

Washington Comprehensive Assessment Program

2011 Assessment System Report



Randy I. Dorn
State Superintendent of
Public Instruction

January 2011

Washington Comprehensive Assessment Program

2011 Assessment System Report

Prepared by
Michael Middleton, Director of Assessment Business and Special Populations
Robin Munson, Ph.D., Director of Student Information

Assessment and Student Information
Office of Superintendent of Public Instruction
Michael Middleton, Director
Robin Munson, Ph.D., Director

Randy I. Dorn
Superintendent of Public Instruction

Ken Kanikeberg
Chief of Staff

Alan Burke, Ed.D.
Deputy Superintendent, K-12 Education

January 2011

Table of Contents

Executive Summary.....	iii
2010 Study Results	1
2011 Changes to the Summative Assessment.....	5
Other Program Initiatives	11
Instructionally Supportive Formative Assessment.....	12
Cost Analysis	15

Executive Summary

The state assessment program experienced numerous changes during the 2009–10 school year and will have similar changes, plus new enhancements, occurring during 2010–11. This summary will provide an overview of the 2010 changes, designs for the 2011 test administration, an introduction to other program initiatives, and a review of formative assessment efforts. More details are found in the main body of the report.

2010 Overview

The first of many changes associated with the 2010 test administration cycle dealt with name recognition. What had been the Washington Assessment of Student Learning, referred to as WASL since the program inception, became a two-named program distinguishing the accountability of testing elementary through middle school grade levels and testing for high school. Assessment names are now the Measurements of Student Progress (MSP) for Grades 3 through 8 and the High School Proficiency Exam (HSPE).

Though the 2009 test was reduced in size (i.e., fewer items) from previous years, the 2010 design was scaled back even further to be in concert with legislative guidance provided in ESSB 5414. Test design implications saw the elimination of the four-point extended response items in reading, mathematics, and science, plus a ratio of one-point multiple choice and completion items¹ to two-point short answer items, such that the points from the two-point items would not exceed 25 percent of the total points on the test.

Reduction of the test size allowed the opportunity to restructure the testing time associated with the annual test administration. Test time for reading, mathematics, and science was reduced to one day per content area from three days for Grades 3 through 5 and two days for Grades 6 through 8 and high school.² Though it was a positive benefit for instructional time, scaling back test days did result in some tension with regard to the length of time students were engaged with the test. We received reports from different school districts around the state that students needed more time than the suggested testing time announced in the state’s testing directions.

The 2010 test administration introduced online testing for state assessments. The Office of Superintendent of Public Instruction (OSPI) established a transition plan for full online testing within a three-year time frame. The first year focused on Grades 6 through 8 reading and mathematics, while the transition plan called for 25 percent student participation for tested materials across the identified grades and content areas. This was achieved through solicitation on a voluntary basis. Though there were isolated instances of technical difficulties, no major flaws were identified. Feedback from most in the field, including survey responses from students, said the overall experience was positive. Based on field response and the context of other states that experienced widespread administration problems with their online programs, OSPI deemed the initial year of online testing a major success.

Finally, the 2010 test administration in mathematics for Grades 3 through 8 required testing on the new mathematics learning standards adopted in 2008. The result of the

¹ In 2010 completion items were only used for mathematics; reading and science will include this item type in the 2011 administration cycle.

² Writing in all grades remains a two-day administration, one prompt per day.

changes associated with test size, testing mode (paper-pencil versus online), and testing to new standards was an extensive series of comparability studies to ensure reported results had connection to previous years' information. Table 1 provides a content-by-grade visual representation of the studies conducted.

2011 Designs

Test administration for 2011 will experience similar changes made in 2010 and new enhancements supporting the goals of the state. Changes will include expansion of online testing to include Grades 4 and 5 and science, with increases in participation in Grades 6 through 8 in the previous years' content areas. The new program enhancement is the introduction to end-of-course testing in high school mathematics, aligned to Algebra 1/Integrated Mathematics 1 and Geometry/Integrated Mathematics 2.

Online testing will expand in Grades 6 through 8 reading and mathematics to a state-wide goal of 80 percent student participation across all testing opportunities, while aiming for 25 percent participation in Grades 4 and 5 reading and mathematics. Additionally, solicitation for participation in the additional content of science for Grades 5 and 8 will aim for 25 percent participation.

Science test administration for Grades 5 and 8 will also experience testing against the new science standards adopted in 2009, while the introduction of the end-of-course mathematics assessment design coincides with initial testing against the new high school learning standards for mathematics adopted in 2008.

Program Initiatives

Other efforts within the assessment program include:

- High School Science end-of-course in Biology for 2012.
- Common Assessments: development work amid two multi-state consortia, with Washington a member of the SMARTER-Balanced development effort.

Instructionally Supportive Formative Assessments

Under the auspice of the agency's Classroom Assessment Integration Director, of which formative assessment is the critical foundation, the following efforts are underway:

- Web-Based Instructionally Supportive Student Assessment Systems: project to identify an interactive tool that districts can use to support teachers in the classroom assessing students' knowledge, skills, and progress toward achievement of learning standards.
- Washington Kindergarten Inventory of Development (WaKIDS): project to identify a quality methodology to determine readiness of students in kindergarten.
- Evaluating the Validity of English Language Proficiency Assessments (EVEA): a research project coordinated with four other states to develop protocols to evaluate the validity in English language proficiency assessments.

2010 Study Results

In the 2009 Assessment Report, it was stated that 46 studies were to be conducted that would demonstrate the degree of validity and reliability retained through earlier re-design efforts of the assessment program. Table 1 provides a visual representation of the studies associated with the specific grade levels and content areas.

Table 1

	Gr 3	Gr 4	Gr 5	Gr 6	Gr 7	Gr 8	HS
READ	Equate 2010 to 2009	Equate 2010 to 2009	Equate 2010 to 2009	Equate 2010 to 2009	Equate 2010 to 2009	Equate 2010 to 2009	Equate 2010 to 2009
				Equate Online to P/P	Equate Online to P/P	Equate Online to P/P	
	Align MSP to Content Stds	Align MSP to Content Stds	Align MSP to Content Stds	Align MSP to Content Stds	Align MSP to Content Stds	Align MSP to Content Stds	Align HSPE to Content Stds
MATH	Build new '10 scale and cuts	Build new '10 scale and cuts	Build new '10 scale and cuts	Build new '10 P/P scale and cuts	Build new '10 P/P scale and cuts	Build new '10 P/P scale and cuts	Equate 2010 to 2009
				Equate Online to P/P	Equate Online to P/P	Equate Online to P/P	
	Align new MSP to Content Stds	Align new MSP to Content Stds	Align new MSP to Content Stds	Align new MSP to Content Stds	Align new MSP to Content Stds	Align new MSP to Content Stds	Align HSPE to Content Stds
	Concordance betw. '09 and '10	Concordance betw. '09 and '10	Concordance betw. '09 and '10	Concordance betw. '09 and '10	Concordance betw. '09 and '10	Concordance betw. '09 and '10	
SCI			Equate 2010 to 2009			Equate 2010 to 2009	Equate 2010 to 2009
			Align MSP to Content Stds			Align MSP to Content Stds	Align HSPE to Content Stds

In recapping those efforts, the fidelity of the program to its long-standing levels of validity and reliability were retained through diligent compliance to procedures and high-caliber judgment and evaluation of numerous national professionals from the field of test measurement.

Alignment to Standards

Per guidance from the U.S. Department of Education peer review process, OSPI employed the services of Alpine Testing Solutions to facilitate the process of alignment studies comparing the 2010 mathematics, reading, and science assessments to the existing state content standards. Thus, the related studies served to validate decisions on the reduction of test length in relation to the alignment characteristics of the previous test versions. For Grades 3 through 8 mathematics, however, the adoption of new content standards and tests, as well as test length reductions, necessitated an initial independent alignment study of these assessments. The primary task of the educators and content specialists, who were selected to participate due to having experience at these grade levels, was the application of professional judgment in evaluating how well each assessment item under the new test designs aligned with the state content standards in terms of content and cognitive complexity.

The alignment study process required two steps:

- Panelists initially made judgments on the cognitive complexity of each Performance Expectation (PE) for math Grades 3 through 8, or Grade Level Expectation (GLE) for high school math only, and all grades for reading and science. Judgments were made independently and then the group discussed one PE or GLE to achieve group consensus. Panels were encouraged to discuss items for where there was no consensus, but were also reminded that consensus did not require unanimous agreement among all members. Conclusions regarding alignment were based on this model of consensus decisions.
- Panelists then focused on the items within the assessments. Panelists rated the cognitive complexity of the item using the same framework applied to the content standards; then identified the PEs or GLEs, if any, to which the item aligned in terms of content. They were asked to evaluate the fit of the items within the PE or GLE using the following scale:
 - ❖ **Complete fit.** The main content required to answer the item correctly is contained in the content standard. If the student gets the item right, this is one relevant piece of information about the student's level of achievement of the content stated in the standard.
 - ❖ **Partial fit.** A significant portion of the content required to answer the item correctly is embodied in the content standard. But there is additional, significant understanding required that is represented by some other content standard. If the student gets the item right, it is because the student has some other significant knowledge that is *not* part of this content standard.
 - ❖ **Slight fit.** There is some relationship between the item content and the content of the content standard, but much more is needed to answer the item correctly. Perhaps only one of several pieces of content required to answer the item correctly is stated in the content standard. Alignment would probably be more complete with some other standards, or it might take several standards to cover the content of the item sufficiently. If an item has a slight fit with one content standard, it may have a slight fit with another as well.
 - ❖ **No fit.** The item does fit any content standard.

Equating

The process of equating is designed to demonstrate the comparability in test difficulties from year-to-year. Having equated tests provides confidence that interpretations of data are consistent against an objective measure of student performance. For purposes of the state's assessment program, equating is conducted with each annual administration for all grades in the content areas of reading, mathematics, and science. The writing assessment, two on-demand essays only, does not provide the test length to fit an equating model.

For 2010, equating was accomplished to support not only the standard paper-pencil mode of test administration, but also for the online administration mode allowing an evaluation of comparability between the two modes (refer to *Equating Between Paper-Pencil and Online Testing* below).

Standard Setting

Due to the adoption of new content standards for MSP mathematics for Grades 3 through 8, the shortening of the overall test length as fewer short-answer constructed response items, and the addition of new item types (completion items), OSPI was compelled to conduct new standard setting activities at the conclusion of the 2010 test administration cycle.

As in previous standard setting efforts, OSPI adopted the modified bookmark method, with an added feature of a contrasting group study conducted prior to the standard setting meeting, and presented as guidance for panelists during standard setting. The modified bookmark method included a presentation of the items from the grade-level test on a range of easiest to hardest based on the item difficulty (p-values). In successive rounds of review, the standard setting panelists deliberate student performance expectations, proposed preliminary cut-scores, and review rationales in context with added evidence/information all designed to solidify the thinking of the group around a final set of cut-score decisions. Additional evidence/information that informed each round of review was:

- Round 1 = Contrasting group guidance
- Round 2 = Item difficulty (p-values)
- Round 3 = Grade-level impact (student performance) data

The contrasting group information was developed through the collection of teachers' evaluations of their own classroom students, as it pertained to one of the student achievement performance-levels, Below Basic, Basic, Proficient and Advanced. When looking at graphs with expected student performance versus true performance, the point of intersection provides focus to potential cut-scores using a basis of educator input.

Once cut-scores were established for each grade, representatives from each grade-level panel were brought together as a committee to address any cross-grade anomalies in the established cut-scores (e.g., proficiency at an earlier grade level was established at higher level of difficulty than the next grade level). The efforts of the committee resulted in a comprehensive set of cut-scores across Grades 3 through 8.

All the associated efforts of the standard setting meetings resulted in the cut-scores approval by the State Board of Education on August 10, 2010.

Equating Between Paper-Pencil and Online Testing

During the 2010 annual test administration, OSPI offered an online testing option for Grades 6 through 8 MSP in reading and mathematics. Participation in the online option was voluntary, resulting in approximately 25 percent of all test materials for these grade-content combinations accessed via the online test platform. To ensure consistent performance interpretation between students who accessed the online test and those who tested via the traditional paper-pencil mode, equating between both test modes was required.

To obtain comparable paper-pencil and online examinee samples for evaluation of mode effects, a match sample process was used to gather a similar set of student data to compare. The sub-sample for the study was selected from all the volunteer online schools to match the paper-pencil test sample used in generating the scoring tables for each content area and grade. The matching criteria used in selecting the study sample were school size, mean scale score for each school, and the gender and ethnic composition of the school.

From the equating analyses it was determined that across the grades, regarding administration via paper-pencil or online, that the two modes held statistic consistency with the following characteristics:

- Less than half a point difference in terms of mean points earned.
- Average item performance differences of less than 0.02 according to p-values (level of difficulty).
- Yielded raw-to-scale score tables producing exactly the same cut scores for 16 of the 18 cuts (Basic, Proficient, Advanced for Grades 6 through 8 across two content tests), and were different by only one raw score point at the two cuts that didn't match.
- Same reliability as each other and in comparison to the 2009 administration results.
- Item fit statistics determined the same, as all were well within acceptable bounds.

Concordance (“Bridging”) Between WASL and MSP

Because these tests are used for state and federal accountability, a bridging study was conducted to provide a translation between the 2009 and 2010 mathematics tests. Bridging studies provide one-time concordance that expresses relationship between previous and the new reporting scales. In the case of Washington’s MSP assessments for Grades 3 through 8 mathematics, the interim results were used to aid conclusions regarding standard setting, and the final concordance tables were used for Annual Yearly Progress (AYP) calculations. The bridging or linking analyses were performed following procedures similar to those used in operational equating (refer to processes in section on Equating above). The sample included in the analyses consisted of the students who were administered the paper-and-pencil version of the mathematics test for each grade.

Using the Rasch Partial Credit Model (PCM), scores on the 2010 mathematics forms were placed on the reporting scale established in 2004 by fixing the item performance parameters (e.g., difficulties and step values) for each item that comprised the anchor set to the items existing bank values. The anchor set used for linking comprised 20 to 22 items and included both internal and external anchors.

Internal anchors reflected the overlap of the old and new standards, and included both multiple choice and short answer items. The internal anchor items were common across all forms of the specific grade level test. The number of internal anchors ranged from 5 to 12 for each form of the test. External anchor items, primarily representing content from the old standards, consisted of multiple choice items and were placed in the five available pilot-item positions. For Grades 4 and 6, there were two linking forms and 10 external anchors; for all other grades, there were three linking forms and 15 external anchors. Table 2 provides the total number of operational items and the distribution of internal and external anchors for each grade.

Table 2

Number of Anchor Items for Mathematics Grades 3-8

Grade	Total Number of Operational Items	Total Number of Anchor Items	Number of Internal Anchor Items	Number of External Anchor Items	Number of Linking Forms
3	30	22	7	15	3
4	30	22	12	10	2
5	30	20	5	15	3
6	35	22	12	10	2
7	35	21	6	15	3
8	35	21	6	15	3

In 2010, the percentage of students meeting the previous performance standard in Grades 3 through 8 mathematics assessments ranges from 53.9 percent for Grade 6 to 66.1 percent for Grade 3. With the exception of Grade 3, a higher percentage of students taking the 2010 mathematics test met the 2009 standard in Grades 4 through 8 (see Table 3). The average increase across all grades is 3.3 percent.

Table 3

Difference in Percentage of Students Meeting Standard

Year	Grade					
	3	4	5	6	7	8
2009	66.3	52.3	61.9	50.9	51.8	50.8
2010	66.1	58.2	65.6	53.9	56.4	53.5
Difference	-0.2	5.9	3.7	3.0	4.6	2.7

2011 Changes to the Summative Assessment

Though many of the characteristics associated with the state’s summative assessments will be similar to the 2010 administration year, two significant adjustments will be unveiled in 2011. First, is the initial administration of end-of-course (EOC) assessments in the content area of high school mathematics, specifically the courses of Algebra 1, Integrated Math 1, Geometry, and Integrated Math 2. Second, is the administration of science assessments in Grades 5 and 8 aligned to the newly adopted learning standards.

End-of-Course Mathematics

For assessment purposes, Algebra 1 and Integrated Math 1 will be linked to assessing first year mathematics, while Geometry and Integrated Math 2 will be linked to assessing second year mathematics, with students being tested according to the course sequence available in their respective districts.

The design of the end-of-course assessments provides for a test of 43 items used for reporting student performance. Table 4 provides a detailed look at item types with item counts projected for a typical test booklet.

Table 4
End-of-Course Exams

Alignment to Content Standards	2011 EOC exams are aligned to new (2008) math content standards.
Testing Time	Administered in classrooms over three consecutive days: 30 minutes for directions/distributing materials, 120 minutes of total testing time.
Variable Test Window	District-determined schedule during last three weeks of school calendar.
Standards Assessed	The Performance Expectations common to Algebra 1 and Integrated Mathematics 1 are assessed for purposes of determining proficiency. The Performance Expectations common to Geometry and Integrated Mathematics 2 are assessed for purposes of determining proficiency. The Performance Expectations unique to each course are assessed for purposes of reporting a strength or weakness.
Test Item Count/Types and Test Points	Each test will consist of 37 operational items used to determine student proficiency. Multiple choice: 29 items; Completion: 5 items; Short answer: 3 items. Total points: 40 points ----- Each test will include six items to assess student strength or weakness on course-specific content. Multiple choice: 3 to 5 items; Completion: 1 to 3 items Total points: 6 points.

The assessable standards, PEs, associated with first and second year mathematics, are linked to the course sequence followed and result in a matrix based on opportunity to learn for a student that is dependent on which course sequence he or she takes. Tables 5 and 6, below, are visual representations of how the performance expectations align to the course sequencing and when students would be expected to learn the associated skills and ultimately assessed on each.

Table 5 highlights the number of assessable PEs that correspond to skills students would learn in first-year or second-year mathematics, regardless of course-sequencing. These performance expectations are the foundation of skills and knowledge students will need to demonstrate proficiency for purposes of AYP and to fulfill graduation requirements. These performance expectations would also be the standards used in providing assessments to students who require testing but have either: (1) taken the course prior to the EOC introduction, or (2) not met standard on a previous mathematics assessment administration. In these latter two student situations, the standards shaded in green comprise what is referred to as the make-up/initial administration, or retake/subsequent administration(s) assessments.

Table 5
Performance Expectations by Course

Shaded Areas Required for Graduation

	Integ Math 1	Integ Math 2	Integ Math 3	
Algebra 1	44 PEs			Alg 1 PEs
Geometry		20 PEs		Geom PEs
Algebra 2				Alg 2 PEs
	Int 1 PEs	Int 2 PEs	Int 3 PEs	ALL PEs

Table 6 highlights the PEs that are linked to course-specific content knowledge but do not overlap with the instructional delivery associated with a first-year and second-year concept in course-sequencing, (i.e., select PEs associated with Algebra 1 are not covered for students in an integrated mathematics curriculum until enrolled in Integrated Math 2). These PEs will be linked to the course-specific version of the EOC assessment and will be used to demonstrate student mastery of course-specific knowledge but will not inform proficiency as it relates to AYP and fulfillment of graduation requirements. Items assessing these PEs will not be included on assessments given to students who would require testing to fulfill the mathematics graduation requirement.

Table 6
Course-Specific Content

Shaded Areas Assessed as Course-Specific Content

	Integ Math 1	Integ Math 2	Integ Math 3	
Algebra 1		Int 2 EOC: 7 Alg 1 EOC: 8		Alg 1 PEs
Geometry	Geo EOC: 7 Int 1 EOC: 7		Geo EOC: 13	Geom PEs
Algebra 2		Int 2 EOC: 12		Alg 2 PEs
	Int 1 PEs	Int 2 PEs	Int 3 PEs	ALL PEs

The EOC assessment is designed for classroom administration, though districts are being provided the flexibility to administer the assessment as best meets the needs of the students, teachers and administration. For test security purposes, there are administration protocols that will require consideration, particularly managing item exposure prior to testing. OSPI will work with districts in determining the best means to meet the needs of all involved.

Under current legislation, students enrolled in classes coded specifically as Algebra 1, Integrated Math 1, Geometry and Integrated Math 2 will be required to be assessed using the

EOC tests. Additionally, per the current legislation, students in the Class of 2013 and beyond who have already taken one or both of the linked courses will be required to test using the make-up or retake assessments for the corresponding courses already completed. Further, students in the Class of 2011 and 2012, who desire to graduate by meeting standard on a mathematics assessment would also have to use the make-up or retake assessments.

Grades 5 and 8 Science Assessments

In spring 2009, the state adopted revised learning standards for science in Grades K–12. Associated with any change to learning standards is the necessity to develop items and tests that align to new learning targets. Since the adoption of the revised learning standards, OSPI has been working through the processes related to item and test development with 2011 being the first administration linked to the new standards for Grades 5 and 8.

Table 7 provides details regarding the item types with item counts projected for a typical test form. Though the item type and count has been adjusted from the 2010 administration design, the total item and raw point count is substantially the same. This should result in the same testing time experienced by students.

Table 7
Science MSP (Grades 5 and 8)

Alignment to Standards	2011 Science MSP aligned to 2009 adopted content standards. New cut scores established by standard setting – Aug 2011
Test Length	Gr 5: Single testing session ~ 75 mins. Gr 8: Single testing session ~ 90 mins.
Test Item Count/Types Test Points	Gr 5: 20 MC, 6 CP, 4 SA Total = 30; Points = 34 Gr 8: 25 MC, 5 CP, 5 SA Total = 35; Points = 40
Test Window	Paper-Pencil: May 2–19, 2011 Online: May 2–June 3, 2011
Online Testing	Statewide participation: Gr 5 – 25%; Gr 8 – 25%

The MSP for science is designed to assess students’ knowledge and skills against the state’s new academic content standards for science in Grades 5 and 8, first released in June 2009. These will be the first tests to assess those standards. The cut scores for performance levels on the tests will be established in August 2011 by the SBE. As established with testing in 2010, the MSP for science has fewer test items than in earlier years of the program and takes fewer days to administer than the WASL. There are no four-point items on the science MSP but new completion items will appear in the 2011 tests.

Retaining Validity and Reliability Studies

In the 2009 Assessment System Report, the relationship between a test’s overall length and its reliability (as tests get shorter, they tend to be less reliable) was introduced. In that report it was explained that diminished reliability due to a shortened test can be mitigated to some extent by good item selection and test form design. With regard to the science MSP, the test length has been adjusted the last two administrations with minimal impact to the reliability elements (Table 8 refers to the dampened effects between the 2009 and 2010 administrations). The resulting

reliability coefficients for the 2009 and 2010 administrations were at or above 0.81, with a median reliability of 0.83 and 0.86. The median change in reliability was only 0.01 between years. The design principles used in stabilizing the reliability of the 2009 and 2010 tests (e.g., reviewing “test characteristic curves” at test build to check for estimated conditional reliabilities) are the same principles applied in designing the 2011 tests.

Table 8

Science			
		Grade 5	Grade 8
2009	# Pts	37	50
	<i>Rel.</i>	<i>.84</i>	<i>.87</i>
2010	# Pts	34	40
	<i>Rel.</i>	<i>.81</i>	<i>.85</i>

Since the end-of-course mathematics assessments will be completely new test constructs, there will be no opportunity to compare the reliability of the test to a previous administration, but inferences drawn from the item performance data accumulated through field testing as compared to the initial operational administration will be evaluated during all the post-administration processing of the tests.

Reliability of a test captures the level of measurement imprecision that is inherent in the particular assessment. This is an important characteristic of a test and industry standards require test developers to attend to this property of their tests. More important than reliability, however, is a test’s validity. Test validity can best be defined as the extent to which decisions or inferences drawn from test results can be counted on as being reasonable and appropriate. In similar context to the information shared last year regarding the transition of the MSP mathematics assessments to new standards, several studies will be required to demonstrate retention of the validity the state has experienced in its assessment program.

Several mechanisms are already in place to monitor the validity of the state’s assessments. Although Washington has used these validity checks for many years, they are now part of the federal Elementary and Secondary Education Act (ESEA) requirements for the quality of state assessment programs known as “Peer Review.” Federal peer review requires each state to submit documentation of the technical quality of its assessments for evaluation by a panel of assessment experts (the “peers”). Washington’s assessment program was determined to be “fully approved” by the United States Department of Education (USED) in August 2008. The changes implemented with the 2011 science tests will require resubmission to USED for the technical quality of the state’s tests. That resubmission will include the results of the 34 “comparability studies” shown in Table 9. All are designed to evaluate the validity of the science MSP.

One of the peer review elements that helps establish test validity and fairness is known as “equating.” Equating is a statistical procedure that assures that the level of knowledge and skills required to meet standard in one year is the same as the knowledge and skills required in any other year. Equating supports a fundamental fairness of a test by establishing that students are held to the same performance standard, regardless of which year they take the test. For all assessments in Grades 3 through 8 reading and mathematics, and high school science, the standard annual equating processes will need to be implemented.

In 2011, equating studies for Grades 5 and 8 science assessments and end-of-course mathematics assessments will not be conducted, as these tests will be new assessments

aligned to new content learning standards. Therefore, an initial standard setting will occur to develop the test scales. All future Grade 5 and 8 science and end-of-course mathematics administrations will be equated back to the 2011 scales.


In a typical year, equated tests can be assumed to carry the same test validity as those used in preceding years. With the redesigned science MSP and end-of-course mathematics assessments, however, validation of the tests' alignment to the content standards must be re-established. Table 9 shows the six alignment studies that are being conducted to evaluate the alignment between the state's academic content standards and the new tests, displayed as "Align MSP to New Content Stds" or "Align EOC to New Content Stds" for science and high school math.

The peer review process requires that the alignment studies be conducted independent of those involved in item and test development, to avoid possible conflicts of interest. As such, an assessment firm has been contracted to conduct the alignment studies independent of both OSPI and Educational Testing Services, the state's contractor for test development. This study represents an important test validation procedure. If scores on the tests are going to be used to judge whether or not students have met the state's academic content standard, then the degree to which the tests are aligned need to be evaluated. Legislative requirement that the tests be redesigned while "retaining the assessment of critical thinking skills" will be judged by the extent the tests align with this requirement.

Table 9

Comparability Studies for 2011 Assessments

	Gr 3	Gr 4	Gr 5	Gr 6	Gr 7	Gr 8	HS
READ	Equate 2011 to 2010	Equate 2011 to 2010	Equate 2011 to 2010	Equate 2011 to 2010	Equate 2011 to 2010	Equate 2011 to 2010	Equate 2011 to 2010
		Equate Online to P/P	Equate Online to P/P	Equate P/P to Online	Equate P/P to Online	Equate P/P to Online	
MATH	Equate 2011 to 2010	Equate 2011 to 2010	Equate 2011 to 2010	Equate 2011 to 2010	Equate 2011 to 2010	Equate 2011 to 2010	
		Equate Online to P/P	Equate Online to P/P	Equate P/P to Online	Equate P/P to Online	Equate P/P to Online	Align Year 1 EOC to New Content Stds
							Align Year 2 EOC to New Content Stds
							Build '11 P/P scale and standard setting
SCIENCE							Equate 2011 to 2010
			Align MSP to New Content Stds			Align MSP to New Content Stds	
			Build '11 P/P scale and standard setting			Build '11 P/P scale and standard setting	
			Equate Online to P/P			Equate Online to P/P	

 reports/studies conducted annually for each grade-content administration

Ten comparability studies are uniquely associated with the new Grades 5 and 8 science tests and the end-of-course mathematics assessments. First, there are the six reports linked to establishing the scales on the new tests. This is implemented through the standard setting procedure. OSPI is required to work in collaboration with the SBE on a standard setting process, with the SBE authorized to set cut scores for the various proficiency levels on the state's tests. Standard setting on the new Grade 5 and 8 science tests and the end-of-course mathematics assessments will be conducted in August 2011, after the tests have been scored.

The test design for the end-of-course mathematics assessments is completely different from the previous HSPE construct. The established cut-scores will be the foundation for equating and validation measures for future test administrations. A new uniform bar will need to be established for the end-of-course mathematics assessments based on the state's proposed accountability plan.

Science assessment, in regard to Grades 5 and 8, will be against the new standards resulting in the adoption of new performance cut-scores. Since the assessments are benchmark indicators of achievement in science skills and knowledge, there is no associated accountability index requiring a uniform bar for AYP reporting and no requirement of the state to determine concordance between the cut-scores from the old test and the new cut-scores.

Finally, six comparability studies will be conducted to equate the scores from the online test administration for Grades 4 and 5 reading and mathematics, and the paper-and-pencil version for Grades 5 and 8 science tests. While more studies will be conducted to equate the reading and mathematics paper-pencil test administration to the online version for Grades 6 through 8. The state's transition to online testing for science in 2011 was referenced earlier (see Table 7). From a validity perspective, it is important to establish that scores on the tests are independent of the mode in which the test is taken. That is, a student should not be advantaged or disadvantaged by taking the test using the online or the paper-and-pencil mode. Some states that have implemented online tests have found that there is no mode effect and can use the same raw-to-scale score tables for both modes. Other states have found that there is a mode effect for their tests and use separate raw-to-scale score tables to ensure fairness. Washington will conduct mode-effect tests by equating the online tests to the paper-and-pencil tests and evaluating the impact of mode on the raw-to-scale tables.

Similar comparability studies regarding mode of testing are also required for Grades 4 and 5 reading and mathematics, as both grade-content combinations will be added to the online administration plan for 2011.

Other Program Initiatives

Online Testing

Statewide test administrations will witness further transitioning to online testing in 2011. In 2010, 25 percent of tested materials in Grades 6 through 8 reading and mathematics were administered online. Participation in 2011 is being encouraged to reach 80 percent amongst the same grade and content areas from 2010, while the state is expecting 25 percent participation in expanded grade and content areas (Grades 4 and 5 in reading and mathematics and Grades 5 and 8 in science). The state's planned expansion of online testing will, in the long run, be more cost effective than paper-pencil testing, and will position Washington for a positive transition to the new common assessments being developed by the SMARTER Balanced Assessment Consortium, of which Washington is a governing state.

OSPI initially estimated all testing to be available online by 2012 but further consideration of the questions involving validity, specific to the writing assessment construct, fairness (linked to graduation requirements for high school testing plus transition to end-of-course testing for 2011 mathematics and 2012 biology), and appropriate preparation, (Grade 3 students accessing online tests), the agency has re-evaluated its transition plan and is weighing a more deliberate plan to coincide with the projected administration timeframe for the new common assessments.

High School Science

Along with the adoption of new science standards linked to the MSP in Grades 5 and 8, new science standards associated with high school learning will be assessed for the first time in spring 2012. But the current comprehensive HSPE format will be replaced by an end-of-course biology assessment. OSPI has adjusted its test development activities since adoption in 2009 in preparation for administration of an end-of-course biology assessment. In a separate report to the Legislature, OSPI has provided recommendations regarding the need and feasibility of additional end-of-course assessments in science and the implications associated with graduation requirements for science.

Common Assessments

As mentioned earlier, Washington is a member-state of the SMARTER Balanced Assessment Consortium (SBAC) in the role of governing state and lead state/fiscal agent. Thirty other states are part of the consortium. The objective of the common assessment program is development of nationally accessible assessment instruments aligned to the Common Core State Standards in English language arts and mathematics. The Common Core State Standards, released in June 2010, are designed to make students exiting high school ready for college or career. The common assessment program is established as a four-year project designed to develop assessment instruments that align with the Common Core State Standards design of college and career readiness. It is the aim of SBAC to develop assessment instruments that support student learning with summative, interim, and formative measures. The consortium has been established since September 2, 2010, per notification of award for the federal Race-to-the-Top Assessment grant and in an official capacity since October 1, 2010 (technical commencement of the grant period).

Instructionally Supportive Formative Assessment

Section 1 of ESSB 5414 states that:

The legislature finds that a statewide student assessment system should improve and inform classroom instruction, support accountability, and provide useful information to all levels of the educational system, including students, parents, teachers, schools, school districts, and the state.

That section continues with an expectation that OSPI, “in consultation with the State Board of Education, shall begin design and development of an overall assessment system that meets the principles and characteristics described in this section.” Those principles and characteristics include “instructionally supportive formative assessments.”

In November 2009, the Classroom Assessment Integration (CA Integration) office was established within the Division of Assessment and Student Information, followed shortly by the formation of the Formative Assessment Technical Advisory Committee (TAC), comprised of individuals with expertise in classroom-based assessment. The members of the Formative Assessment TAC are:

- Dr. Barbara Plake, Emeritus, University of Nebraska
- Dr. James Popham, Emeritus, University of California, Los Angeles
- Dr. Joe Ryan, Emeritus, Arizona State University
- Dr. Michael Trevisan, Washington State University
- Nancy Skerritt, Assistant Superintendent for Teaching and Learning Tahoma School District
- Brian Rick, District Assessment Coordinator, Bellingham School District

The Formative Assessment TAC participated in a statewide symposium on formative assessment that engaged a broad array of stakeholders from around the state, to help set the priority actions needed for the initial implementation of formative assessments. The symposium identified four priority goal areas:

Goal 1: Development of Common Understanding of Formative Assessment Processes.

OSPI is utilizing professional development focused on ensuring internal and statewide clarity of purpose and language around the process of formative assessment. Through partnership with Education Northwest and Dr. Margaret Heritage of UCLA's National Center for Research on Evaluation, Standards, and Student Testing, OSPI staff along with our Educational Service District partners are receiving professional development on the formative assessment process throughout the 2010–11 school year.

Goal 2: Creation of Professional Development Materials and Delivery of Training.

In conjunction with Goal 1 objectives, an online formative assessment course is being developed and will be available to teachers statewide in February. Supportive instructional resource materials that support learning the formative assessment process within a professional learning community setting are near completion and will be available to teachers in the winter of the 2010–11 school year.

Goal 3: Creation of Web-based Instructional Support Materials

As the broader context of the SBAC work around formative assessment has become clearer, the CA Integration office has aligned formative assessment work around the development of web-based instructional support materials and formative assessment processes for use with classroom instructional units designed around the standards. This work is in synchronization with SBAC formative assessment workgroup goals. Additionally, CA Integration office staff will participate as a member of the formative assessment workgroup. To date, instructional units in mathematics for Grades 1, 4, and 7 have been developed and will be available to the field in January 2011.

Goal 4: Piloting of Web-Based Instructionally Supportive Student Assessment Systems

Web-Based Instructionally Supportive Student Assessment Systems

A Request for Proposals (RFP) to pilot an instructionally supportive student assessment system (ISSAS) and formative professional development materials was initiated in March 2010. The CA Integration office collaborated in this effort with OSPI's Division of District and School Improvement and Accountability (DSIA). Sixty schools representing approximately 17,000 students were set to pilot these systems with support and coordination from the CA Integration

office. Eighty-three schools representing 25,000 students for DSIA were also set to pilot these systems.

In response to recently announced budget reduction efforts, and the awarding of the SBAC grant of \$150M to create an aligned formative, interim, and summative assessment system aligned to the common core standard by 2014–15, OSPI chose to freeze the Education Legacy funds associated with this pilot. DSIA, using allocated federal funds, has moved ahead with using one of the two web-based instructionally supportive student assessment systems in their mandated work with the lowest performing schools. The second web-based instructionally supportive student assessment system identified will be piloted on a smaller scale through means other than state funding.

The development of formative assessments that are linked to the state's new mathematics learning standards, scheduled to be available in the fall of 2011, have also been put on hold given the current fiscal challenges.

WaKIDS

In June 2010, the CA Integration office initiated efforts associated with development of the Washington Kindergarten Inventory of Developing Skills (WaKIDS) program, a measure of incoming kindergarten student readiness for the learning environment of regular schooling. Nearly 3,000 children from more than 120 classrooms around the state have participated to date in a voluntary pilot program aimed at bringing families, teachers, and early learning providers together to support each child's learning and transition into public schools.

Evaluating the Validity of English Language Proficiency Assessments (EVEA)

In October 2009, OSPI was awarded a competitive Enhanced Assessment Grant by the federal Department of Education's Institute for Educational Studies (IES) to investigate ways to enhance the validity of the state's English language proficiency test, the Washington Language Proficiency Test II (WLPT-II). This grant was awarded to Washington as the lead state, in partnership with four other states Idaho, Oregon, Montana, Indiana, and several research institutions.

During the first month, Washington began studying the issues around validity with English language proficiency assessments. The underlying premise of the study and work efforts involved the development of validity arguments common to all participating states, as well as those specific to each individual state. The study plan, facilitated by edCount of Washington, D.C., has OSPI and its research partner, UCLA's National Center for Research on Evaluation, Standards, and Student Testing, constructing methodologies that will determine the level of validity and reliability associated with English language proficiency assessments, both in general terms and specific to the program employed by Washington. These efforts link to specific guidance embedded in ESSB 5414 to:

- (g) Be culturally, linguistically, and cognitively relevant, appropriate, and understandable to each student taking the assessment.
- and
- (i) Provide a way to analyze the assessment results relative to characteristics of the student such as, but not limited to, English language learners, gender, ethnicity, poverty, age and disabilities.

The CA Integration office's participation in the activities of this 18-month grant has greatly expanded the knowledge and sophistication of the staff to issues of test validity for students in diverse populations and has been applied to the formative assessment work.

Cost Analysis

An element of this annual report is to provide cost analyses of program re-design. Though the testing program will experience further changes as it adopts new standards for science and new test constructs for mathematics end-of-course, the analysis of the cost impacts from these changes was previously included in the 2009 report. The contracted supports to the assessment program will not experience any cost changes due to re-designs.

The one potential cost issue is an implementation issue directly linked to the graduation requirements associated with the mathematics end-of-course administration. Because current law states that students in the Class of 2013 and beyond will be required to meet standard on both Year 1 and Year 2 end-of-course assessments, and a large percentage of students in the Class of 2013 and beyond have completed one or more of the courses associated with the end-of-course assessments, there will be a surge of testing in this first year of assessing through the end-of-course format.

These additional students were not included in cost estimates discussed as part of the transition to end-of-course testing and will need to be addressed through added funding, adjustment to the graduation requirements, or a combination of both. Under the existing rules, the cost estimates are projected at \$1.7 million. Modifying the graduation requirements for the Class of 2013 and 2014 to passing one of two end-of-course assessments could reduce the extra costs by approximately \$600,000.

Office of Superintendent of Public Instruction
Old Capitol Building
P.O. Box 47200
Olympia, WA 98504-7200

For more information about the contents
of this document, please contact:
Michael Middleton, OSPI
E-mail: michael.middleton@k12.wa.us
Phone: (360) 725-6434

To order more copies of this document,
please call 1-888-59-LEARN (1-888-595-3276)
or visit our Web site at <http://www.k12.wa.us/publications>

Please refer to the document number below for quicker service:
10-0040

This document is available online at:
<http://www.k12.wa.us/>

This material is available in alternative format upon request.
Contact the Resource Center at (888) 595-3276, TTY (360) 664-3631

OSPI provides equal access to all programs and services without discrimination based on sex, race, creed, religion, color, national origin, age, honorably discharged veteran or military status, sexual orientation including gender expression or identity, the presence of any sensory, mental, or physical disability, or the use of a trained dog guide or service animal by a person with a disability.

For questions and complaints of alleged discrimination, contact:

Title IX/Section 504 Coordinator
Equity and Civil Rights Director
P.O. Box 47200
Olympia, WA 98504-7200
(360) 725-6162



Office of Superintendent of Public Instruction
Old Capitol Building
P.O. Box 47200
Olympia, WA 98504-7200
2011