

ALIGNMENT STUDY
2006
PROCEDURAL MANUAL

Meagan Karvonen
Western Carolina University

Pat Almond
Behavioral Research and Teaching

Gerald Tindal
University of Oregon

DRAFT

August 8, 2006

ALIGNMENT STUDY PROCEDURAL MANUAL

This manual documents the procedures that will be used by Dillard Research Associates to conduct the alignment study in August 2006. This manual describes the procedures for the analysis of alignment between the Alternate Content Standards in reading, writing, and math, grades 3-8 and 11. Procedures for calculating alignment indices and methods for describing alignment between state Content Standards and Alternate Content Standards will be described in the final report.

Rating of alignment between standards and assessment will be conducted using a slightly modified version of Webb's alignment method. Judgments about the content and cognitive demand represented in the state's Alternate Content Standards and in the Alternate Assessment will be made by state teachers, and alignment indices will be calculated based on those ratings.

Rater Training

Rater training will take place during the first two-hour block of on-site time. Training will begin with an overview of alignment, the purposes of the study, and the responsibilities of the raters. Raters will discuss definitions and examples for cognitive complexity (depth of knowledge), first as a whole group, then in dyads by subject area. Outcomes from both whole group and dyadic ratings will be discussed as a whole group in order to clarify distinctions and create decision rules. Training on judgments about content match will be conducted in whole group format, using clarifying examples across the three subject areas to highlight procedures for determining primary and secondary matches.

If persistent inconsistencies across raters are found during the actual coding stages, targeted retraining may also be conducted.

Rating Procedures

Team Composition

Each content area and grade level will be reviewed by a team of three raters, including two content area experts and one special educator. For each rating task (e.g., 6th grade math performance events), each content expert on the team will initially rate every item independently. The special educator's role will be to consult with both of the content experts to "translate" terms that may be unfamiliar to the non-special educator. For example, content experts may need help understanding terminology related to functional applications, assistive technology, or instructional practices unique to special education.

If ratings on some items require resolution through consensus (see "reliability"), the special educator will participate with the content experts to arrive at a consensus rating.

General Sequence

1. Across subject areas, all teams will begin by rating complexity, or depth of knowledge (DOK) on the Alternate Benchmarks within the Alternate Content Standards for their grade levels and subject area(s).
2. Teams will code (1) primary content matches, (2) secondary content matches, and (3) depth of knowledge on each Alternate item.
 - In mathematics raters will work through each grade level, completing ratings of all three Alternate Assessment formats at one grade level before moving forward. Teacher Observation of Academic Skills will be first in the sequence, followed by Performance Events and Collections of Student Work.
 - Language arts experts will be responsible for rating both reading and writing assessments. They will cycle through the grade levels by rating Teacher Observation of Academic Skills and Performance Events within a grade level and subject area (e.g., 7th grade reading). Once those two formats are completed across the grades and subjects, raters will then code Collections of Student Work in a similar sequence.

What is Rated

Webb's method requires ratings of Alternate Assessment items on two dimensions: content match and cognitive demand, also referred to as complexity or depth of knowledge (DOK). Content matches will include identification of a single primary "hit" – the Alternate Benchmark that best matches the PAWS-Alt item. Raters will also be asked to identify as many secondary "hits" as they think are appropriate. Secondary hits are Alternate Benchmarks that seem to be somewhat connected to the Alternate Assessment item, but not as clearly as the primary hit.

Depth of Knowledge will be measured using a 6-point scale developed by Karvonen et al. (2006). This scale is a hybrid of Webb's four levels and Bloom's taxonomy, but extended downward in order to be sensitive to the range of cognitive demand required of students eligible to participate in alternate assessments. Each point on the scale has several associated verbs to help raters distinguish between levels of complexity.

Differences across Formats

While the same general procedures and rating criteria will apply across all PAWS-Alt formats, slight differences will occur in the procedures for each format. Sample coding forms are included in the appendix.

Teacher Observation of Academic Skills: Raters will refer directly to TOAS documents and make ratings based solely on the item listed.

Performance Event: Raters will first review one of the teacher scripts to become familiar with the assessment in general. They will then set the documents aside and make ratings based on the teacher prompt and student response listed in the coding form.

Collection of Student Work: Raters will code primary and secondary hits as well as DOK for the skill lifted from the data collection form. Once they have completed that stage, they will also make supplemental judgments about the supporting evidence in the collection. (These supplemental ratings are intended as formative

information for the state Department of Education and are not included in Webb's alignment indices.)

Reliability

Within each team (triad), all data points will be double coded by the two content area specialists. Once ratings are completed for an entire assessment form in one grade level, a DRA representative will compare ratings to identify discrepancies. Where exact agreement is not found for primary hits or adjacent agreement is not found for DOK, items will be resubmitted to the triad for discussion and consensus by the group on a final rating. Exact agreement for primary hits and exact and adjacent agreement for DOK will be calculated on initial ratings. If initial review of each coding form reveals consistent discrepancies, the consensus process will also include retraining and refinement of coding rules to improve future consistency.

APPENDIX: SAMPLE CODING FORMS

Teacher Observation of Academic Skills
Collection of Student Work
Performance Events

DRAFT

Coding Form: Items to Alternate Benchmarks

Subject <i>Circle One</i>			Grade <i>Circle One</i>							Rater ID
Reading	Writing	Mathematics	3	4	5	6	7	8	11	

Item #	Primary	Secondary (list all)	DOK
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			

Coding Form: Collection of Student Work to Alternate Content Standards
Example

Subject <i>Circle One</i>			Grade <i>Circle One</i>							Rater ID
Reading	Writing	Mathematics	3	4	5	6	7	8	11	

Collection ID: 4-1

Skill: Student will attend to a story then look at the last picture from the story.

Primary match to Alternate Standard <i>(list number, or write "X" if can't match)</i>	
Secondary matches to Alternate Standard <i>(list all that apply, or write "X" if can't match)</i>	
Depth of Knowledge (DOK) <i>(code 1-6 or X if too vague to rate)</i>	

Supporting Evidence #1:

- | | | |
|---|-----|----|
| 1. Aligns with alternate benchmark? | YES | NO |
| 2. Aligns with skill on data collection form? | YES | NO |
| 3. DOK: _____ | | |

Supporting Evidence #2:

(____ Check here if evidence #2 is identical to #1; otherwise, rate the evidence on the three criteria below)

- | | | |
|---|-----|----|
| 1. Aligns with alternate benchmark? | YES | NO |
| 2. Aligns with skill on data collection form? | YES | NO |
| 3. DOK: _____ | | |

**Reading Performance Events
Excerpted Example**

Reading Grade 7

Rater ID: _____

<i>Item</i>	<i>Question/Stimulus</i>	<i>Student Response</i>	<i>Primary AltBench Required</i>	<i>Secondary Alt Benchmarks</i>	<i>DOK Required</i>
1	Say: What is the story mostly about?	The student identified/indicated that the story was mostly about Will going on a field trip, Will and dinosaurs, or Will winning a prize (accepted any other correct answer).			
2	Say: What kind of dinosaur skeleton did Will and his class see first?	The student indicated that Will and his class saw a flying dinosaur.			
3	Say: What was the special prize in the story?	The student identified/indicated that going to Dinosaur Camp was the special prize.			
4	Say: Why is Will excited to go to the museum?	The student indicated that Will likes dinosaurs.			
5	Say: How did Will win the special prize at the museum?	The student indicated that Will's ticket was chosen in the drawing, that Will had the right ticket, or any correct response that identifies the ticket.			