973 |
| Science | Physics-Conduct | 0.998 |
| Science | Physics-Representations | 1.000 |
| Science | Physics-Conclusions | 1.000 |
| Science | Physics-Communications | 1.000 |
| Science | Physics-Concept | 1.000 |

| Discipline | Passport | Cronbach's Alpha Coefficient |
|---|---|---|
| Fine & Performing Arts | Fame, Fortune, Feast or Famine | 0.989 |