

**Summary of Alaska Pilot
Alternate Assessment
January to June 2006**

Submitted by

Gerald Tindal
Dillard Research Associates
June, 2006

Dillard Research Associates
32534 Dillard Road
Eugene, OR 97405

geraldt@dillardresearchassociates.com
Dillardresearchassociates.com

Table of Contents

Pilot Web Connections

Process – Concurrent with Testing

1. February Training Evaluation
2. Bias Review Summary
3. Minutes from Meetings
4. Access Problems Noted in Emails & Individual Reports

Process – Follow-up to Testing

5. Teacher Evaluation Survey
6. Focus Group Minutes and Summary/Analysis
7. Protocol Analysis – Rescoring and Comments
8. Policy Frequently Asked Questions

Outcomes – Results

9. Demographics of Pilot Participants
10. Web Site Proficiency Results
11. Reading Data
12. Writing Data
13. Mathematics Data
14. Language Survey

Alignment – Cross Walk (Phase 1)

Pilot Web Connections

The December pilot of web connectivity was conducted with 18 individuals and 11 respondents; generally, the majority of respondents did not report serious problems, some log in problems were noted as well as long video loading times. Five more teachers reported similar issues after the pilot training.

Process – Concurrent with Testing

1. February Training Evaluation

A 14 item survey was completed immediately after the training session in February (workshop evaluation). A number of issues were rated on satisfaction: preparation, organization, rapport, response to comments, effort, contribution, access to materials, relevance, proficiency, evaluation, knowledge and skill, ability to communicate, and overall teaching of the workshop. On all of these dimensions, the average rating was about 4 (good) on the 5-point scale from the 25 teachers responding to the survey. All comments were positive.

2. Bias Review Summary

A bias review panel met immediately after the initial training workshop. A total of 44 comments were made that resulted in this many changes to the stimulus materials (the items presented to students, the wording of the teacher, or the scoring of responses).

3. Minutes from Meetings

A total 14 conference calls were made throughout the winter and spring. The issues focused mostly on training, then on the pilot, preparation for the technical advisory committee, and finally, on the next steps (FAQ and alignment with grade level expectations).

4. Access Problems Noted in Emails & Individual Reports

The pilot process was analyzed using three sources of document: (a) a list of issues compiled by Aran Felix from EED, (b) a list of issues compiled by Missy Gorden from Fairbanks, and (c) emails that had been sent during the pilot itself (February through May 12). Each comment was transcribed into an excel spreadsheet; the total was 79 general comments that spawned 87 issues in the following topics:

Training Module issues	31
Mathematics Assessment issues	18
Reading Assessment issues	14
Web Connection issues	10
Video Problems	7
Writing Assessment issues	7

The frequency of issues by source is listed below to better understand if they arose from a few individuals. Clearly, the vast majority was identified from a careful scrutiny by EED, from the conference call, and from one teacher who took great care to document the issues. Otherwise, most teachers identified a few issues only.

<u>Source</u>	<u>No. Issues</u>
Aran Felix	19
Conference Call	17
Missy Gordon	12
Pat Almond	5
Margie Fahlsey	4
Stacey Street	3
Terri Robbins	2
Darilynn Caston	2
Nan Koentopp	3
Brian Gong	2
Kris Selman	2
Elizabeth Galloway	1
Rachel Galbraith	1
Rachel Kokoske	1
Regina Feliciano	1
Kris Selman	1
Jill Besse	1
Fran Maiuri	1
Aaron Glasgow	1

These comments have been ‘essentialized’ to form 6 separate files that are being used to (a) make changes in a newly developed handbook or (b) make changes directly into the web site.

Process – Follow-up to Testing

5. Teacher Evaluation Survey

As part of the pilot, teachers were sent a survey in addition to student and administrator protocols. A total of 54 responses were made (one third [16 students] for each of reading, writing, and mathematics). This survey included 9 statements to which they logged degrees of disagreement to agreement and then reported on the time to administer, score, enter data, and note any comments. All critical issues were judged as agree to strongly agree (ease of materials access, ability to print materials, quality of copies, organization for administration, use of accommodations, attentiveness of the student, pace of the administration, ease of scoring, and the tests’ reflection of student skill). The test took about 40 minutes to administer, 40 minutes to score, and 40 minutes to enter data.

6. Focus Group Minutes and Summary/Analysis

A conference call was held on May 11 with approximately 12 teachers. The response was very positive in most aspects of the test administration and scoring. Obvious problems were noted about the video download times and some web access problems. There was general agreement that the online training was effective when supplemented (preceded) by face-to-face training. A number of issues were raised about the extremely low functioning students as well as about accommodations.

In the end, the focus group comments were sorted into the following topics. *Administration*

included use of accommodations, pilot administration, directions, general comments, and scoring. *Online system* topics involved connectivity, student descriptions, video download. *Test development* topics included bias review, use of the checklist, and content review. Finally, *training topics* included 2 days with the computer lab, face-top-face training, online training, practice tests, proficiency examinations, and feedback on scoring.

7. Protocol Analysis – Rescoring and Comments

All student protocols were re-scored and notes taken of the common mistakes. Every type of mistake was made that could be made: inappropriate recording to incorrect scoring.

8. Frequently Asked Questions

This FAQ includes issues raised during the pilot that (a) need to be addressed by DRA (involving training or test administration and scoring), (b) EED (involving policy and regulations, or (c) both DRA and EED.

Outcomes – Results

9. Demographics of Pilot Participants Students

A total of 48 districts were involved in the pilot, including students in grades 3-11 (evenly distributed). There were slightly more males (20) than females (15), with 13 not reporting gender.

10. Web Site Proficiency Results

When all the tasks and trials are multiplied in the training process, a total 1,081 instances resulted in proficiency with 34 resulting in non-proficiency. Most tasks in reading could be mastered in 1-2 trials by most teachers; several required 3-4 trials and tasks 4 and 5 required the most number of trials to gain proficiency for several teachers. Most tasks in writing could be mastered within 1 and 2 trials. Again, tasks 4 and 5 required 3-4 trials for several teachers. In mathematics, most teachers mastered most tasks in 1-2 trials, though several tasks required 3-4 trials for a number of teachers and tasks 2, 4, 17, and 19 required many trials for a few teachers.

11. Reading Data

The average performance for the students taking the various reading tests was generally above 50% correct with the number of students participating through the test gradually reducing (e.g. 25 students took task 1 and 13 took task 12). Ironically, for the 27 students who read passages, the lowest average reading production was with the easy passage (22), while students read nearly 43 words for the medium passage and 48 words for the difficult passages. With all three passages, a large standard deviation reflected a great amount of between student variance. Most students took the test in the standard condition (not administered inappropriate or proficient).

12. Writing Data

Again, student performance was above 50% for the 10 tasks in the writing portion of the assessment; the amount of variation (size of the standard deviation) was not as great in writing as it was in reading. Only slight decreasing participation occurred through the tasks (from 24 with the first task to 21 taking the last task). The vast majority of students performed in the upper ranges of each item and task. Most students took the standard administration.

13. Mathematics Data

On the 22 mathematics tasks, students performed on average above 50% with variation among students quite reasonable (comparing the standard deviation to the mean). Considerably fewer students participate in the latter tasks (fewer than 15) compared to the number completing the early tasks (typically more than 20). On most tasks, students performed in the middle to upper reaches. S in reading and writing, most students took the standard administration of the mathematics tasks.

14. Language Survey

The student language survey instrument was administered to 6 students who were judged to have anywhere from 1-8 expressive words. Approximately 4-5 of them could use beginning language to meet basic needs, greet others, seek attentions, express preferences, express emotions, label objects, acknowledge people, make introductions, relay information, obtain permission, describe action, describe events, attend to others, follow a single step command, recognize others, recognize objects or actions, know when to act, recognize location, recognize attributes as well how and why and finally, follow complex directions.

Alignment – Cross Walk (Phase 1)

In the last section of the report, the extended grade-level expectations have been cross-walked with the current extended assessment (pilot version from 2006) for both reading and mathematics. Each subject domain has two sections: (a) a list of grade-level expectations that has either a reference to a current task that addresses it OR a brief description (name) of a new task that has been developed, and (b) a mock-up of all the new tasks that have been developed to cross-walk over to the grade level expectation. Note that the list of grade level expectations are color-coded to reference similarities (the same color) across the grade levels (vertically aligning them) so that the same task (in either the current extended assessment or the new tasks) can be used to address multiple grade-level expectations.