

May 1, 2007

Catherine Freeman
Deputy Assistant Secretary
U.S. Department of Education
400 Maryland Avenue SW
Washington, DC 20202-0001

Dear Deputy Assistant Freeman:

This letter describes additional evidence that Oregon is providing in response to the U.S. Department of Education's Standards and Assessment decision letter dated June 22, 2006.

We want to affirm that Oregon is committed to a high-quality and technically sound assessment system that can serve the needs of its stakeholders. An essential component of such a system is the presence of clear communication regarding technical quality. We believe that Oregon has been able to leverage the requests for additional information from the peer review to substantially improve the documentation of the Oregon Statewide Assessment system such that it demonstrates the technical adequacy and innovation inherent in the system, meets the needs of the peer review, and provides added value for all of Oregon's stakeholders.

Oregon asks the U.S. Department of Education to recognize that this has been a particularly difficult year for Oregon. As has been well publicized, Oregon's computer-based testing contract was terminated in January. Since then we have expanded the paper and pencil testing solution (approximately 600,000 tests) over the past two months to ensure each student had the opportunity to participate in Oregon's assessment system. Despite these significant challenges, Oregon is pleased to present a full collection of evidence that addresses each of the elements of the June 22, 2006 letter from the USED.

Oregon is very pleased that American Institutes for Research (AIR) is our new testing contractor and testing under this contract will begin in the fall of 2007. Under the new contract, Oregon is on-track to implementing its full computer-based assessment system with a number of improvements to the adaptive engine and reports that will assist in creating meaningful assessment data that will support accountability, instruction and student achievement. Some of these improvements are described in sections 4.2

System Improvements

Per our previous communications with the USED, Oregon has implemented some important global improvements to its system since the peer review to address issues identified by USED:

- Oregon re-convened a Technical Advisory Committee (TAC) including national assessment experts in the fields of Computer Based Testing, Longitudinal Growth Modeling, Special Education and LEP, test administration, and performance assessments. There have been three

meetings of the committee with first meeting of this committee held on April 21 - 22, 2006, November 16-17, 2006, and April 14, 2007. The TAC has provided exceptional guidance regarding improving the technical adequacy of Oregon's system.

- Oregon is pleased to note that it contracted with AIR to create a set of technical manuals that describes the Oregon Statewide Assessment (Doc 2.1 – 2.6). The peer reviewers clearly indicated that they had difficulty locating information easily and suggested this clarification strategy. We are optimistic that the development of these manuals will also increase the confidence of our stakeholders and make review of the assessment system accessible and transparent. The technical manuals consist of eight volumes describing annual procedures and results, test development, academic achievement standards setting, reliability and validity evidence, and score reporting and interpretation. The Alternate Assessment (volumes 7 and 8 – Doc 2.7, 2.8) will be completed in summer, 2007 after the data are analyzed from the first administration in spring 2007.
- Oregon will not include targeted down assessments, modified assessments or any assessment that does not result in a valid score in 2006-07 AYP calculations. Further, the CLRAS (i.e. life skills assessments) is no longer offered as part of the Oregon assessment system. Finally, juried assessments will not be included in the 2006-07 AYP calculations nor in subsequent AYP calculations until such a time that Oregon is able to demonstrate that they are comparable with the knowledge and skills assessments.

The chart below provides a summary of the changes to the Oregon assessment as a result of the peer review.

Revised Vision of the Relationship between Oregon Content Standards and Oregon's Assessments

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9-12	targeted-down assessments
General Assessments	Aligned	Aligned	Aligned	Aligned	Aligned	Aligned	Aligned	Removed from AYP calculations
Extended (1%)	Linked to grade level content standards with depth added to provide access for students previously served by CLRAS			Linked to grade level content standards with depth added to provide access for students previously served by CLRAS				
CLRAS	Removed from AYP Calculations							

Form Comparability

For the general assessment, Oregon recognizes the importance of form comparability. Per the reviewers' request, we have attempted to provide sufficient information in the technical manual to provide assurances of comparability among Oregon's forms. The issue is complicated and required a collection of information investigating the issue from a variety of perspectives. As part of our approach, we identified two major areas of comparability that need to be addressed. First, we address the content aspects of comparability (consistency of content, depth of knowledge and difficulty). Second we address issues of scale comparability (i.e. item calibrations, differential item functioning and measurement invariance). The following table summarizes the evidence provided.

Peer Review Section	Request for Additional Evidence	Content Comparability	Scale Comparability
3.5	Additional evidence supporting the comparability of Spanish and Russian side-by-side translations with English versions	<ul style="list-style-type: none"> • Report on Analysis of Translation Accuracy (Document 6.2) • Translation review tracking forms (Document 6.3) • Test Specifications and Blueprints (Documents 3.1.1 through 3.2.7) 	<ul style="list-style-type: none"> • Analysis of DIF for Dual-Language Mathematics Assessments (Document 6.1) • Analysis of Measurement Invariance between English and Dual-Language Mathematics Assessments (Document 6.4).
4.2	Additional evidence that adaptive tests are comparable to each other and paper/pencil versions at the achievement levels, restricted to grade level content, and matched to detailed grade level test blueprints.	<ul style="list-style-type: none"> • Test Specifications and Blueprints (Documents 3.1.1 through 3.3.7) • Alignment of Oregon Content Standards and Oregon Assessments (Document 5) 	<ul style="list-style-type: none"> • Comparability of Student Scores Obtained from Paper and Computer Administrations (Document 4.1)
4.3	Additional evidence that documents consistency of strand content among the paper-and-pencil and computer –adaptive versions that includes detailed test form construction rules and test maps.	<ul style="list-style-type: none"> • Test Specifications and Blueprints (Documents 3.1.1 through 3.3.7) 	<ul style="list-style-type: none"> • Display of a random sample of individual student assessments administered in 2006-07 (Document 7) • Graphical display of the content distribution of items for all assessments administered via the computer adaptive test in 2005-06 (Document 7)
4.4	Additional evidence supporting the comparability of paper-and-pencil and computer-adaptive test difficulties between school years.	<ul style="list-style-type: none"> • Test Specifications and Blueprints (Documents 3.1.1 through 3.3.7) 	<ul style="list-style-type: none"> • Analytic and Psychometric Services: Technical Report II (AIR, October 2006) (Document 11) • Analysis of Operational Parameters over Two School Years 2005-06 and 2006-07 (Document 8)

2.1: Additional evidence including approved, re-established academic achievement standards that show alignment to the State's grade level content standards with technical and stakeholder participation

Per the requirement of the reviewers, Oregon conducted a standard setting session in December 2006 for grades 3-8 and 10 for Mathematics and Reading/Literature and for grade spans 3-5, 6-8 and grade 10 for Science. Oregon contracted with CTB/McGraw Hill to prepare and implement the standard setting session using CTB's proprietary bookmarking system. We are extremely pleased with the results and are confident in their validity. A comprehensive description of the process and outcome is described in the Standard Setting Technical Report (Doc 1.1). The Oregon State Board of Education adopted the achievement standards at their March, 2007 meeting (Doc 1.2.).

The academic achievement standards include four levels of achievement: Does Not Yet Meet Standard, Nearly Meets Standard, Meets Standard, and Exceeds Standard. Cut scores that differentiate among the four achievement levels were set through the use of the bookmark standard setting procedure. This procedure has been used in over half of the states, and it has been used previously in the Oregon Statewide Assessment System. The details of the standard setting process that includes evidence to support the procedural validity of the process can be located in the technical report prepared by CTB/McGraw-Hill (Doc 1.1). Achievement level descriptors that describe the competencies within each achievement level are found in each grade level's test blueprints (Doc 3.1.1-3.3.7).

As described in the standard setting technical documentation, the standard setting was conducted based on the current grade level content standards and incorporated a broad representation of stakeholder participation. A total of 210 panelists convened in Portland, Oregon on December 11, 12 and 13, 2006 to engage in the standard setting activity. When the committee was recruited, the ODE took care to invite participants who had a variety of backgrounds (gender, ethnicity, district size, regional location, and educational specialty such as LEP and Special Education) and roles (i.e. parents, teachers, administrators and representatives from businesses and the community). In addition, panelists with experience in instruction in grades (4, 6, and 7) were asked to serve on the panels for grades 3, 5, 8, and CIM to ensure that the panels were comprised of participants with expertise across at each grade. Pages 12-14 of the standard setting technical report show the diverse stakeholders who developed the academic achievement standards.

Panelists set the bookmarks on grades 3, 5, 8, and 10, while grades 4, 6 and 7 were interpolated as described in (Doc 1.1, page 21). The panelists were asked to review the interpolations based on their understanding of progression of achievement standards and their experience with the interpolated grades.

The academic achievement standards reflect the knowledge, skills, and abilities that students should know and be able to do as documented in Oregon's academic content standards for each grade. The bookmark standard setting procedure that is detailed in Document 1.1 includes achievement level descriptors. The standard setting procedure description along with the achievement level descriptors explicate the relationships between the current content standards and the current achievement standards.

At the end of the session, panelists were surveyed regarding the process that was used and the quality of outcomes (Doc 1.1, page 424). Overall, 83% of the participants either Agreed or Strongly Agreed and 10.1% either Disagreed or Strongly Disagreed with the statement "Overall, I am satisfied with my group's final bookmarks."

In addition, ODE engaged in an extensive external public review process after the standard setting to determine the reasonableness and validity of the bookmarked and interpolated cut scores. A website used to inform stakeholders about the standard setting session including a video documenting the entire standard setting process can be found at <http://www.ode.state.or.us/search/page/?id=849>. Of the 246 attendees of these 15 workshops around the state, 91% felt that they were fairly or very confident that the process appropriately placed the cut scores.

A follow-up survey was also sent to participants in the standard setting process to determine whether their satisfaction level maintained several months after the process. In the follow-up survey, 85% of respondents agreed or strongly agreed with the results of the standard setting.

Future Improvements

- As part of Oregon's current contract with WestEd, Oregon will establish an appropriate metric to determine whether or not a change in the content standard revision cycle (i.e. an interim change in the content standards) should prompt a change in assessment (i.e., with revised tests, achievement standards, and opportunity for system adjustment).

2.2: Additional evidence including approved, re-established alternate academic achievement standards appropriately linked to Oregon's content standards

Per the requirement of the reviewers, Oregon has revised its alternate assessment for the 2006-07 school year. The assessment is performance tasked based, focused on Academic knowledge and skills and is linked to grade level content standards. Administration of the assessment will be completed by April 30th and schools will have until May 15th to finish submitting their results to the state. Changes to the statewide assessment system were announced to a gathering of special education administrators in October 2006. The presentation summarizing the announcement is available in (Doc 9, section 2.1) or via the following link http://www.ode.state.or.us/teachlearn/testing/admin/alt/ea/updates/200607/navigating_10132006.ppt.

Blueprints for the Extended Assessment System are included in Document 9.1.1 technical document. This document details the linking process used to develop the items as linked to grade level content standards.

An independent alignment study was conducted on February 28-March 2, 2007 by Lindy Crawford, Ph.D. and Marilee McDonald from the University of Colorado at Colorado Springs and is described in (Doc 9.4.1). The study included a content match and analysis of cognitive demand, also referred to as complexity or depth of knowledge (DOK). The ratings were applied to the relationship of the items contained in the Oregon Extended Assessment to the academic grade-level content standards. Panelists were recruited and enlisted by the University of Oregon Behavioral Research and Teaching to represent subject matter expertise (reading/literature, science, mathematics, or special education) and to represent the state geographically. Twenty-seven different teachers attended the alignment workshops across the three days.

Results of this evaluation indicated that the extended assessments were successful in measuring a large percentage of the content standards within each grade level (i.e., high categorical concurrence). Standards had a higher level of representation on tests than did the individual objectives. On average, 75% of all standards within a particular content area and across three grade levels were assessed on the alternate assessment. The extended assessment was also found to have an equal distribution of items across content standards (this was found to be no less than .80 within each content standard, which is quite strong considering Webb's criteria that .70 indicates adequate balance). Finally, the

study concluded that the test items reviewed appeared to be designed effectively for testing a population of students with significant cognitive disabilities. Specifically, items were grouped within tasks and were either arranged hierarchically with the most difficult item presented last or were designed to assess fairly discrete skills (such as in mathematics) that could be demonstrated in isolation of an understanding of all of the content presented within the task. The teachers participating in the alignment study commented on this, agreeing that this better enables the tests to capture what students with significant disabilities actually know and are able to demonstrate.

Achievement standard setting for the alternate assessment will be conducted over the course of a three-day standard setting session occurring over May 21st and June 4th-5th. Documentation outlining the recruitment process and state's overview of the Bookmarking standard setting plan is included in this submission (Doc 9.1, section 1.1) and this standard setting will be conducted in a similar manner to the process undertaken for the general assessment in December (as noted above) . These alternate achievement standards will be presented to the State Board of Education for formal adoption in July, 2007. Descriptor development is in progress and will be reviewed by the Board on May 17, 2007.

Future Improvements

- Conduct achievement standard setting for the alternate assessment (May 21st, June 4 and 5, 2007)
- A technical manual documenting the extended assessment will be available in Summer 2007 after the psychometric data are analyzed.
- Due to the challenges in developing and administering a new assessment within a single year, Oregon created 2 grade groups (i.e. elementary and middle/high school). However, Oregon intends to create a separate assessment for middle and high school respectively for the 2007-08 school year. Further, because this is the first year of the administration of the assessment, there will likely be improvements to the system implemented in 2007-08 for the elementary version of the assessment based on psychometric analyses and review of the 2006-07 data. For these reasons, Oregon anticipates conducting a standard setting session for all 3 grade groups in 2007-08.
- Because Oregon's extended assessment now assesses actual grade level standards (rather than alternate content standards), there was some concern noted in the alignment study that the assessment may be considered too difficult. This will be reviewed based on stakeholder feedback and empirical data. If warranted by this review, Oregon will make adjustments in preparation for the 2007-08 assessment and the subsequent standard setting required.

2.3: Additional evidence that all students are tested on academic content standards, not just on life skills

- Oregon will not include results from CLRAS (life skills) in 2006-07 AYP calculations.
- Oregon has revised its Extended assessment to include additional depth and breadth of content (see section 2.2)
- Oregon contracted for an independent review of the extended assessment to ensure that tasks used in its extended assessments are linked appropriately to grade level content standards (see section 2.2).

3.1: Document the consistency in achievement level definitions for adaptive and paper/pencil modes

Qualitative Comparability

Oregon is pleased to provide detailed grade level test blueprints (Doc 3.1.1 – Doc 3.3.7). Although there are substantial revisions throughout these documents, some of the highlights include

- A description of the process used to ensure annual alignment of tests to grade level content standards
- The inclusion of additional exemplar items to better illustrate the eligible content
- The addition of achievement level descriptors resulting from the most recent academic achievement standards setting session (December, 2006)
- A comprehensive description of the procedures used to develop tests including the extensive role of stakeholders in the processes
- Comprehensive test blueprints that describe allocation of items by strand and difficulty for paper and computer adaptive tests
- A plan to implement item writing and test specifications that include explicit expectations for distribution of cognitive complexity
- Notation regarding emergent changes to the assessment system:
 - As described above, Oregon had to implement emergency adjustments to its assessment system given contractual disagreements with its testing contractor that resulted in contractor termination of the statewide assessments mid-way through the 2006-07 testing window. Rather than dramatically changing the test blueprints to reflect this one-time departure from established assessment practice, we have described the established blueprints and noted in the blueprints when and how the 2006-07 administration differed from typical years.
 - Under a new contract, Oregon is on-track to implementing its full computer-based assessment system in 2007-08

We note that the reviewers are specifically concerned about the degree to which Oregon's computer based tests are restricted to grade-level content standards and we have several processes in place that ensure that this is the case:

- In prior years, each grade level test was administered out of individual item pools that were aligned to grade-level content standards and it was not possible for the computer to select outside of the appropriate grade-level items. In 2007-08, a more elegant solution will be used to restrict the tests to grade level content. Oregon's new adaptive engine will include a grade level constraint within the adaptive algorithm itself (in contrast to the previous practice of constructing approximately 48 separate item pools; One pool for each content/grade/opportunity combination) so that for all ability levels, only grade level items will be selected. A description of the specifications for this new adaptive engine is provided as (Doc 4.2).
- Beginning in 2006-07 Oregon discontinued administering tests to students that were aligned to lower grade levels (i.e. challenge down); such tests are invalid and are not included in calculations of participation or performance.
- Oregon's tests are comprised of test items written by Oregon teachers to align to specific grade level standards.

We are pleased to note that Oregon contracted with WestEd to conduct an independent alignment study of its entire item operational item bank (Doc 5). WestEd analyzed the categorical concurrence, depth of knowledge, range of knowledge and balance of representation across 3,626 Reading/literature, 2,616 Mathematics and 1,161 Science items for a total of 7,403 operational items. The study found that virtually all (99% or more) of the items in Oregon's item pools have partial or strong alignment to Oregon's grade level content standards. The study did note several key issues to which Oregon must attend:

- There were differences between the test blueprints and the allocation by strand within item pools. However, Oregon asks the reviewers to note that WestEd's independent analysis of Oregon's assessments to Oregon's test specifications was based on WestEd's coding of the content standard assessed by Oregon's items at *the item pool level*. This result is expected given that final administration of tests to students may not be exactly the same proportion of content delivered in actual tests to students as is available in item pools. As an alternate view, Oregon conducted a separate evaluation of the adherence of actual tests administered via Oregon's computer-based test to the blueprints. See (Doc 8).
- Oregon did not have an a priori allocation of items by depth of knowledge nor was the judged distribution of depth of knowledge consistent across grades
- A very small number of items that were found not to align to grade level standards

Oregon's plans for addressing these issues are listed below under Future Improvements.

Scale Comparability

Oregon completed a comprehensive analysis regarding the comparability of administration methods by comparing the standard mode of assessment (i.e., the computer-based adaptive test) with the paper and pencil form that is created for and administered to an increasingly small number of students and schools who need paper tests to access the system (Doc 4.1).

This study explored the comparability of scores on the state Math, Reading and Science tests across the computer-based adaptive test and the paper form of the same test. While the computer and paper tests are built to the same grade-level test specifications, monitoring for administration method effects ensures that unanticipated effects of test mode can be identified and mitigated to have minimal impact on test scores.

Results suggest that average scores and standard deviations are quite similar across administration methods at all grades and subjects. Although the differences were quite small (less than half a scale score point), third graders tended to show slightly larger average differences. None of the differences were statistically significant, and all scores correlated highly with each other.

This study provides evidence that scores are comparable across TESA and paper delivery specifically, and together with other comparability studies the state has completed, contributes to demonstrating the validity and technical adequacy of the entire system.

Future Improvements

- AIR will conduct simulations to demonstrate the likely distribution of items by strand and difficulty to ensure that the adaptive engine and item pools are creating tests are aligned to grade level and otherwise also consistent with the test blueprints
- AIR will independently review and document the consistency between Oregon's tests (both computer-based and paper) created by Oregon content specialists and Oregon's test specifications prior to the operational administration. This process will ensure continuing alignment of all paper forms in the future.
- Oregon will replicate the analysis described in (Doc 8) to monitor the content administered via the computer-based tests periodically throughout the year.
- The more advanced and elegant adaptive engine (Doc 4.2) in combination with larger item pools will allow tighter content constraints (i.e. alignment to test specifications across strand) while improving measurement precision for the composite score and strand scores across all ability levels.

3.2 Additional evidence supporting the comparability of tests based on the extended content standards and the Extended Career and Life Role Assessment (CLRAS)

- Oregon will not include results from CLRAS (life skills) test.
- Oregon has only one alternate assessment in 2006-07 that is based on academic content standards

3.3: Additional evidence supporting the comparability of the plain language and regular test forms modes

- As of July 1, 2004, plain language was not offered as a separate form. Rather, all of Oregon's math and science tests exclusively contain items written according to universal design principles, including plain language.

3.4: Additional evidence supporting the comparability of the Juried assessments and the Knowledge and Skills Tests

- As of October 2006, Oregon has suspended the Juried Assessment for the purposes of AYP calculations. Oregon hopes to reinstate the Juried assessments at such a time that it can demonstrate comparability with the Knowledge and Skill Tests.

3.5: Additional evidence supporting the comparability of Spanish and Russian side-by-side translations with English versions

Qualitative Comparability

Oregon contracts with an independent third party vendor that specializes in translations to translate each of its Mathematics and Science items into Spanish and Russian. Upon receiving these translations, Oregon employs bilingual Oregon educators to review each translated item and make any corrections necessary for accuracy and common usage. Based on the reviewers' questions regarding the comparability of the translated tests, Oregon now tracks each change recommended by the bilingual educators. The tracking sheet includes the number of words that were changed and a notation if the change occurs in the stem, graphic or responses and denotes whether the change is a word addition, deletion, or correction. Copies of the translation review tracking sheets are provided as Doc 6.3. Because such detailed documentation is not available for items translated prior to 2006-07, Oregon contracted with AIR to conduct a study of translation accuracy to provide additional evidence regarding the quality of the translations. The results of this study are included as Doc 6.2. Finally, as noted in our test specifications for Mathematics and Science (Docs 3.1.1 – Docs 3.2.7) the dual language assessments are developed with the same test specifications as are the English-only assessments.

Scale Comparability

Oregon demonstrates scale comparability by virtue of a collection of evidence. While we would prefer to demonstrate comparability for both Spanish and Russian dual language forms, the number of students in Oregon who take the Russian forms is extremely limited.

Number of Tests Administered in 2005-06 in Russian, Spanish and English

Grade	Russian		Spanish		English	
	Mathematics	Science	Mathematics	Science	Mathematics	Science
3	28		3,050		75,364	
4	36		2,115		71,089	
5	31	28	1,957	963	77,783	56,193
6	23		1,527		72,981	
7	12		1,231		75,091	
8	21	13	1,176	539	83,747	66,752
CIM (9-12)	43	8	1,635	1047	113,058	73,854

Given the small number of Russian tests administered, it is very unlikely we would have the statistical power to detect even extremely large differences between English and Russian forms. However, Oregon does use the same process for developing and administering the Russian forms that it uses for the Spanish forms. Therefore, in the absence of alternatives, we suggest it is reasonable to extend conclusions regarding comparability to both languages based on the evidence Oregon provides for the comparability of English and Spanish forms. The independent translation review conducted by AIR (Doc 6.2) also provides evidence of comparability of the Russian mathematics tests.

The description of a series of comprehensive analyses (Doc 6.1) demonstrates the comparability of the English-only and the dual-language (English-Spanish) forms at the item level. To do this, we frame the difference in forms as a type of differential item functioning (DIF) and treat the item parameters derived from the English form as the referent group and item parameters derived from the dual language form as the focal group. We analyze the differences in the parameters using three approaches, Mantel-Haenszel procedure, Logistic Regression procedure and examination of the Item Characteristic Curves. The first two approaches allow for a statistical test of the disadvantages and advantages that the group of interest has on an item when compared to the rest of the population. The graphs of the ICC curves assist in describing the specific relationship between student ability levels and differences in item difficulty.

Results from the analysis show the DIF displayed in Oregon's dual language tests is well within the bounds established by results found in previous research on translated test items. DIF by itself does not provide conclusive evidence of non-comparability. Items that exhibit moderate to high DIF (using the Educational Testing Service criteria) should be examined by content experts with linguistic competence in both languages to determine if the item translation is at fault or if the difference in performance may be attributed to differences in students' opportunity to learn. This type of analysis is described in the decision matrix that takes into account both the DIF analysis and independent translation review (Document 6.1, Appendix A).

Multi-group confirmatory factor analyses (CFA) were used to evaluate whether construct invariance could be established between the English-only and dual-language (English-Spanish) versions of the Mathematics tests, the results are given in Document 6.3. The evaluation incorporated a rigorous evaluation by constraining factor loadings, means, and residual variances to be equal across both groups. This methodology provides the strongest possible evaluation of score comparability. Results of the analyses indicate there is strong evidence to support a determination of "Strict Invariance" between the English-only and dual-language forms for grades 6-8 and 10 and evidence for invariance of most model parameters in grades 3-5. There was marginal evidence for differences in strand score means between the two forms in grades 3-5. These differences may well be due to construct relevant differences between form groups including sample size, local testing decisions, demographics, level of English proficiency, opportunity to learn and numerous other concomitant variables. Given the non-

random assignment to group, results of the analyses suggest a high degree of score comparability across forms.

Future Improvements

- Replicate these analysis on the dual language Science assessments administered in 2007-08
- Refine the confirmatory factor analyses to investigate the possible sources of differences in the means and variances
- Work with stakeholders to conduct a controlled accommodation study that incorporates dual language forms
- Implement in 2007-08 a data review board that will assist in evaluating sources of DIF

4.1 Additional evidence for each assessment, including alternate assessments, that documents the standard setting process with descriptions of the selection of judges, methodology employed, and final results.

- Oregon conducted a comprehensive standard setting for reading/literature, mathematics and science in December, 2006 (Doc 1.1). Please see section 2.1 of this letter and volume 3 of the technical manual (Doc 2.3) for an additional description of the standard setting process
- Oregon revised its Extended Assessment to include additional depth and breadth of content (see 2.2) and Oregon will conduct a standard setting for the revised Extended Assessment (May 21st, June 4 and 5, 2007, see Section 2.2)

4.2 Additional evidence that adaptive tests are comparable to each other and paper/pencil versions at the achievement levels, restricted to grade-level content, and matched to detailed grade level test blueprints.

- Please see 3.1 and 3.5 and the following documentation previously described: description of the grade level blueprints (Docs 3.1.1 – 3.3.7), the previous and future versions of Oregon's computer-based adaptive engine (Doc 4.2), the alignment study conducted by WestEd (Doc 5), and Oregon's study of computer-based test content as administered to students (Doc 7)
- As described in 3.1, Oregon demonstrates the comparability between the state computer-based adaptive assessment and the fixed-form paper test across tests administered on separate occasions. This evaluation is the most stringent evaluation of the adaptive engine. In this case, potential sources of construct irrelevant variance included both mode of administration and item selection. If the adaptive test were functioning differently for students (i.e. if adaptive tests were not comparable to each other) the paper form would likely also not be comparable to the adaptive form. In contrast, Oregon's study demonstrates that for an individual student, the ability estimates derived from the adaptive test are not significantly impacted by mode of administration, and that the use of an adaptive test results give similar estimates of ability as does a paper administration of test, even across multiple administrations.
- As described in 3.5 Oregon presents strong evidence of measurement invariance for mathematics between the English-only form and the Spanish Dual-Language form. Recall that while both are computer-based, Oregon's English tests are adaptive while the Spanish dual-language form is fixed-form. In addition to the evidence provided earlier describing the comparability of computer-based adaptive vs. paper administration methods, this analysis also provides evidence of the comparability between computer-based adaptive and fixed-form tests. The finding of measurement invariance (Doc 6.4) and reasonable DIF (Doc 6.1) provides additional evidence of the comparability of adaptive tests to each other.
- Beginning with 2007-08, performance of the adaptive algorithm will be automatically monitored and reported on, providing on-going documentation that on-line tests are being administered as designed. Consistent performance of the algorithm will provide evidence that the adaptive tests are

- 1) comparable to each other across tests administered within year, across year, and for all achievement levels, 2) that tests are administered according to detailed test blueprints, and 3) are restricted to grade-level content.
- Finally, Oregon had planned a study based on a complex design provided by AIR to be administered in the middle of March, 2007. The study would have provided an ability estimate for a student based on a fixed computer test as well as an adaptive computer based test within a single testing session and the correlation of ability estimates derived from this study would have provided substantial evidence of comparability. However, Oregon's computer-based testing contract was terminated prior to the implementation of the study.

Future Improvements

- Oregon will implement the AIR adaptive engine study in 2007-08 and will report to USED the results
- Oregon will periodically replicate the studies for a random sample of grades within each content area to ensure continued comparability across the computer-based adaptive and paper tests
- Oregon will leverage a data review board to help evaluate and monitor administration mode as a source of DIF

4.3 Additional evidence that documents consistency of strand content among the paper-and-pencil and computer-adaptive versions that includes detailed test form construction rules and test maps

- Please see 3.1 and 3.5 and the following documentation previously described: description of the grade level blueprints (Docs 3.1.1 – 3.3.7), the previous and future versions of Oregon's adaptive engine (Doc 4.2). The alignment study conducted by WestEd (Doc 5), and Oregon's study of tests administered to students (Doc 7)

4.4 Additional evidence supporting the comparability of paper-and-pencil and computer adaptive test difficulties between school years

- For both paper and pencil tests and computer-based tests, the item banking process contributes to whether or a not a test is comparable across years. Item banks are built to the same specifications each year, as are the specifications from which individual tests are constructed as described in the technical manual. Oregon included a description of its scaling process in its technical manual (Doc 2.1 -Volume 1, 4.1.2)
- Information on the stability of item parameters over time as evidence of comparability of computer-adaptive test difficulties between school years is provided in (Doc 8). Based on a pilot study conducted by American Institutes for Research (Doc 11), the present analysis extends the analysis to cover school years 2005-06 and 2006-07 at all grades for reading and mathematics. Items used in two successive years were evaluated to investigate the stability of item difficulty estimates. Similar to the AIR study, this additional analysis replicates the findings that the items are very consistent over time.
- For 2006-07, Oregon had anticipated less than 1% of its tests being completed via paper and pencil and therefore planned to use a virtually identically form as was used in 2005-06. Once the data are scored from the 2006-07 May administration, Oregon will replicate the above study for paper and pencil forms.

Future Improvements

- Annually include in our technical manual correlations of item calibrations between current and prior years for both computer-based and paper and pencil tests

- Leverage a data review board and technical advisory committee to identify items that should be retired based on item exposure or item performance. In addition the committees will identify items that should be recalibrated based on item performance and suggest appropriate remedies in the item writing process based on the findings from the committees

4.5 Documentation that CLRAS (life skills) scores do not count for AYP either alone or in combination with extended assessment scores.

- Oregon will not include results from CLRAS (life skills) in 2006-07 AYP calculations.

4.6 Documentation that supports the reliability and validity of alternate assessments

- Six levels of review were conducted on the Extended Assessment items as an initial determination of validity based on accuracy of the items, alignment (linkage) to the standards, and appropriateness for the purpose of the assessment (content) and for the populations involved. The six levels of review were:
 - Initial Content (Panel) Review
 - Standards coverage and vertical alignment (internal department) review
 - Secondary Content and Appropriateness (teacher/user) review
 - Sensitivity review for students who are Deaf and Blind
 - Alignment study
 - Final Content and Population overview
- Technical documentation for the development of Oregon’s alternate assessment is included in Document 9.3.1. This document articulates the development and design process for the Extended Assessments, including scoring and administration justifications and support as well as discussions of studies designed in support of initial validity arguments such as content and alignment reviews, pilot testing, sensitivity reviews, and special education reviews. Technical support is also included in the Alignment study included in Document 9.4.1 that describes the link between the items as developed for this population and the state’s grade level content standards developed per subject area. Technical studies in support of the reliability for this new assessment are proposed for August 2007 and will include inter-rater reliability studies using video administration, split half reliability studies, internal consistency studies, cross-task consistency
- Further documentation is also available in the administration manuals included as components of the assessment (Document 9.5.1 & 9.5.2). Individual subject area administration manuals are included in Document 9.6.1 – 9.6.7. Details and agendas reflecting the content review and final review processes are included in Document 9.7.1.
- The state’s requirement for the inclusion of accommodations in a student’s instructional program is included in assessment and accommodation trainings that are conducted in conjunction with the assessment. When trainings are not being conducted, this intent is reflected consistently in the accommodations table as a recurring footnote that states: “The intent of this accommodation is not to be restricted to students on an IEP, but to otherwise allow for consistency between instruction and assessment” The Accommodations Tables are available at <http://www.ode.state.or.us/teachlearn/testing/manuals/tables/asmtwraccomtable0607.pdf>

5.1 Document the alignment of 3-8 and high school assessments in reading/language arts and mathematics with academic content standards and with the re-established academic achievement standards.

- As described in section 2.1 Oregon contracted with WestEd to conduct an independent alignment study of its entire item operational item bank (Doc 5). WestEd analyzed the categorical concurrence, depth of knowledge, range of knowledge and balance of

representation. Further, Oregon conducted an extensive field review of the proposed achievement standards (also described in section 2.1).

5.2 Document the alignment/linkages of the Oregon alternate assessments to the State's academic content standards and to re-established academic achievement standards.

- Per the requirement of the reviewers, Oregon has revised its alternate assessment for the 2006-07 school year. The assessment is performance task based, focused on academic knowledge and skills, and is linked to the grade-level content standards. Additional information is provided above in section 2.2.

7.1 Document that achievement level descriptions appear on all student/parent reports

- Revised reports will include the achievement level description for each student. Drafts of these reports are included in (Doc 10.1). In addition, draft reports that will be made available computer-based for teachers use and to have available for parents use are provided in (Doc 10.2).

7.2 Document the existence of parent reports for extended assessments and for CLRAS

- Parents of students who take the extended assessment will receive reports in the same format as do parents of students who take the general assessment (Doc 10.1) with the exception being that the alternate achievement level descriptors will appear for students taking the alternate assessment. Although there were specific parent reports related to the alternate assessment, they will need to be revised in summer, 2007.

7.3 Document that alternate assessment performance ratings are tied to NCLB achievement levels used for reporting

- As described in the response to 2.2 above, achievement standard setting for the alternate assessment will be conducted over the course of a three-day standard setting session occurring over May 21st, June 4th and June 5th. Documentation outlining the recruitment process and state's overview of the Bookmarking standard setting plan is included in this submission (Doc 9.1, section 1.1) and this standard setting will be conducted in a similar manner to the process undertaken for the general assessment in December (as noted above) . These alternate achievement standards will be presented to the State Board of Education for formal adoption in July, 2007. Descriptor development is in progress and will be reviewed by the Board on May 17th 2007.

Oregon has incorporated the recommendations of the peer reviewers to make substantial improvements in Oregon's technical documentation and processes for its assessment system. We would appreciate the opportunity to discuss any elements of concern regarding this response such that we may work together to ensure that Oregon's stakeholders have access to a high quality assessment system that meets the criteria for the standards and assessment review.

Sincerely,

Anthony Alpert
Director of Assessment, Oregon Department of Education

Cc: Superintendent Castillo
Governor Kulongoski
Ed Dennis
Patrick Burk
David Harmon
Doug Kosty