

June 28, 2010

David Harmon  
U.S. Department of Education  
400 Maryland Avenue SW  
Washington, DC 20202-0001

Dear Dr. Harmon:

This letter describes the evidence that Oregon is providing that demonstrates the technical quality of its Grade 3 Spanish Reading/Literature assessment. We appreciate the opportunity to receive technical assistance regarding this submission and look forward to working with you and others at the Department to ensure that Oregon provides its stakeholders with a Grade 3 Spanish Reading/Literature assessment that addresses the requirements of the law and meets their unique needs.

Please let me know if I can provide any additional information or answer any questions you might have.

Thank you very much for your support,

Tony Alpert  
Director of Assessment  
Oregon Department of Education

## **Oregon's Grade 3 Spanish Reading/Literature Assessment: Evidence of Compliance with NCLB**

### **Assessment of Grade 3 Reading/Literature Content Standards in a Spanish Native-Language Assessment**

The No Child Left Behind Act (NCLB) Section 1111(b)(3)(C)(ix-x) authorizes eligible students to test in their language of origin to increase the validity of the accountability designations that use the data resulting from required assessments. Absent native language assessments, SEAs have greater uncertainty as to whether student performance is the result of content knowledge or proficiency in the English Language.

Six overarching Common Curriculum Goals (CCGs) provide the framework for Oregon's Grade 3 Reading/Literature assessment. After a review of these goals and the standards that give them further specificity, Oregon felt justified in moving forward to develop a Grade 3 Spanish Reading/Literature assessment for students who might benefit from such an assessment. In reaching this conclusion, Oregon considered each of the following factors:

- Oregon's Reading content standards are not restrictive in terms of the language in which a student reads. The standards address measuring comprehension of informative and literary text, but do not specify that the text be written in English.
- As a common practice, Oregon tries to eliminate construct-irrelevant variance, such as English language proficiency, when measuring the skills set out in the content standards.
- Bilingual educators across the state have pointed out that the information provided by the Spanish reading assessment would be more valuable for Spanish-speaking students and their teachers than the information provided by the English assessment.

In summer 2009, Oregon collaborated with the U.S. Department of Education to convene a panel of experts to advise Oregon on the development of a Grade 3 Spanish Reading/Literature assessment aligned to Oregon's reading content standards that meets the requirements of NCLB.

The priority of the committee was to determine the method of test design that would achieve rigorous comparability between the English and Spanish versions of the test while still meeting the needs of stakeholders. The committee discussed two options. The first option discussed was to create a test comprised of only authentic Spanish passages. However, creating a completely new test posed at least 2 threats to comparability. The first threat to validity was that without common passages and items, Oregon would likely only be able to use social moderation during standard setting to determine comparability. While social moderation can be acceptable, there is limited quantitative evidence that can be collected to support that the resulting cut-scores are comparable. The second threat identified was that creating a separate new test would pose challenges in creating a new separate item pool that was sufficient in size to support an adaptive test and was similar in composition to the English item pools.

The committee determined that the alternate option of translating existing English passages into Spanish created the most likely opportunity for demonstrating comparability. The committee recommended the following approach to the project:

1. Establish clear decision rules for determining which passages are ineligible for translation. In addition, establish clear protocols for translation such that difficulty, complexity, and meaning are maintained as much as possible
2. Use Oregon's well established process to engage an expert translation contractor to translate the passages and associated items. Consistent with Oregon's process engage expert local educators to review the translations based on the established protocols.
3. Allow districts to administer the test to students who would receive a more accurate assessment score via a test provided in Spanish. However, until the USED approved the test for use in AYP designations, all students would also take the English version of the reading assessment. Since Oregon's item banks are of sufficient size to support multiple testing opportunities, students would be able to take the English test up to two times in addition to the Spanish test without seeing items repeatedly.
4. Oregon would demonstrate comparability based on the following pieces of evidence
  - a. Alignment of items to content standards
  - b. Accuracy of translation
  - c. Consistency of tests with the English reading assessment blueprint as described by allocation of content and reporting of results
  - d. Consistency in the underlying measured construct as demonstrated by multi-group confirmatory factor analysis
  - e. Consistency in the difficulty item parameters as calibrated among the English and Spanish versions respectively
5. In 2010-11 Oregon would field test items on the English reading test that would incorporate authentic Spanish passages that were translated into English prior to field testing.

### **Background on the General Reading/Literature Assessment**

Oregon administered a Reading/Literature assessment to all students in the benchmark grades of 3, 5, 8, and 10 from 1991 until 2004. From 2005 to 2010, Oregon has administered a Reading/Literature assessment to all students in grades 3 – 8 and 10. In February 2010, Oregon received permission from the U.S. Department of Education to change the grade of accountability for high school from grade 10 to grade 11. Starting in 2010-2011, Oregon will administer the Reading/Literature assessment to all students in grade 11 instead of grade 10. In December 2006 Oregon conducted an alignment and comparability study and additional standard setting session to address process concerns that were raised during the peer review for Oregon's Mathematics and Reading/Literature assessments. In 2007, Oregon completed negotiations with American Institutes for Research (AIR) for administration of Oregon's general education online assessments through a secure online testing system (OAKS Online). In 2009-2010, over 99% of general education Reading/Literature assessments were administered online. In 2010-2011, only students with visual impairments who need the Braille version of the assessment will take the Reading/Literature assessment in paper/pencil Braille format. All other students will take the Reading/Literature assessment online via OAKS Online.

Please be aware that although we have tried to be consistent in our vocabulary, there are several changes that have recently occurred that might cause confusion absent a brief overview of Oregon terms.

### **Summary of Evidence Submitted in Response to Standards and Assessments Peer Review Guidance**

The remainder of this letter highlights the evidence and cites documents provided to the U. S. Department of Education supporting the use of Oregon's Spanish 3<sup>rd</sup> grade Reading/Literature assessment under NCLB. The discussion is organized around the critical elements listed in the Peer Review Guidance.

### 3.5

**If its assessment system includes various instruments (e.g., the general assessment in English and either a native-language version or simplified English version of the assessment), how does the State demonstrate comparable results and alignment with the academic content and achievement standards?**

Oregon offers the following evidence demonstrating that Oregon's English and Spanish Reading/Literature assessments are comparable:

- Alignment of Oregon Content Standards and Oregon Third Grade Spanish Reading Tests (Doc.1)
- Alignment to Oregon Reading Content Standards: Grade 3 Spanish Reading Pilot, 2009-10 (Doc. 2)
- Summary of Score Comparability Analyses Report (Doc. 3)
- Reading Test Specifications and Blueprints (Grade 3) (Doc. 8)

#### Qualitative Evidence of Comparability

In 2007, ODE contracted with WestEd to conduct an independent alignment study analyzing the degree of alignment between Oregon's English Reading/Literature assessment item pool and Oregon's Reading/Literature content standards. Six overarching Common Curriculum Goals (CCGs) provide the framework for Oregon's Grade 3 Reading/Literature assessment:

- Determine the meaning of new words by applying knowledge of word relationships and context clues.
- Find, understand, and use specific information in a variety of texts across the subject areas to perform a task.
- Demonstrate general understanding of grade-level informational text across the subject areas.
- Demonstrate general understanding of grade-level literary text.
- Develop an interpretation of grade-level informational text across the subject areas.
- Develop an interpretation of grade-level literary text.

For Grade 3, WestEd found that 99% of the items analyzed had a strong alignment to a strand. About 25% of the items that are included on Oregon's Grade 3 Spanish Reading/Literature pilot assessment were included in WestEd's review in their English version and subsequently translated into Spanish. The remaining items were written after 2007 and translated to Spanish (Doc 1: Alignment of Oregon Content Standards and Oregon Third Grade Spanish Reading Tests). A description of Oregon's item development process is located in Section 6 of Oregon's Technical Manual Volume 2: Test Development (Doc. 5.1, p. 13).

To confirm that those items from the Grade 3 Spanish Reading/Literature item pool written after 2007 were written following the same process reviewed by WestEd, Oregon's Technical Advisory Committee advised Oregon to supplement the WestEd analysis with an alignment analysis of a random sample of these translated items. This item-level analysis focused on the issue of categorical concurrence (the extent to which each item measured its intended CCG and content standard). In 2010, Oregon contracted with the Educational Policy Improvement Center (EPIC) to conduct an independent alignment study of the items that were included in the Grade 3 Spanish

Reading/Literature assessment operational item bank (Doc 2: Independent Item Alignment Analysis by EPIC). EPIC analyzed the categorical concurrence of a random sample of 25 Spanish Reading/Literature items. The study found that all 25 items strongly aligned to Oregon's grade level content standards.

All 169 items in the Grade 3 Spanish Reading/Literature assessment have both an English and Spanish version of the items. The items have been reviewed for content alignment in both their English and Spanish form, and the translation between the two languages have been checked to determine if they are comparable.

The process for creating the Spanish Reading/Literature assessment was reviewed by an external expert who affirms that the process is sound and should continue to create as products assessments that are aligned to Oregon's content standards (Doc. 1).

### Quantitative Evidence of Comparability

As part of the submission of evidence for its Grade 3 Spanish Reading/Literature assessment, Oregon has completed a confirmatory factor analysis based on the 2009-10 assessment.

Multi-group confirmatory factor analyses (CFA) were used to evaluate whether construct invariance could be established between the English and Spanish versions of the reading assessment, the results are given in Document 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 evidence to support a determination of at least "Metric Invariance" sometimes know as "Weak Invariance" between the English-only and dual-language forms. Metric invariance is used to describe that the subscales of the reading assessment are providing similar magnitudes of information in comprising the total reading/literature scale score. The next level of invariance (i.e., "strong" invariance) would require that the means of the groups taking the Spanish and English versions respectively were equal (Vandenberg & Lance, 2000; see Doc 3). However, that is not the case for these populations given the non-random assignment to groups, results of the analyses suggest a high degree of score comparability across forms. According to the literature, it is necessary stop this hierarchal testing of sequential models of invariance after one of the steps is not affirmed.

Although, the CFA was only able to support "Metric Invariance," this affirms that users are able to use the results of the assessments to make comparable decisions about scores since they are on the same scale as the English Reading test and have the same factor structure. The report Summary of Score Comparability Analyses (Doc 3) includes the complete description of methodology and results.

For additional evidence of comparability of the native-language reading assessment based on differential item functioning (DIF), please refer to Critical Element 4.3 (c) below.

#### **4.1**

**For each assessment, including all alternate assessments, has the State documented the issue of validity (in addition to the alignment of the assessment with the content standards), as described in the *Standards for Educational and Psychological Testing* (AERA/APA/NCME, 1999), with respect to all of the following categories:**

**(a) Has the State specified the purposes of the assessments, delineating the types of uses and decisions most appropriate to each? and**

The Grade 3 Spanish Reading/Literature assessment has the same purposes as the rest of the OAKS, found in Volume 5 of Oregon's Technical Manual (Doc 5.2: Technical Report Volume 5: Test Administration, Section 3, p. 2):

Oregon's Assessments of Knowledge and Skills (OAKS) are summative assessments, which are assessments of learning generally carried out at the end of an instructional period. Summative assessments are typically used for program accountability and to assign achievement level scores to students. Summative assessments are not designed as diagnostic tools for student placement or as formative assessments. Given the specific focus and purpose of summative assessments, the OAKS can only be used as part of a collection of evidence regarding the academic needs of individual students. The primary purpose of the OAKS is to ascertain the achievement level of individual students and compare that achievement with the Achievement Standards established by the State Board of Education.

Additional context regarding the purpose of OAKS can be found in Volume 1 of Oregon's Technical Manual (Doc 5: Technical Report Volume 1: Annual Report)

The Grade 3 Spanish Reading/Literature native-language assessment will increase access to OAKS by providing Spanish-speaking 3<sup>rd</sup> grade students with an opportunity to demonstrate their proficiency in Oregon's Reading/Literature content standards in Spanish. Consistent with the federal law, only the native language test scores from LEP students who first enrolled in a U.S. School within the last 5 years will be incorporated in the AYP designations. Test scores from students who don't meet these criteria will be invalidated and will not count toward participation or performance calculations. In addition to LEP students, some districts with Spanish language immersion programs have chosen to administer the Grade 3 Spanish Reading/Literature to its 3<sup>rd</sup> grade students as one of their three Reading/Literature testing opportunities as part of their program evaluation (even though these test scores will be invalidated for state use as described above)(Doc 5.1: Technical Report Volume 2: Test Development, p. 22).

**(b) Has the State ascertained that the assessments, including alternate assessments, are measuring the knowledge and skills described in its academic content standards and not knowledge, skills, or other characteristics that are not specified in the academic content standards or grade-level expectations? And**

Oregon offers the following evidence demonstrating that the Grade 3 Spanish Reading/Literature assessment is aligned to Oregon's content standards:

- Reading Test Specifications and Blueprints (Grade 3) (Doc. 8)
- An Alignment of Oregon Content Standards and Oregon Third Grade Spanish Reading Tests (Doc.1)
- Alignment to Oregon Reading Content Standards: Grade 3 Spanish Reading Pilot, 2009-10 (Doc. 2)

Oregon's item development process is described in detail in the Reading Test Specifications and Blueprints (Grade 3) (Doc. 8).

In addition to analyzing Oregon's general item development process described in the Reading Test Specifications and Blueprints (Grade 3) (Doc. 8), the Alignment of Oregon Content Standards and Oregon 3<sup>rd</sup> Grade Spanish Reading/Literature assessment (Doc.1) provides a discussion of the specific item selection and translation process followed by Oregon in developing the Grade 3 Spanish Reading/Literature assessment.

This report contends that because Oregon adhered to the same item development process followed in developing those items found by WestEd to align to Oregon's content standards in 2007, the items from the Grade 3 Spanish Reading/Literature assessment should also align.

The 2007 WestEd alignment analysis of Oregon's items found a high categorical concurrence between the items and the content standards for all subjects, including Grade 3 Reading/Literature, as well as an appropriate range of knowledge and breadth of knowledge in the item bank. Oregon's content standards have remained the same since 2007 and the development and updating of the item bank has followed the same procedures used to create the bank used for WestEd's study. Consequently, the current item bank should have the same alignment characteristics. Because the Grade 3 Spanish Reading/Literature item bank represents a subset of the Grade 3 English Reading/Literature item bank, the content of the items in the Spanish bank are parallel to the content of the items in the English bank. As the table below shows, the distribution of items by content standard for the Grade 3 Spanish Reading/Literature assessment is very similar to the one for the English Reading/Literature item pool, indicating that the Spanish test should assess the same breadth of content as the English version of the test (Doc 1, p. 12).

Common Curriculum Goal <sup>1</sup>	Content Standard	English Item Count	Percent of Items in English Pool	Spanish Item Count	Percent of Items in Spanish Pool
VOCA	1.1	48	9%	12	7%
VOCA	1.2	100	18%	31	18%
READ	2.1	20	4%	12	7%
READ	2.2	35	6%	12	7%
READ	2.3	16	3%	7	4%
DGU	3.1	54	10%	12	7%
DGU	3.2	23	4%	4	2%
DGU	3.3	8	1%	4	2%
DGU	3.4	10	2%	3	2%
DGU	3.5	26	5%	9	5%
DGU	3.6	42	8%	18	11%
DAI	4.1	33	6%	7	4%
DAI	4.2	45	8%	10	6%
DAI	4.3	27	5%	8	5%
DAI	4.4	26	5%	7	4%
DAI	4.5	13	2%	3	2%
DAI	4.6	28	5%	10	6%
Total (% do not add to 100 due to rounding)		<b>554</b>	100%	<b>169</b>	100%

<sup>1</sup> Common Curriculum Goals are: Vocabulary (VOCA), Read to Perform a Task (READ), Demonstrate General Understanding (DGU), and Develop an Interpretation (DAI). For detailed definitions of the reading content standards, see Doc 2, Appendix A.

The Alignment to Oregon Reading Content Standards: Grade 3 Spanish Reading Pilot, 2009-10 (Doc. 2) tested the premise raised in the Alignment of Oregon Content Standards and Oregon Third Grade Spanish Reading Tests (Doc.1) by analyzing the categorical concurrence of a random sample of 25 Spanish Reading/Literature items. This report concludes that all 25 items strongly aligned to Oregon's grade level content standards, supporting the argument that Oregon's item writing process continues to maintain a high degree of alignment to Oregon's content standards.

During test administration, the Grade 3 Spanish Reading/Literature test blueprint guides the OAKS Online adaptive test delivery system to ensure that students are presented with items aligned to each of the Grade 3 Reading/Literature content standards. As with more traditional test blueprints, OAKS Online uses the following types of rules to determine which items are presented to each student during a given testing opportunity:

- Length of the test
- Content areas to be covered and acceptable range of items within each content area
- Acceptable range of item difficulty for the specified grade level
- Items that cannot appear on the same test
- Number and location of field-test items, if applicable

The Spanish Reading/Literature item test pools were designed to be equivalent to the English Reading/Literature item pools in proportion of content and breadth of difficulty so that students can meet the standards regardless of which items the test engine presents to them. Item selection is limited to items written for the specified grade level and is constrained to represent the test specifications, ensuring the appropriate representation of each SRC and coverage of the specified breadth and depth. Each item pool contains approximately 500 items by grade and content area, a sufficient number of items to ensure students are presented with a test representing the breadth and depth identified in the test specifications and content standards, regardless of the item difficulty. Section 8.3 of Oregon's Annual Report Volume 2: Test Design (Doc 5.1) provides greater detail on how the OAKS Online Adaptive Item Selection Algorithm works to ensure that only items that measure the knowledge and skills described in Oregon's content standards are assessed.

**(c) Has the State ascertained that its assessment items are tapping the intended cognitive processes and that the items and tasks are at the appropriate grade level? and**

Oregon's Reading/Literature test pool is designed so that each test includes items that encompass a range of depth of knowledge and level of difficulty. The three depth of knowledge levels (also referred to as cognitive complexity) used to design and describe Oregon's multiple choice test items in Reading/Literature are:

- Recall/Literal Comprehension: Item requires student to recall a fact, information, or procedure from the passage.
- Skill and Concept: Item requires a student to use a skill or concept, or to infer information not directly stated in the passage.
- Strategic Thinking: Item requires a student to use analytical skills to support a conclusion about the author's purpose or craft.

Oregon's Reading/Literature test specifications dictate the distribution of items by intended cognitive process through the test blueprint. Particular reading strands, by the nature of the skill being assessed, will elicit student responses at a specific level of cognitive complexity. This relationship is described in the bullets below:

- The content standards included under the SRCs of Reading to Perform A Task and Demonstrate General Understanding items are generally at the

- Recall/Literal Comprehension level since the content standards require that the correct response can be found directly stated in the text;
- Vocabulary and Develop an Interpretation items are generally at the Skill and Concept level since correct responses to these items are implicit in the text;
- Examine Content and Structure Items (not included in the Grade 3 content standards) are generally at the Strategic Thinking level since they require the reader to critically analyze features of the text and make determinations about the author's purpose and craft.

The following crosstab shows the relationship between Oregon's score reporting categories, content standards, and the separately coded cognitive complexity levels for the items used in the Spanish reading pilot.

**Relationship between Strand and Level of Complexity – 3<sup>rd</sup> Grade Spanish Reading Items**

Count		Level of Complexity			Total
		Recall/ Literal Comp.	Skill & Concept	Strategic Thinking	
Strand	Demonstrate General Understanding	40	6	4	50
	Develop an Interpretation	0	40	5	45
	Reading to Perform a Task	22	9	0	31
	Vocabulary	2	41	0	43
Total		64	96	9	169

To indirectly control for the presentation of items at various levels of cognitive complexity, Oregon structures its Reading/Literature item bank to explicitly code its items to the above knowledge levels (Doc 8). In addition, the 2007 alignment study by WestEd verified an inherent depth of knowledge in Oregon's Content Standards (Doc. 1, p. 14).

Oregon plans to jointly control item selection in both dimensions (content standard and cognitive complexity) in the 2011-12 school year.

**(d) Has the State ascertained that the scoring and reporting structures are consistent with the sub-domain structures of its academic content standards (i.e., are item interrelationships consistent with the framework from which the test arises)? and**

The scoring and reporting structures for Spanish Reading/Literature are the same as those for the English Reading/Literature assessment. The Sample Individual Student Report (ISR) and Class Roster Report illustrate this scoring and reporting structure (Doc. 13).<sup>2</sup> The assessment domain for Grade 3 Reading/Literature is carefully described in the Test Specifications and Blueprints (Doc 8). Structurally, the Reading/Literature

<sup>2</sup> Although the sample ISR and Class Roster Reports display student scores for the Grade 3 Spanish Reading/Literature native-language assessment, these reports do not currently indicate that the test was administered in Spanish. The student's test results for the Spanish administration appear on p. 3 of Doc. 13.

Common Curriculum Goals are associated with a Score Reporting Category, its component eligible content standards, and a sample assessment question that serves as an exemplar to provide additional information regarding the eligible content. In addition, the Test Specifications and Blueprints describe the relative proportion of the test that is comprised by each score reporting category and the achievement level descriptors that are based on the associated eligible content and were central to the standard setting process.

**(e) Has the State ascertained that test and item scores are related to outside variables as intended (e.g., scores are correlated strongly with relevant measures of academic achievement and are weakly correlated, if at all, with irrelevant characteristics, such as demographics)? and**

Evidence of convergent and discriminant validity of the Spanish reading test is provided in the following table. Of the 855 third graders who participated in the Spanish pilot, 847 also were tested with the English reading test and the OAKS mathematics assessment. A significant number (635) also took the Oregon English language proficiency assessment.

The Spanish reading scores had a 0.556 correlation with mathematics scores, providing evidence of higher discriminant validity than the English reading correlation with mathematics, 0.685.

The correlation between Spanish and English reading scores was moderate, at 0.625. This might be expected due to varying degrees of proficiency in the two languages among the sample of students taking the tests.

Correlations with the English language proficiency test show that language proficiency is more strongly correlated with reading in English (0.731) than with reading in Spanish (0.601). This finding provides support for the intent behind the Spanish reading assessment: that it should provide a measure of reading skill that is independent of the student's background in the English language.

**Correlation Matrix of Scores on Four Tests Taken by Participants in the Spanish Reading Pilot**

		Spanish Reading	English Reading	English Language Proficiency	Mathematics
Spanish Reading	Pearson Correlation	1	.625**	.601**	.556**
	Sig. (2-tailed)		.000	.000	.000
	N	847	847	635	845
English Reading	Pearson Correlation	.625**	1	.731**	.685**
	Sig. (2-tailed)	.000		.000	.000
	N	847	847	635	845
English Language Proficiency	Pearson Correlation	.601**	.731**	1	.582**
	Sig. (2-tailed)	.000	.000		.000
	N	635	635	640	638
Mathematics	Pearson Correlation	.556**	.685**	.582**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	845	845	638	849

\*\* . Correlation is significant at the 0.01 level (2-tailed).

Although the probability of student success on the test is likely to be associated with key indicators of socio-economic status, Oregon does attempt to systematically review each item for differential performance by subgroup to ensure that items function similarly across all subgroups. For a description of Oregon's use of DIF analysis for translated items, please see Oregon's Annual Technical Report, Volume 1, Section 2.4 (Doc. 5). A specific description of Oregon's use of DIF analysis to evaluate the Grade 3 Spanish Reading/Literature assessment is included in the Summary of Score Comparability Analyses (Doc. 3).

**(f) Has the State ascertained that the decisions based on the results of its assessments are consistent with the purposes for which the assessments were designed? and**

Following administration of the Grade 3 Spanish Reading/Literature assessment in spring 2010, Oregon surveyed administrators, teachers, and test coordinators for their opinions and suggestions regarding the assessment. A summary of the survey results is contained in Doc 7: Oregon Spanish Reading (Grade 3) Survey Feedback. Overall, the survey results indicate that the Grade 3 Spanish Reading/Literature assessment is serving its intended purpose. In particular, the results reveal the following conclusions about how Oregon school districts are using this assessment:

- School districts are predominantly administering this assessment to LEP students enrolled in English Language Development programs. Although Oregon does not restrict the administration of this assessment to LEP students, Oregon will limit the use of scores from this assessment for AYP accountability to LEP students.
- School districts would like to use the results from this assessment for both accountability purposes and to determine areas of instructional need.
- The vast majority of students felt either equally comfortable or more comfortable taking this assessment compared to taking the English Reading/Literature assessment.
- School districts perceived this assessment to be equally or more valid than the English Reading/Literature assessment at assessing Spanish-speaking students' reading skills. Comments reflect the feeling that, for Spanish-speaking students, the English Reading/Literature assessment is more a measure of the students' English proficiency than of their reading skills (Doc 7, p. 10).

**(g) Has the State ascertained whether the assessment produces intended and unintended consequences?**

A variety of means are used to monitor the consequences of Oregon's assessment system, including: (1) soliciting input from a wide variety of stakeholder groups, (2) meetings of the state Assessment Advisory Committee, and (3) extensive stakeholder/public input during the development of content standards and achievement standards.

#### 4.2

**For each assessment, including all alternate assessments, has the State considered the issue of reliability, as described in the *Standards for Educational and Psychological Testing* (AERA/APA/NCME, 1999), with respect to all of the following categories:**

- (a) Has the State determined the reliability of the scores it reports, based on data for its own student population and each reported subpopulation? and**
- (b) Has the State quantified and reported within the technical documentation for its assessments the conditional standard error of measurement and student**

**classification that are consistent at each cut score specified in its academic achievement standards? and**

As described in the Score Interpretation Guide, (Doc 9: Oregon's Technical Report Volume 6: Interpretive Guide), Oregon reports the Standard Error of Measurement at the student level for the total score and for each reported strand score.

The reliability of the Spanish reading assessment is comparable to the English version, as shown in the following descriptive statistics for conditional standard errors of measurement. The data are from tests administered to students who took both language versions of the reading assessment.

**Comparison of Conditional SEMs by Test Language**

**All Test-Takers**

		Spanish SEM	English SEM
N	Valid	847	847
	Missing	0	0
Mean		3.5021	3.5417
Median		3.4200	3.4300
Std. Deviation		.26242	.38078
Percentiles	20	3.3300	3.3360
	40	3.3900	3.3900
	60	3.4700	3.4780
	80	3.6100	3.6500

Because the Spanish and English reading tests were administered by the same adaptive algorithm, the reliability data are quite uniform across languages. The slight degradation of standard errors for students taking the English reading test may be due to some students attempting the test without having sufficient levels of English proficiency. The following tables show conditional standard errors of measurement for selected subpopulations.

**Comparison of Conditional SEMs by Test Language**

**Hispanic Subgroup**

		Spanish SEM	English SEM
N	Valid	796	796
	Missing	0	0
Mean		3.5057	3.5073
Median		3.4200	3.4200
Std. Deviation		.26747	.27592
Percentiles	20	3.3340	3.3300
	40	3.3900	3.3800
	60	3.4800	3.4620
	80	3.6200	3.6200

**Comparison of Conditional SEMs by Test Language**

**Economically Disadvantaged Subgroup**

		Spanish SEM	English SEM
N	Valid	780	780
	Missing	0	0
Mean		3.5061	3.5125
Median		3.4200	3.4200
Std. Deviation		.26515	.28574
Percentiles	20	3.3320	3.3300
	40	3.3900	3.3800
	60	3.4800	3.4700
	80	3.6200	3.6200

By design, the Spanish assessment was administered to a relatively homogenous group of 855 students, which limits the opportunity to examine reliability comprehensively for all subpopulation in the pilot administration. Data on classification consistency and reliability by subpopulation for the general reading assessment may be found in the Technical Report, Volume 4, Reliability and Validity, Doc 15.

Additional data on classification consistency for the Spanish and English reading tests using the latent distribution method developed by Guo (2006) is provided in Doc 3. (See tables 8 and 9 on page 20.)

**(c) Has the State reported evidence of generalizability for all relevant sources, such as variability of groups, internal consistency of item responses,**

**variability among schools, consistency from form to form of the test, and inter-rater consistency in scoring?**

For tests developed according to item response theory, internal consistency of responses is measured by item fit statistics. Good fit means that responses to the item are determined by overall proficiency in the subject and that, when proficiency is taken into account, there is no covariance among items. That is, when an item fits the model, responses to the item are a function of subject matter knowledge, not of construct irrelevant knowledge. The State examines Rasch Winsteps fit statistics using industry standard criteria for item rejection. Tests and forms constructed from items that fit the model yield scores that are comparable across years and populations. See Annual Report, Volume 1, Section 4.2 (Doc 5) for a detailed description of the use of item fit statistics.

**4.3**

**Has the State ensured that its assessment system is fair and accessible to all students, including students with disabilities and students with limited English proficiency, with respect to each of the following issues:**

**(a) Has the State ensured that the assessments provide an appropriate variety of accommodations for students with disabilities? and**

Consistent with the Individuals with Disabilities Act of 2004, all students with disabilities are eligible to participate in the OAKS. The student's Individualized Education Program team, which includes the student's parents or guardian, makes the decision regarding the most appropriate method for a student with disabilities to participate in testing. Each student must be considered individually and not merely on the basis of the student's disability category (Doc 5.2, Section 3.2.4).

To support students with disabilities taking the standard OAKS administration, Oregon publishes an Accommodation Manual and Tables containing a comprehensive list of state-approved accommodations that students may use when testing. Accommodations appearing on these tables have been approved by the Oregon Accommodations Panel and do not compromise the learning expectations, construct, grade-level standard, or measured outcome of OAKS. In Oregon, any student is eligible to use an accommodation, including both students with and without disabilities, and testing with state-approved accommodations is considered a standard administration (Doc 5.2, Section 3.2.2).

**(b) Has the State ensured that the assessments provide an appropriate variety of linguistic accommodations for students with limited English proficiency? and**

LEP students are given equal opportunity to participate in the OAKS, with districts choosing the appropriate testing option for each LEP student individually for each assessment on the basis of what is in the best interest of the student (Doc 5.2, Section 3.2.3).

LEP students may use accommodations from the Accommodations Manual identified by the school district as appropriate for the individual student and assessment (Doc 11: 2009-10 Oregon Accommodations Manual). Examples of accommodations designed to support LEP students in taking OAKS include translated English/Spanish versions of OAKS Math, Science, and Social Sciences (Doc 5.2, Section 3.2.2). In addition, Oregon has developed the Grade 3 Spanish Reading/Literature assessment; pending approval by the U.S. Department of Education, Oregon proposes that Grade 3 LEP students be able to use this assessment to meet AYP testing requirements.

**(c) Has the State taken steps to ensure fairness in the development of the assessments? and**

Oregon's item writers are typically Oregon teachers who have received training in item construction, are familiar with test specifications, and have demonstrated skill in writing items that pass both content and sensitivity panel review. To ensure fairness and sensitivity of Oregon's test items, Oregon specifically instructs its item writers to ensure that items reflect the diversity of Oregon students; avoid emotionally-charged issues such as death, violence, drug and alcohol abuse, criminal activities, or the occult; and are free of ethnic, gender, political, and religious bias (Doc 5.1, Section 6.1). After items are written they are reviewed by Oregon's Assessment Sensitivity Panel which examines items for sources of bias (Doc 10: Oregon Administrative Rule # 581-022-0620: Test Development). Specifically, the Sensitivity Panel ensures that items:

- present racial, ethnic, and cultural groups in a positive light
- do not contain controversial, offensive, or potentially upsetting content
- avoid content familiar only to specific groups of students because of race or ethnicity, class, or geographic location
- aid in the elimination of stereotypes
- avoid words or phrases that have multiple meanings (Doc 5.1, Section 6.2)

After items have been administered, a differential item functioning (DIF) analysis is conducted. Items exhibiting DIF are referred to content specialists for further review (Doc 5.1, Section 8.1).

Oregon conducts DIF analyses in two contexts: (1) to detect potential bias affecting scores for each reported subpopulation and (2) to examine the fairness of bilingual assessments in mathematics and science. See Oregon's Technical Manual Volume 1: Annual Report (Doc 5, Sections 2.4 and 4.1.2) for a detailed discussion of DIF analyses. In addition, the Summary of Score Comparability Analysis Report (Doc 3) specifically describes the findings of the DIF analysis for the 2009-10 administration of the Grade 3 Spanish Reading/Literature assessment.

**(d) Does the use of accommodations and/or alternate assessments yield meaningful scores?**

Yes. Oregon defines an accommodation as a practice or procedure in presentation, response, setting, and timing or scheduling that, when used in an assessment, provides equitable access to all students. The Accommodations Manual and Tables provide a comprehensive list of state-approved accommodations that students may use when testing (Doc. 11). Accommodations appearing on these tables have been approved by the Oregon Accommodations Panel and do not compromise the learning expectations, construct, grade-level standard, or measured outcome of OAKS. Oregon continually assesses the needs of its students and addresses those needs as they arise. Accommodations are recommended, evaluated, and made available on an ongoing basis by ODE through a formal review process involving the Oregon Accommodations Panel. The Accommodations Panel uses a combination of policy, judgments, and research to ensure that accommodations provide valid scores that allow all students to demonstrate what they know and can do. (Doc 5.2, Section 3.2.2).

**4.4**

**When different test forms or formats are used, the State must ensure that the meaning and interpretation of results are consistent.**

**(a) Has the State taken steps to ensure consistency of test forms over time?**

Tests are composed of items linked to the scale, so that scores are comparable across time. To ensure that the OAKS Online item pools remain stable over time, each pool contains a percentage of items, typically approximately 80%-90%, that have been previously used operationally and are psychometrically sound (Doc 5.1, Section 8.2). Each OAKS Online content assessment is linked to the scale using bank values for each item as anchors. In adaptive testing, calibrating each item to the scale is crucial to

assuring that test scores retain consistent meaning across time. To compute the difficulty of new items, Oregon uses a sparse matrix calibration method using all bank items together as an anchor. Furthermore, Oregon's field testing process is designed to ensure consistency of score meaning over time (Doc 5, Section 4.1).

**(b) If the State administers both an online and paper and pencil test, has the State documented the comparability of the electronic and paper forms of the test?**

Oregon only offers the Grade 3 Spanish Reading/Literature assessment online.

**4.5**

**Has the State established clear criteria for the administration, scoring, analysis, and reporting components of its assessment system, including all alternate assessments, and does the State have a system for monitoring and improving the on-going quality of its assessment system?**

Oregon's describes OAKS test administrator training requirements, test scheduling and administration procedures, and security procedures in Oregon Administrative Rule # 581-022-0610 (Doc 12) and in its Technical Report Volume 5: Test Administration (Doc 5.2). Oregon specifies the scoring, reporting, and quality control procedures for American Institutes for Research, the contractor for OAKS Online, in the contract statement of work, detailed requirements documents, and weekly meeting summaries.

Oregon solicits feedback on the quality and effectiveness of its assessment system through: (1) quarterly meetings of the Assessment Advisory Committee, (2) contracted studies and evaluations of the system by measurement experts, (3) annual meetings of the Technical Advisory Committee, (4) meetings of the assessment and content panels that are charged with keeping the assessments aligned with the content standards, and (5) frequent communication with district testing coordinators as issues come up. The assessment contractors also play a critical role in maintaining the quality of the assessment system (Doc 5.1, Appendix).

**5.1**

**Has the State outlined a coherent approach to ensuring alignment between each of its assessments, or combination of assessments, based on grade-level achievement standards, and the academic content standards and academic achievement standards the assessment is designed to measure?**

Oregon manages the alignment of its assessments and academic content standards by controlling every step of test development, from the development of the academic content standards, development of test specifications, to item writing, content and bias review, field testing, review of item performance, setting of achievement standards, and test form development. The "Life of an Item" diagrams in the Reading/Literature Test Specifications and Blueprints (Grade 3) (Doc 8, Appendix D) show the major steps in the process of test development. These steps are described in greater detail in Oregon's Technical Report Volume 2: Test Development (Doc 5.1). The test specifications created prior to item writing and test construction describe how the Reading/Literature content standards will be assessed at grade 3. The relative weight of sub-domains is indicated, and sample items illustrate the type of skill measured (Doc 8).

Item writing for the Grade 3 Reading/Literature assessment is carried out by Oregon teachers, managed by the ODE Language Arts assessment specialist. To ensure quality item writing, potential item writers are asked to submit an application that includes three items they have written, indicating the targeted content standard.

Oregon provides blueprints in the item writing workshop that lay out the item format and number of questions to be written for each piece of eligible content and associated depth

of knowledge. Items are reviewed for alignment with the content standards (at both the strand and sub-strand levels) during the item writing workshop through a peer-editing process. After items are word-processed and entered into the ODE's Item Management System, they are formally reviewed by content panels for alignment with grade-specific content standards and the assigned level of cognitive demand. A separate Sensitivity Panel then reviews items for potential bias (Doc 5.1, Section 6.2).

The report "Alignment of Oregon Content Standards and Oregon Grade 3 Spanish Reading Tests" (Doc 1) provides an overview of the process followed by Oregon in developing the Grade 3 Spanish Reading/Literature assessment, including the translation process and the criteria used by Oregon to support alignment. This report contends that because Oregon adhered to the same item development process followed in developing those items found by WestEd to align to Oregon's content standards in 2007, the items from the Grade 3 Spanish Reading/Literature assessment should also align.

The report "Alignment to Oregon Reading Content Standards: Grade 3 Spanish Reading Pilot, 2009-10" (Doc. 2), which supplements the analysis in Doc 1, found that a random sample of 25 items chosen from the Spanish Reading/Literature pool strongly aligned to Oregon's grade level content standards, supporting the argument that Oregon's item writing process continues to maintain a high degree of alignment to Oregon's content standards.

## 5.2

**Are the assessments and the standards aligned comprehensively, meaning that the assessments reflect the full range of the State's academic content standards? Are the assessments as cognitively challenging as the standards? Are the assessments and standards aligned to measure the depth of the standards? Does the assessment reflect the degree of cognitive complexity and level of difficulty of the concepts and processes described in the standards?**

Item writing for the state assessment includes an analysis of cognitive complexity. Oregon uses the following codes based on Bloom's taxonomy: Recall, Skill/Concept (Basic Application), Strategic thinking and Extended thinking. While Oregon does therefore consider cognitive complexity as part of item writing and item review, it does not yet include cognitive complexity as part of the test design and administration process. Oregon plans to incorporate this additional dimension by the 2011-12 school year.

Because of the close relationship between the assessed reading strands and the levels of cognitive complexity, the writing of state assessment items requires writers to design items at all three identified levels of knowledge (Recall/Literal Comprehension, Skill and Concept, and Strategic Thinking). By requiring that items from each of the reading strands be present in certain percentages on the test assures that students are presented with items from all levels of cognitive complexity. (Also see the discussion under Critical Element 4.1 (c) above.)

## 5.3

**Are the assessments and the standards aligned in terms of both content (knowledge) and process (how to do it), as necessary, meaning that the assessments measure what the standards state students should both know and be able to do?**

The 2007 WestEd alignment analysis of Oregon's items found a high concurrence between the items and the content standards for all subjects, including Grade 3 Reading/Literature, as well as an appropriate range of knowledge and breadth of knowledge in the item bank. Oregon's content standards have remained the same since

2007 and the development and updating of the item bank has followed the same procedures used to create the bank used for WestEd's study. Consequently, the current item bank should have the same alignment characteristics. Because the Grade 3 Spanish Reading/Literature item bank represents a subset of the Grade 3 English Reading/Literature item bank, the content of the items in the Spanish bank are parallel to the content of the items in the English bank (Doc 1, p. 12).

Oregon's Reading/Literature test specifications inherently control the distribution of items by intended cognitive process through the test blueprint. Particular reading strands, by the nature of the skill being assessed, will elicit items at a specific level of cognitive complexity. (Doc 8). In addition, the 2007 alignment study by WestEd verified an inherent depth of knowledge in Oregon's Content Standards (Doc. 1, p. 14).

## 5.5

### **Do the assessments yield scores that reflect the full range of achievement implied by the State's academic achievement standards?**

Oregon's Reading Test Specifications and Blueprints (Doc 8) provide:

- 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

In addition, an analysis of statistics on enacted tests by common curriculum goal (CCG) and language appears in the Alignment of Oregon Content Standards and Oregon Third Grade Spanish Reading Tests (Doc. 1, pp. 17-18).

## 5.7

### **What ongoing procedures does the State use to maintain and improve alignment between the assessments and standards over time?**

Oregon maintains alignment through a continuous improvement process that links item writing through an evaluation of completed tests. The process begins each year with assessment content specialists analyzing the item bank and reviewing content converge in terms of item depth and breadth as established by the content standards and test specifications. Based on a gap analysis between item pools and test specifications, Oregon plans its item writing sessions. The item writing sessions engage Oregon educators in a process of writing items that are specifically aligned to Oregon standards based on the needs identified in the gap analysis. The item writing process is described in greater detail in Oregon's Technical Report Volume 2: Test Development (Doc 5.1). The combination of the pool analysis and completed tests provides a comprehensive view of the alignment between the tests and the standards.

## 6.3

### **What guidelines does the State have in place for including all students with limited English proficiency in the tested grades in the assessment system?**

#### **(a) Has the State made available assessments, to the extent practicable, in the language and form most likely to yield accurate and reliable information on what these students know and can do?**

LEP students are given equal opportunity to participate in the OAKS, with districts choosing the appropriate testing option for each LEP student individually for each assessment on the basis of what is in the best interest of the student (Doc 5.2, Section 3.2.3).

LEP students may use accommodations from the Accommodations Manual identified by the school district as appropriate for the individual student and assessment (Doc. 11). Examples of accommodations designed to support LEP students in taking OAKS include translated English/Spanish versions of OAKS Math, Science, and Social Sciences (Doc

5.2, Section 3.2.2). In addition, Oregon has developed the Grade 3 Spanish Reading/Literature assessment; pending approval by the U.S. Department of Education, Oregon proposes that Grade 3 LEP students be able to use this assessment to meet AYP accountability requirements.

## **2.1**

**Has the State formally approved/adopted challenging academic achievement standards in reading/language arts and mathematics for each of grades 3 through 8 and for the 10-12 grade range? These standards were to be completed by school year 2005-2006.**

and

## **7.3**

**How has the State provided for the production of individual interpretive, descriptive, and [non-clinical] diagnostic reports that indicate relative strengths and instructional needs?**

**(a) Do these individual student reports express results in terms of the State's achievement standards rather than numerical values such as scale scores or percentiles?**

Oregon provides reports that contain scale scores, score ranges based on the standard error of measurement, achievement levels, and achievement level descriptors. Oregon provides strand data at the individual and classroom level to support identifications of student strengths and instructional needs (Doc. 13).<sup>3</sup>

The academic achievement standards adopted by the State Board of Education are applicable to both the English and Spanish versions of the reading assessment. (Doc 6 provides a description of the standard setting procedures.) The evidence presented in the current submission shows qualitative and quantitative comparability of the reading tests in both languages and, as such, is intended to support the use of the same cut scores and achievement level descriptors.

**(b) Do these individual student reports provide information for parents, teachers, and principals to help them understand and address a student's specific academic needs? Is this information displayed in a format and language that is understandable to parents, teachers, and principals, for example through use of descriptors that describe what students know and can do at different performance levels? Are the reports accompanied by interpretive guidance for these audiences?**

The combined individual student report is an efficient means by which students, parents, and teachers can see a snapshot of student performance in an easy to understand format. These reports (along with the other individual student reports) can be printed on-demand to provide the information stakeholders need in a convenient and economical manner. Volume 6 of Oregon's Technical Reports (Doc. 9) provides an interpretive guide for stakeholders to better understand and use the reports.

---

<sup>3</sup> Although the sample ISR and Class Roster Reports display student scores for the Grade 3 Spanish Reading/Literature native-language assessment, these reports do not currently indicate that the test was administered in Spanish. The student's test results for the Spanish administration appear on p. 3 of Doc. 13.

**(c) How has the State ensured that these individual student reports will be delivered to parents, teachers, and principals as soon as possible after the assessment is administered?**

Students receive their test scores immediately upon completing an online test. Data from the online tests are loaded into ODE's system within 2 days and are thereafter immediately accessible through the various reporting systems. In addition, Oregon Administrative Rule # 581-022-1670: Individual Student Assessment, Recordkeeping, and Reporting requires school districts to report at least annually to parents or guardians of all students in grades K-12 regarding their student's scores on all state assessments (Doc. 14).