Oregon Department of Education

2015–2016 Technical Report

Oregon's Alternate Assessment System

Peer Review Documentation: Critical Elements 1-6

Oregon's Alternate Assessment System Technical Report: Peer Review Documentation: Critical Elements 1-6

This technical report is one of a series that describes the development of Oregon's Statewide Assessment System. The complete set of volumes provides comprehensive documentation of the development, procedures, technical adequacy, and results of the system.

TABLE OF CONTENTS

Peer Review Critical Elements Reference Tables

2016 Critical Elements	
Critical Element 1	Statewide system of standards and assessments
Critical Element 2	Assessment system operations
Critical Element 3	Technical quality - validity
Critical Element 4	Technical quality - other
Critical Element 5	Inclusion of all students
Critical Element 6	Academic achievement standards and reporting

2016 Critical Elements	
Critical Element 1 - Statewide system of standards and assessments	
1.1 State adoption of academic content standards for all students	The State formally adopted challenging academic content standards for all students in reading/language arts, mathematics and science and applies its academic content standards to all public elementary and secondary schools and students in the State.
1.2 Coherent and rigorous academic content standards	The State's academic content standards in reading/language arts, mathematics and science specify what students are expected to know and be able to do by the time they graduate from high school to succeed in college and the workforce; contain content that is coherent (e.g., within and across grades) and rigorous; encourage the teaching of advanced skills; and were developed with broad stakeholder involvement.
1.3 Required Assessments	The State's assessment system includes annual general and alternate assessments (based on grade-level academic achievement standards or alternate academic achievement standards) in: Reading/language arts and mathematics in each of grades 3-8 and at least once in high school (grades 10-12); Science at least once in each of three grade spans (3-5, 6-9 and 10-12).

1.4 Policies for including all students in assessments	The State requires the inclusion of all public elementary and secondary school students in its assessment system and clearly and consistently communicates this requirement to districts and schools. For students with disabilities, policies state that all students with disabilities in the State, including students with disabilities publicly placed in private schools as a means of providing special education and related services, must be included in the assessment system; For English Learners: A) Policies state that all English learners must be included in the assessment system, unless the State exempts a student who has attended schools in the U.S. for less than 12 months from one administration of its reading/ language arts assessment; B) If the State administers native language assessments, the State requires English learners to be assessed in reading/language arts in English if they have been enrolled in U.S. schools for three or more consecutive years, except if a district determines, on a case-by-case basis, that native language assessments would yield more accurate and reliable information, the district may assess a student with native language assessments for a period not to exceed two additional consecutive years.
1.5 Participation Data	The State's participation data show that all students, disaggregated by student group and assessment type, are included in the State's assessment system. In addition, if the State administers end-of-course assessments for high school students, the State has procedures in place for ensuring that each student is tested and counted in the calculation of participation rates on each required assessment and provides the corresponding data.

Critical Element 2 - Assessment system operations	
2.1 Test Design and Development	The State's test design and test development process is well-suited for the content, is technically sound, aligns the assessments to the full range of the State's academic content standards, and includes: A) Statement(s) of the purposes of the assessments and the intended interpretations and uses of results; B) Test blueprints that describe the structure of each assessment in sufficient detail to support the development of assessments that are technically sound, measure the full range of the State's grade-level academic content standards, and support the intended interpretations and uses of the results; C) Processes to ensure that each assessment is tailored to the knowledge and skills included in the State's academic content standards, reflects appropriate inclusion of challenging content, and requires complex demonstrations or applications of knowledge and skills (i.e., higher-order thinking skills); D) If the State administers computer-adaptive assessments, the item pool and item selection procedures adequately support the test design.
2.2 Item Development	The State uses reasonable and technically sound procedures to develop and select items to assess student achievement based on the State's academic content standards in terms of content and cognitive process, including higher-order thinking skills.
2.3 Test Administration	The State implements policies and procedures for standardized test administration, specifically the State: A) Has established and communicates to educators clear, thorough and consistent standardized procedures for the administration of its assessments, including administration with accommodations; B) Has established procedures to ensure that all individuals responsible for administering the State's general and alternate assessments receive training on the State's established procedures for the administration of its assessments; C) If the State administers technology-based assessments, the State has defined technology and other related requirements, included technology-based test administration in its standardized procedures for test administration, and established contingency plans to address possible technology challenges during test administration.
2.4 Monitoring test administration	The State adequately monitors the administration of its State assessments to ensure that standardized test administration procedures are implemented with fidelity across districts and schools.

2.5 Test Security	The State has implemented and documented an appropriate set of policies and procedures to prevent test irregularities and ensure the integrity of test results through: A) Prevention of any assessment irregularities, including maintaining the security of test materials, proper test preparation guidelines and administration procedures, incident-reporting procedures, consequences for confirmed violations of test security, and requirements for annual training at the district and school levels for all individuals involved in test administration; B) Detection of test irregularities; C) Remediation following any test security incidents involving any of the State's assessments; D) Investigation of alleged or factual test irregularities.
2.6 Systems for protecting data integrity and privacy	The State has policies and procedures in place to protect the integrity and confidentiality of its test materials, test-related data, and personally identifiable information, specifically: A) To protect the integrity of its test materials and related data in test development, administration, and storage and use of results; B) To secure student-level assessment data and protect student privacy and confidentiality, including guidelines for districts and schools; C) To protect personally identifiable information about any individual student in reporting, including defining the minimum number of students necessary to allow reporting of scores for all students and student groups.
Critical Element 3 - Technical quality - validity	
3.1 Overall validity, including validity based on content	The State has documented adequate overall validity evidence for its assessments, and the State's validity evidence includes evidence that the State's assessments measure the knowledge and skills specified in the State's academic content standards, including: A) Documentation of adequate alignment between the State's assessments and the academic content standards the assessments are designed to measure in terms of content (i.e., knowledge and process), the full range of the State's academic content standards, balance of content, and cognitive complexity; B) If the State administers alternate assessments based on alternate academic achievement standards, the assessments show adequate linkage to the State's academic content standards in terms of content match (i.e., no unrelated content) and the breadth of content and cognitive complexity determined in test design to be appropriate for students with the most significant cognitive disabilities.

3.2 Validity based on cognitive processes	The State has documented adequate validity evidence that its assessments tap the intended cognitive processes appropriate for each grade level as represented in the State's academic content standards.
3.3 Validity based on internal structure	The State has documented adequate validity evidence that the scoring and reporting structures of its assessments are consistent with the sub-domain structures of the State's academic content standards on which the intended interpretations and uses of results are based.
3.4 Validity based on relations to other variables	The State has documented adequate validity evidence that the State's assessment scores are related as expected with other variables.
Critical Element 4 - Technical quality - other	
4.1 Reliability	The State has documented adequate reliability evidence for its assessments for the following measures of reliability for the State's student population overall and each student group and, if the State's assessments are implemented in multiple States, for the assessment overall and each student group, including: A) Test reliability of the State's assessments estimated for its student population; B) Overall and conditional standard error of measurement of the State's assessments; C) Consistency and accuracy of estimates in categorical classification decisions for the cut scores and achievement levels based on the assessment results; For computer-adaptive tests, evidence that the assessments produce test forms with adequately precise estimates of a student's achievement.
4.2 Fairness and accessibility	The State has taken reasonable and appropriate steps to ensure that its assessments are accessible to all students and fair across student groups in the design, development and analysis of its assessments.
4.3 Full performance continuum	The State has ensured that each assessment provides an adequately precise estimate of student performance across the full performance continuum, including for high- and low-achieving students.
4.4 Scoring	The State has established and documented standardized scoring procedures and protocols for its assessments that are designed to produce reliable results, facilitate valid score interpretations, and report assessment results in terms of the State's academic achievement standards.

4.5 Multiple assessment forms	If the State administers multiple forms within a content area and grade level, within or across school years, the State ensures that all forms adequately represent the State's academic content standards and yield consistent score interpretations such that the forms are comparable within and across school years.
4.6 Multiple versions of an assessment	If the State administers assessments in multiple versions within a content area, grade level, or school year, the State: A) Followed a design and development process to support comparable interpretations of results for students tested across the versions of the assessments; B) Documented adequate evidence of comparability of the meaning and interpretations of the assessment results.
4.7 Technical analyses and ongoing maintenance	The State has a system for monitoring and maintaining, and improving as needed, the quality of its assessment system, including clear and technically sound criteria for the analyses of all of the assessments in its assessment system (i.e., general assessments and alternate assessments).

Critical Element 5 - Inclusion of all students	
5.1 Procedures for including SWDs	The State has in place procedures to ensure the inclusion of all public elementary and secondary school students with disabilities in the State's assessment system, including, at a minimum, guidance for IEP Teams to inform decisions about student assessments that: A) Provides clear explanations of the differences between assessments based on grade-level academic achievement standards and assessments based on alternate academic achievement standards, including any effects of State and local policies on a student's education resulting from taking an alternate assessment based on alternate academic achievement standards; B) States that decisions about how to assess students with disabilities must be made by a student's IEP Team based on each student's individual needs; C) Provides guidelines for determining whether to assess a student on the general assessment without accommodation(s), the general assessment with accommodation(s), or an alternate assessment; D) Provides information on accessibility tools and features available to students in general and assessment accommodations available for students with disabilities; E) Provides guidance regarding selection of appropriate accommodations for students with disabilities; F) Includes instructions that students eligible to be assessed based on alternate academic achievement standards may be from any of the disability categories listed in the IDEA; G) Ensures that parents of students with the most significant cognitive disabilities are informed that their student's achievement will be based on alternate academic achievement standards and of any possible consequences of taking the alternate assessments resulting from district or State policy (e.g., ineligibility for a regular high school diploma if the student does not demonstrate proficiency in the content area on the State's general assessments); H) The State has procedures in place to ensure that its implementation of alternate academic achievement standards for students with the most significant cognitive disabilities promotes student access to the general curriculum.

5.2 Procedures for including ELs	The State has in place procedures to ensure the inclusion of all English learners in public elementary and secondary schools in the State's assessment system and clearly communicates this information to districts, schools, teachers, and parents, including, at a minimum: A) Procedures for determining whether an English learner should be assessed with accommodation(s); B) Information on accessibility tools and features available to all students and assessment accommodations available for English learners; C) Guidance regarding selection of appropriate accommodations for English learners.
5.3 Accommodations	The State makes available appropriate accommodations and ensures that its assessments are accessible to students with disabilities and English learners. Specifically, the State: A) Ensures that appropriate accommodations are available for students with disabilities under IDEA and students covered by Section 504; B) Ensures that appropriate accommodations are available for English learners; C) Has determined that the accommodations it provides (i) are appropriate and effective for meeting the individual student's need(s) to participate in the assessments, (ii) do not alter the construct being assessed, and (iii) allow meaningful interpretations of results and comparison of scores for students who need and receive accommodations and students who do not need and do not receive accommodations; D) Has a process to individually review and allow exceptional requests for a small number of students who require accommodations beyond those routinely allowed.
5.4 Monitoring test administration for special populations	The State monitors test administration in its districts and schools to ensure that appropriate assessments, with or without appropriate accommodations, are selected for students with disabilities under IDEA, students covered by Section 504, and English learners so that they are appropriately included in assessments and receive accommodations that are: A) Consistent with the State's policies for accommodations; B) Appropriate for addressing a student's disability or language needs for each assessment administered; C) Consistent with accommodations provided to the students during instruction and/or practice; D) Consistent with the assessment accommodations identified by a student's IEP Team or 504 team for students with disabilities, or another process for an English learner; E) Administered with fidelity to test administration procedures.

Critical Element 6 - Academic achievement standards and reporting	
6.1 State adoption of academic achievement standards for all students	The State formally adopted challenging academic achievement standards in reading/language arts, mathematics and in science for all students, specifically: A) The State formally adopted academic achievement standards in the required tested grades and, at its option, also alternate academic achievement standards for students with the most significant cognitive disabilities; B) The State applies its grade-level academic achievement standards to all public elementary and secondary school students enrolled in the grade to which they apply, with the exception of students with the most significant cognitive disabilities to whom alternate academic achievement standards may apply; C) The State's academic achievement standards and, as applicable, alternate academic achievement standards, include: (a) At least three levels of achievement, with two for high achievement and a third for lower achievement; (b) descriptions of the competencies associated with each achievement level; and (c) achievement scores that differentiate among the achievement levels.
6.2 Achievement standard setting	The State used a technically sound method and process that involved panelists with appropriate experience and expertise for setting its academic achievement standards and alternate academic achievement standards to ensure they are valid and reliable.
6.3 Challenging and aligned academic achievement standards	The State's academic achievement standards are challenging and aligned with the State's academic content standards such that a high school student who scores at the proficient or above level has mastered what students are expected to know and be able to do by the time they graduate from high school in order to succeed in college and the workforce. If the State has defined alternate academic achievement standards for students with the most significant cognitive disabilities, the alternate academic achievement standards are linked to the State's grade-level academic content standards or extended academic content standards, show linkage to different content across grades, and reflect professional judgment of the highest achievement standards possible for students with the most significant cognitive disabilities.

6.4 Reporting	The State reports its assessment results, and the reporting facilitates timely, appropriate, credible, and defensible interpretations and uses of results for students tested by parents, educators, State officials, policymakers and other stakeholders, and the public, including: A) The State reports to the public its assessment results on student achievement at each proficiency level and the percentage of students not tested for all students and each student group after each test administration; B) The State reports assessment results, including itemized score analyses, to districts and schools so that parents, teachers, principals, and administrators can interpret the results and address the specific academic needs of students, and the State also provides interpretive guides to support appropriate uses of the assessment results; C) The State provides for the production and delivery of individual student interpretive, descriptive, and diagnostic reports after each administration of its assessments that: 1) Provide valid and reliable information regarding a student's achievement; 2) Report the student's achievement in terms of the State's grade-level academic achievement standards (including performance-level descriptors); 3) Provide information to help parents, teachers, and principals interpret the test results and address the specific academic needs of students; 4) Are available in alternate formats (e.g., Braille or large print) upon request and, to the extent practicable, in a native language that parents can understand; 5) The State follows a process and timeline for delivering individual student reports to parents, teachers, and principals as soon as practicable after each test administration.

Overview

This document provides updated technical adequacy documentation for the Oregon Extended Assessment (ORExt), which is Oregon's alternate assessment based on alternate academic achievement standards (AA-AAAS). The documentation includes test design and development, technical characteristics of the instruments, and its use and impact in providing proficiency data on grade level state standards as part of the mandates from the Every Student Succeeds Act of 2015 (ESSA).

The ORExt assessments were redesigned in 2014-15, including a recently developed vertical scale in Grades 3-8 in English language arts and mathematics to support eventual determinations of student growth over time. The test is aligned to new *Essentialized Standards* (ES) that are part of comprehensive *Essentialized Assessment Frameworks* (EAFs) that have been written at three levels of complexity (low, medium, and high). The ES have been linked to grade level content and expectations, but systematically reduced in terms of depth, breadth, and complexity (RDBC). All ORExt items employed in the 2015-16 ORExt administration were developed in 2014-15; all items are aligned to the new ES. Oregon general and special education teachers have reviewed all test items for: 1) alignment to the EAFs, 2) accessibility for students with significant cognitive disabilities, 3) sensitivity, and 4) bias. All operational items met the established criteria. See *Sections 1.1, 1.2, 6.1,* and *6.3* for additional information related to the comprehensive grade level standards to ES linkage, as well as alignment of items to the ES.

The test design supports student access, including access to read aloud for directions and prompts, presentation of one item per page, and items designed at three levels of complexity where the low level complexity items include graphic and/or object support. For assessors, the scoring process has also been simplified, with answers being either correct (1) or incorrect (0). Partial credit is no longer part of the scoring metric for the ORExt. In addition, the one item per page format not only increases student ability to focus attention, but also reduces the burden on assessors to mask items that are not being tested. The field appears to have been appreciative of the redesign, particularly the Essentialized Standards and new access and efficiency features.

In addition to developing and reviewing/editing over 5,000 new items, conducting an operational field test, and developing a vertical scale, the development of a new ORExt required that new AAAS be developed and approved. Comprehensive Standard Setting meetings were conducted on June 15-17, 2015, which were then approved by the Oregon State Board of Education on June 25, 2015, including new achievement level descriptors (ALDs) and cut scores for the assessments. Comprehensive Annual Measureable Objective (AMO) reports were finalized on July 10, 2015. Our efforts shifted this year to focus on developing curricular and instructional resources.

Critical Element 1: Statewide System of Standards and Assessments

1.1 State Adoption of Academic Content Standards for All Students

The Oregon State Board of Education (SBE) adopted new, challenging academic content standards, the Common Core State Standards (CCSS), in English language arts and mathematics in Grades K-12 on October 28, 2010 (http://www.ode.state.or.us/news/announcements/announcement.aspx?ID=6156). These CCSS are utilized for all students in Oregon's public schools. Oregon was actively involved in the development of the CCSS, as the Oregon Department of Education (ODE), the Educational Enterprise Steering Committee (EESC), Oregon's Education Service Districts, and school district representatives provided feedback on the draft CCSS standards.

Similarly, the SBE adopted the Next Generation Science Standards (NGSS) on March 6, 2014 (http://www.ode.state.or.us/news/announcements/announcement.aspx?=9699). The NGSS establish learning targets for all students in Oregon's public schools in Grades K-12. The ODE and the Oregon Science Content and Assessment Panel provided direct feedback related to the NGSS. The NGSS are being phased in over time instructionally, so students are being assessed relative to the Oregon Science (ORSci) standards that were adopted in 2009.

The table below provides examples of essentialized standards in grades 5, 8, & 11 in the subject areas of English language arts (ELA), mathematics, and science. Complete EAF spreadsheets are available at the link provided here, as well. In the right column are designations for estimated difficulty of an item: L (low), M (medium), and H (high).

http://www.brtprojects.org/publications/training-modules

See *Appendix 1.1* for a User Guide that explains the development process and intended uses for the EAFs.

GRADE 5

Area	Cluster	Standard	Sub-Standard	Essential-ized Standard	L/M/H Descriptors
Reading Standards for Literature K–5	1. Key Ideas and Details	3. Compare and contrast 2 or more characters, settings, or events in a story or drama, drawing on specific details in the text (e.g., how characters interact).	None	Identify a character, setting, or event in a story read to student.	L - Sentence of 7 words or less that contains 1 character, setting, or event read to student. M - 2 short sentences that contain 1 character, setting, or event read to student. H - 2 medium sentences that contain 1 character, setting, or event read to student.
Math	Number & Operations in Base Ten	1. Understand the place value system.	1. Recognize that in a multi-digit number, a digit in one place represents 10 times as much as it represents in the place to its right and 1/10 of what it represents in the place to its left.	Use place value to compare numbers that are multiples of 10 and ones' versus tens' place and .5.	L - identify multiples of 10: 10, 20, 30, 40, 50, 60. M - identify the relation between the place values for the double-digit numbers 11, 22, 33, 44, 55. H - identify which number is in the ten's place and one's place.

Area	Cluster	Standard	Sub-Standard	Essential-ized Standard	L/M/H Descriptors
Science*	Matter and Its Interactions	**NGSS Standard:** Measure and graph quantities to provide evidence that regardless of the type of change that occurs when heating, cooling, or mixing substances, the total weight of matter is conserved.	**OR Science Standards:** 5.3S.1 Based on observations and science principles, identify questions that can be tested, design an experiment or investigation, and identify appropriate tools. Collect and record multiple observations while conducting investigations or experiments to test a scientific question or hypothesis. 5.3S.2 Identify patterns in data that support a reasonable explanation for the results of an investigation or experiment and communicate findings using graphs, charts, maps, models, and oral and written reports.	Measure and/or compare the weight of different types of matter.	L - Measure the weight/mass of common objects in various phases of matter using pictures of such objects (i.e., an object on a scale that weighs 3 pounds); M - Compare the weight/mass of common objects in various stages of matter using pictures of such objects (e.g., a balloon weighs less than a rock or glass of water) - Choose the correct tool to measure the weight/mass of objects; H - Compare the weight/mass of common objects in various phases of matter using graphs and data.

Note. The science essentialized standards are dually-linked to both NGSS and Oregon Science standards, respectively. Both general education standards are thus listed for science in these EAF tables.

GRADE 8

Area	Cluster	Standard	Sub-Standard	Essentialized Standard	L/M/H Descriptors
Reading Standards for Literature 6–12	2. Craft and Structure	6. Analyze how differences in the points of view of the characters and the audience or reader (e.g., created through the use of dramatic irony) create such effects as suspense or humor.	None	Identify the narrator or a character in a story read to student.	L - 3 sentences that contain 2 characters or narrators read to student. M - Paragraph of 4 sentences that contains 2 characters or narrators read to student. H - Paragraph of 5 sentences that contains 2 characters or narrators read to student.
Math	Statistics & Probability	1. Investigate patterns of association in bivariate data.	3. Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept. For example, in a linear model for a biology experiment, interpret a slope of 1.5 cm/hr as meaning that an additional hour of sunlight each day is associated with an additional 1.5 cm in mature plant height.	Compare rates using slower/less, faster/more, same (mph, beats per second, $ per hour, $ per lb).	L - identify faster rate using (0-20). M - identify slower , faster, or same rate using (21-50). H - identify slower, faster, or same rate using (51-100).
Science	Energy	**NGSS Standard:** Plan an investigation to determine the relationships among the energy transferred, the type of matter, the mass, and the change in the average kinetic energy of the particles as measured by the temperature of the sample.	**OR Science Standards** 8.2P.2 Explain how energy is transferred, transformed, and conserved.	Recognize temperature as a measure of how hot or cold matter is, and that heat is transferable.	L - Recognize the difference between hot and cold (e.g., objects, outside); M - Recognize that hot and cold are related to measures of temperature, including changes in temperature; H - Identify examples of heat transfer, and how such transfer might be minimized/maximized (e.g., wearing a coat to stay warm).

GRADE 11

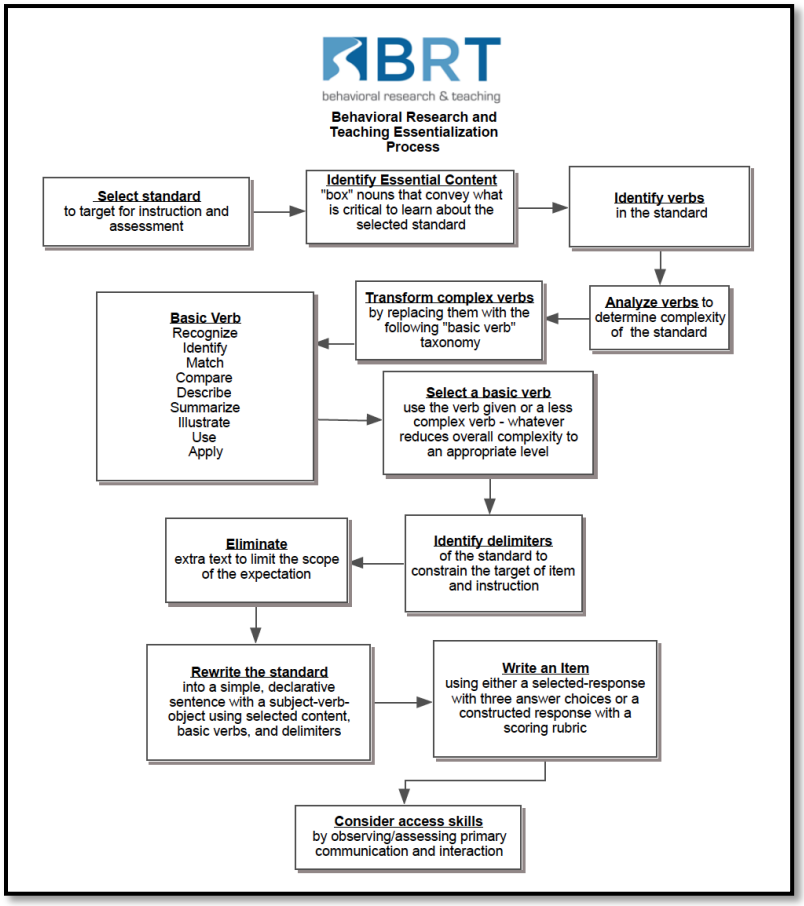
Area	Cluster	Standard	Sub-Standard	Essentialized Standard	L/M/H Descriptors
Reading Standards for Literature 6–12	2. Craft and Structure	4. Determine the meaning of words and phrases as they are used in the text, including figurative and connotative meanings; analyze the impact of specific word choices on meaning and tone, including words with multiple meanings or language that is particularly fresh, engaging, or beautiful. (Include Shakespeare as well as other authors.)	None	Identify the meaning of figurative, connotative, or words with 2 or more meanings.	L - Paragraph of 4 sentences read to student. M - Paragraph of 5 sentences read to student. H - 2 paragraphs read to student
Math	Expressing Geometric Properties with Equations	2. Use coordinates to prove simple geometric theorems algebraically	7. Use coordinates to compute perimeters of polygons and areas of triangles and rectangles, e.g., using the distance formula.	Identify the perimeter of triangles, squares, rectangles, and pentagons.	L - identify perimeter of triangles with side lengths (1-5). M - identify the perimeter of squares and rectangles with side lengths (1-10). H - identify the perimeter of pentagons with side lengths (1-20).

Area	Cluster	Standard	Sub-Standard	Essentialized Standard	L/M/H Descriptors
Science	Earth's Systems	Develop a model to illustrate how Earth's internal and surface processes operate at different spatial and temporal scales to form continental and ocean-floor features.	H.1E.2 Describe the structure and composition of Earth's atmosphere, geosphere, and hydrosphere. H.2E.1 Identify and predict the effect of energy sources, physical forces, and transfer processes that occur in the Earth system. Describe how matter and energy are cycled between system components over time. H.2E.2 Explain how Earth's atmosphere, geosphere, and hydrosphere change over time and at varying rates. Explain techniques used to elucidate the history of events on Earth.	Identify different (geoscience) processes that shape the Earth including associated Earth features. (S08ESS2.2)	L - Identify conditions that lead to specific types of surface weathering (i.e., with water, ice, or wind as vehicle - Which shows water erosion? - a river, pond or volcano); M - Identify geoscience processes that shape local geographic features (e.g., earthquakes, volcanoes, meteorites/craters - Which is an example of volcanism? – pictures of a volcano, river, rain); H - Extend M-level questions by linking features to the geoscience process (e.g., Which type of erosion process likely led to the canyon? - river, rain, wind; Which feature is associated with recent volcanism? - island, valley, river).

1.2 Coherent and rigorous Academic Content Standards

The CCSS, ORSci, and NGSS define what students in Oregon should know and be able to do by the time they graduate from high school. These CCSS, which were developed by national stakeholders and education experts, have been determined to be coherent and rigorous by researchers at the Fordham Institute (see *Appendix 1.2*). They were also developed with wide stakeholder involvement, particularly here in Oregon. The new ORExt is linked directly to the content in the CCSS in English language arts (reading, writing, & language) and mathematics. The ORExt is dually linked to the ORSci as well as the NGSS. The NGSS are widely accepted by most relevant science instruction organizations as reflective of rigorous and coherent science concepts (see http://www.nextgenscience.org/support-scientific-engineering-and-education-communities).

The new *Essentialized Assessment Frameworks* (EAFs) are publicly available at the link on page 15 above in the *Essentialized Assessment Frameworks* section. A User Guide is provided to instruct educators regarding the intended uses of the *Essentialized Standards* (ES), including the development of Present Levels of Academic Achievement and Functional Performance (PLAAFP) and Individualized Education Program (IEP) goals and objectives. The basic essentialization process employed to generate essentialized standards and write aligned items for the ORExt is outlined below. The process can also be used to support the development of curricular and instructional materials, founded in research-based pedagogy.



1.3 Required Assessments

The ORExt assessments were administered in the 2015-16 school year in ELA and math in Grades 3-8 and once in Grade 11; science is assessed in Grades 5, 8, & 11. This assessment plan meets the requirements for grade level assessment in Grades 3-8 and once in high school (Grades 10-12) for ELA and mathematics, while science is assessed once in the 3-5 grade band, once in the 6-9 grade band, and once in the 10-12 grade band:

Content Area	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
English language arts	X	X	X	X	X	X	X
Mathematics	X	X	X	X	X	X	X
Science			X			X	X

1.4 Policies for Including All Students in Assessments

Originally, Oregon statute required that all students participate in statewide assessments, with exceptions allowed for district-approved parent request for assessment waivers (parent opt-out requests) related to student disability or religious beliefs (*see* Oregon Administrative Rule, OAR § 581-022-0612)

> Exception of Students with Disabilities from State Assessment Testing
> (1) For the purposes of this rule a "student with a disability" is a student identified under the Individuals with Disabilities Education Act, consistent with OAR chapter 581, division 015, or a student with a disability under Section 504 of the Rehabilitation Act of 1973.
> (2) A public agency shall not exempt a student with a disability from participation in the Oregon State Assessment System or any district wide assessments to accommodate the student's disability unless the parent has requested such an exemption.

However, House Bill 2655 established a Student Bill of Rights on January 1, 2016, which permitted parents or adult students to annually opt-out of Oregon's statewide summative assessments, pursuant to OAR § 581-022-1910.

The Governor published a memorandum for Superintendents, Principals, and District Test Coordinators related to the change (see *Appendix 1.4.1*).

The expectation that all students in the assessed grades participate, including students with disabilities, is elaborated clearly and pervasively across all guidance documents. For example in the Oregon Test Administration Manual (TAM), where it states that, "All students enrolled in grades 3–8 and in high school must take the required Oregon Statewide Assessments offered at their enrolled grade, including students re-enrolled in the same grade as in the prior year, unless the student receives a parent-requested exemption..." (see *Appendix 1.4.2*, p. 39).

1.4A English Learners

English learners are included as appropriate in Oregon's statewide assessment system. (see *Appendix 1.4A.1*, pp. 121-123). The Smarter Balanced assessment directions are translated into multiple languages and available via the Oaks portal. OAR 581-022-0620 (2) requires ODE to provide translated OAKS assessments for populations at or above 9% in grades K-12 within three years after the school year in which the language exceeds the threshold (see *Appendix 1.4A.2*). In addition, the accommodations available to students who participate in the ORExt include translation into the native language, where appropriate (see *Appendix 2.3A1,* pp. 37-44).

1.4B Native Language Assessments

The ORExt is not administered in a native language format, though it can be translated into a student's home language.

1.5 Participation Data

Oregon's participation data indicate that most students in the tested grade levels are included in our assessment system. The students with disabilities subgroup did not meet minimum participation requirements in 2014-15, the most current data available at the time of this report, in reading or math, with rates at 93.6% and 93.2%, respectively. Documentation of this requirement is provided within the Annual Performance Report, Indicator B3, which is submitted to the United States Department of Education's (USED's) Office of Special Education Programs (OSEP). Participation and performance summaries are provided below. Additional information regarding state performance is published in the 2014-15 State Report Card (see *Appendix 1.5*, pages 1-6 for student and teacher demographics and pages 15-47 for assessment information).

Participation Details

Participation Target: 95%

English Language Arts Student Group	2014-15 Counts		Participation Rate
	Number of Participants	Number of Non-participants	
All Students	288,843	12,203	95.9
Economically Disadvantaged	166,037	5,521	96.8
English Learners[1]	37,852	591	98.5
Students with Disabilities	43,168	2,939	93.6
Underserved Races/Ethnicities	78,971	2,491	96.9
American Indian/Alaska Native	4251	181	95.9
Native Hawaiian/Pacific Islander	2,067	62	97.1
Black/African American	6522	459	93.4
Hispanic/Latino	66,131	1,789	97.4
Asian	11,832	358	97.1
White	182,127	8,561	95.5
Multi-Racial[2]	15,913	793	95.3

Mathematics Student Group	2014-15 Counts		Participation Rate
	Number of Participants	Number of Non-participants	
All Students	287,668	13,432	95.5
Economically Disadvantaged	165,364	6,219	96.4
English Learners[1]	37,783	672	98.3
Students with Disabilities	42,959	3,143	93.2
Underserved Races/Ethnicities	78,696	2,780	96.6
American Indian/Alaska Native	4,232	203	95.4
Native Hawaiian/Pacific Islander	2,057	71	96.7
Black/African American	6,468	517	92.6
Hispanic/Latino	65,939	1,989	97.1
Asian	11,785	413	96.6
White	181,348	9,367	95.1
Multi-Racial[2]	15,839	872	94.8

Notes:

1. English Learners is the new term for students identified as having a language other than English. ODE made this change to be consistent with other state reports and guidance.
2. Multi-Racial does not include students who reported Hispanic Ethnicity – these students are all reported under Hispanic. See http://www.ode.state.or.us/news/announcements/announcement.aspx?=4630 for more information.

Detailed AMO reports for Oregon public schools are available online at:
http://www.ode.state.or.us/data/reportcard/reports.aspx.
Additional Performance and Participation Data available online at:
http://www.ode.state.or.us/apps/BulkDownload/BulkDownload.Web/default.aspx.

Critical Element 2 – Assessment System Operations

2.1 Test Design and Development

The test specifications document that describes our approach to assessment and test design for the ORExt is published in *Appendix 2.1*. The document includes our approach to reducing the depth, breadth, and complexity (RDBC) of grade level content standards, an overview of the essentialization process and EAF documents, the planned test design for the ORExt, test development considerations, sample test items, item specifications, and universal tools/designated supports/accommodations. No new items were developed in 2015-16, so the 2014-15 test specifications are the most current available.

2.1A ORExt Purpose

The stated purpose of the ORExt is to provide the state technically adequate student performance data to ascertain proficiency on grade level state content standards for students with significant cognitive disabilities. A long-term goal of the program is to also provide information regarding annual student growth related to these content standards over Grades 3-8, as measured by vertically scaled assessments in ELA and mathematics. The results of the assessment are currently reported in comparison to four performance levels: Level 1, Level 2, Level 3, and Level 4. Levels 3 and 4 denote a proficient level of performance, while Levels 1 and 2 denote performance that is not proficient. BRT and ODE developed a scaled score interpretation guide to assist stakeholders in interpreting the meaning of the scaled scores generated by the ORExt, supported by the state's achievement level descriptors. This guidance is published in *Appendix 2.1A*.

2.1B ORExt Test Blueprint

Appendix 2.1B includes the entire test blueprint for the ORExt, as conveyed by the balance of representation across content areas and domains. Field-testing is conducted each year in order to support the continuous improvement of test functioning. However, items are selected to maintain this balance of representation. Oregon teachers validated the content of the assessment, agreeing with the standards that were and were not selected to develop the Essentialized Standards to which the ORExt test items are aligned.

2.1C Test Development Processes

The test development process implemented for the ORExt is conveyed in *Appendix 2.1C*, including standard selection and validation, item development, item review, review of all Oregon teacher feedback and updating of items, and scaling and item selection. The appendix articulates the process that used to generate the materials with comma separated value files used to create item templates that fed into Adobe InDesign© through a data merge. Final test packages are reviewed for accuracy and content and then disseminated via secure file transfer to Oregon Qualified Assessors.

2.1D Computer-Adaptive Considerations

The ORExt is not a computer-adaptive instrument, so these concerns do not apply.

2.2 Item Development

The item development process followed is elaborated in *Appendix 2.2.1*, which is the PowerPoint used in training all Oregon item writers. The item development process was structured with the following steps. Item writers were first oriented to the student population, as the pool of item writers included both content and special education experts. The Essentialization Process used to RDBC grade level standards was then modeled so writers would understand how the item alignment targets, the Essentialized Standards, were generated. Lecture, guided practice, and independent practice activities and follow-up discussion ensured comprehension of the process. BRT staff developed exemplar items for every Essentialized Standard, varying the complexity from Low (L) to Medium (M) to High (H) levels of difficulty to convey the different performance expectations at each level. The balanced vertical scaling design provided an overall form-to-form and grade-to-grade level framework for the test formation process once items were developed (see *Appendix 2.2.2*). Sample items are provided in *Appendix 2.2.3* for stakeholder reference, demonstrating the format and style of typical items on the ORExt.

2.3 Test Administration

The ORExt assessments are administered according to the administration, scoring, analysis, and reporting criteria established in the ORExt General Administration Manual (see *Appendix 2.3*). Important updates to the testing process are distributed via the Assessment and Accountability Updates listserve, as well (see http://www.ode.state.or.us/search/page/?id=1303). ODE uses this system to communicate information that is relevant for the statewide assessment system, including the ORExt. Announcements are sent to the listserv by email and are also posted to the ODE website. The standardization of test administration is supported by a comprehensive training process described below in Section 2.3B.

2.3A Administration and Accommodations

The state has ensured that appropriate universal tools, designated supports, and accommodations are available to students with disabilities and students covered by Section 504 by providing guidance and technical support on accommodations (see *Appendices 2.3A.1* and *2.3A.2*). Guidelines regarding use of the accommodations for instructional purposes are included in the document, as all students are expected to receive test accommodations that are consistent with instructional accommodations.

Accommodations are built into the flexibility provided by the ORExt test though they have not yet been researched for the ORExt. However, annual training and proficiency testing efforts related to becoming a qualified assessor and/or qualified trainer for the ORExt support standardized use of available accommodations that are not already part of the test design. Based on annual analyses, results demonstrate that student performance varies according to their abilities and not construct-irrelevant factors, such as sex, race, or ethnicity (See Section 4.2).

The state has ensured that appropriate accommodations are available to students with limited English proficiency by providing guidance and technical support on

accommodations (see *Appendix 2.3A.1*) Communication systems for this student population are limited; exposure to multiple languages can make a student's communication system more complex. The ORExt uses universal design principles and simplified language approaches in order to increase language access to test content for all students. In addition, directions and prompts may be translated/interpreted for students in their native language.

An analysis of accommodated versus non-accommodated administrations is needed in order to demonstrate that the provision of language accommodations is not providing any advantage to students with limited English proficiency, nor any disadvantage to other participants. This type of analysis should be feasible once accommodations information is collected during data entry, which is currently planned for the 2016 administration.

The Oregon Extended assessments can be administered using both Large Print and Braille (contracted and non-contracted) versions, as well. Oregon has ensured that the Oregon Extended assessments provide an appropriate variety of accommodations for students with disabilities. The state has provided guidance on accommodations in presentation, response, setting, and timing in the *Accommodations Manual 2013-14: How to Select, Administer, and Evaluate Accommodations for Oregon's Statewide Assessments* (see *Appendix 2.3A.2*). The Oregon Extended assessments are also designed according to universal design principles and utilize a simplified language approach (see *Appendix 2.3A.3*).

In the 2013-2014 school year, the state developed a training and proficiency program for sign language interpretation of its assessments and has updated the site annually since that time. The training process (http://lms.brtprojects.org) included videos of interpreters administering items to students, materials that support appropriate administration (i.e., transcripts and PowerPoint slides that supplement the video administrations and the current ODE accommodations manual), and proficiency testing to support standardized interpretation for Oregon's assessments, including the ORExt. A 15-item proficiency test was administered, with an 80% required for passing (12/15 items correct). In 2015-16, the site was used to train 46 participants. One participant took two attempts to pass the proficiency test. The overall average score on the proficiency test was 97%.

The ORExt assessments provide an appropriate variety of linguistic accommodations for students with limited English proficiency. They also use a simplified language approach in test development in order to reduce language load of all items systematically (see *Appendix 2.3A3*). Any given student's communication system may include home signs, school signs, English words, and Spanish words, for example. With the exception of independendent reading of prompts, the ORExt assessment can be translated or interpreted by a Qualified Assessor (QA) working with an interpreter in the student's native language. QAs are allowed to translate/interpret the test directions. QAs can adapt the assessment to meet the needs of the student, while still maintaining standardization due to systematic prompts and well-defined answers.

2.3B Comprehensive Training System

Comprehensive information for ongoing training for all qualified assessors (QAs) and Qualified Trainers (QTs) is provided in *Appendices 2.3B.1-2.3B.8*. Training and QA/QT proficiency is determined annually via an online distribution and assessment system located at https://or.k12test.com. This website hosts all resources and information needed to administer, score, report, and interpret the results from the ORExt. The website also includes proficiency assessments that are required for all QAs and QTs who may administer the ORExt. QTs are directly trained by ODE and BRT staff as part of a train the trainers model. QTs then provide direct trainings for new QAs in their respective regions.

The Oregon Department of Education (ODE) provided four direct statewide trainings for new Qualified Trainers (QTs) and Qualified Assessors (QAs) in face-to-face regional trainings. The schedule for the regional trainings, as well as relevant training information, is provided below:

Date	Who/Team	Location
11-3-2015	Team: Brad Lenhardt & Gerald Tindal Contact: Mary Apple mary.apple@imesd.k12.or.us	IMESD Pendleton, OR
11-5-2015	Team: Brad Lenhardt & Gerald Tindal Contact: Sharon Meeuwsen sharon_m@nwresd.k12.or.us	NWESD Hillsboro, OR
11-09-2015	Team; Brad Lenhardt & Gerald Tindal Contact: Eleni Boston eleni.boston@wesd.org	Willamette ESD Salem, OR
11-10-2015	Team: Brad Lenhardt & Steve Jonas Contact: Pam Wurzell pam_wurzell@soesd.k12.or.us	SOESD- Medford, OR
11-17-2015	Team: Brad Lenhardt & Gerald Tindal Contact: Catherine Halliwell-Templin Catherine.halliwell-templin@hdesd.org	HDESD- Redmond, OR

Only trained Qualified Assessors (QAs) can administer the Oregon Extended assessment. Qualified Assessors who also receive direct instruction from ODE and BRT may become

Qualified Trainers (QTs) who are certified to train local staff using the train-the-trainers model. Training for new assessors must be completed on an annual basis. Assessors who do not maintain their respective certifications for any given year must re-train if they choose to enter the system again.

The tables below contain data from the Oregon Extended Assessment Training and Proficiency Website (http://or.k12test.com/). All assessors need to complete some form of training each year to retain their status for administering the Extended Assessments.

New assessors, or returning assessors who needed further training again in 2015-16, were required to pass four proficiencies with a score of 80% or higher. These four proficiencies were in Administration, English Language Arts, Mathematics, and Science. Returning QAs or QTs for the 2015-16 school year only needed to pass a Refresher Proficiency, again with a score of 80% or higher. The tables below contain data on the number of assessors (participants) in each of the four proficiencies, as well as the Refresher Proficiency. Included in the data is the number of attempts needed to attain a passing score as well as the average passing score of the participants.

An analysis of the Oregon Extended Assessment Training and Proficiency Website showed 601 Assessors in-Training, 1,103 Qualified Assessors, and 129 Qualified Trainers.

294 Test Participants – Administration Proficiency

Number of Participants	Percentage of Participants	Attempts to Pass	Average Passing Score
202	68.7%	1	90.6%
72	24.5%	2	87.8%
18	6.1%	3	94.7%
2	<1%	4	94.0%

292 Test Participants – English Language Arts Proficiency

Number of Participants	Percentage of Participants	Attempts to Pass	Average Passing Score
282	96.6%	1	91.0%
9	3.1%	2	88.9%
1	<1%	3	100%

290 Test Participants – Mathematics Proficiency

Number of Participants	Percentage of Participants	Attempts to Pass	Average Passing Score
289	99.7%	1	96.6%
1	<1%	2	85.0%

288 Test Participants – Science Proficiency

Number of Participants	Percentage of Participants	Attempts to Pass	Average Passing Score
288	100.0%	1	93.0%

938 Test Participants – Refresher Proficiency

Number of Participants	Percentage of Participants	Attempts to Pass	Average Passing Score
897	95.6%	1	92.1%
41	4.4%	2	90.3%

Data from the training and proficiency website reflect greater test administrator familiarity with the new format of the ORExt, with a higher number of assessors only needing to pass the Refresher Proficiency. Administration Proficiency continued to be the most challenging to new assessors, but all were able to pass with the majority needing one or two attempts. With revised guidance for IEP teams who are making assessment participation decisions by the Oregon Department of Education, there were fewer students participating in the ORExt this year compared to last. There were also 255 fewer Qualified Assessors and 10 fewer Qualified Trainers compared to last year.

Evaluations are collected at each QT training in November. The results reflect general approval, but also suggest areas of improvement that ODE and BRT work on for subsequent trainings/subsequent years, as appropriate. QT evaluations this year included several positively worded statements regarding the quality of training rated on a scale where 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, and 4 = Strongly Agree. The first section evaluated the state-level information and the knowledge of the ODE presenters, the participants' level of comfort with the training provided, the participants' ability to carry this training and materials back to train district staff, and the overall utility of the training. Seventy-one percent of participants strongly agreed with these statements, 25% agreed, and 2% disagreed and strongly disagreed, respectively. In the second section, participants were asked to evaluate the BRT trainers and their guidelines regarding how to use the training and proficiency website and related resources. Sixty-five percent of participants strongly agreed with these statements, 31% agreed, and 2% disagreed and strongly disagreed, respectively. Overall, these results demonstrate that participants felt that the training was high quality and they felt confident that they could train their staff upon return to their respective districts with the knowledge and resources gained.

In addition, all technical assistance questions that we receive from the field as part of our HelpDesk are documented. The log of the technical assistance provision is reviewed each month, as well as annually, in order to determine what aspects of our assessment system need further clarification or improvement. The HelpDesk log is published in *Appendix 2.3B.9*.

Oregon monitors the quality of its system in several ways in order to support continuous improvement. In terms of the assessment quality, item statistics are reviewed each year

and items that are not functioning as intended are removed and replaced by better functioning field-test items.

In 2014-15, items were reviewed in two phases, first using classical test theory (CTT) and second using Rasch analyses. All items flagged as a result of the statistical reviews were analyzed, item-by-item, by a team of measurement and content experts at BRT. Not all flagged items were removed, as several did not have apparent design flaws. Considerations regarding domain representation as well as item difficulty range also were considered during the review process. We also employed different decision rules for unique items versus horizontally- or vertically-scaled anchor items. It was important in many cases to maintain anchor items. Items with clear design flaws were removed from subsequent analyses and reporting. The following flagging criteria were employed:

> **CTT:** A unique item was flagged if it had a *p*-value of .10 or lower, .90 or higher, or a point biserial < .15. Anchor items were flagged if they had a *p*-value of .10 or lower or .95 and higher on all forms or a point biserial < .45 on any form.
> **Rasch:** Unique items were flagged if their outfit mean square values were between 0 and .25 or > 1.5. Anchor items were flagged if their outfit mean square values were < .5, > 1.8 for horizontal items, or > 2.0 for vertical anchor items.

Out of a total of 5,929 items developed in 2014-15, 166 were removed (2.8%).

We also implement a consequential validity study each year that surveys QAs and QTs regarding the academic and social consequences of the OREXt, both intended and unintended. The Consequential Validity report is published in *Appendix 2.3B.10*. ODE and BRT staff review the results of the survey annually to determine what program improvements are needed. A summary of the results is provided below.

ODE implemented a research survey program to address the need to document the consequences, both intended and unintended, of the OREXt Assessments. The research questions have been framed based upon current consequential validity approaches for alternate assessments in the literature, as well as issues that are of specific value in Oregon. The survey included 295 respondents. This was 20% of the solicited respondents, who were all Qualified Assessors (QAs) and Qualified Trainers (QTs) in the or.k12test.com database. The sample was 86% female and represented all regions of the state, as well as age ranges. The survey included a range of quantitative and qualitative components.
The quantitative results demonstrate that QAs and QTs continue to feel that the OREXt test items were easy to administer and score (59% Strongly Agree/ 47% Agree). They also felt that the items were accessible for students who participated (21% Strongly Agree/ 47% Agree). QAs and QTs felt marginally positive about the educational impacts of the OREXt and marginally negative about its social impacts. The results demonstrate that the OREXt content area assessments generally require up to one hour to administer. In addition, of the SWSCD who participate in the OREXt, approximately 29% come from culturally and linguistically diverse backgrounds. However, on 5% of those students demonstrate higher academic communication skills in a language other than English.

The qualitative results revealed four areas in which educators appreciated the ORExt and four areas of needed improvement. QAs and QTs said that they appreciated: 1) the assessment's efficiency (i.e., more streamlined administration, ease of administration, easier to give and score online, online materials distribution); 2) overall item and test design (i.e., one item per page, visual supports, scoring protocol and student materials design, accessibility of test questions); and, 3) data entry and scoring system was easier and more clear, and 4) the availability of the ORora.

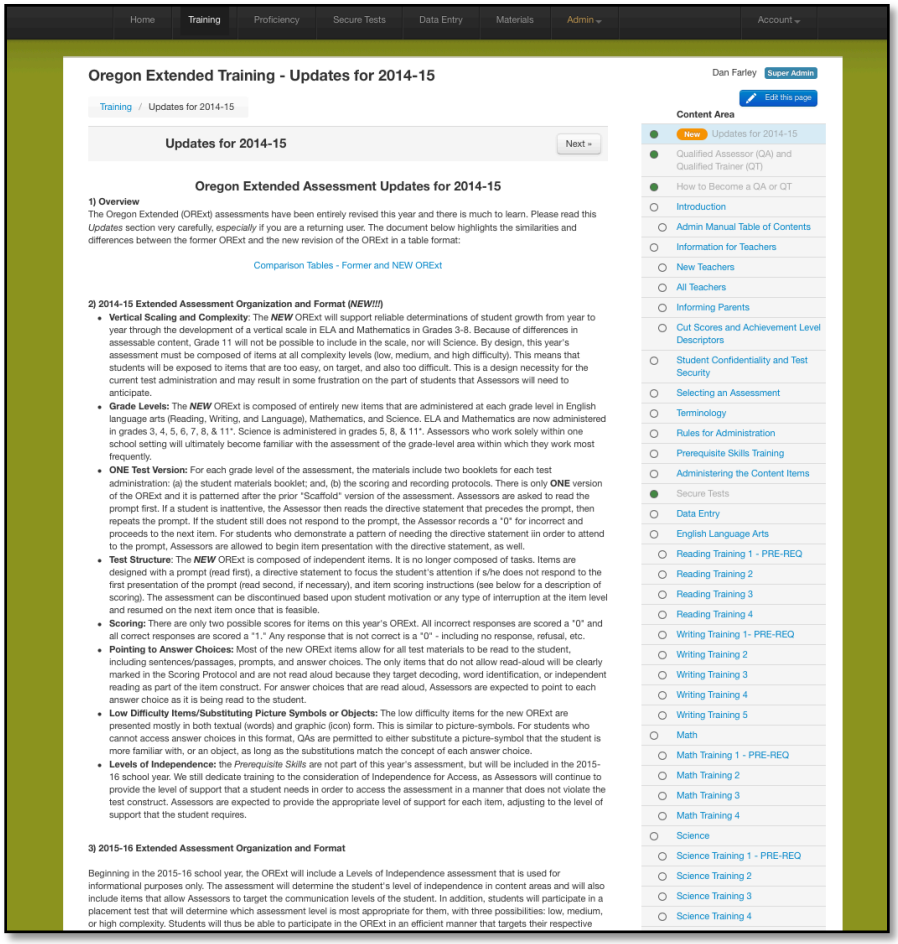
Teachers recommended the following areas of improvement, not all of which are actionable: 1) ELA assessment needs to be more difficult, 2) a functional skills assessment needs to be developed, 3) new items for very low functioning students should be developed, and 4) a more comprehensive ORora. Complete results from the survey can be found in *Appendix 2.3B.10*.

2.3C Technology-based Assessments

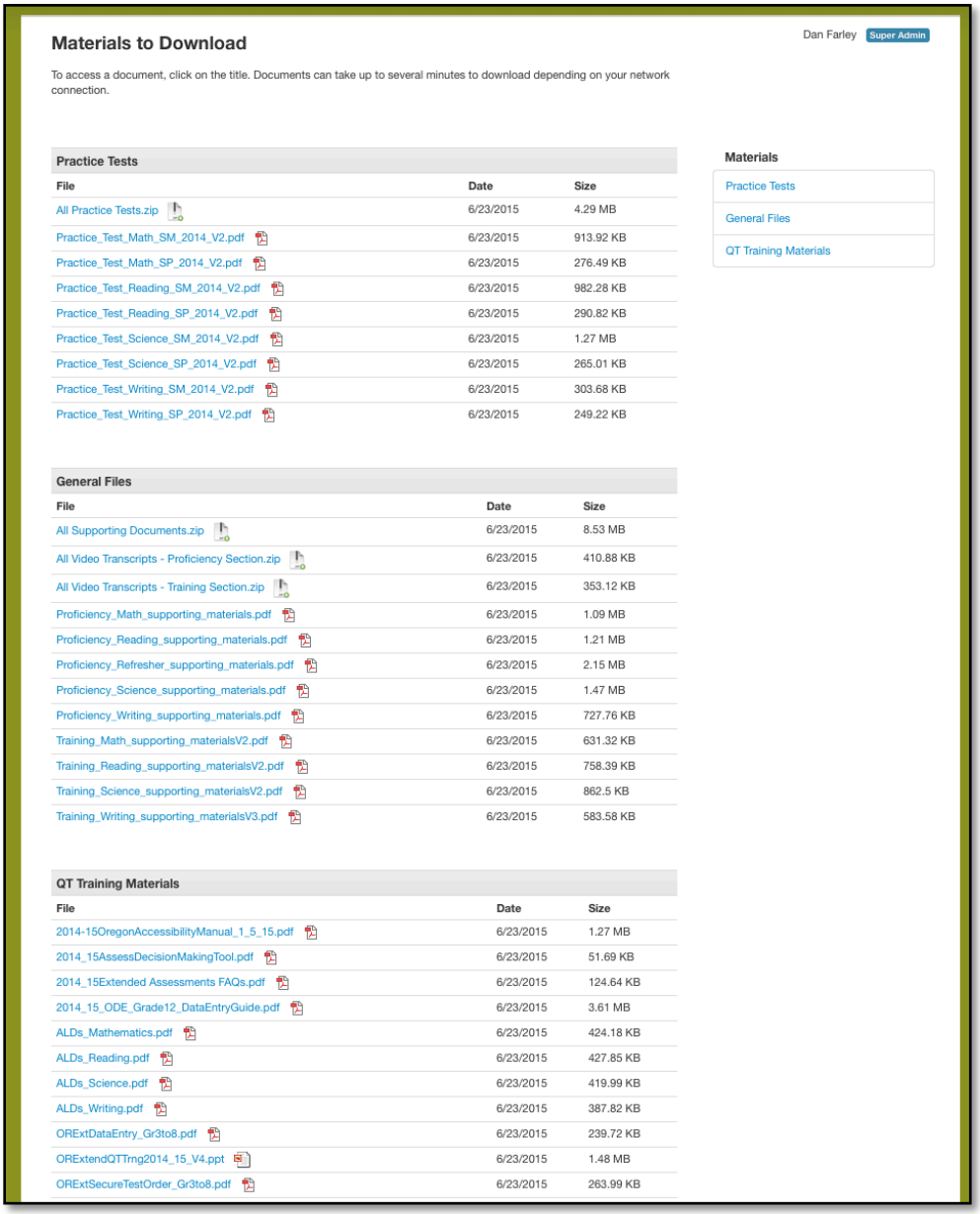
The ORExt is not currently administered in a technology-based platform. A pilot tablet administration study was conducted in the spring 2016, however. A complete report of the results of the study, including the research plan and the lessons learned, is provided in *Appendix 2.3C*. In short, this year's tablet administration study demonstrated that the writing items and audio files need further refinement. Practice items in a tablet format also need to be provided for stakeholders to prepare for future tablet-based administrations. Next year, in year two of the three-year tablet administration development process, we will address issues related to database communication systems to ensure data security and accurate data storage and access. In year three, the pilot tablet administration will be scaled up to the district level. It is hoped that ODE will be able to provide a comprehensive, well-supported tablet administration for students who are able to participate in this modality by the spring of 2019. The paper/pencil version will also continue to be available throughout this timeframe.

2.4 Monitoring Test Administration

The ODE maintains a rigorous training system to support standardized test administration for the ORExt, located at https://or.k12test.com (secure website, but see screenshot on below for an example of training content).



The or.k12test.com website includes a training section that addresses any systems updates, the process for becoming a Qualified Assessor or Qualified Trainer, student eligibility expectations, student confidentiality and test security, test administration and scoring expectations, examples of appropriate and inappropriate administration (video), supporting student access to items without violating the test construct, content area trainings that demonstrate how to administer items in ELA, Math, and Science (video, with supporting test materials), and how to access secure tests and complete data entry. Information for QAs, QTs, and parents regarding the OREXt is also provided, as are all necessary support materials. For QAs, these materials include practice tests to prepare both themselves and students for the annual assessment and all of the training materials used on the website. In addition to these materials, QTs have access to all training materials necessary to provide annual training to QAs in their purview (see screenshot below):



Materials to Download

Dan Farley Super Admin

To access a document, click on the title. Documents can take up to several minutes to download depending on your network connection.

Practice Tests

File	Date	Size
All Practice Tests.zip	6/23/2015	4.29 MB
Practice_Test_Math_SM_2014_V2.pdf	6/23/2015	913.92 KB
Practice_Test_Math_SP_2014_V2.pdf	6/23/2015	276.49 KB
Practice_Test_Reading_SM_2014_V2.pdf	6/23/2015	982.28 KB
Practice_Test_Reading_SP_2014_V2.pdf	6/23/2015	290.82 KB
Practice_Test_Science_SM_2014_V2.pdf	6/23/2015	1.27 MB
Practice_Test_Science_SP_2014_V2.pdf	6/23/2015	265.01 KB
Practice_Test_Writing_SM_2014_V2.pdf	6/23/2015	303.68 KB
Practice_Test_Writing_SP_2014_V2.pdf	6/23/2015	249.22 KB

Materials

Practice Tests
General Files
QT Training Materials

General Files

File	Date	Size
All Supporting Documents.zip	6/23/2015	8.53 MB
All Video Transcripts - Proficiency Section.zip	6/23/2015	410.88 KB
All Video Transcripts - Training Section.zip	6/23/2015	353.12 KB
Proficiency_Math_supporting_materials.pdf	6/23/2015	1.09 MB
Proficiency_Reading_supporting_materials.pdf	6/23/2015	1.21 MB
Proficiency_Refresher_supporting_materials.pdf	6/23/2015	2.15 MB
Proficiency_Science_supporting_materials.pdf	6/23/2015	1.47 MB
Proficiency_Writing_supporting_materials.pdf	6/23/2015	727.76 KB
Training_Math_supporting_materialsV2.pdf	6/23/2015	631.32 KB
Training_Reading_supporting_materialsV2.pdf	6/23/2015	758.39 KB
Training_Science_supporting_materialsV2.pdf	6/23/2015	862.5 KB
Training_Writing_supporting_materialsV3.pdf	6/23/2015	583.58 KB

QT Training Materials

File	Date	Size
2014-15OregonAccessibilityManual_1_5_15.pdf	6/23/2015	1.27 MB
2014_15AssessDecisionMakingTool.pdf	6/23/2015	51.69 KB
2014_15Extended Assessments FAQs.pdf	6/23/2015	124.64 KB
2014_15_ODE_Grade12_DataEntryGuide.pdf	6/23/2015	3.61 MB
ALDs_Mathematics.pdf	6/23/2015	424.18 KB
ALDs_Reading.pdf	6/23/2015	427.85 KB
ALDs_Science.pdf	6/23/2015	419.99 KB
ALDs_Writing.pdf	6/23/2015	387.82 KB
ORExtDataEntry_Gr3to8.pdf	6/23/2015	239.72 KB
ORExtendQTTrng2014_15_V4.ppt	6/23/2015	1.48 MB
ORExtSecureTestOrder_Gr3to8.pdf	6/23/2015	263.99 KB

In addition, monitoring and reporting related to test administration issues for the ORExt is addressed via general ODE reporting systems. Information regarding this process can be located in the general assessment system Peer Review evidence submission.

2.5 Test Security

2.5A Prevention of Assessment Irregularities

Test security policies and consequences for violation are addressed in the Test Administration Manual on an annual basis (see *Appendix 1.4.2,* p. 23-28). These policies include test material security, proper test preparation guidelines and administration procedures, consequences for confirmed violations of test security, and annual training requirements at the district and school levels for all individuals involved in test

administration. Consequences for adult-initiated test irregularities may be severe, including placing teaching licenses in jeopardy (see *Appendix 1.4.2,* p. 31-33).

2.5B Detection of Test Irregularities

The ODE utilizes a localized monitoring system where school test coordinators oversee building-level administration by trained, Qualified Assessors, and report to centralized district test coordinators, who are then responsible for reporting any confirmed violations to ODE. Improprieties are defined as adult-initiated or student-initiated and investigated accordingly (see *Appendix 1.4.2,* p. 29-33).

2.5C Remediation Following Test Security Incidents

ODE's alternate assessment program manager investigates and remediates substantiated test security incidents for the OREXt by working with district test coordinators. Additional information regarding this process can be located in the general assessment system Peer Review evidence submission.

2.5D Investigation of Test Irregularities

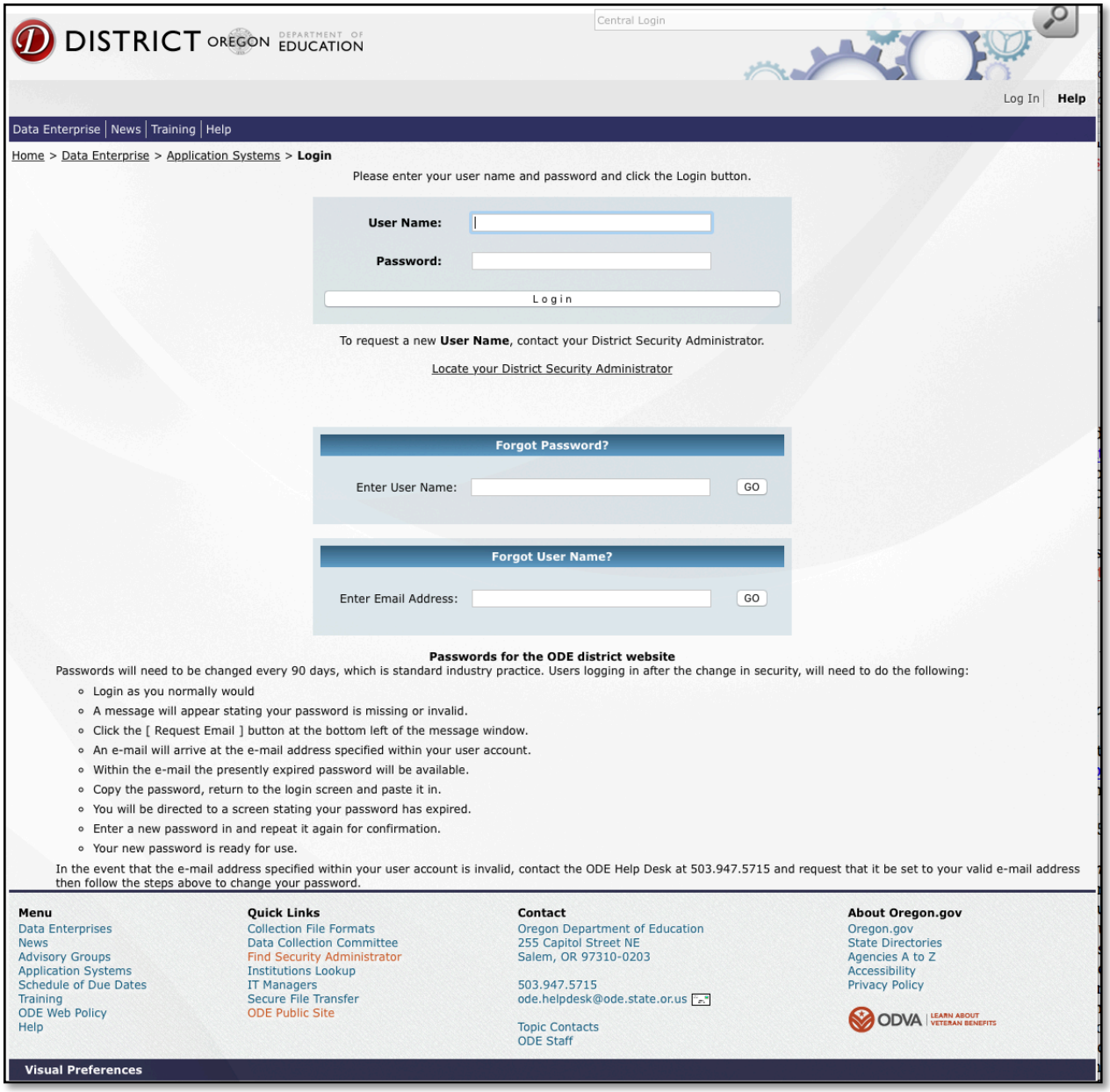
School and district test coordinators conduct initial investigations into all alleged test irregularities. Once reported to ODE, all alleged test irregularities are investigated in consultation with district test coordinators and the test vendor, as appropriate (see *Appendix 1.4.2,* p. 28-33). In the event that a test irregularity is determined to be factual, consequences are determined based upon contextual issues that are brought to light during the investigation. Additional information regarding this process can be located in the general assessment system Peer Review evidence submission.

2.6 Systems for Protecting Data Integrity and Privacy

2.6A Integrity of Test Materials

Test materials for the OREXt are maintained throughout development, dissemination, and administration via multiple mechanisms. All items under development are stored in secure file servers managed by Behavioral Research & Teaching at the University of Oregon, the test vendor for the OREXt. Item reviews necessary to provide alignment, bias, and sensitivity information are conducted online using the secure Distributed Item Review (DIR) platform http://brtitemreview.com (secure website, but see *Appendix 3.1B* for a system overview).

For the 2015-2016 school year, all secure test distribution and data entry was hosted by ODE's secure file transfer system, which is a password-protected test distribution and data entry system located at https://district.ode.state.or.us/apps/login/ (secure website, but see screenshot below for reference). A data entry guide is provided in *Appendix 2.6*.



Additional information regarding test security can be located in the general assessment system Peer Review evidence submission.

2.6B Secure Student-Level Assessment Data

Student level data is protected by relevant training and through a secure data system in which all data entry is conducted online using password-protected, secure procedures on the https://or.k12test.com or https://district.ode.state.or.us/apps/login/ websites, as identified above. Only trained users with a vested educational interest who have signed test security agreements are authorized to access to online data entry systems. See *Appendix 2.6* for additional data entry expectations for 2015-16.

2.6C Protecting Personally Identifiable Information

All confidential, personally identifiable student information is protected by policy and supported by training (see *Appendix 1.4.2,* p. 27). The minimum number of students necessary to allow reporting of students and student subgroups varies by rating (i.e., achievement, growth, graduation, and school size) by level (i.e., school/district/state), and by number of years of assessment data available. For example, to receive an achievement rating, schools must have at least 40 tests for the two most recent school years in reading or mathematics. Alternatively, small schools receive an achievement rating if they have at least 40 tests over the most recent four years. If a school does not have at least 40 tests over a four-year period, they will not receive an achievement score (see *Appendix 2.6C)*. Similar rules are applied to student subgroups, including students with disabilities, English learners, and students from diverse racial/ethnic backgrounds (see *Appendix 2.6C*, p. 7).

Critical Element 3 – Technical Quality: Validity

3.1 Overall Validity, Including Validity Based on Content

As elaborated by Messick (1989)[1], the validity argument involves a claim with evidence evaluated to make a judgment. Three essential components of assessment systems are necessary: (a) constructs (what to measure), (b) the assessment instruments and processes (approaches to measurement), and (c) use of the test results (for specific populations). Validation is a judgment call on the degree to which each of these components is clearly defined and adequately implemented.

Validity is a unitary concept with multifaceted processes of reasoning about a desired interpretation of test scores and subsequent uses of these test scores. In this process, we want answers for two important questions. Regardless of whether the students tested have disabilities, the questions are identical: (1) How valid is our interpretation of a student's test score? and (2) How valid is it to use these scores in an accountability system? Validity evidence may be documented at both the item and total test levels. We use the *Standards*[2] (AERA et al., 2014) in documenting evidence on content coverage, response processes, internal structure, and relations to other variables. This document follows the essential data requirements of the federal government as needed in the peer review process.[3] The critical elements highlighted in Section 4 in that document (with examples of acceptable evidence) include (a) academic content standards, (b) academic achievement standards, (c) a statewide assessment system, (d) reliability, (e) validity, and (f) other dimensions of technical quality.

In this technical report, data are presented to support the claim that Oregon's AA-AAAS provides the state technically adequate student performance data to ascertain proficiency on grade level state content standards for students with significant cognitive disabilities – which is its defined purpose. The AA-AAAS are linked to grade level academic content, generate reliable outcomes at the test level, include all students, have a cogent internal structure, and fit within a network of relations within and across various dimensions of content related to and relevant for making proficiency decisions. Sample items that convey the design and sample content of ORExt items are provided in *Appendix 2.2.3*.

The assessments are administered and scored in a standardized manner. Assessors who administer the ORExt are trained to provide the necessary level of support for appropriate test administration on an item-by-item basis. There are four levels of support outlined in training: full physical support, partial physical support, prompted support, and no support. Items were designed to document students' skill and knowledge on grade level academic content standards, with the level of support provided designed not to interfere with the

---

[1] Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education.

[2] American Educational Research Association (AERA), American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing.* Washington, DC: AERA.

[3] U. S. Department of Education (2015). *Peer Review of State Assessment Systems: Non-Regulatory Guidance for States for Meeting the Requirements of the Elementary and Secondary Education Act of 1965, as Amended.*

construct being measured. Only one test administration type is used for the ORExt, patterned after the former *Scaffold* version of the assessment. Assessors administer the prompt and if the student does not respond, the Assessor reads a directive statement designed to focus the student's attention upon the test item and then repeats the prompt. If the student still does not respond, the Assessor repeats the prompt as needed and otherwise scores the item as incorrect and moves on to the next item. Training documentation is provided in *Appendices 2.3B.1-2.3B.8*.

Given the content-related evidence that we present related to test development, alignment, training, administration, scoring, the reliability information reflected by adequate coefficients for tests, and, finally, the relation of tests across subject areas (providing criterion-related evidence), we conclude that the alternate assessment judged against alternate achievement standards allows valid inferences to be made on state accountability proficiency standards.

3.1A Alignment Between AA-AAAS and Academic Content Standards

Our foundation of validity evidence from content coverage for the ORExt comes in the form of test specifications (see *Appendix 2.1*) and test blueprints (see *Appendix 2.1B*). Among other things, the *Standards* (AERA et al., 2014)[2] suggest specifications should "define the content of the test, the proposed test length, the item formats…" (Standard 4.2, p. 85).[2]

All items are linked to grade level standards and a prototype was developed using principles of universal design[4] with traditional, content-referenced multiple-choice item writing techniques[5]. The most important component in these initial steps addressed language complexity and access to students using both receptive, as well as expressive, communication. Additionally, both content breadth and depth were addressed. We developed one test form for the ORExt that utilizes a scaffold approach. This approach allows for students with very limited attention to access test content, while the supports are not utilized for students who do not need this support.

We developed the test iteratively by developing items (see *Appendix 2.2.1,* which conveys our item writer training materials), piloting them, reviewing them, and editing successive drafts. We used a combination of existing panels of veteran teachers who have worked with the Oregon Department of Education (ODE) in various advising roles on testing content in general and special education, using the same processes and criteria, as well as the introduction of newer teachers who are qualified as we proceed to remain relevant. Behavioral Research and Teaching (BRT) personnel conducted the internal reviews of content. After the internal development of prototype items, all reviews then involved Oregon content and special education experts with significant training and K-12 classroom experience.

---

[4] Johnstone, C., Thompson, S., Bottsford-Miller, N., & Thurlow, M. (2008). Universal design and multimethod approaches to item review. *Educational Measurement, 27*(1), 26-36.

[5] Halydyna, T. , & Rodriquez, M. C. (2013). *Developing and validating test items*. New York, NY: Routledge.

The ORExt incorporates continuous improvement into its test design via field-testing in all content areas on an annual basis, with an average of 25% new items. These items are compared to operational items based on item functioning and test design factors, generating data used to replace items on an annual basis, incorporating the new items that fill a needed gap with regard to categorical concurrence, or provide for a wider range of functioning with regard to complexity levels: low – medium – high, comparable to Webb's DOK (see *Section 3.1A*).

BRT employed a multi-stage development process in 2014-15 to ensure that test items were linked to relevant content standards, were accessible for students with significant cognitive disabilities, and that any perceived item biases were eliminated. The item review process included 51 reviewers with an average of 22 years of experience in education. The ORExt assessments have been determined to demonstrate strong linkage to grade level academic content, overall. Full documentation of the linkage study is provided in *Appendix 3.1A.* No item development was required in 2015-16.

Because the assessments demonstrate sufficient to strong linkage to Oregon's general education content standards and descriptive statistics demonstrate that each content area assessment is functioning as intended, it is appropriate to deduce that these standards define the expectations that are being measured by the Oregon Extended assessments.

The Oregon Extended assessments yield scores that reflect the full range of achievement implied by Oregon's alternate achievement standards. Evidence of this claim is found in the standard setting documentation submitted in *Section 6.2*. Standards were set for all subject areas on June 15-17, 2015. Standards included achievement level descriptors and cut scores, which define Oregon's new alternate achievement standards (AAS). The State Board of Education officially adopted the AAS on June 25, 2015.

3.1B AA-AAAS Linkage to General Content Standards

Complete results of the analysis of the linkage of the new *Essentialized Assessment Frameworks*, (EAF), composed of *Essentialized Standards* (ES), to grade level CCSS in English language arts and mathematics and linked to ORSci and NGSS in science, are presented in *Section 3.1A*. The claim is that the ES are sufficiently linked to grade level standards, while the ORExt items are aligned to the ES. In addition to presenting linkage information between grade level content standards and the ES, the linkage study presents alignment information related to the items on the new ORExt in comparison to the ES. Extended assessments have been determined to link sufficiently to grade level academic content standards. Field test items are added each year based on item alignment to standards.

The Oregon Extended assessments link to grade level academic content, as reflected in the item development process. Oregon also had each operational item used on the Oregon Extended assessment evaluated for alignment by as part of the comprehensive linkage study (see *Section 3.1A*). The professional reviewers included both special and general education experts, with content knowledge and experience in addition to special education expertise. Reviewers were trained by synchronous webinars on linkage/alignment, as well

as item depth, breadth, and complexity and then completed their ratings online via BRT's Distributed Item Review (DIR) website (see *Appendix 3.1B* for an overview). Mock linkage ratings were conducted in order to address questions and ensure appropriate calibration. Reviewers rated each essentialized standard on a 3-point scale (0 = no link, 1= sufficient link, 2= strong link) as it related to the standard the test developers had defined for that essentialized standard. Items were evaluated, in turn, based upon their alignment to the essentialized standard on a 3-point scale (0 = insufficient alignment, 1 = sufficient alignment, 2 = strong alignment). Adequate linkage was defined as being rated a 1 or 2 by at least two raters. Additional comment was requested for any essentialized standard or item whose linkage was rated 0. Essentialized standards or items that did not meet this criterion were either revised to meet the criterion, or were not utilized for the operational assessment.

3.2 Validity Based on Cognitive Processes

Evidence of content coverage is concerned with judgments about "the extent to which the content domain of a test represents the domain defined in the test specifications" (AERA et al., 2014, Standard 4.12, p. 89)[7]. As a whole, the OREXt is comprised of sets of items that sample student performance on the intended domains. The expectation is that the items cover the full range of intended domains, with a sufficient number of items so that scores credibly represent student knowledge and skills in those areas. Without a sufficient number of items, the potential exists for a validity threat due to construct under-representation (Messick, 1989)[6].

The OREXt assessment is built upon a variety of items that address a wide range of performance expectations rooted in the CCSS, NGSS, and ORSci content standards. The challenge built into the test design is based first upon the content within each standard in English language arts, mathematics, and science. That content is RDBC in a manner that is verified by Oregon general and special education teachers to develop assessment targets that are appropriate for students with the most significant cognitive disabilities. Our assessments utilize universal design principles in order to include all students in the assessment process, while effectively challenging the higher performing students. For students who have very limited to no communication and are unable to access even the most accessible items on the OREXt, an Observational Rating Assessment (ORA) was developed that will be implemented in 2015-16. The ORA is completed by teachers and documents the student's level of communication complexity (expressive and receptive), as well as level of independence in the domains of attention/joint attention and mathematics. A complete report of ORora results from 2015-16 is provided in *Appendix 5.1D.*

Fifty-one reviewers analyzed all OREXt items for bias, sensitivity, accessibility to the student population, and alignment to the *Essentialized Standards*. A total of 21 reviewers were involved in the English language arts item reviews. An additional 21 reviewers were involved in the Mathematics item reviews. Science employed nine reviewers. Reviewers were organized into grade level teams of two special educators and one content specialist.

---

[6] Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education.

Substantive evidence that has been documented suggests that the ORExt items are tapping the intended cognitive processes and that the items are at the appropriate grade level through the linkage/alignment studies documented above, including reviews of linkage, content coverage, and depth of knowledge. A comprehensive report of the item review process is available in *Appendix 3.1A.*

3.3 Validity Based on Internal Structure (Content and Function)

The Oregon Extended assessments reflect patterns of emphasis that are supported by Oregon educators as indicated by the following three tables that highlight the balance of standard representation by grade level for English language arts, mathematics, and science on the ORExt. The representation ratios can be calculated by dividing the standards by the total within each respective column. For example, in Grade 3 Reading, approximately 25% of the items are in the Reading Standards for Literature domain, as that domain has 4 written Essentialized Standards (ES) out of the total of 16 (4/16 = 25%).

The test blue prints below directly correspond to the number of ES written in each domain within the Essentialized Assessment Frameworks (EAF) spreadsheets. There are additional grade level standards addressed by the ES, as some ES link to multiple grade level content standards. However, the blueprints below reflect only the written ES and are thus an underrepresentation of the breadth of grade level content addressed by the ORExt.

English Language Arts

Domain	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
RF	2	2	2				
RI	4	4	4	5	5	5	5
RL	4	4	4	5	5	5	5
WR	4	4	4	4	4	4	4
LA	2	2	2	2	2	2	2
TOTAL	**16**	**16**	**16**	**16**	**16**	**16**	**16**

Note. RF = Reading Standards: Foundational Skills. RI = Reading Standards for Informational Text. RL = Reading Standards for Literature. WR = Writing. LA = Language.

Mathematics

Domain	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
OAT	7	4	3				
NBT	2	6	8				
NOF	3	8	6				
MED	8	5	4				
GEO	2	3	2	3	3	4	7
RPR				3	2		
TNS				9	7	2	
EXE				6	2	6	
STP				5	6	3	5
FUN						4	7
NAQ							2
ALG							2
TOTAL	**22**	**26**	**23**	**26**	**20**	**19**	**23**

Note. OAT = Operations and Algebraic Thinking. NBT = Numbers and Operations in Base Ten. NOF = Numbers and Operations – Fractions. MED = Measurement and Data. GEO = Geometry. RPR = Ratio and Proportional Relationships. TNS = The Number System. EXE = Expressions and Equations. STP = Statistics and Probability. FUN = Functions. NAQ = Numbers and Quantities. ALG = Algebra.

Science

Domain	Grade 5	Grade 8	Grade 11
LFS	4	9	8
PHS	4	7	9
ESS	4	6	6
ETS	2	2	
TOTAL	**14**	**24**	**23**

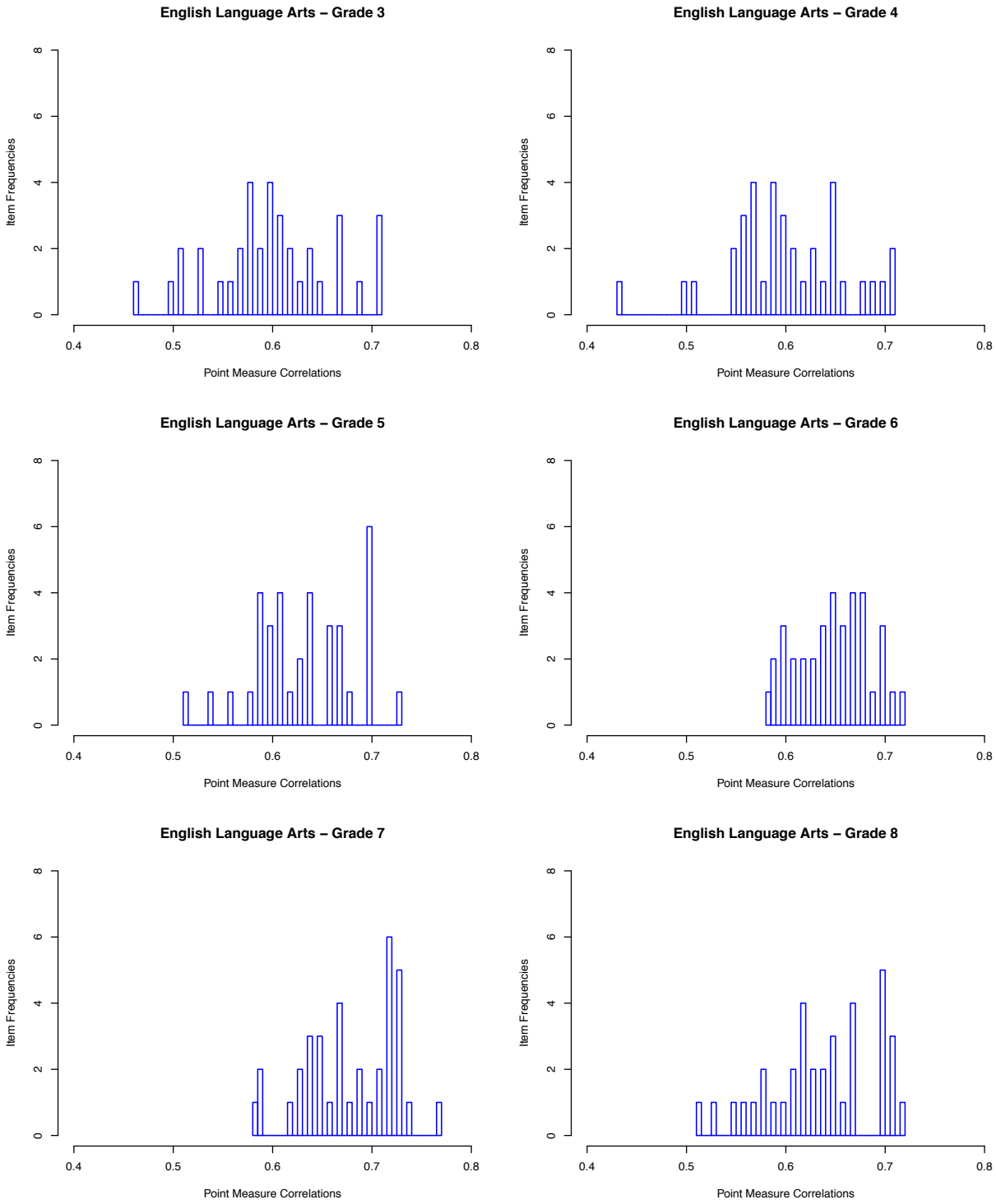
Note. LFS = Life Science Standards. PHS = Physical Sciences. ESS = Earth and Space Sciences. ETS = Engineering, Technology, and Applications.

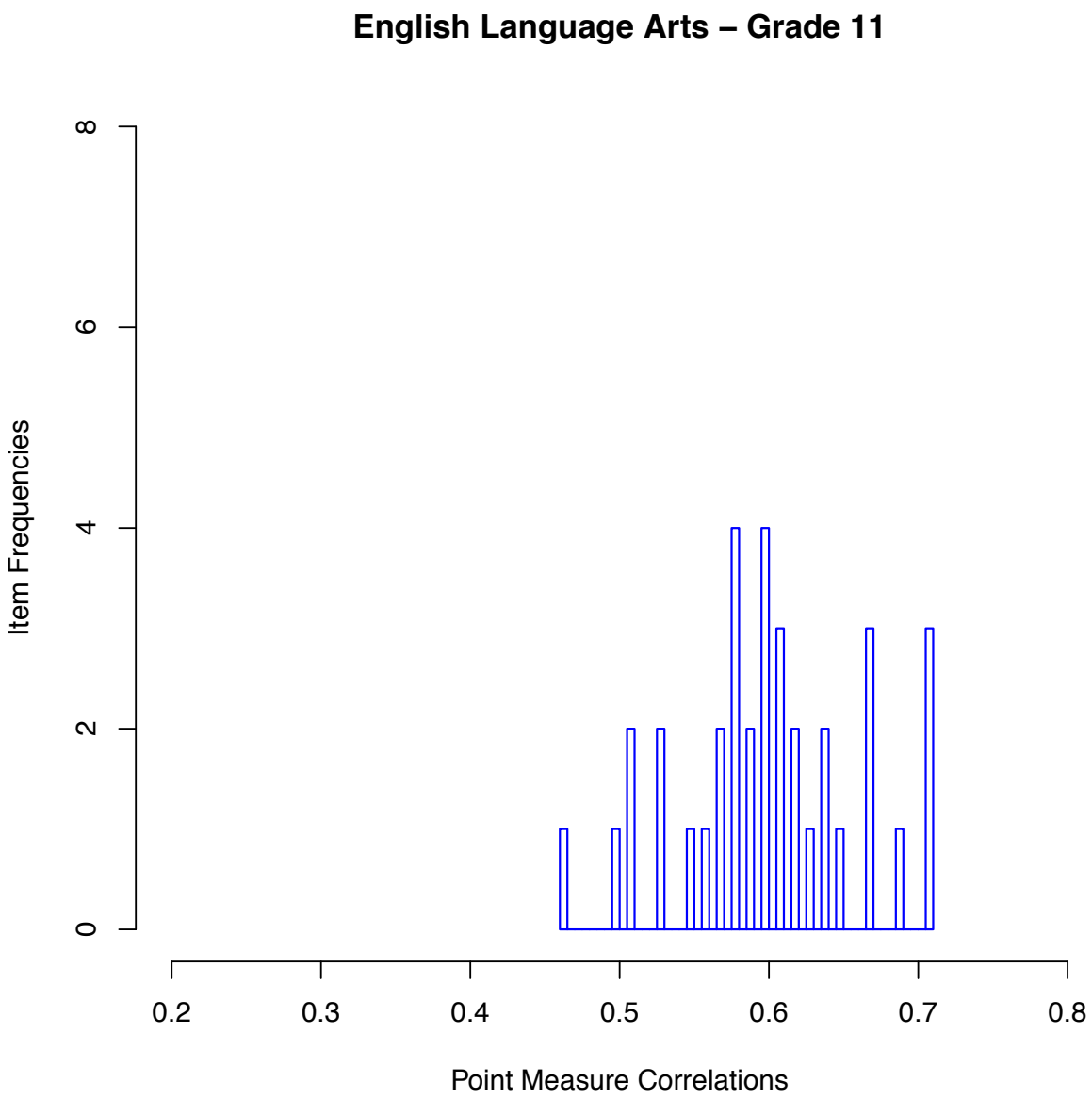
The primary purpose of the ORExt assessment is to yield technically adequate performance data on grade level state content standards for students with significant cognitive disabilities in English language arts, mathematics, and science at the test level. All scoring and reporting structures mirror this design and have been shown to be reliable measures at the test level (see *Section 4.1*). The process of addressing any gaps or weaknesses in the system is accomplished via field-testing (see *Section 3.1A*).

Distributions of point measure correlations and outfit mean square statistics for operational items are provided below, by content area and grade. Point measure correlations display how the item scores correlate with the latent overall score, while outfit

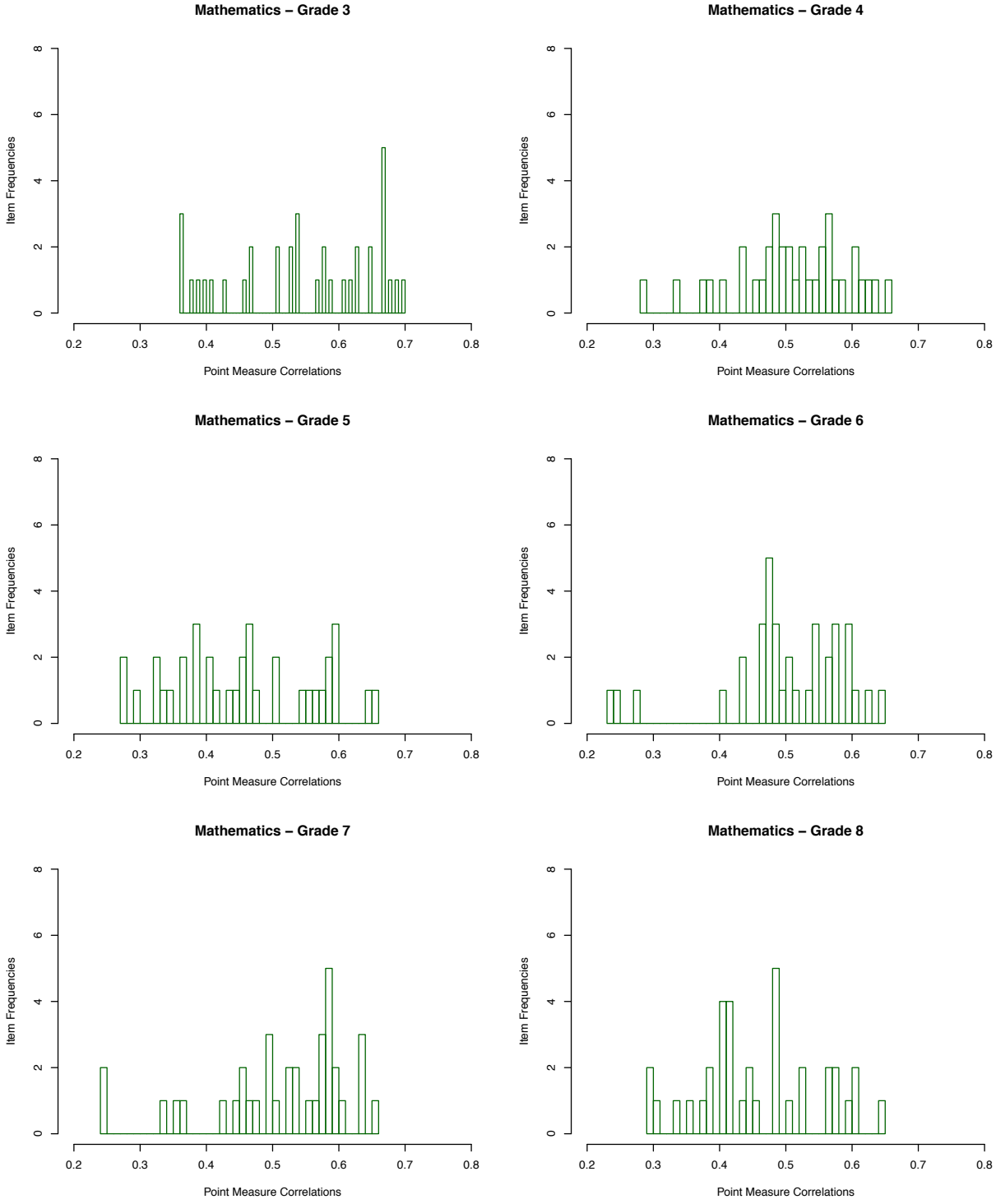
mean square statistics closer to 1.0 denote minimal distortion of the measurement system. All items included in the 2015-16 operational assessment are represented. Point measure correlations in ELA range from 0.43 to 0.76.

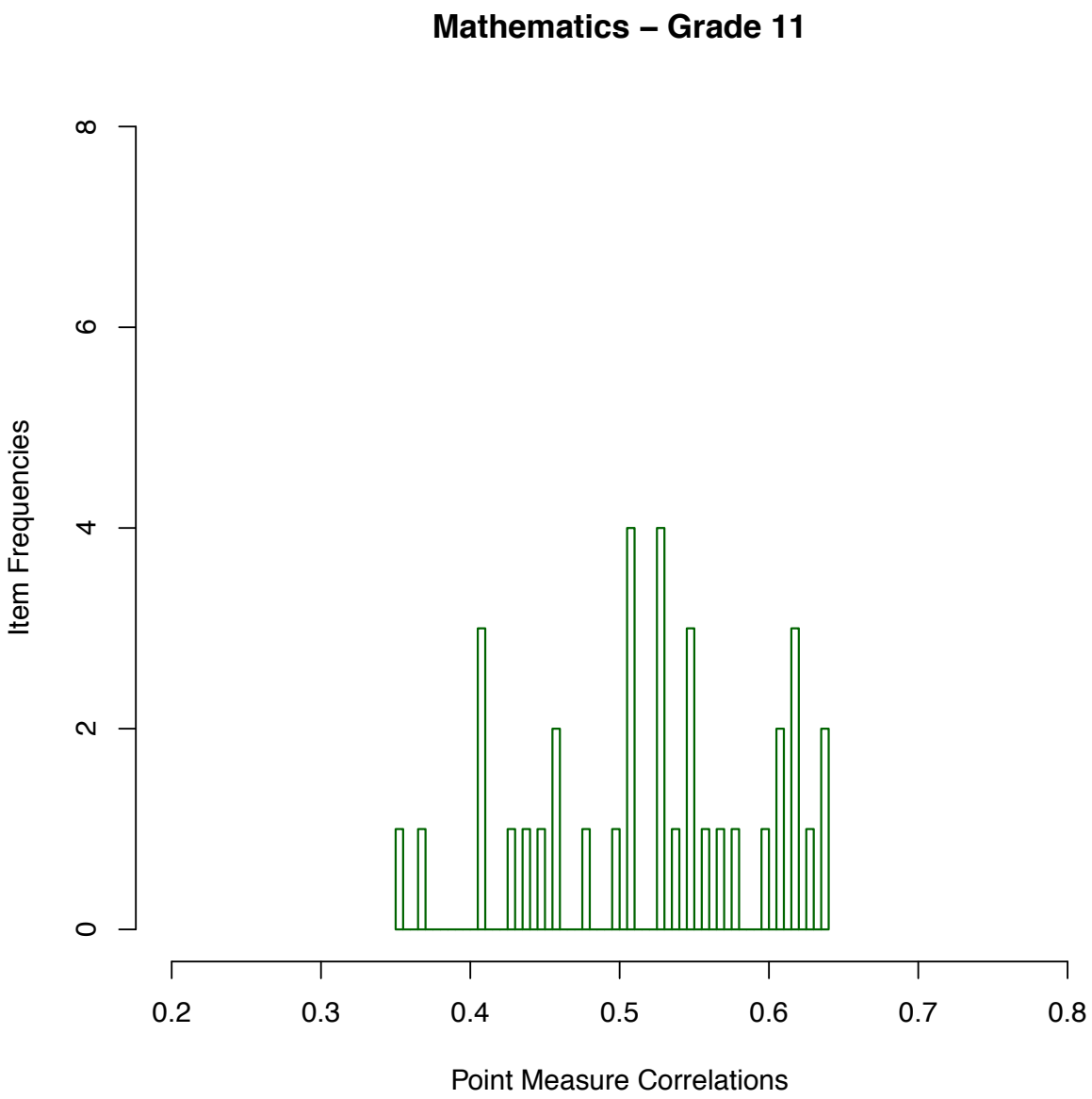
Point Measure Correlations



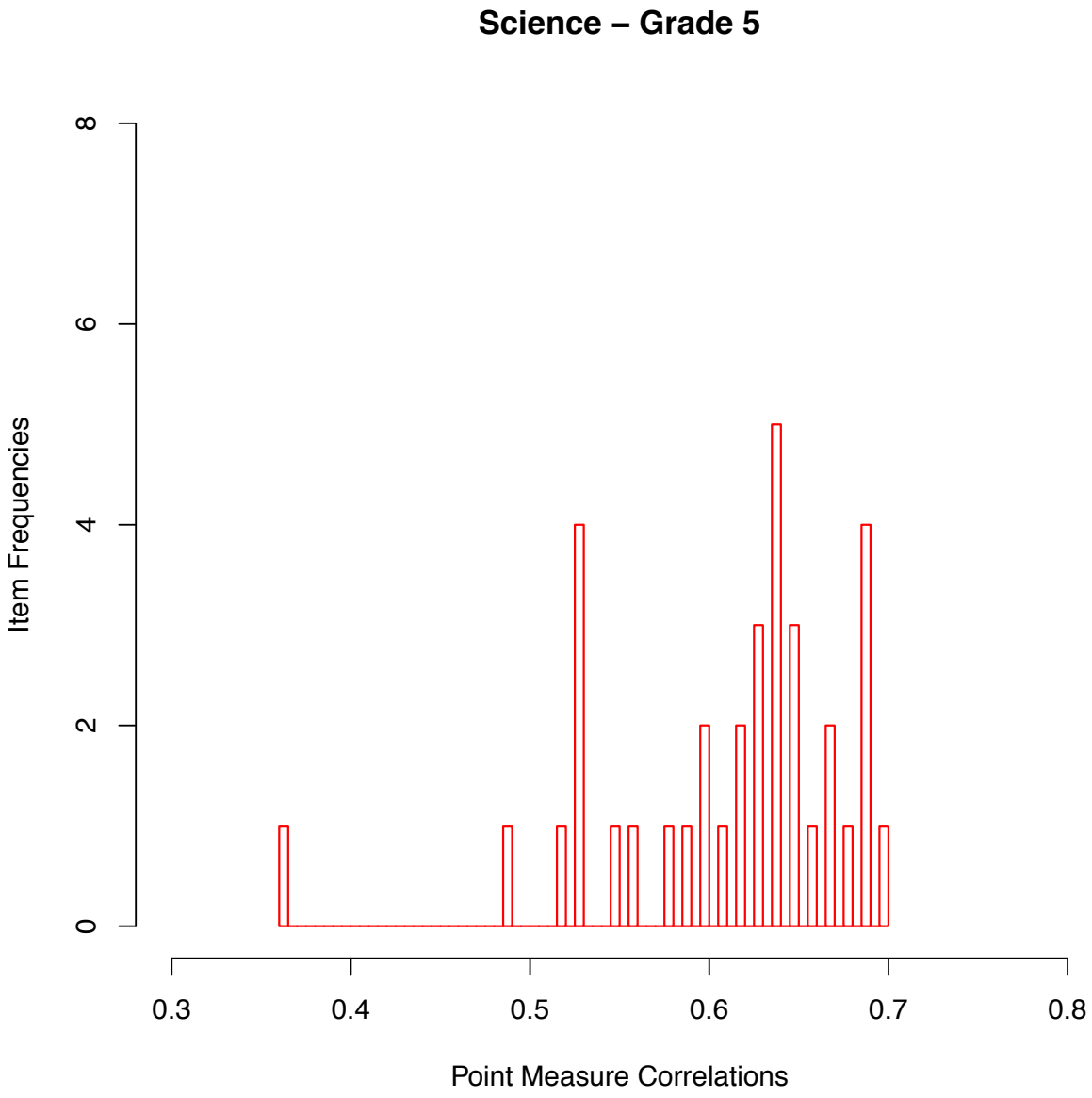

English Language Arts – Grade 11
Item Frequencies
0
2
4
6
8
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Point Measure Correlations

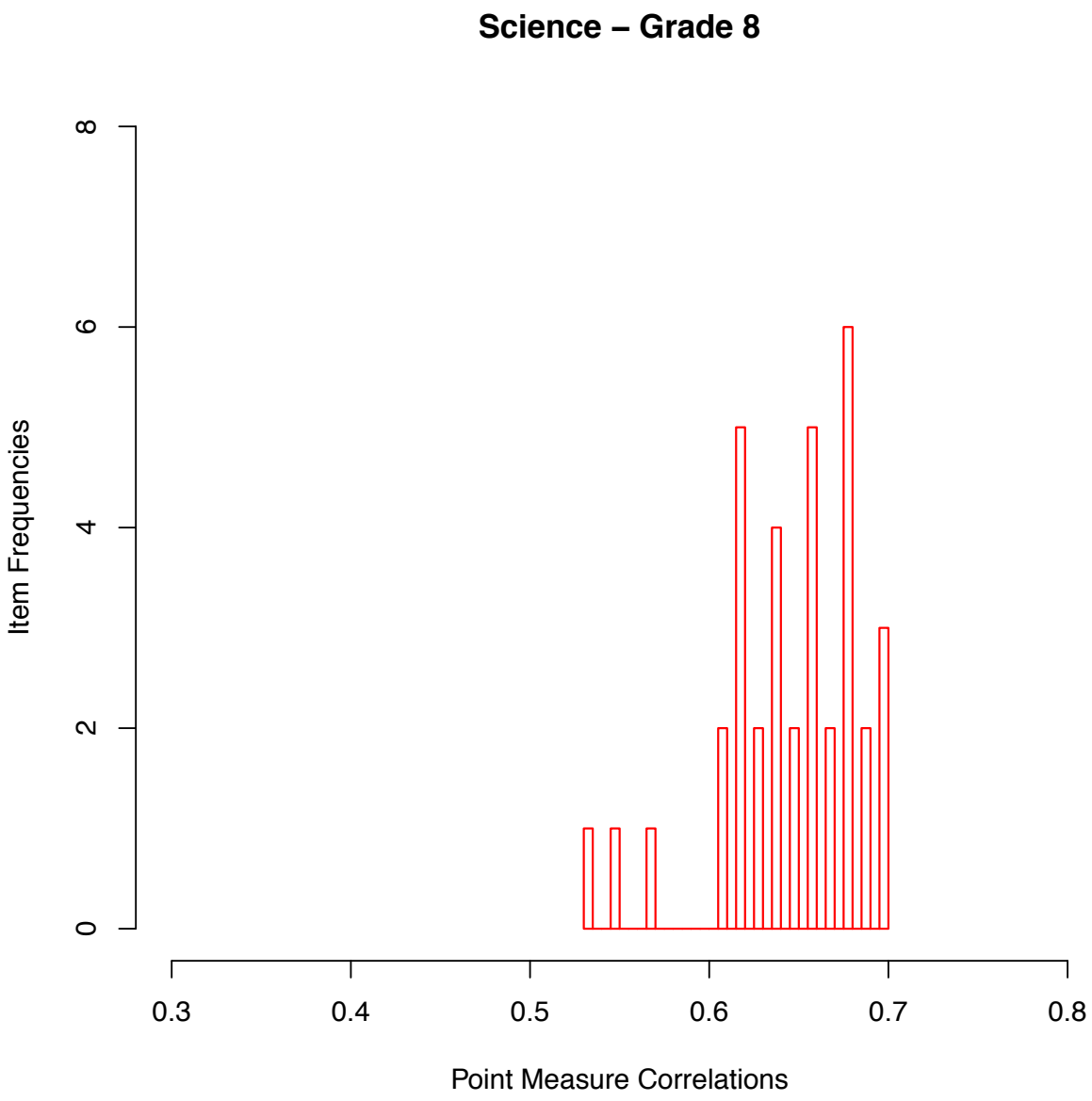
Point-measure correlations in mathematics range from 0.25 to 0.70.

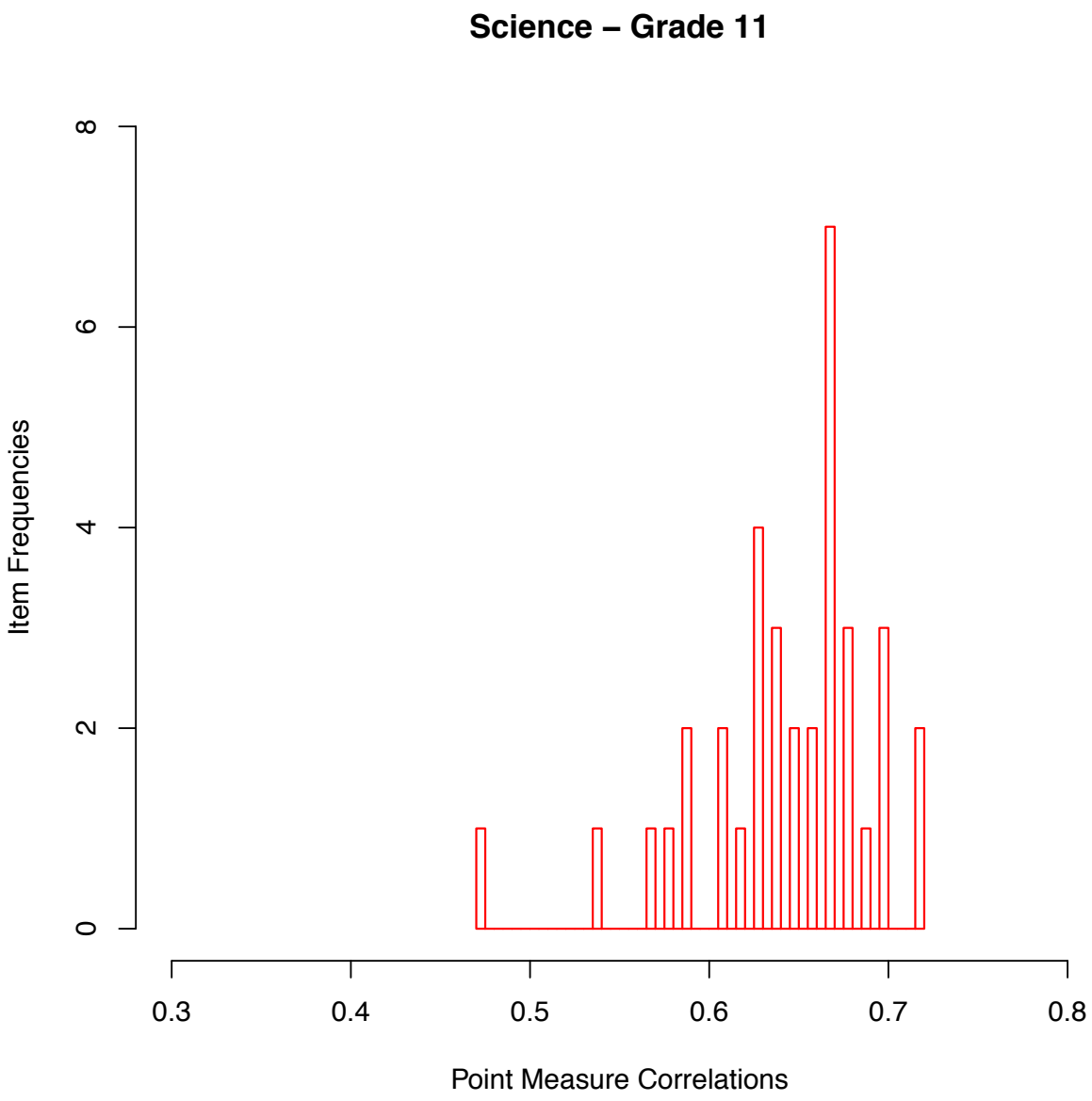



Mathematics – Grade 11
Item Frequencies
0
2
4
6
8
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Point Measure Correlations

Point-measure correlations in science range from 0.35 to 0.72.

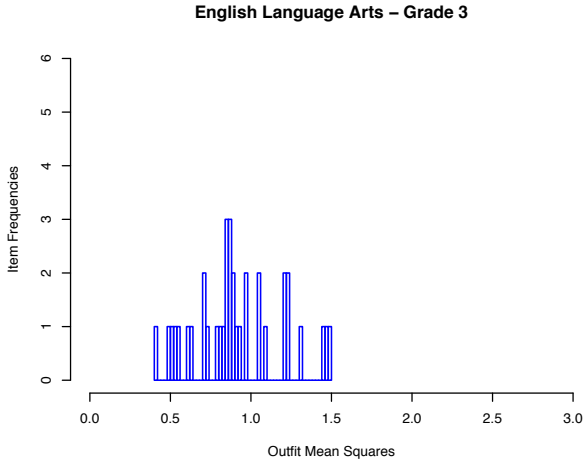


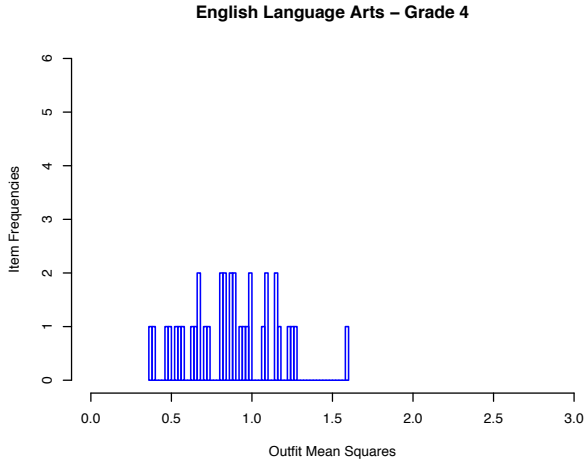

Science – Grade 8
Item Frequencies
0
2
4
6
8
0.3
0.4
0.5
0.6
0.7
0.8
Point Measure Correlations

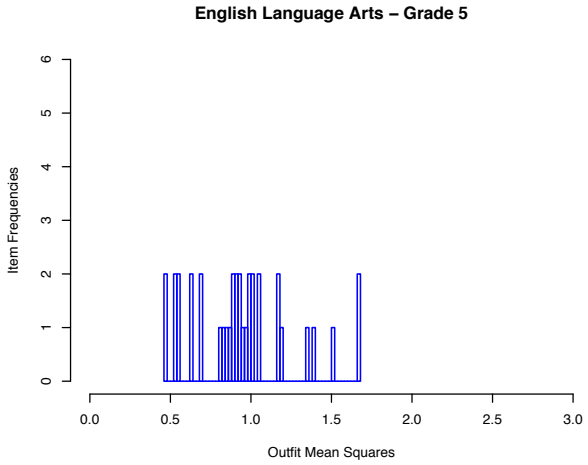

Science – Grade 11
Item Frequencies
0
2
4
6
8
0.3
0.4
0.5
0.6
0.7
0.8
Point Measure Correlations

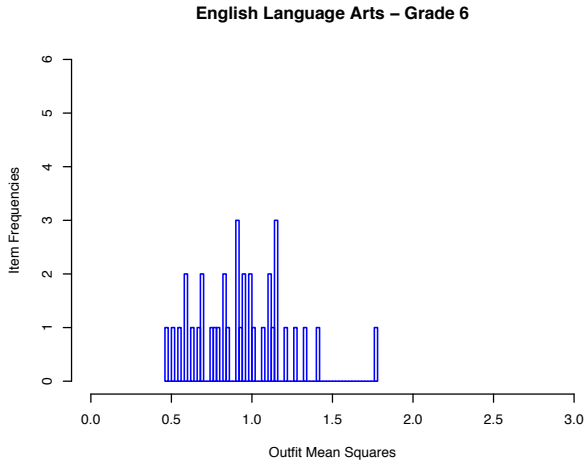
Outfit Mean Square Distributions

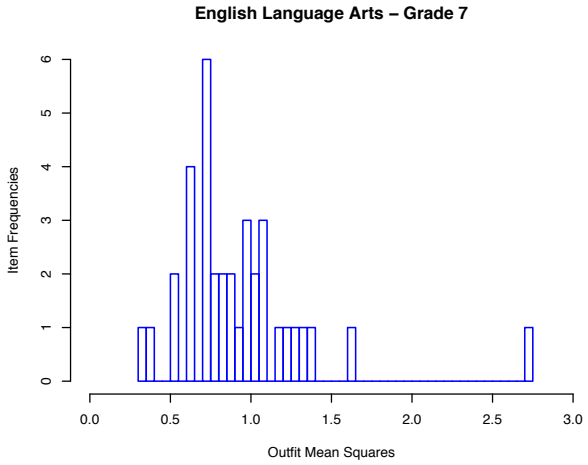
Outfit mean square values below 1.0 demonstrate that values are too predictable and perhaps redundant, while values above 1.0 indicate unpredictability. Items above 2.0 are deemed insufficient for measurement purposes and flagged for replacement. While most OMS values in ELA were between 0.5 and 1.5, some values were above 2.0 and will be reviewed.

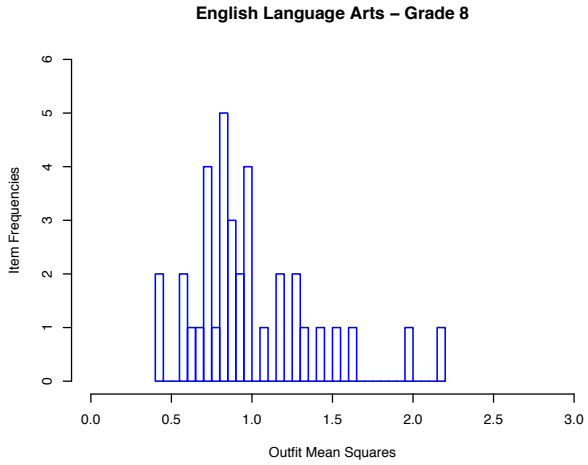


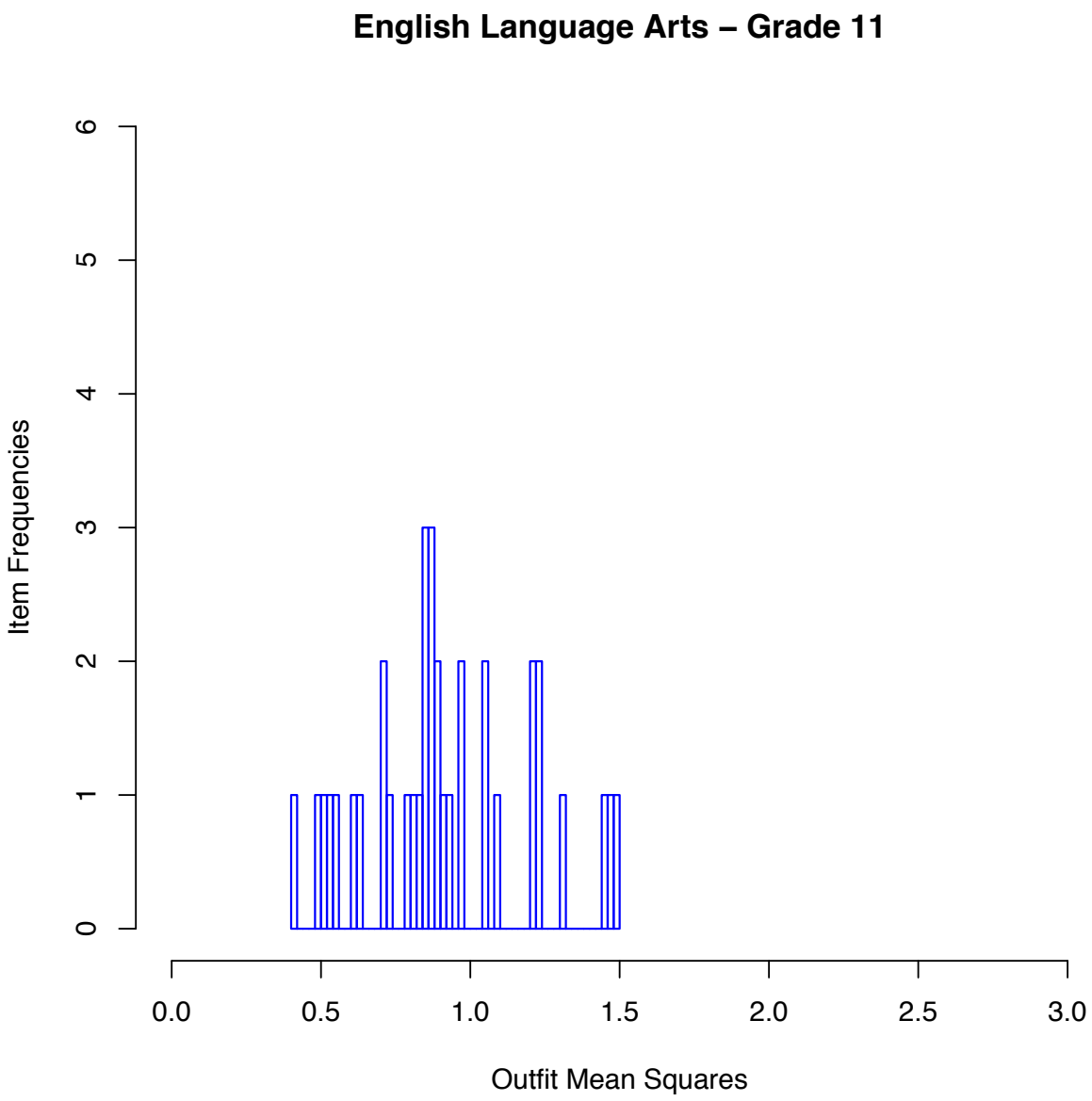




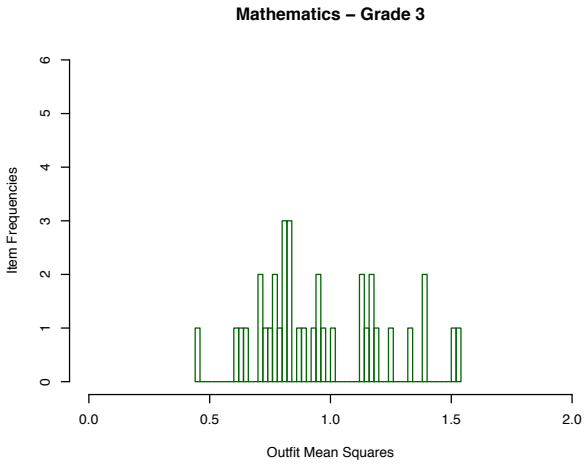


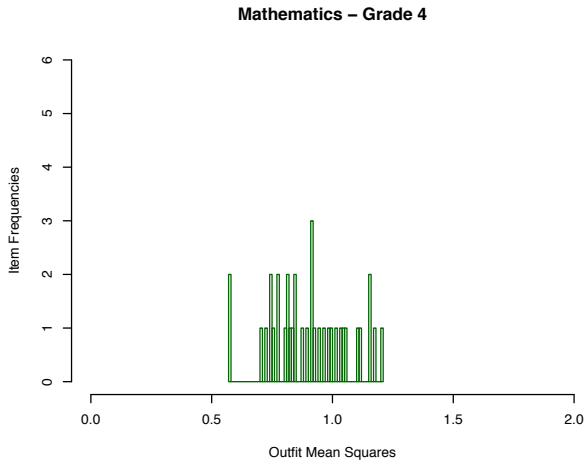


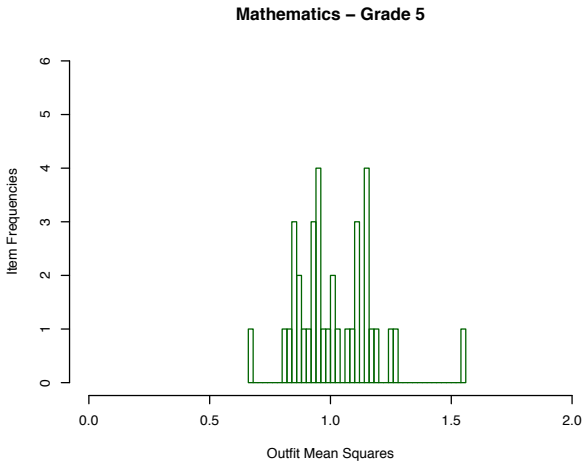


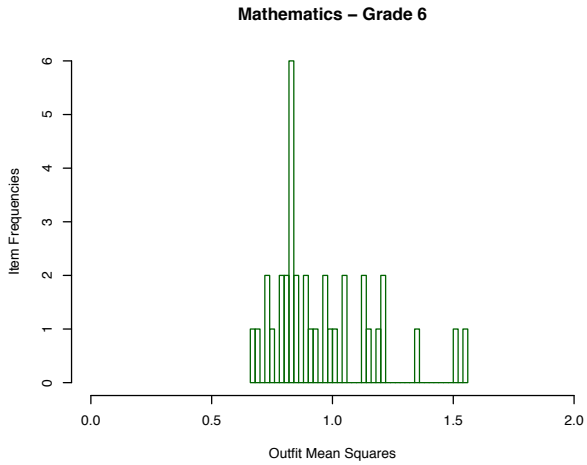

English Language Arts – Grade 11
Item Frequencies
0
1
2
3
4
5
6
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Outfit Mean Squares

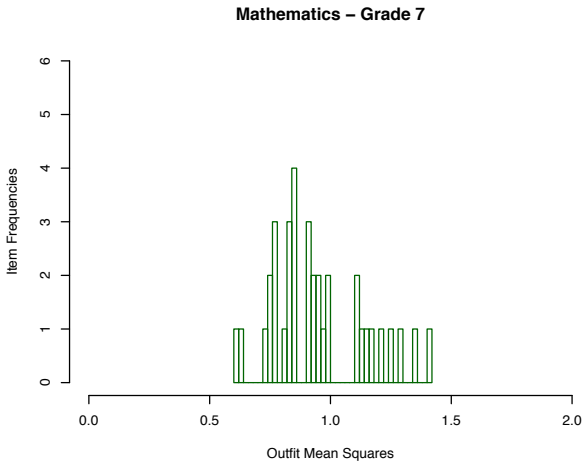
With the exception of a single item in Grade 11, which will be replaced, mathematics OMS values ranged from .40 to 1.6, demonstrating that the items are performing within expected ranges.

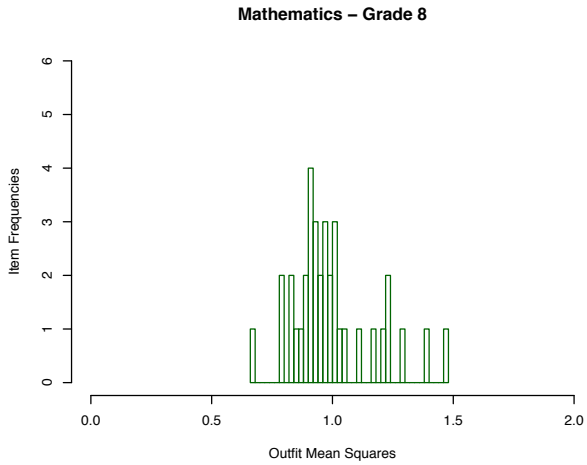


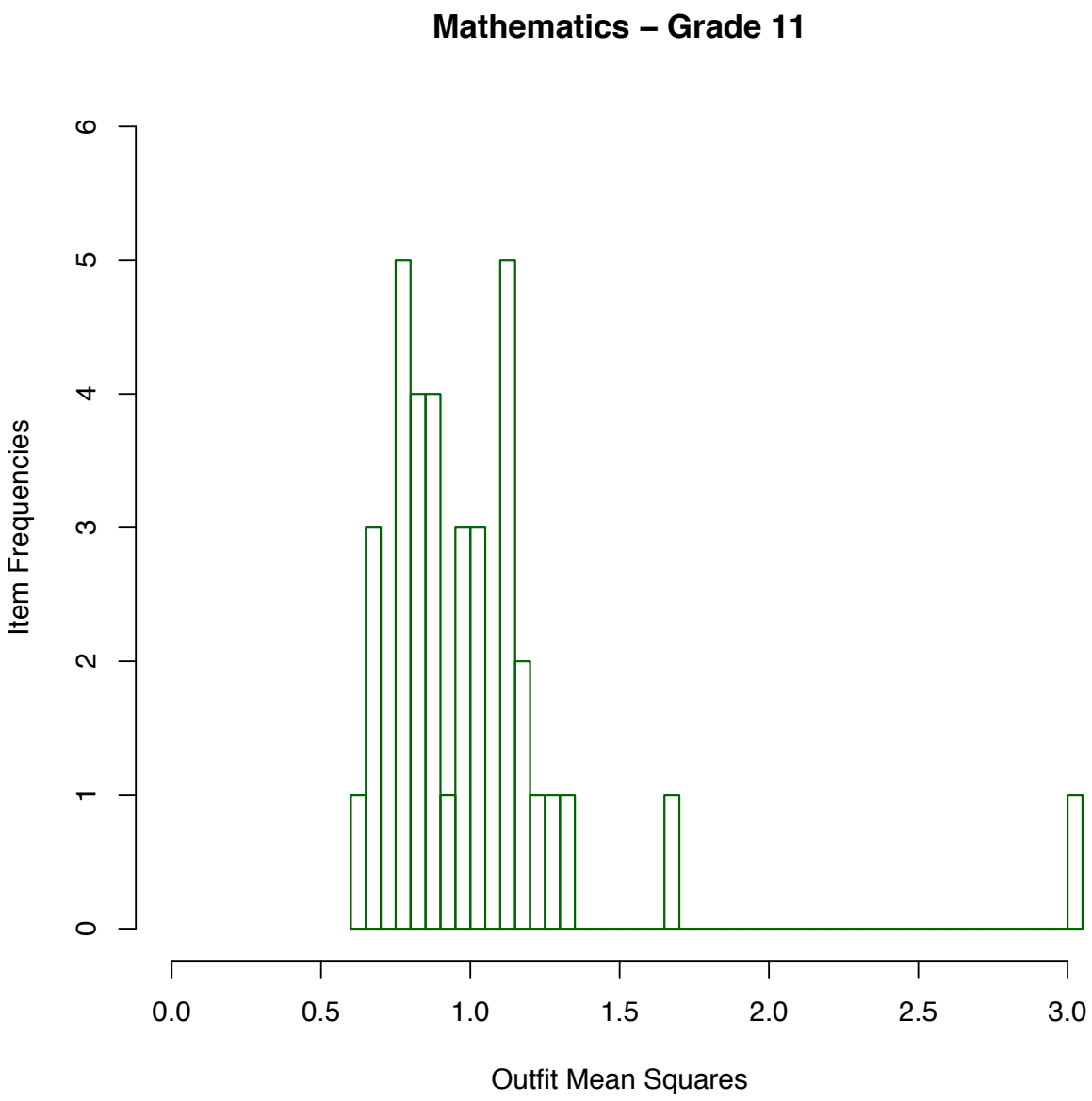




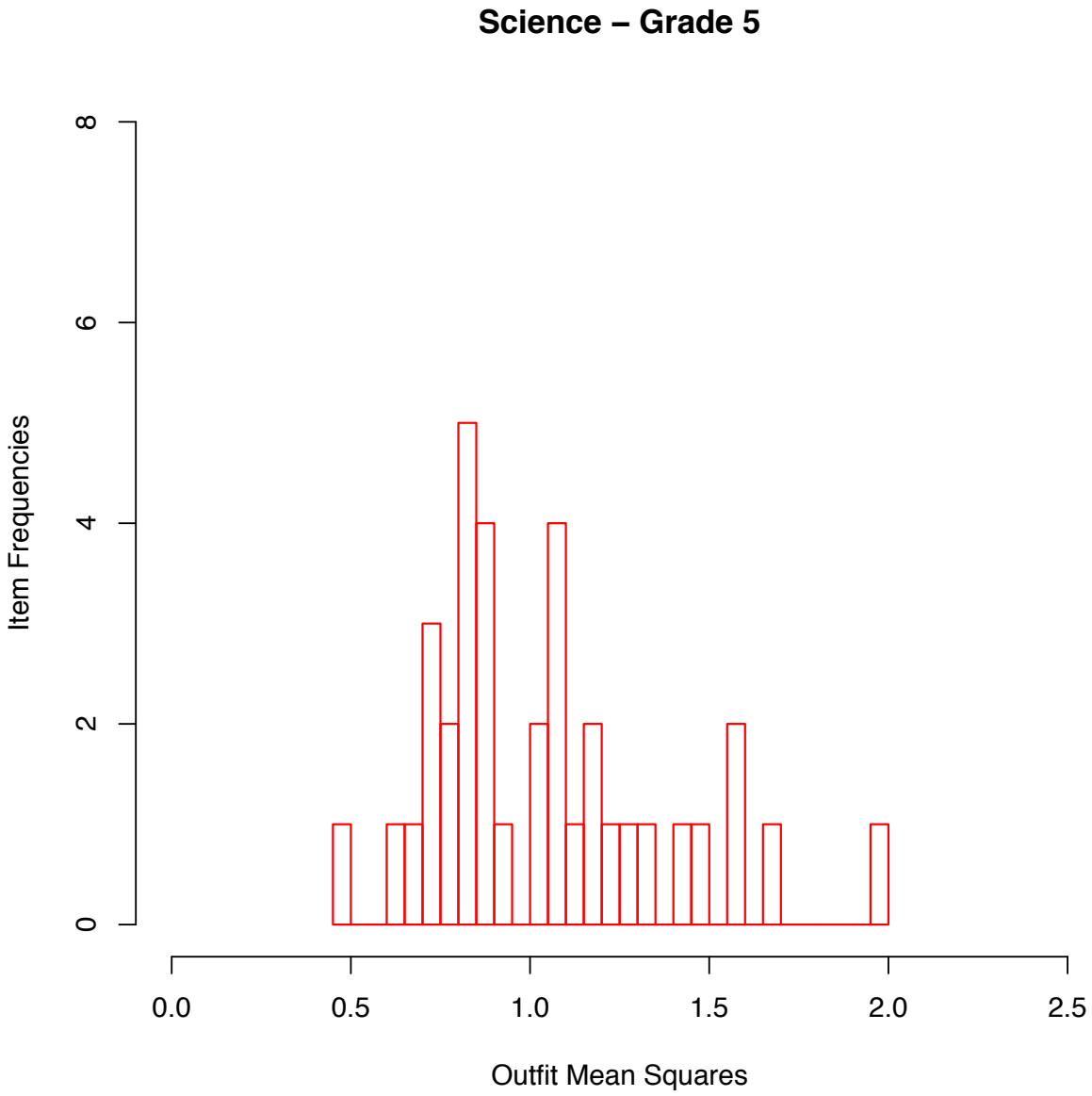


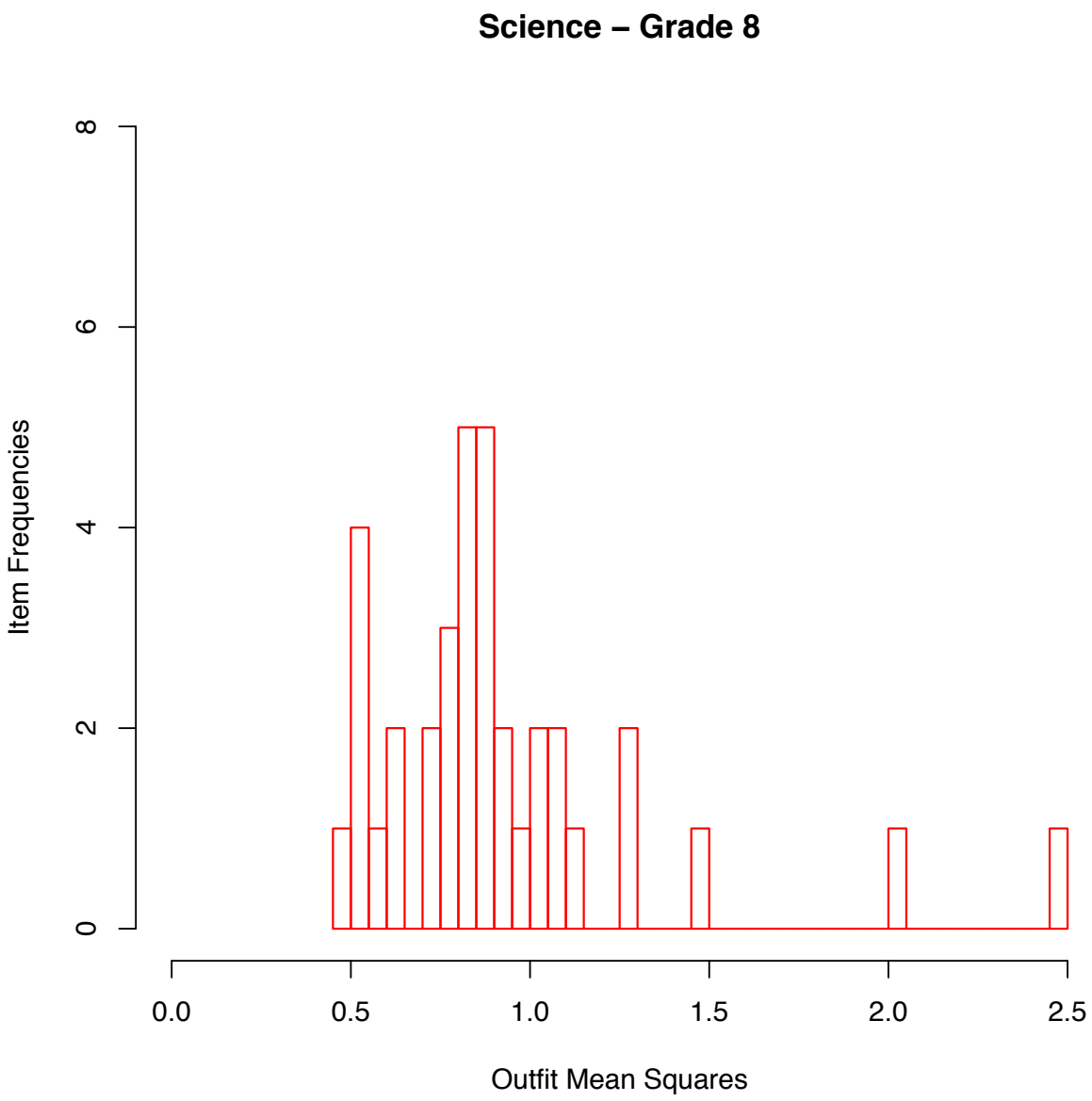


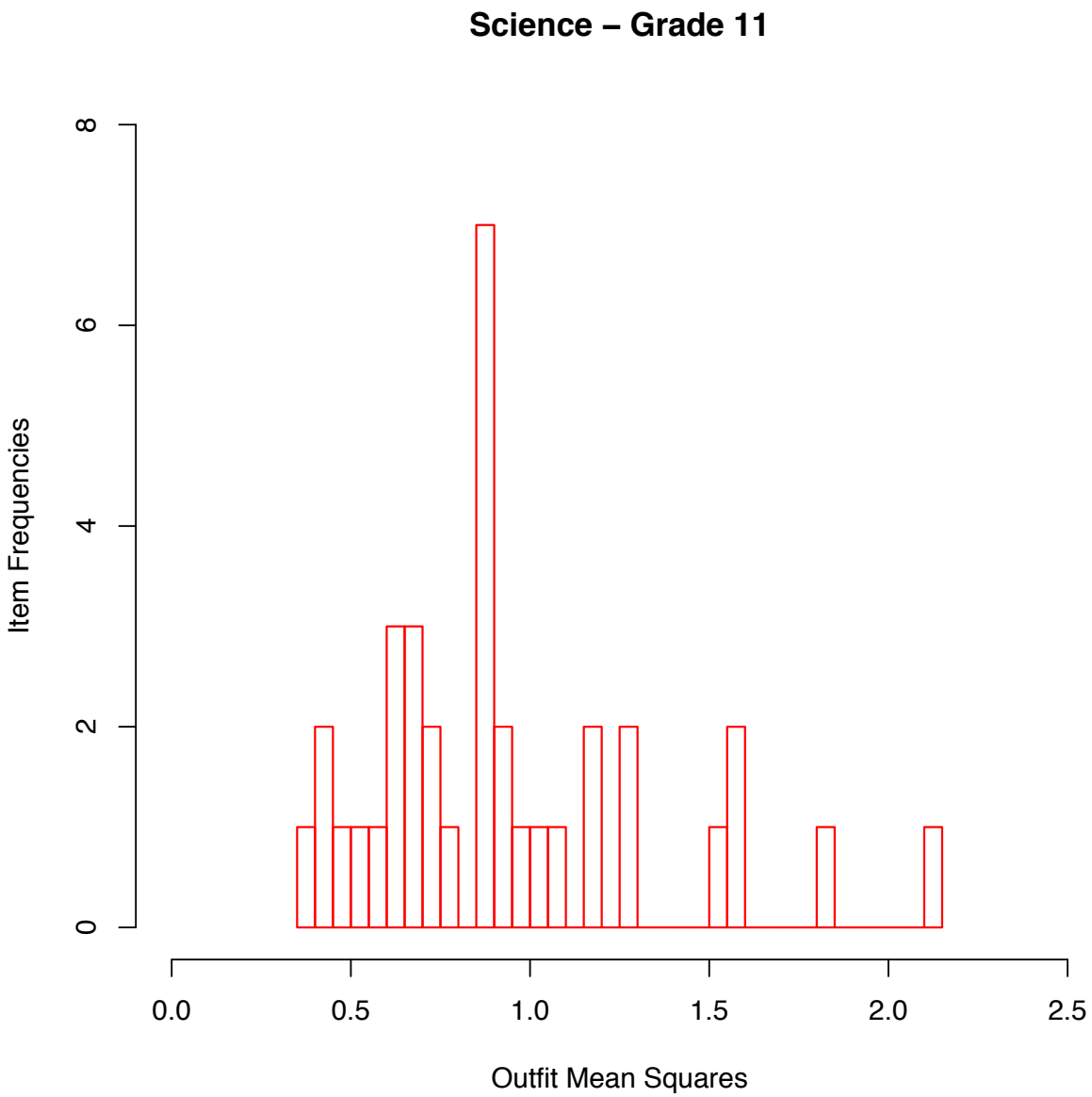



Mathematics – Grade 11
Item Frequencies
0
1
2
3
4
5
6
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Outfit Mean Squares

Science OMS values ranged from .40 to 2.5, demonstrating that the items are performing within expected ranges with a few exceptions that will be reviewed.

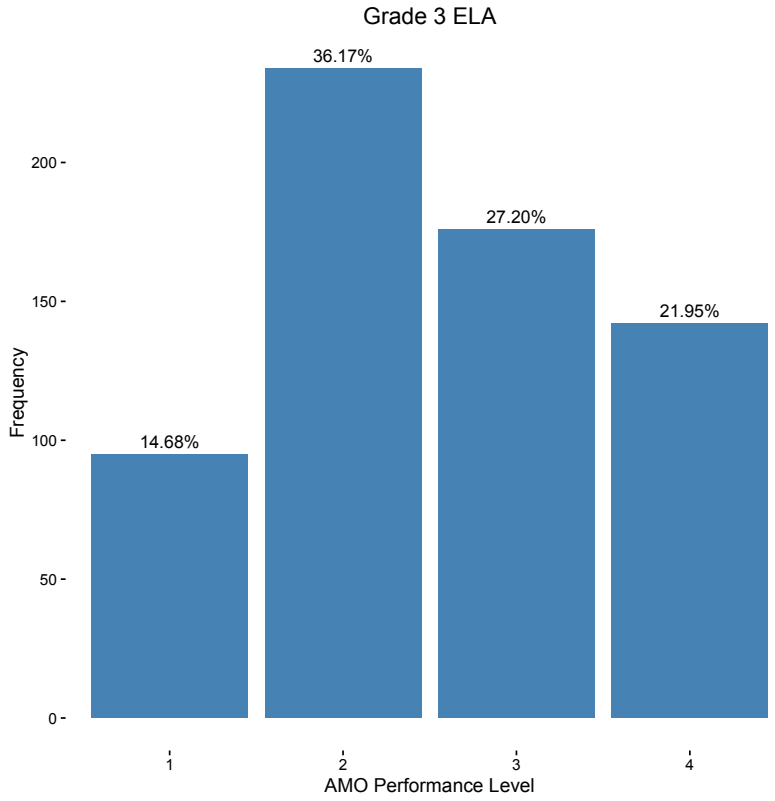


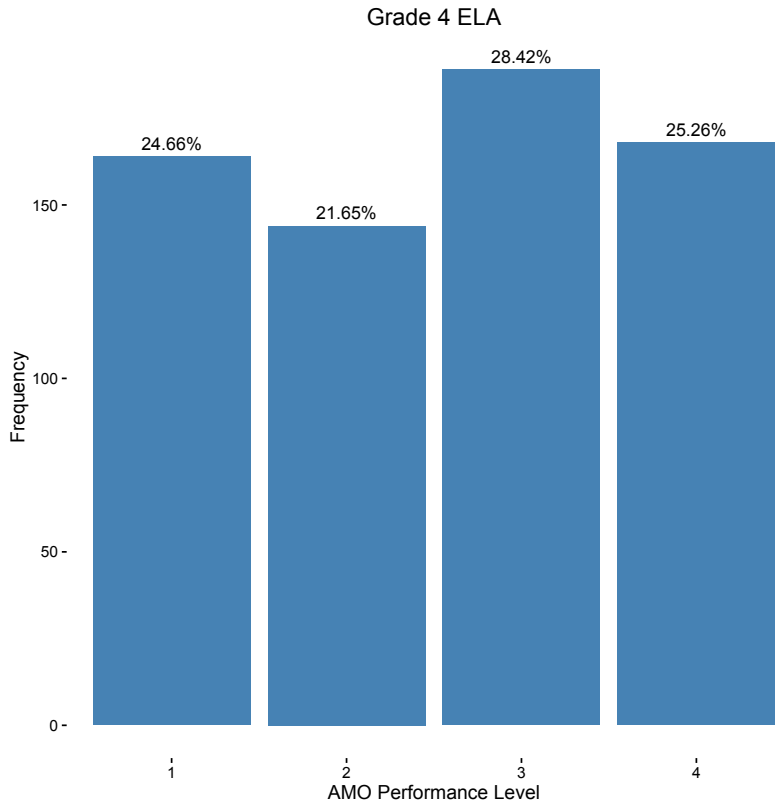

Science – Grade 8
Item Frequencies
0
2
4
6
8
0.0
0.5
1.0
1.5
2.0
2.5
Outfit Mean Squares

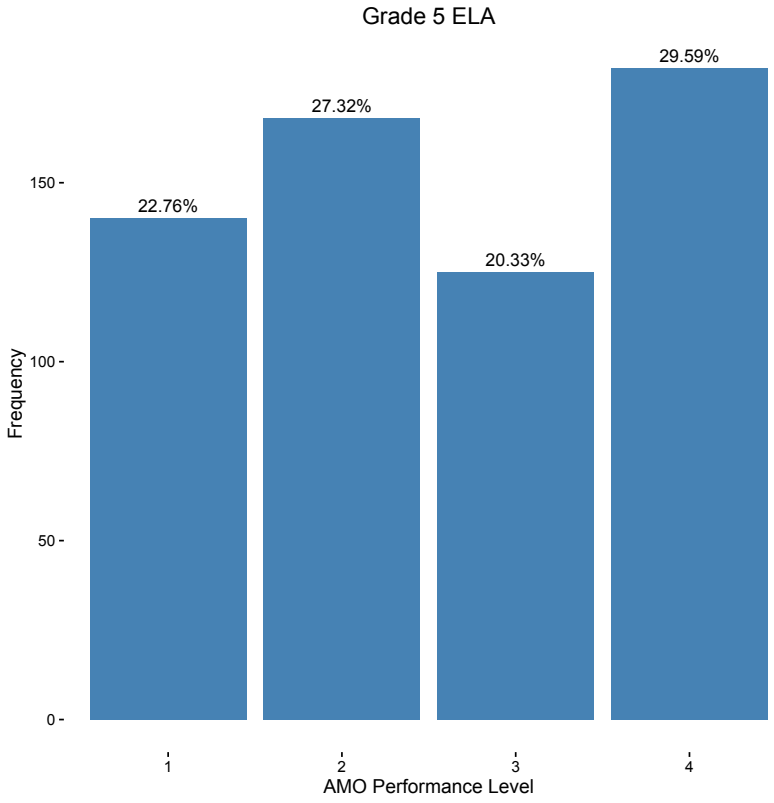


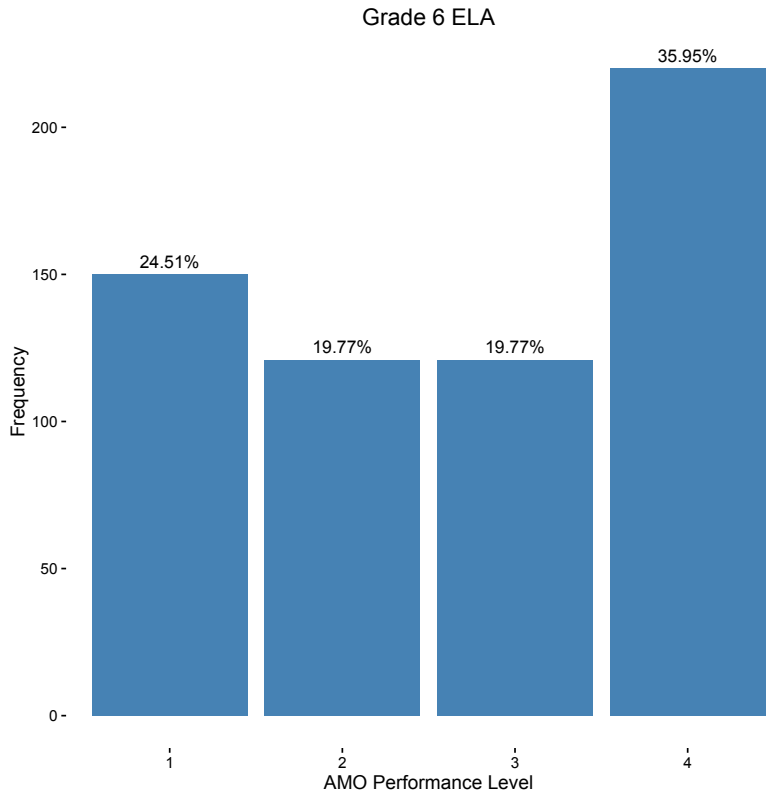
Annual Measureable Objectives Frequencies & Percentages

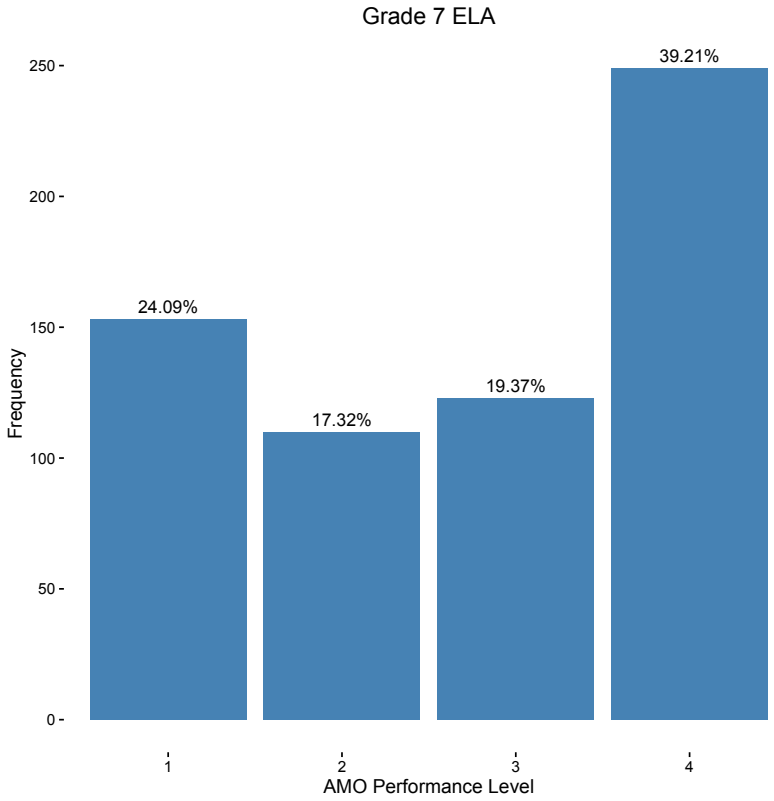
Annual Measurable Objective (AMO) calculations were conducted based upon student performance on the OREXt tied to the vertical scale using Rasch modeling. Though there are expected decreases in Level 4 due to a shifting test population, overall results are largely consistent with 2014-15, with approximately 50% of students with significant cognitive disabilities achieving proficiency across grades and content areas. ELA results are presented in blue, mathematics in dark green, and science in red.

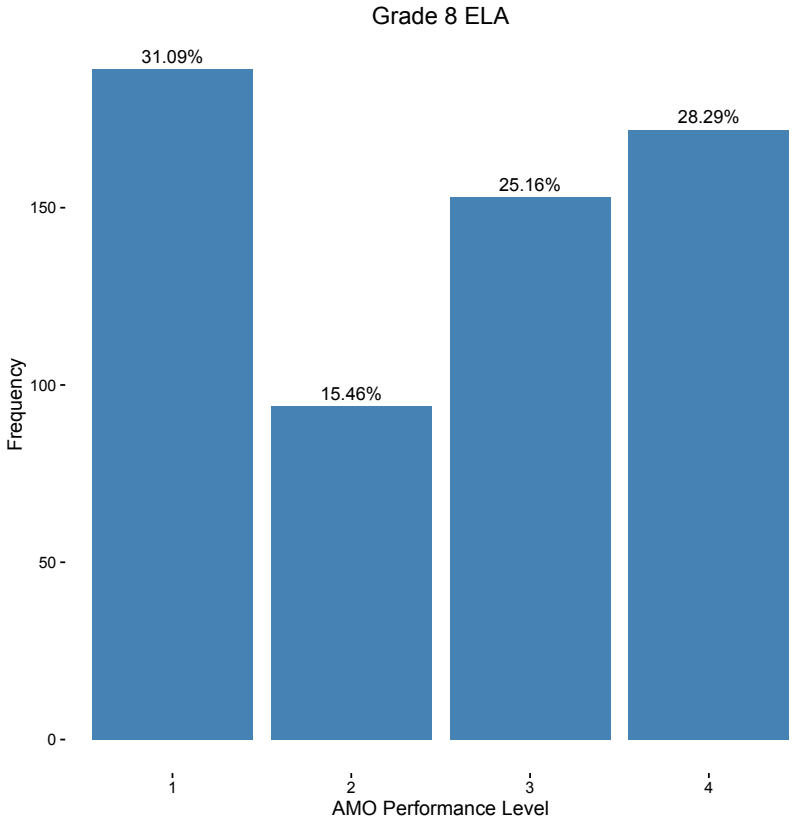

Grade 3 ELA
36.17%
27.20%
21.95%
14.68%
200
150
100
50
0
Frequency
1
2
3
4
AMO Performance Level

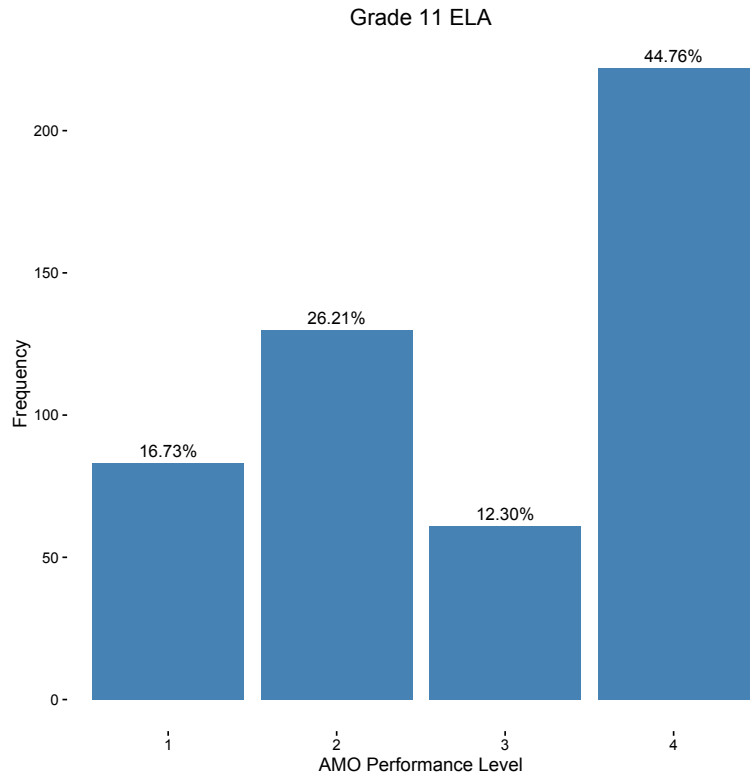

Grade 4 ELA
28.42%
25.26%
24.66%
21.65%
150
100
50
0
Frequency
1
2
3
4
AMO Performance Level

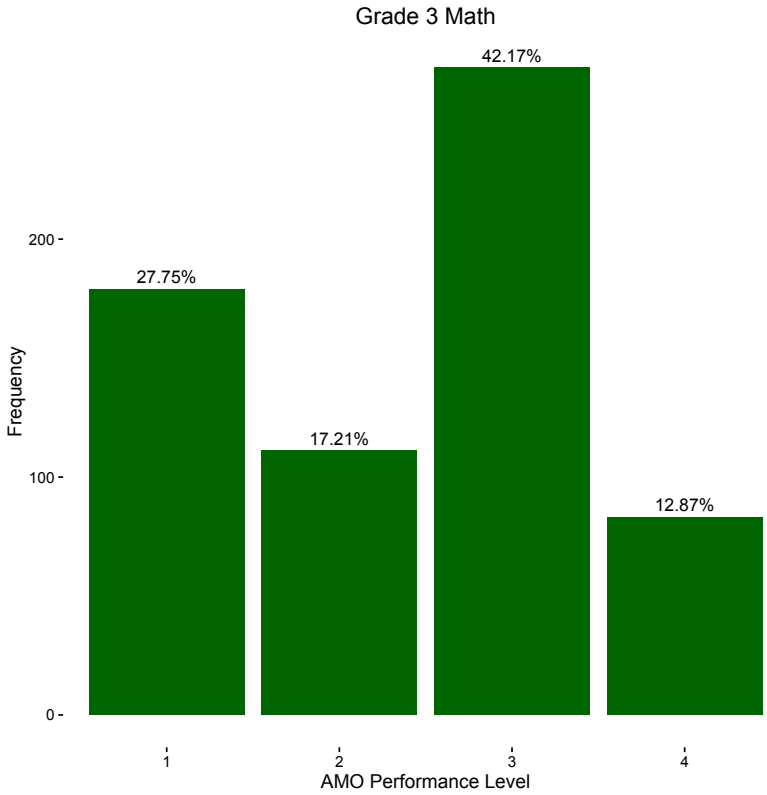

Grade 5 ELA
29.59%
27.32%
22.76%
20.33%
150
100
50
0
Frequency
1
2
3
4
AMO Performance Level

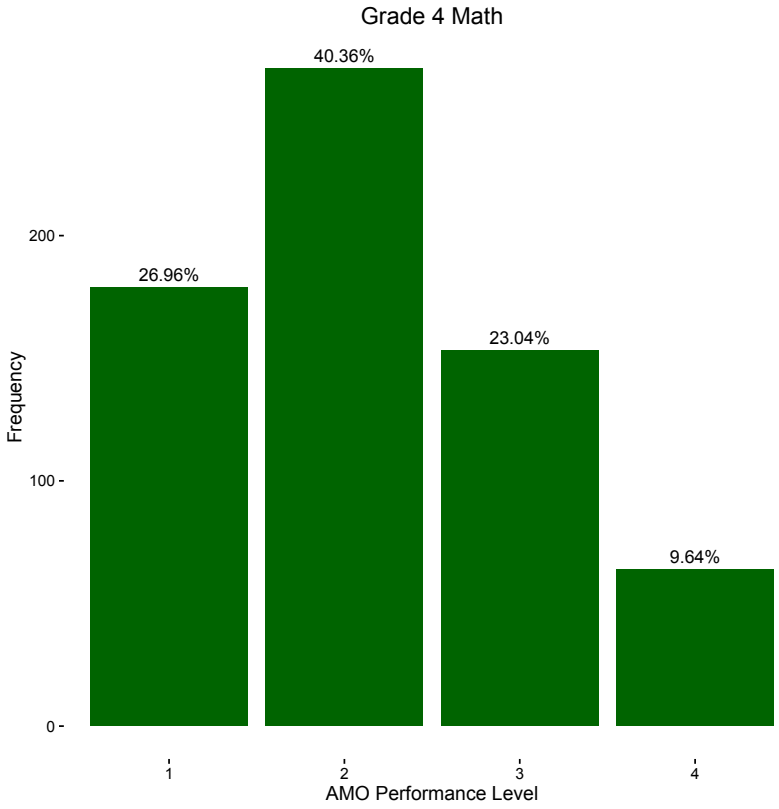

Grade 6 ELA
35.95%
24.51%
19.77%
19.77%
200
150
100
50
0
Frequency
1
2
3
4
AMO Performance Level

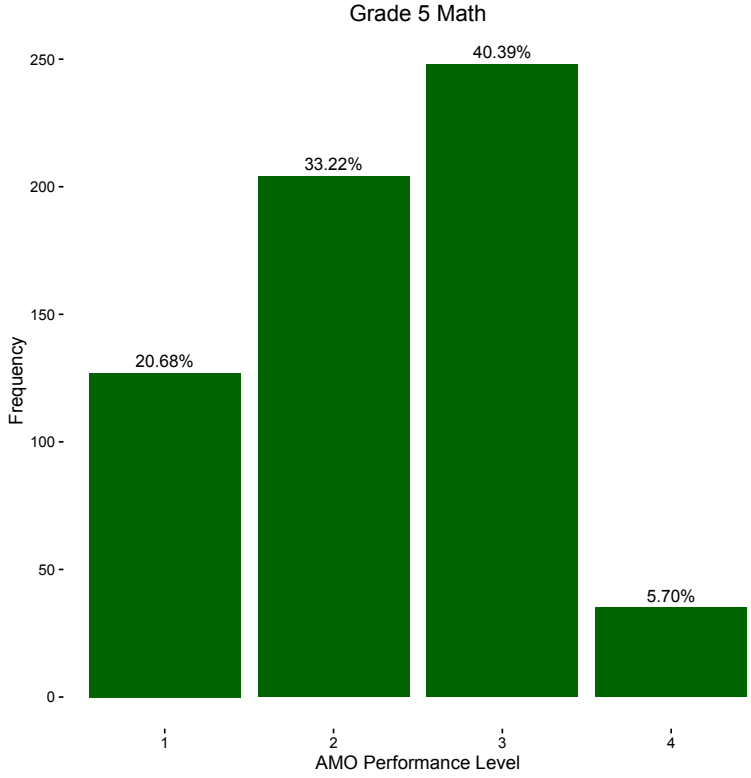

Grade 7 ELA
Frequency
250
200
150
100
50
0
24.09%
17.32%
19.37%
39.21%
1
2
3
4
AMO Performance Level

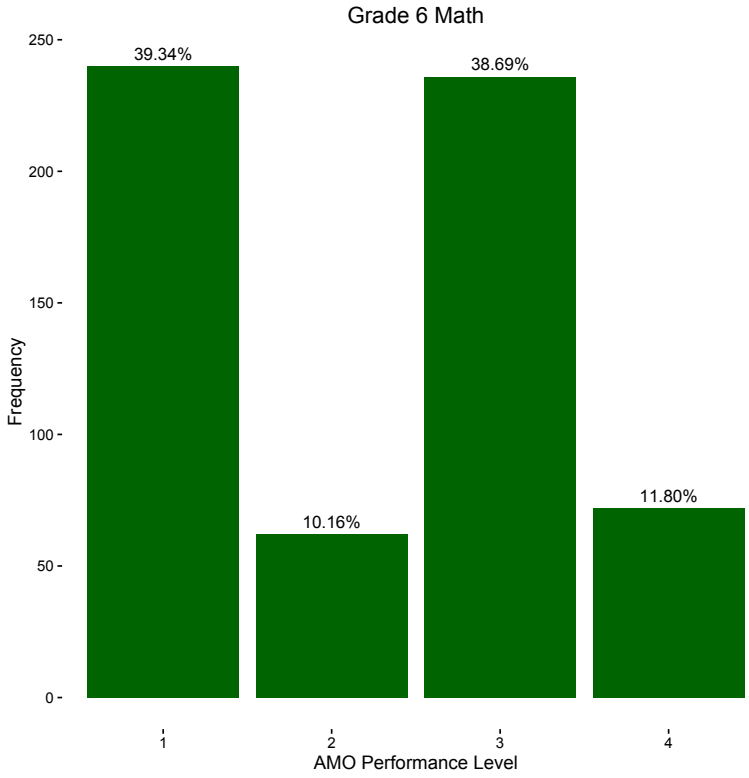

Grade 8 ELA
Frequency
150
100
50
0
31.09%
15.46%
25.16%
28.29%
1
2
3
4
AMO Performance Level

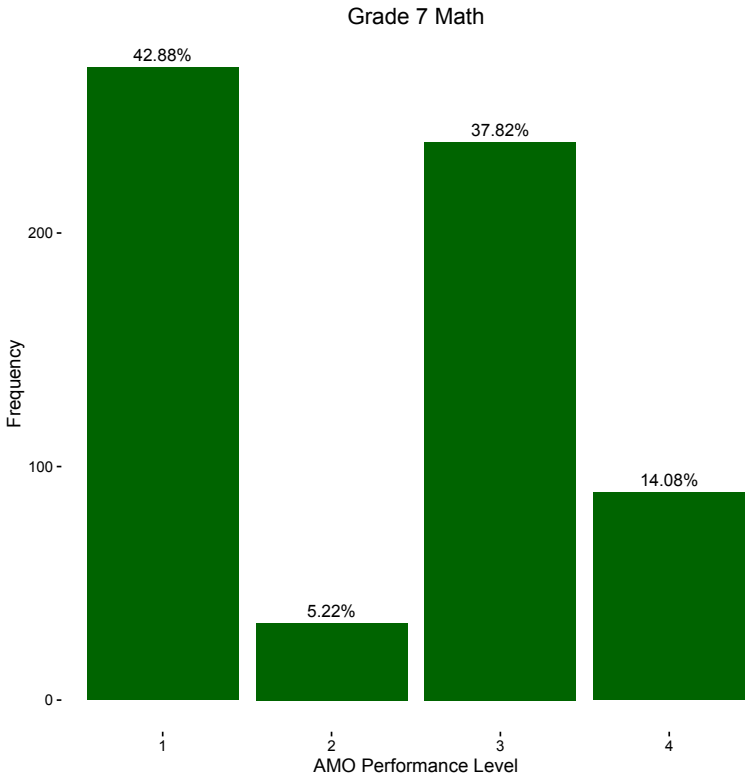

Grade 11 ELA
44.76%
26.21%
16.73%
12.30%
200
150
100
50
0
Frequency
1
2
3
4
AMO Performance Level

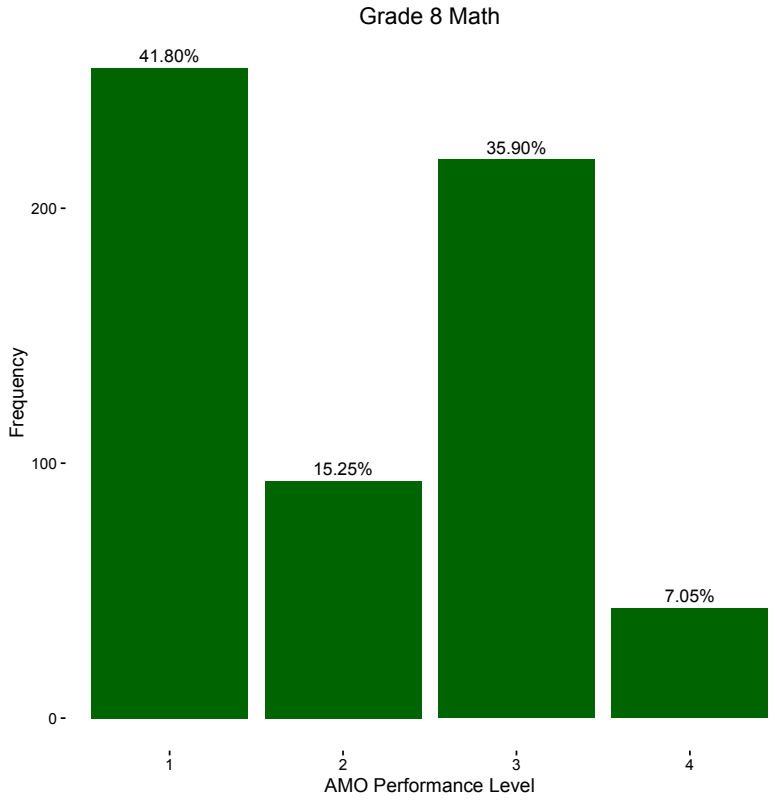

Grade 3 Math
42.17%
27.75%
17.21%
12.87%
200
100
0
Frequency
1
2
3
4
AMO Performance Level

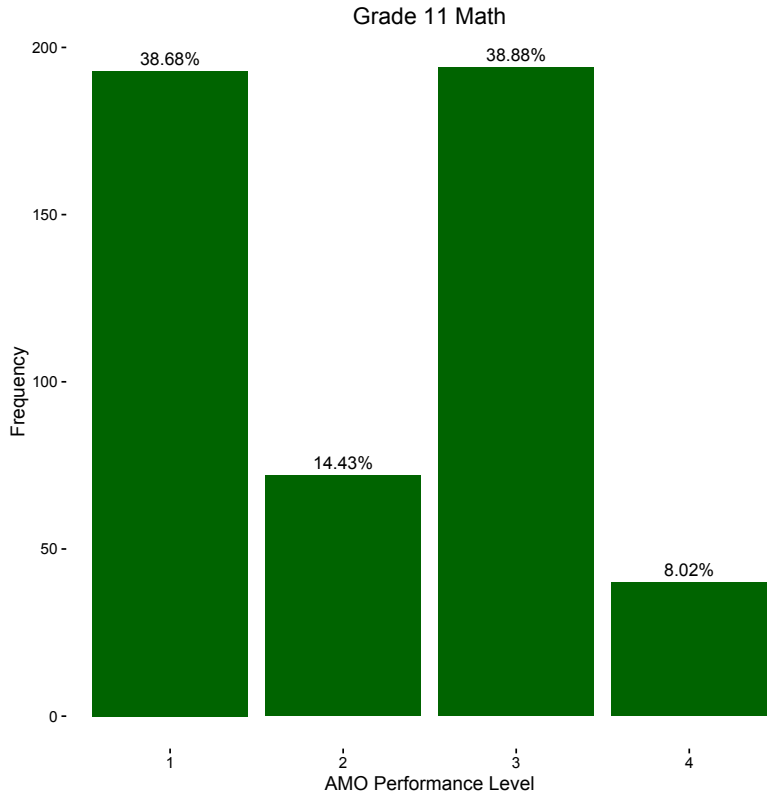

Grade 4 Math
40.36%
26.96%
23.04%
9.64%
200
100
0
Frequency
1
2
3
4
AMO Performance Level

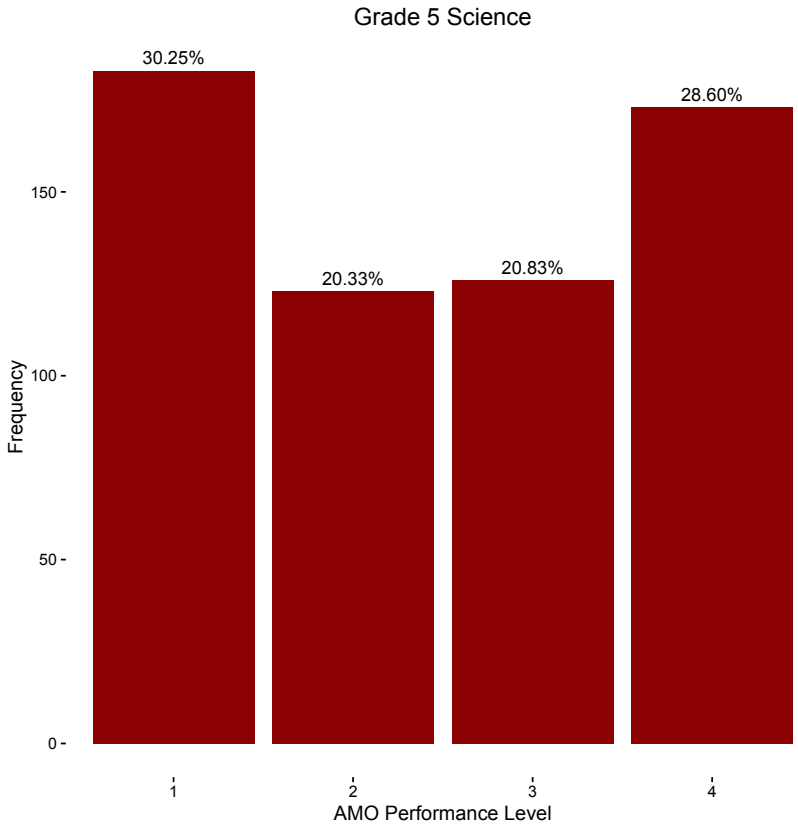

Grade 5 Math
40.39%
33.22%
20.68%
5.70%
250
200
150
100
50
0
Frequency
1
2
3
4
AMO Performance Level

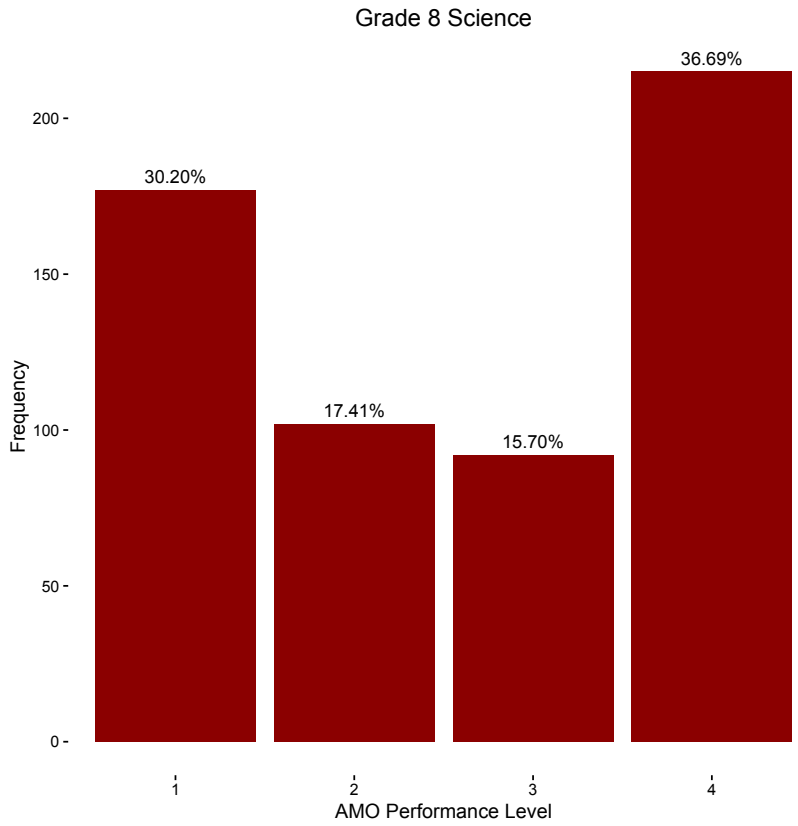

Grade 6 Math
250
200
150
100
50
0
Frequency
39.34%
10.16%
38.69%
11.80%
1
2
3
4
AMO Performance Level

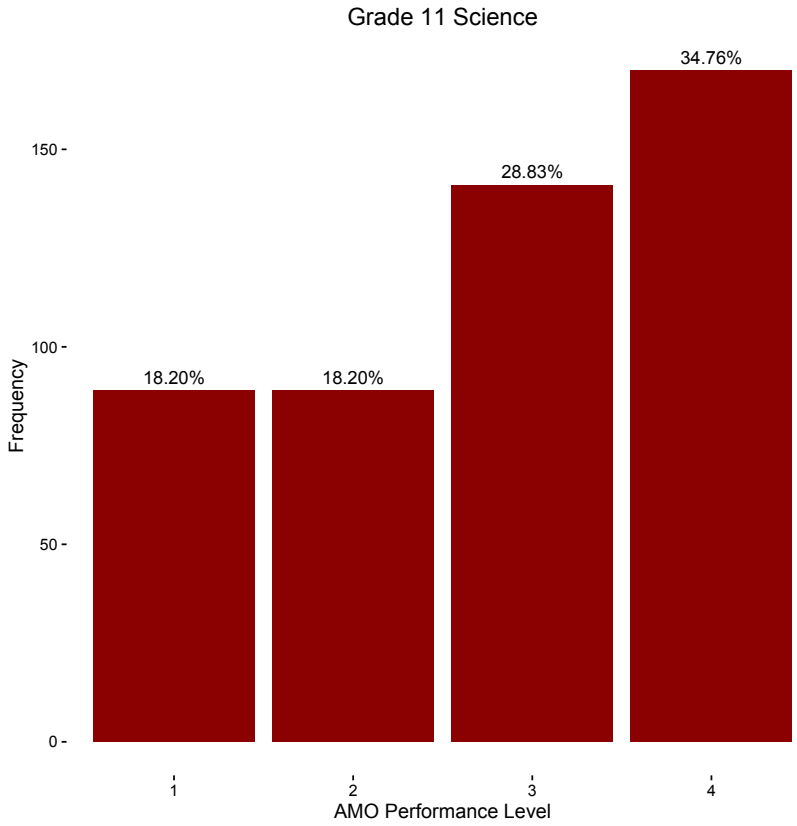

Grade 7 Math
200
100
0
Frequency
42.88%
5.22%
37.82%
14.08%
1
2
3
4
AMO Performance Level


Grade 8 Math
41.80%
35.90%
15.25%
7.05%
200
100
0
Frequency
1
2
3
4
AMO Performance Level


Grade 11 Math
38.68%
38.88%
14.43%
8.02%
200
150
100
50
0
Frequency
1
2
3
4
AMO Performance Level


Grade 5 Science
30.25%
20.33%
20.83%
28.60%
150
100
50
0
Frequency
1
2
3
4
AMO Performance Level


Grade 8 Science
36.69%
30.20%
17.41%
15.70%
200
150
100
50
0
Frequency
1
2
3
4
AMO Performance Level



3.4 Validity Based on Relations to Other Variables

Perhaps the best model for understanding criterion-related evidence comes from Campbell and Fiske (1959)[7] in their description of the multi-trait, multi-method analysis [we translate the term 'trait' to mean 'skill']. In this process (several) different traits are measured using (several) different methods to provide a correlation matrix that should reflect specific patterns supportive of the claim being made (that is, provide positive validation evidence). Sometimes, these various measures are of the same or similar skills, abilities, or traits, and other times they are of different skills, abilities, or traits. We present data that quite consistently reflect higher relations among items **within** an academic subject than **between** academic subjects. We also present data in which performance on items is totaled within categories of disability, expecting relations that would reflect appropriate differences (see Tindal, McDonald, Tedesco, Glasgow, Almond, Crawford, & Hollenbeck, 2003).[8]

---

[7] Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multi-trait, multi-method matrix. In W. A. Mehrens & R. L. Ebel (Eds.), *Principles of educational and psychological measurement: A book of selected readings* (pp. 273-302). Chicago, IL: Rand McNally & Company.

[8] Tindal, G., McDonald, Tedesco, M., Glasgow, A., Almond, P., Crawford, L., & Hollenbeck, K. (2003). Alternate assessments in reading and math: Development and validation for students with significant disabilities. *Exceptional Children*, 69(4), 481-494.

Analyses Within and Across Subject Areas

We conducted correlational analyses to explore the validity of the OREXt. We first describe the purpose of the analysis, as well as our anticipated results. We then discuss our observed results before concluding with an overall evaluative judgment of the validity of the test.

In the correlational analysis, we explore the correlations among students' total scores across subject areas. The purpose of the analysis was to investigate how strongly students' scores in one area were related to students' scores in other subject areas. If the correlations were exceedingly high (e.g., above .90), it would indicate that the score a student receives in an individual subject has less to do with the intended construct (i.e., reading) than with factors idiosyncratic to the student. For example, if all subject areas correlated at .95, then it would provide strong evidence that the tests would be measuring a global student-specific construct (i.e., intelligence), and not the individual subject constructs. We would expect, however, that the tests would correlate quite strongly given that the same students were assessed multiple times. Therefore, we would expect moderately strong correlations (e.g., 0.7) simply because of the within-subject design. Idiosyncratic variance associated with the individual student is thus captured.

Correlational Analyses Results

Full results of the correlation analysis by content area and grade level are reported below. The results are significant, yet the overall correlations across content areas suggest that we are indeed measuring different, though strongly related constructs, with between-test scaled score correlations ranging from 0.82 to 0.89.

Grade 3 Content Area Correlations

***n* - size**	**Content Area**	**ELA RIT Correlation**	**Math RIT Correlation**
644	ELA	–	.88*
	Math		–

* $p < .001$

Grade 4 Content Area Correlations

***n* - size**	**Content Area**	**ELA RIT Correlation**	**Math RIT Correlation**
662	ELA	–	.86*
	Math		–

* $p < .001$

Grade 5 Content Area Correlations

***n* - size**	**Content Area**	**ELA RIT Correlation**	**Math RIT Correlation**	**Science RIT Correlation**
601	ELA	–	.84*	.86*
	Math		–	.82*
	Science			–

* $p < .001$

Grade 6 Content Area Correlations

***n* - size**	**Content Area**	**ELA RIT Correlation**	**Math RIT Correlation**
608	ELA	–	.84*
	Math		–

* $p < .001$

Grade 7 Content Area Correlations

***n* - size**	**Content Area**	**ELA RIT Correlation**	**Math RIT Correlation**
629	ELA	–	.85*
	Math		–

* $p < .001$

Grade 8 Content Area Correlations

***n* - size**	**Content Area**	**ELA RIT Correlation**	**Math RIT Correlation**	**Science RIT Correlation**
580	ELA	–	.86*	.88*
	Math		–	.85*
	Science			–

* $p < .001$

Grade 11 Content Area Correlations

***n* - size**	**Content Area**	**ELA RIT Correlation**	**Math RIT Correlation**	**Science RIT Correlation**
485	ELA	–	.85*	.89*
	Math		–	.87*
	Science			–

* $p < .001$

Results of the correlation analysis within English language arts (ELA:Reading:Writing) are reported below and suggest high correlations between ELA and Reading, as expected, from .96 to .98. Writing is correlated with ELA from .92 to .95 and with reading from .83 to .90.

English Language Arts Subscore Correlations

***n*-size**	**Grade Scaled Score**	**ELA RIT Correlation**	**Reading RIT Correlation**	**Writing RIT Correlation**
647	3 ELA	–	.96*	.93*
	3 Reading		–	.85*
	3 Writing			–
665	4 ELA	–	.96*	.92*
	4 Reading		–	.84*
	4 Writing			–
615	5 ELA	–	.96*	.93*
	5 Reading		–	.83*
	5 Writing			–
612	6 ELA	–	.97*	.94*
	6 Reading		–	.87*
	6 Writing			–
635	7 ELA	–	.98*	.94*
	7 Reading		–	.89*
	7 Writing			–
608	8 ELA	–	.98*	.93*
	8 Reading		–	.88*
	8 Writing			–
496	11 ELA	–	.98*	.95*
	11 Reading		–	.90*
	11 Writing			–

* $p < .001$

The ORExt assessments appear to be measuring separate constructs, as intended, indicated by the correlations. No unexpected test functioning statistics are present based on student characteristics that should not be related, such as gender and ethnicity. Student performance appears to be primarily related to item difficulty and not the result of construct irrelevant aspects that have been reviewed.

Critical Element 4 - Technical Quality: Other

4.1 Reliability

Test reliability can be viewed through several lenses, all of which document how consistently an assessment performs across occasions, contexts, and raters[9]. Typical strategies for addressing reliability include documention of internal consistency, split-half reliability, and test-retest reliability. If multiple forms are implemented, test form reliability documentation is also requisite. The implementation plan for the ORExt includes initial documentation of internal consistency (Cronbach's alpha). The 2015-16 technical report will include internal consistency estimates, split-half reliability analyses, as well as a small test-retest assessment of reliability comparisons by means of our pilot tablet administration study. There is only one test form for the ORExt, so test form comparisons are not possible.

4.1A Test Reliability

Marginal reliability results (true score variance/ true score variance + error variance) demonstrate that the tests are quite reliable at the total test level. Full reliability statistics for each of the operational tests administered this year are provided below. These results demonstrate that the total test reliabilities were quite high, ranging from .87 to .91. Each table below provides the content area, grade, and the marginal reliabilities. All test forms were composed of 36 operational and 12 embedded field-test items.

English Language Arts

The test reliabilities for ELA were in the high range, from .87 to .91.

Grade	Marginal Reliability
3	0.91
4	0.91
5	0.91
6	0.88
7	0.89
8	0.90
11	0.87

Mathematics

The test reliabilities for mathematics were in the high range, from .89 to .91.

Grade	Marginal Reliability
3	0.91
4	0.91
5	0.90
6	0.90

[9] American Educational Research Association (AERA), American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing.* Washington, DC: AERA

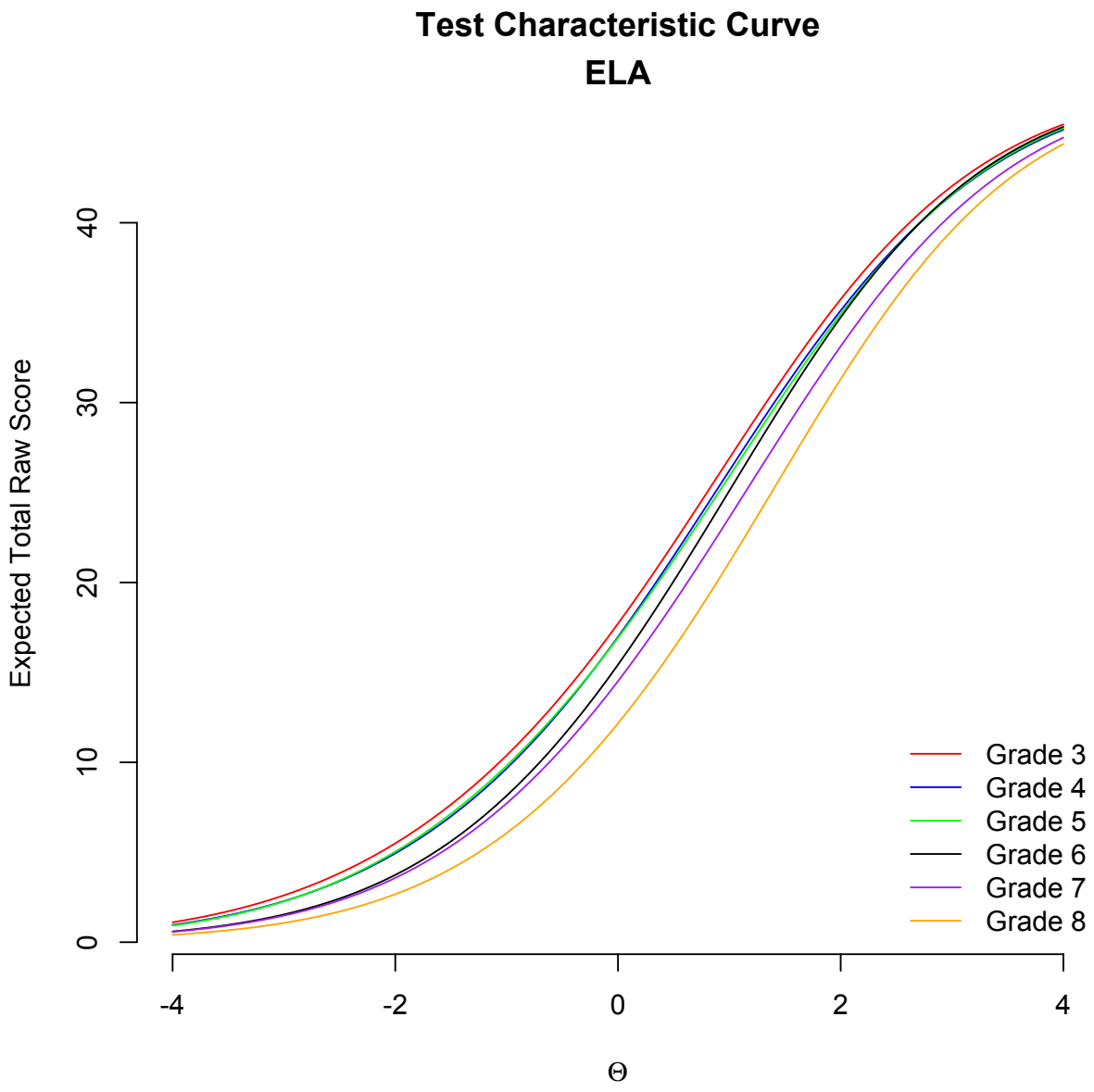
Grade	Marginal Reliability
7	0.90
8	0.89
11	0.90

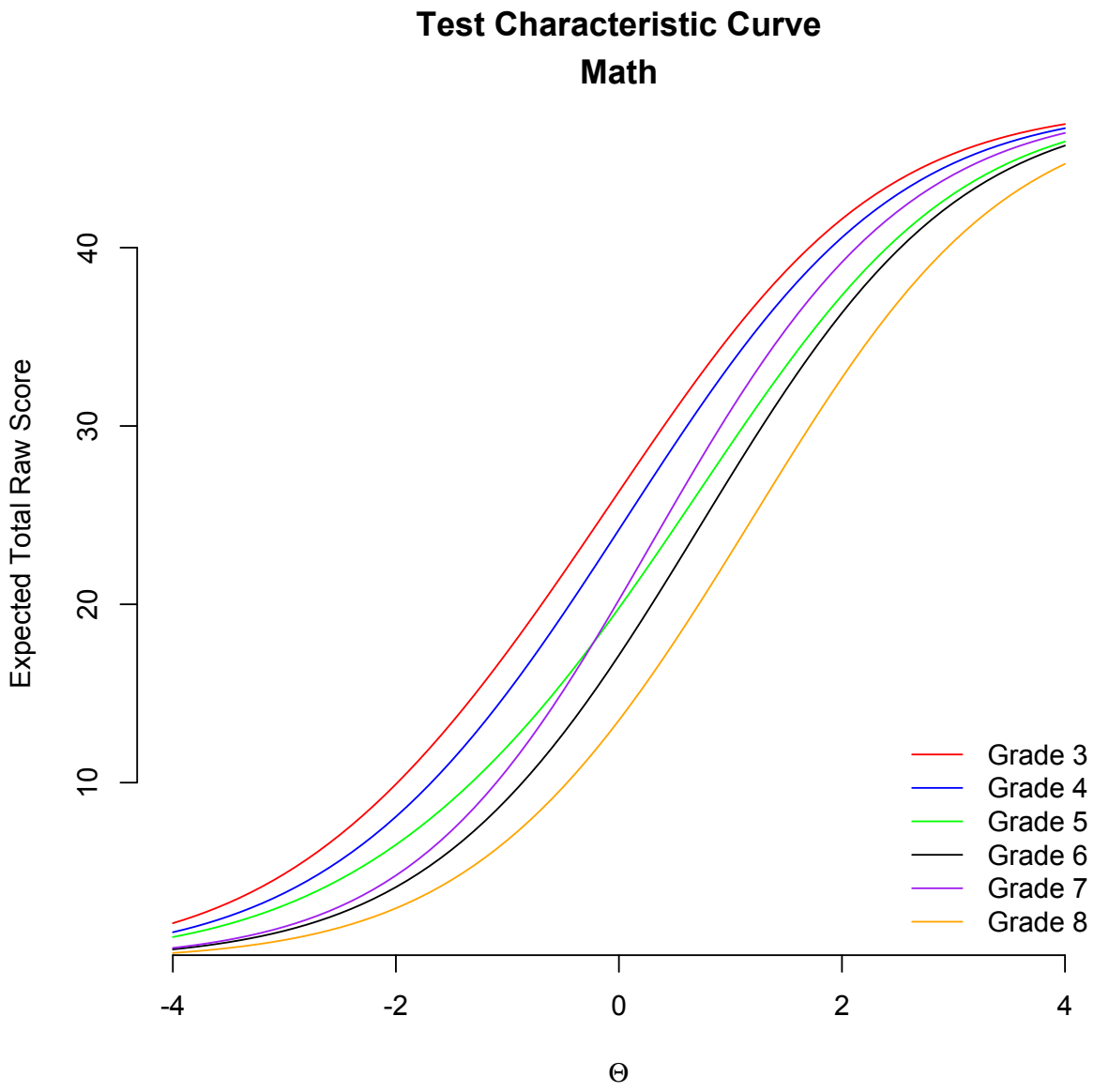
Science
The test reliabilities for science were in the high range, from .88 to .91.

Grade	Marginal Reliability
5	0.91
8	0.88
11	0.89

Validation of ORExt Vertical Scales

The test characteristic curves (TCCs) for the grade-level assessments in ELA and mathematics demonstrate incrementally increasing growth and test demands across Grades 3-8, with the exception of Grade 7 mathematics. The Grade 7 mathematics assessment will be revised to be more difficult across all complexity levels for next year's administration to address this discrepancy. Grade 11 and science tests are not vertically scaled; TCCs are thus not presented for Grade 11 or science.


Test Characteristic Curve
ELA
Expected Total Raw Score
0
10
20
30
40
-4
-2
0
2
4
Θ
Grade 3
Grade 4
Grade 5
Grade 6
Grade 7
Grade 8


Test Characteristic Curve
Math
Expected Total Raw Score
10
20
30
40
-4
-2
0
2
4
Θ
Grade 3
Grade 4
Grade 5
Grade 6
Grade 7
Grade 8

4.1B Overall and Conditional Standard Errors of Measure

The average SEM associated with each cut score for 2015-16 student data are presented in the table below, supported by a KEY. The SEMs decreased in almost all cases compared to last year, suggesting that the measures are more reliable when student eligibility is more strictly controlled. See *Section 4.2* below for means and standard deviations by grade and subject area.

SEM = Standard Error of Measure associated with the cut score to the left; averaged to the tenths' place.
Level 1 = Does Not Yet Meet (not included as the lowest level of proficiency)
Level 2 = Nearly Meets
Level 3 = Meets
Level 4 = Exceeds

English Language Arts

Grade	Level 2	SEM	Level 3	SEM	Level 4	SEM
3	**192**	4.3	**213**	3.8	**228**	5.5
4	**200**	3.8	**213**	3.9	**228**	5.4
5	**202**	3.9	**220**	4.2	**232**	5.8
6	**205**	3.6	**220**	4.1	**233**	6.2
7	**208**	3.7	**222**	4.2	**236**	6.3
8	**213**	3.6	**224**	4.0	**236**	5.7
11	**899**	3.8	**920**	4.5	**927**	5.6

Mathematics

Grade	Level 2	SEM	Level 3	SEM	Level 4	SEM
3	**192**	3.9	**201**	3.9	**218**	5.1
4	**193**	3.8	**206**	3.9	**219**	5.2
5	**193**	4.2	**206**	3.8	**220**	4.4
6	**204**	3.6	**208**	3.6	**222**	4.6
7	**207**	3.6	**209**	3.6	**223**	4.7
8	**208**	3.7	**212**	3.7	**226**	4.5
11	**901**	3.6	**907**	3.8	**922**	5.1

Science

Grade	Level 2	SEM	Level 3	SEM	Level 4	SEM
5	**506**	3.6	**517**	4.0	**530**	5.5
8	**810**	3.7	**820**	4.3	**831**	6.2
11	**901**	3.6	**914**	3.9	**929**	5.7

4.1C Classification Accuracy & Consistency

Results from the 2015-16 OREXt test administration were compared to the 2014-15 results using Rudner's classification index (Rudner, 2005). Results closer to 1.0 indicate the likelihood that a student was appropriately classified as proficient or not proficient (accuracy) and the likelihood that the student would be classified in the same category

given an additional test administration. The calculation utilizes item difficulty and theta value distributions, as well as related standard errors of measurement, to generate probabalistic estimates based on one test administration. Complete results, generated from the *cacIRT* package in *R*, are provided below. Results denote very high levels of classification accuracy and consistency.

Test Classification Accuracy

Grade	English Language Arts	Mathematics	Science
3	0.94	0.93	-
4	0.95	0.93	-
5	0.95	0.90	0.95
6	0.95	0.92	-
7	0.95	0.91	-
8	0.94	0.90	0.94
11	0.94	0.92	0.95

Test Classification Consistency

Grade	English Language Arts	Mathematics	Science
3	0.92	0.90	-
4	0.93	0.90	-
5	0.92	0.87	0.93
6	0.92	0.89	-
7	0.92	0.88	-
8	0.92	0.86	0.92
11	0.92	0.89	0.93

The ORExt is not a computer-adaptive instrument so estimate precision documentation based upon that test design is not provided.

4.2 Fairness and Accessibility

The state has taken steps to ensure fairness in the development of the assessments, including an analysis of each test item by Oregon teachers not only for linkage to standards, but also for access, sensitivity, and bias (see *Appendix 3.1A)*. In addition, we reviewed test functioning as relevant to race/ethnicity and disability subgroups. This process increases the likelihood that students are receiving instruction in areas reflected in the assessment, and also that the items are not biased toward a particular demographic or sub-group.

The full ethnic and disability demographics for students taking the ORExt are reported below. Students ethnicity/race was reported in seven categories: (a) Asian, (b) American Indian/Alaskan Native, (c) Black or African-American, (d) Native Hawaiian or Other Pacific Islander, (e) Hispanic, (f) White, or (g) Multi-ethnic. The majority of students were reported as White (54-65%). These results are largely consistent with the demographics reported for the general assessments, though percentages taking the ORExt are slightly

higher for most students of color and generally lower for students who are Asian or White (see *Appendix 4.2*).

Race – Ethnicity Percentages and Totals by Content Area and Grade Level

English language arts

	Grade						
Ethnicity-Race	3	4	5	6	7	8	11
Asian	0.04	0.04	0.03	0.05	0.04	0.03	0.04
American Indian/ Alaska Native	0.02	0.02	0.02	0.03	0.02	0.03	*
Black or African-American	0.03	0.03	0.04	0.04	0.04	0.03	0.04
Native Hawaiian or Other Pacific Islander	*	*	*	*	*	*	*
Hispanic	0.28	0.26	0.29	0.22	0.24	0.25	0.20
White	0.57	0.60	0.54	0.60	0.60	0.61	0.65
Multi-ethnic	0.06	0.05	0.07	0.06	0.05	0.04	0.05
Total *n*-sizes	612	637	598	598	620	592	474

Note. *$n < 10$; percentages may not add to 1.0 due to rounding.

Mathematics

	Grade						
Ethnicity-Race	3	4	5	6	7	8	11
Asian	0.04	0.04	0.03	0.05	0.04	0.03	0.04
American Indian/ Alaska Native	0.02	0.02	0.02	0.03	0.02	0.03	*
Black or African-American	0.03	0.03	0.04	0.04	0.03	0.03	0.04
Native Hawaiian or Other Pacific Islander	*	*	*	*	*	*	*
Hispanic	0.28	0.26	0.29	0.22	0.24	0.25	0.20
White	0.57	0.61	0.54	0.61	0.61	0.61	0.64
Multi-ethnic	0.06	0.05	0.07	0.06	0.05	0.04	0.05
Total *n*-sizes	610	636	597	596	617	594	476

Note. *$n < 10$; percentages may not add to 1.0 due to rounding.

Science

	Grade		
Ethnicity-Race	5	8	11
Asian	0.03	0.03	0.04
American Indian/ Alaska Native	0.02	0.02	0.02
Black or African-American	0.04	0.03	0.03
Native Hawaiian or Other Pacific Islander	*	*	*
Hispanic	0.29	0.25	0.20
White	0.55	0.61	0.65
Multi-ethnic	0.07	0.05	0.05
Total *n*-sizes	588	571	466

Note. *$n < 10$; percentages may not add to 1.0 due to rounding.

Student reported exceptionalities included Intellectual Disability (ID), Hearing Impairment (HI), Visual Impairment (VI), Deaf-Blindness (DB), Communication Disorder (CD), Emotional Disturbance (ED), Orthopedic Impairment (OI), Traumatic Brain Injury (TBI), Other Health Impairment (OHI), Autism Spectrum Disorder (ASD), and Specific Learning Disability (SLD). The majority of students who participated in the ORExt were students with ID (27-49%) and students with ASD (23 -34%), followed by students with OHI (10 - 16%). ODE policy for 2015-16 changed to require students who participate in the OREXt to take the assessment in all relevant content areas. There is thus very little change in terms of participation percentages across content areas, as evidenced by the total *n*-sizes per grade level displayed below.

Exceptionality Percentages By Content Area and Grade Level

English language arts

	Grade						
Category	3	4	5	6	7	8	11
ID 10	0.27	0.31	0.28	0.35	0.38	0.48	0.49
HI 20	*	*	*	*	*	*	*
VI 40	*	*	*	*	*	*	*
DB 43	*	*	*	*	*	*	*
CD 50	0.14	0.11	0.08	0.07	0.03	0.04	0.04
ED 60	0.02	0.03	0.03	0.04	0.04	0.03	0.04
OI 70	0.03	0.03	0.02	0.04	0.03	0.02	0.04
TBI 74	*	*	*	*	*	*	*
OHI 80	0.16	0.16	0.15	0.14	0.13	0.10	0.11
ASD 82	0.29	0.28	0.34	0.28	0.28	0.26	0.23
SLD 90	0.08	0.06	0.07	0.08	0.08	0.05	0.04
Total *n*-sizes	612	637	598	598	620	592	474

Note. *$n < 10$; percentages may not add to 1.0 due to rounding.

Mathematics

	Grade						
Category	3	4	5	6	7	8	11
ID 10	0.27	0.31	0.28	0.35	0.38	0.48	0.49
HI 20	*	*	*	*	*	*	*
VI 40	*	*	*	*	*	*	*
DB 43	*	*	*	*	*	*	*
CD 50	0.13	0.11	0.08	0.06	0.03	0.04	0.04
ED 60	0.02	0.03	0.03	0.04	0.04	0.03	0.04
OI 70	0.03	0.03	0.03	0.04	0.03	0.02	0.04
TBI 74	*	*	*	*	*	*	*
OHI 80	0.16	0.15	0.15	0.14	0.13	0.10	0.11
ASD 82	0.29	0.28	0.34	0.28	0.29	0.26	0.23
SLD 90	0.08	0.07	0.07	0.08	0.07	0.05	0.04
Total *n*-sizes	610	636	597	596	617	594	476

Note. *n < 10; percentages may not add to 1.0 due to rounding.

Science

	Grade		
Category	5	8	11
ID 10	0.29	0.48	0.48
HI 20	*	*	*
VI 40	*	*	*
DB 43	*	*	*
CD 50	0.08	0.04	0.04
ED 60	0.03	0.03	0.04
OI 70	0.03	0.02	0.03
TBI 74	*	*	*
OHI 80	0.15	0.11	0.11
ASD 82	0.34	0.26	0.23
SLD 90	0.07	0.06	0.04
Total *n*-sizes	588	571	466

Note. *n < 10; percentages may not add to 1.0 due to rounding.

Observed Means and Standard Deviations

The following tables provide information regarding observed means and standard deviations by content area and grade level. The Grade 3-8 English language arts and mathematics scaled scores are centered on 200, while all Grade 11 scores are centered on 900 (to reinforce that they are not on the vertical scale). Science is centered on 500 at Grade 5 and centered on 800 at Grade 8. The vertically scaled scores generally convey incremental gains in achievement across grade levels, though the results suggest small losses appearing at Grade 8 in ELA. These scales were selected to clearly determine whether scores are on the same scale and also to differentiate among the statewide

assessments in use to avoid confusion (i.e., SBA, OAKS, OExt, ELPA, KA). The general pattern is that RIT scores decreased from 2014-15 to 2015-16. This decrease is attributed not to the scale, nor to deceleration of growth, but to the substantive shift in the tested student population as a result of ODE eligibility guidelines.

2014-15 RIT Scores

	ELA		**Math**		**Science**	
Grade	Mean	SD	Mean	SD	Mean	SD
3	219.3	24.6	201.5	20.8		
4	222.8	23.6	204.8	19.8		
5	224.9	25.0	205.3	18.1	517.6	25.6
6	226.3	24.0	207.7	17.7		
7	226.4	25.0	207.9	19.0		
8	225.4	24.1	207.8	17.3	822.1	25.8
11	922.5	28.5	903.8	21.1	920.8	27.7

2015-16 RIT Scores

	ELA		**Math**		**Science**	
Grade	Mean	SD	Mean	SD	Mean	SD
3	210.3	23.0	197.6	20.2		
4	212.3	22.9	198.1	18.7		
5	217.1	24.5	201.2	17.2	514.2	22.1
6	220.1	25.5	204.8	17.6		
7	223.6	28.9	205.4	19.0		
8	221.2	24.8	206.7	17.2	819.0	25.6
11	920.7	27.7	902.3	20.0	918.0	24.9

Observed Means Reported by Sex

The following tables provide information regarding average student performance by grade level and sex (Female/Male) in each of the content areas assessed on the OExt. Significant differences based on a two sample *t*-test are noted in Grades 4 and 5 in mathematics and Grade 5 in science.

English Language Arts

	Grade						
Sex	**3**	**4**	**5**	**6**	**7**	**8**	**11**
Female	207.7	209.9	216.5	219.6	224.9	223.8	921.1
Male	211.2	213.5	217.6	220.4	223.2	220.2	919.9

Note. *$p < .05$

Mathematics

	Grade						
Sex	**3**	**4**	**5**	**6**	**7**	**8**	**11**
Female	195.1	194.6*	198.0*	203.3	203.9	207.4	900.6
Male	198.3	200.0*	202.6*	205.4	206.3	206.6	902.8

Note. *$p < .05$

Science

	Grade		
Sex	**5**	**8**	**11**
Female	510.3*	819.9	916.1
Male	515.9*	818.9	918.7

Note. *$p < .05$

Observed Means Reported by Race

The following tables provide information regarding average student performance by grade level and race (Non-White/White) in each of the content areas assessed on the OREXt. Significant differences are noted by two sample *t*-tests in ELA Grade 7,

English Language Arts

	Grade						
Race	**3**	**4**	**5**	**6**	**7**	**8**	**11**
Non-White	206.7	209.0	217.7	217.1	218.1*	220.6	917.7
White	210.6	212.8	217.2	220.8	224.9*	221.7	920.8

Note. *$p < .05$

Mathematics

	Grade						
Race	**3**	**4**	**5**	**6**	**7**	**8**	**11**
Non-White	194.6	195.7	201.9	203.9	203.0	207.3	899.7
White	197.6	198.5	201.0	204.9	206.0	206.8	902.5

Note. *$p < .05$

Science

	Grade		
Race	**5**	**8**	**11**
Non-White	515.0	816.7	913.6
White	514.0	819.7	918.5

Note. *$p < .05$

Observed Means Reported by Disability Status

The following tables provide information regarding average student performance by grade level and disability category in each of the content areas assessed on the OREXt. Students with SLD were generally the highest performing group, though students with CD and ED performed higher at certain grade levels/content areas. The lowest performing group is consistently students with OI, followed by students with ID or ASD, depending upon grade level.

English Language Arts

Disability	**Grade**						
Code	**3**	**4**	**5**	**6**	**7**	**8**	**11**
ID	205.5	206.5	211.7	218.1	220.1	220.7	916.5
HI	*	*	*	*	*	*	*
VI	*	*	*	*	*	*	*
DB	*	*	*	*	*	*	*
CD	220.0	223.3	230.7	233.0	234.3	240.9	940.7
ED	221.6	230.6	233.4	234.1	242.9	239.2	943.3
OI	166.1	187.8	196.8	195.8	195.2	195.5	896.5
TBI	*	*	*	*	*	*	*
OHI	207.7	210.3	219.4	226.5	231.2	226.7	930.1
ASD	209.8	213.1	214.3	213.5	219.3	214.3	916.8
SLD	227.6	231.5	236.3	238.5	247.4	240.1	940.6

Note. *$n < 10$

Mathematics

Disability	**Grade**						
Code	**3**	**4**	**5**	**6**	**7**	**8**	**11**
ID	193.9	193.0	197.3	202.9	202.5	206.6	898.6
HI	*	*	*	*	*	*	*
VI	*	*	*	*	*	*	*
DB	*	*	*	*	*	*	*
CD	207.1	208.3	210.4	212.4	213.5	216.3	915.0
ED	211.7	215.8	212.1	216.1	220.0	217.4	915.0
OI	161.1	179.2	187.1	187.5	182.1	191.2	883.2
TBI	*	*	*	*	*	*	*
OHI	195.1	197.0	202.4	207.6	208.5	209.6	911.3
ASD	195.0	198.0	199.5	201.3	204.3	202.3	900.7
SLD	214.4	214.6	214.3	218.2	221.5	220.6	918.3

Note. *$n < 10$

Science

Disability	**Grade**		
Code	**5**	**8**	**11**
ID	509.6	819.0	913.9
HI	*	*	*
VI	*	*	*
DB	*	*	*
CD	529.5	837.3	934.2
ED	531.5	843.1	938.3
OI	495.8	792.1	904.9
TBI	*	*	*
OHI	517.2	826.0	930.7
ASD	509.4	810.1	912.3
SLD	536.2	838.6	937.4

Note. *$n < 10$

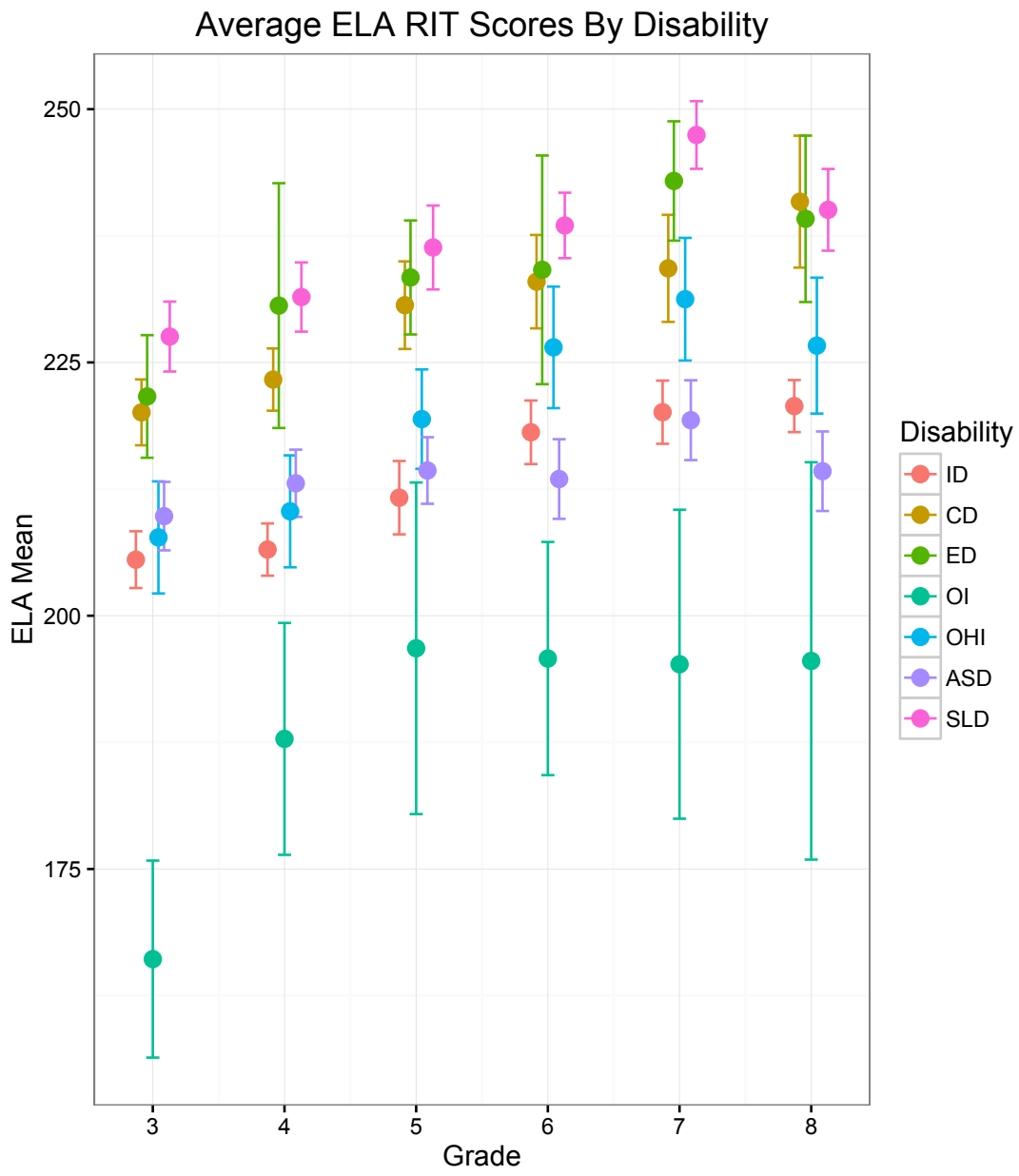
Graphs of Observed Means By Disability

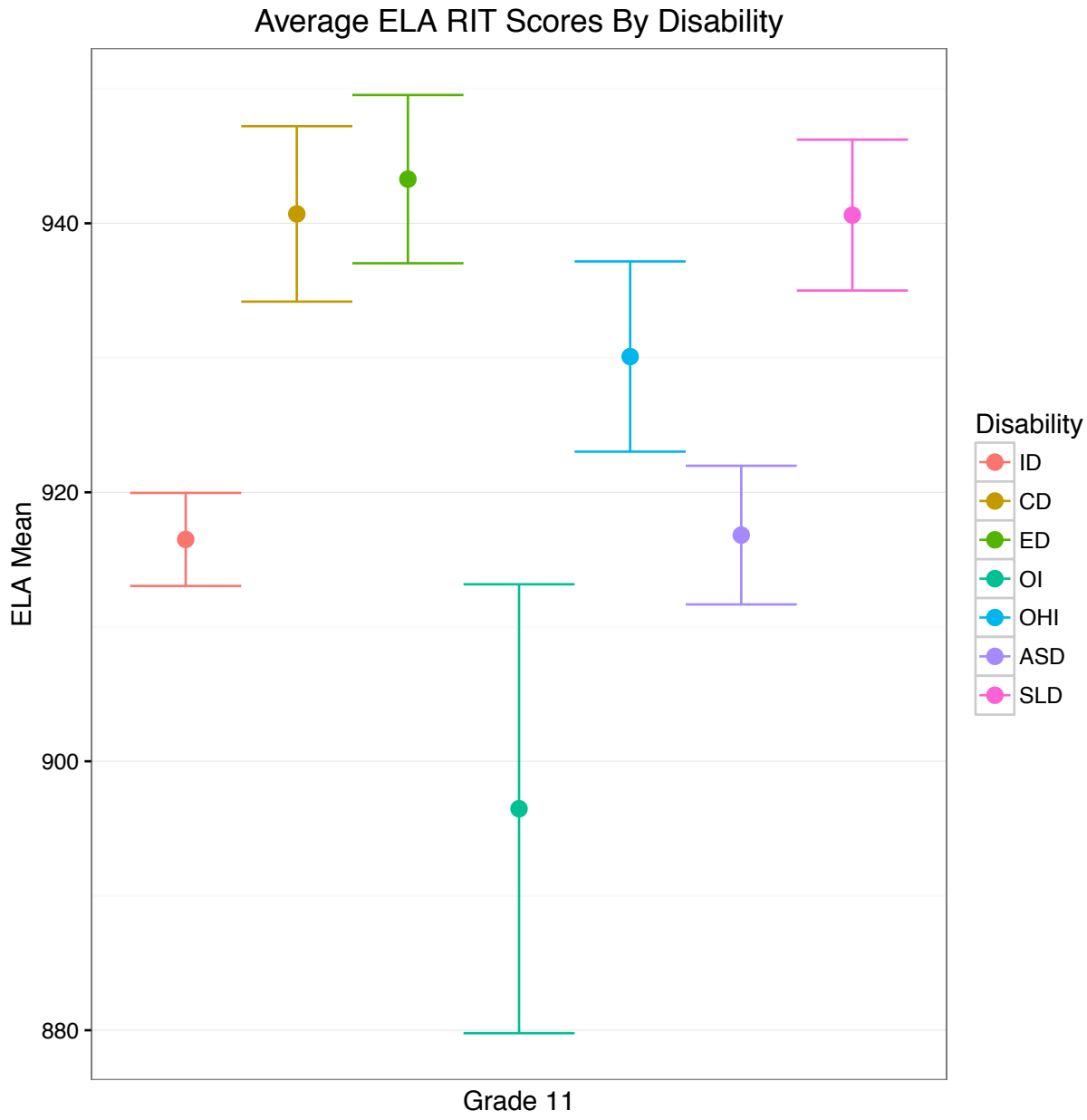
The graphs below convey information similar to that shared above in graphic form.

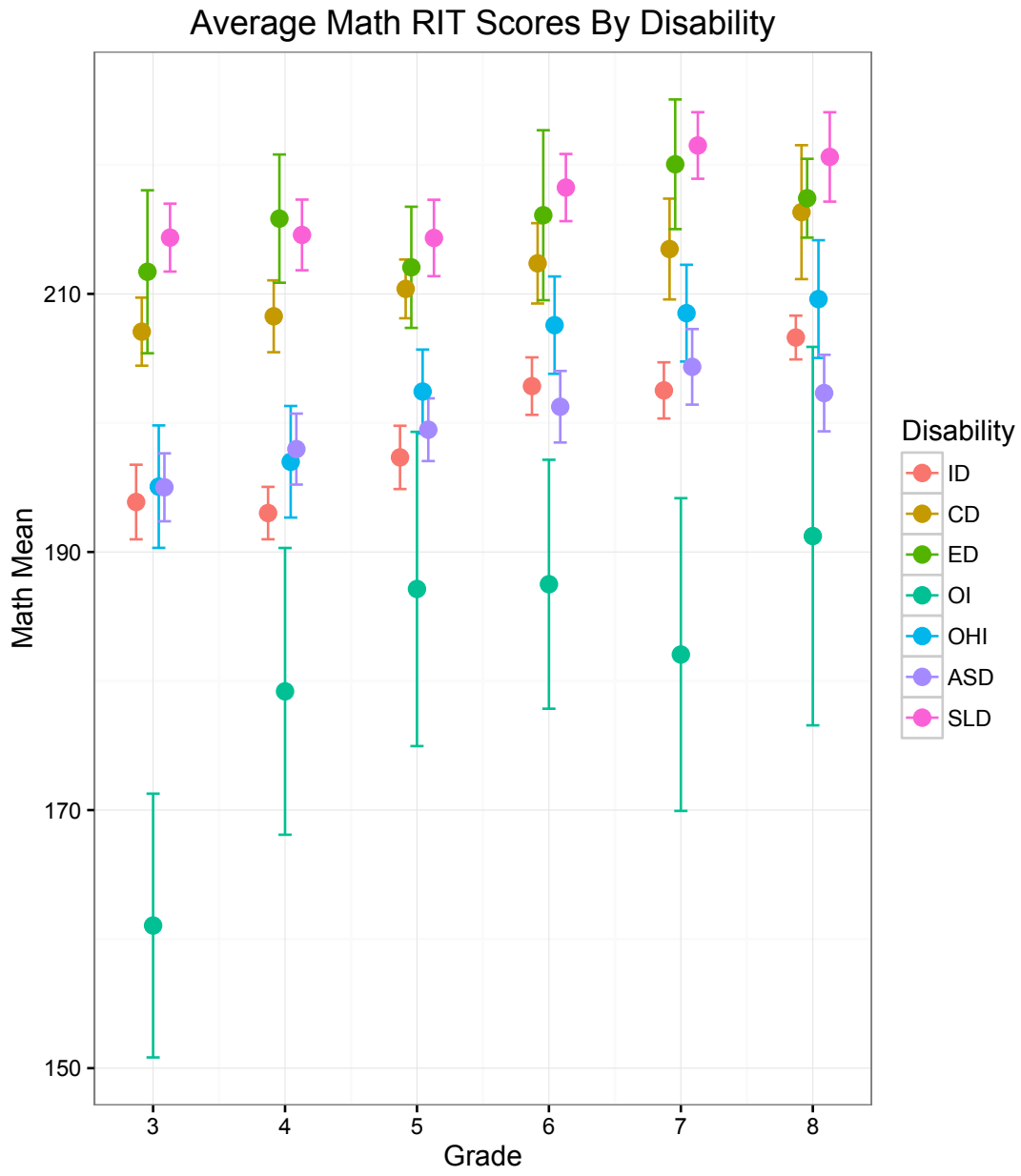
The graphics include 95% confidence interval error bars, so determining which subgroups performed in a manner that is significantly better than others is readily apparent by looking at the location of the error bars. Error bars that do not overlap in terms of the *y*-scale are significantly different. Only students who generally had more than 10 members at each grade level are reported. This required the removal of graphs for students in the HI, VI, DB, and TBI categories.

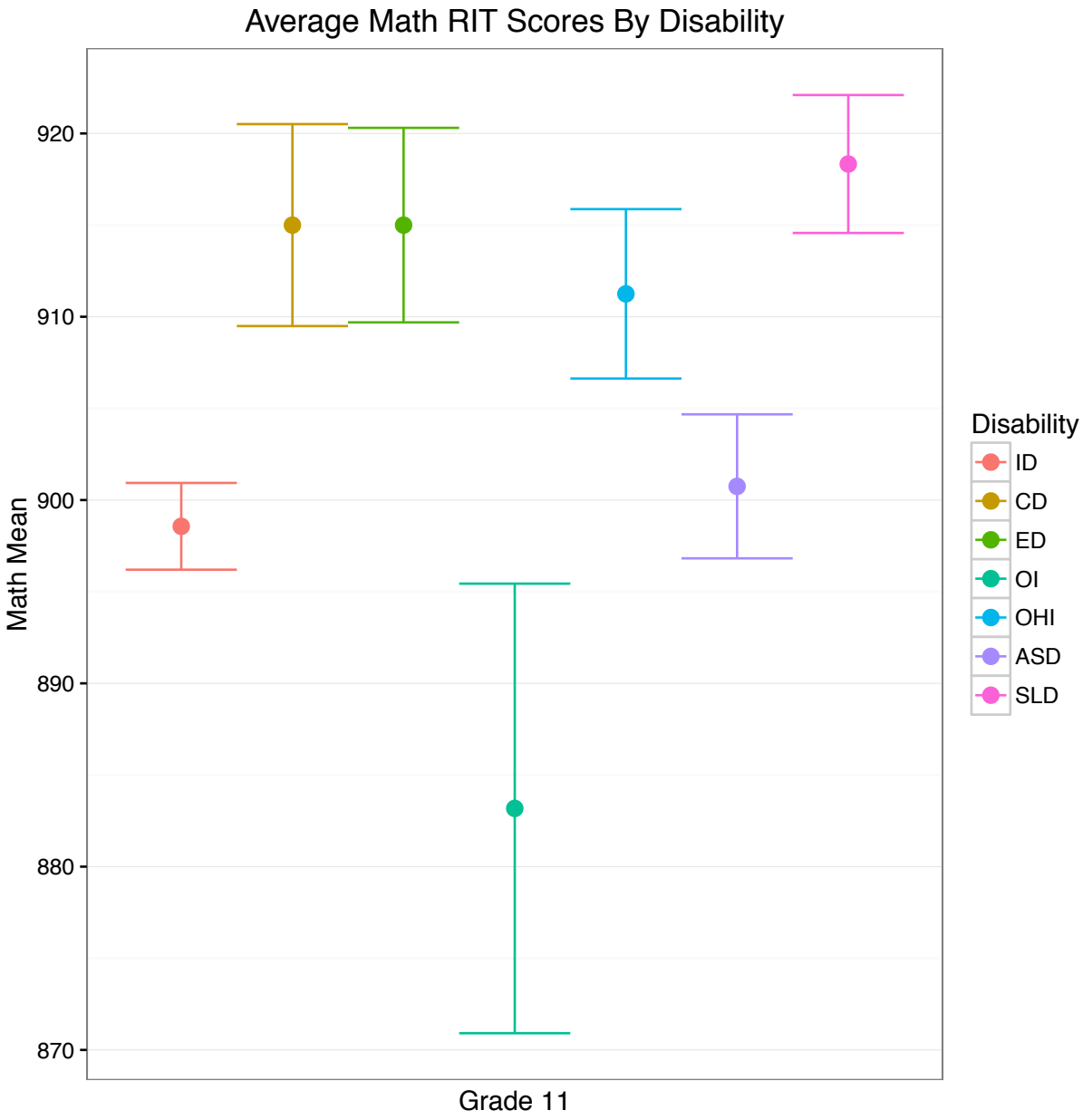
Students with OI are again the lowest performing group, being significantly outperformed by all other subgroups. Students with SLD are consistently outperforming most peers, with students with ED and CD performing at similarly high levels.

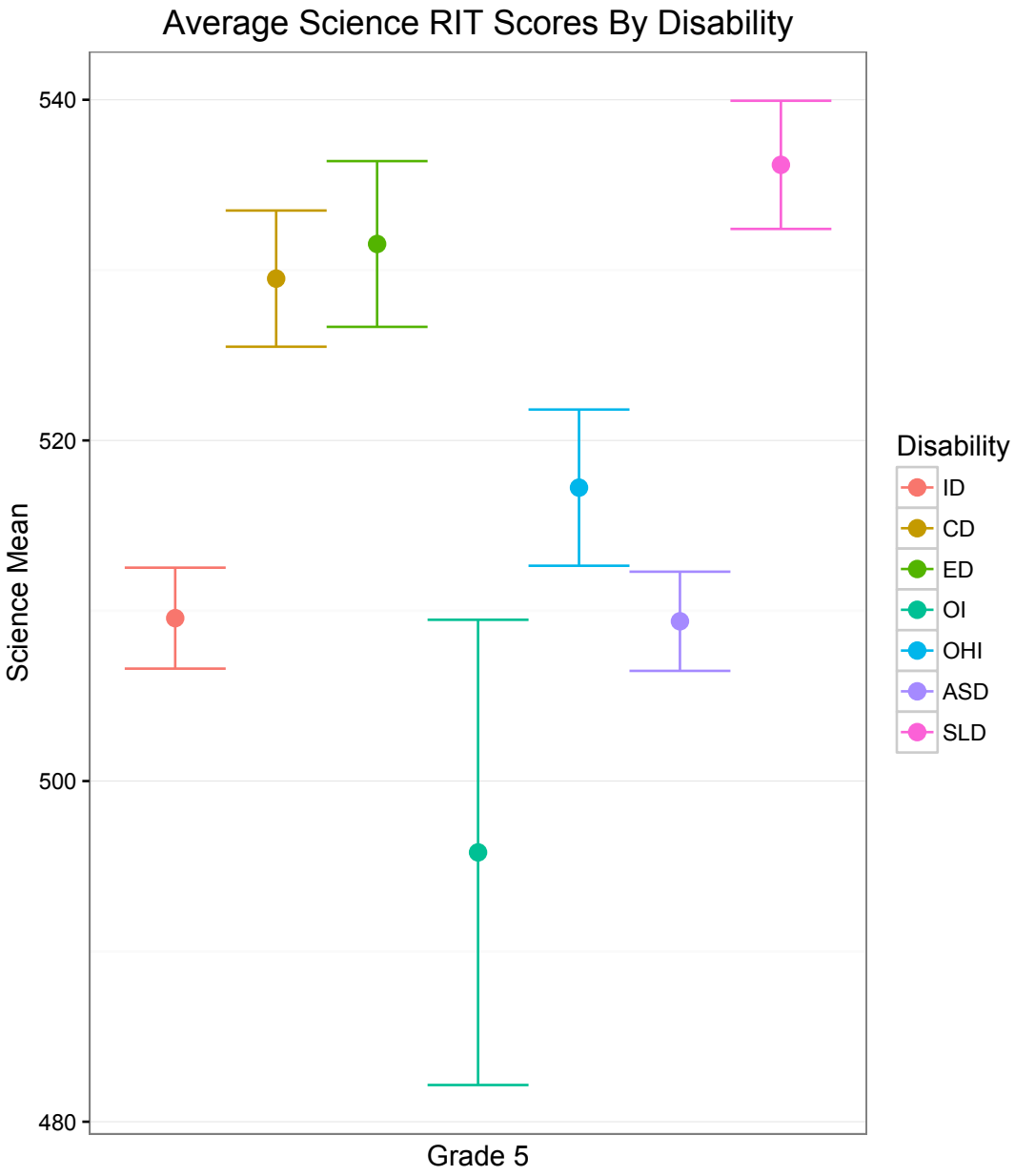
Students with OI are consistently the lowest performing group, which led to concerns regarding test accessibility. However, the results fo this year's consequential validity study demonstrate that the OI label is insufficient to fully describe the severity and range of concomitant disabilities that students whose primary label is OI conveys (see *Appendix 2.3B.10*).

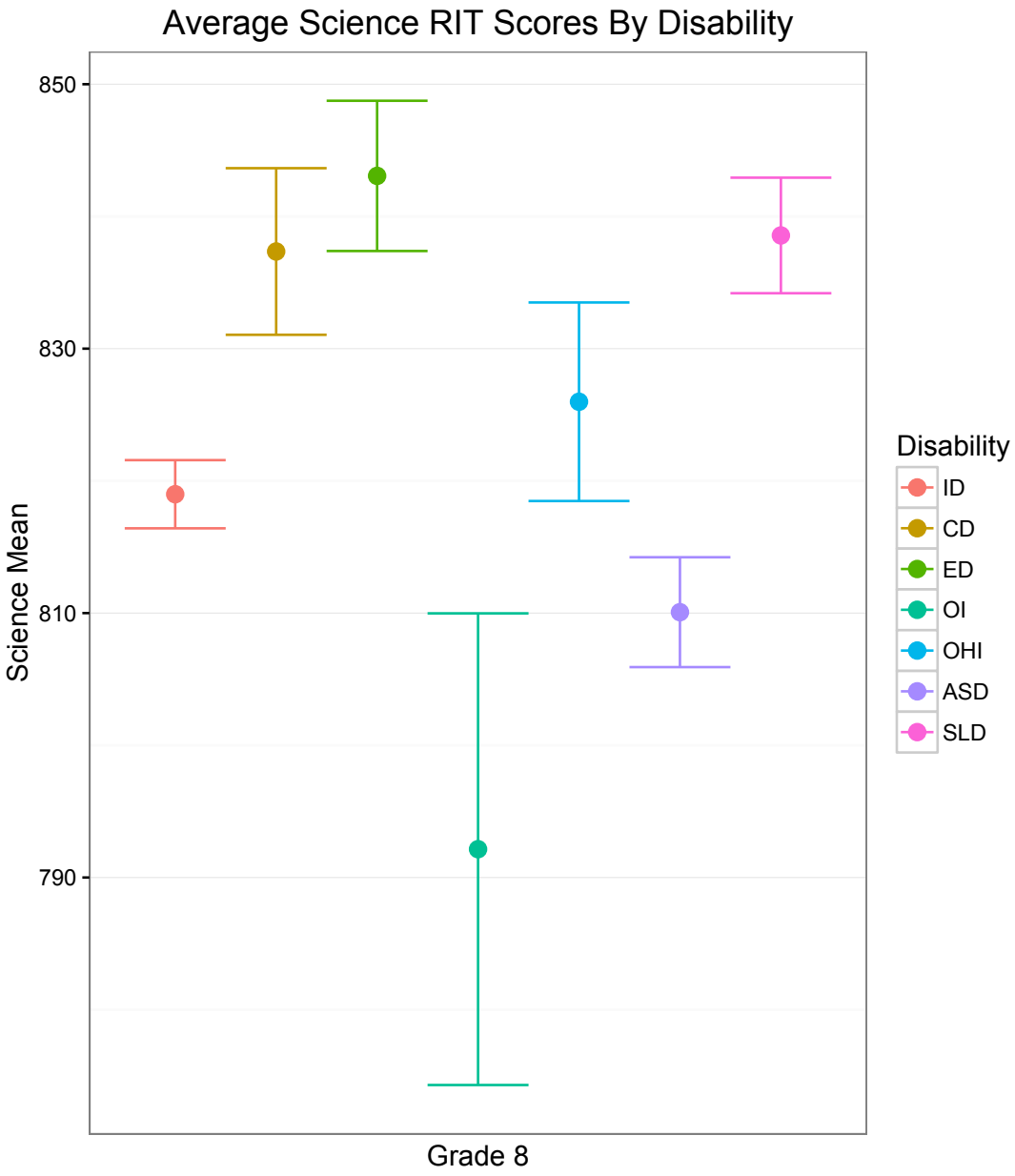

Average ELA RIT Scores By Disability
ELA Mean
250
225
200
175
3
4
5
6
7
8
Grade
Disability
ID
CD
ED
OI
OHI
ASD
SLD

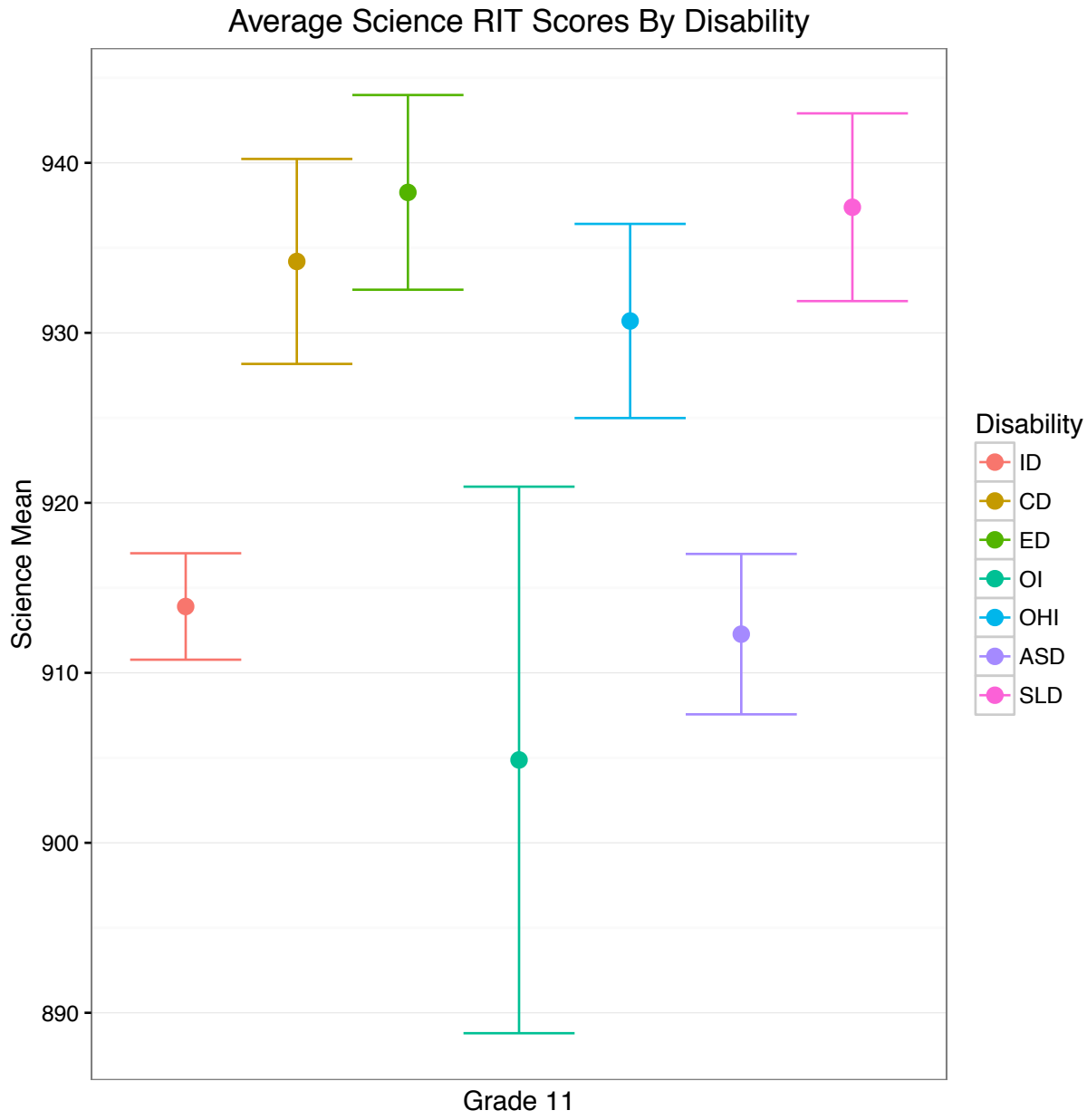

Average ELA RIT Scores By Disability
ELA Mean
940
920
900
880
Grade 11
Disability
ID
CD
ED
OI
OHI
ASD
SLD


Average Math RIT Scores By Disability
Math Mean
210
190
170
150
3
4
5
6
7
8
Grade
Disability
ID
CD
ED
OI
OHI
ASD
SLD


Average Math RIT Scores By Disability
Math Mean
920
910
900
890
880
870
Grade 11
Disability
ID
CD
ED
OI
OHI
ASD
SLD


Average Science RIT Scores By Disability
Science Mean
540
520
500
480
Grade 5
Disability
ID
CD
ED
OI
OHI
ASD
SLD


Average Science RIT Scores By Disability
850
830
810
790
Science Mean
Grade 8
Disability
ID
CD
ED
OI
OHI
ASD
SLD



4.3 Full Performance Continuum

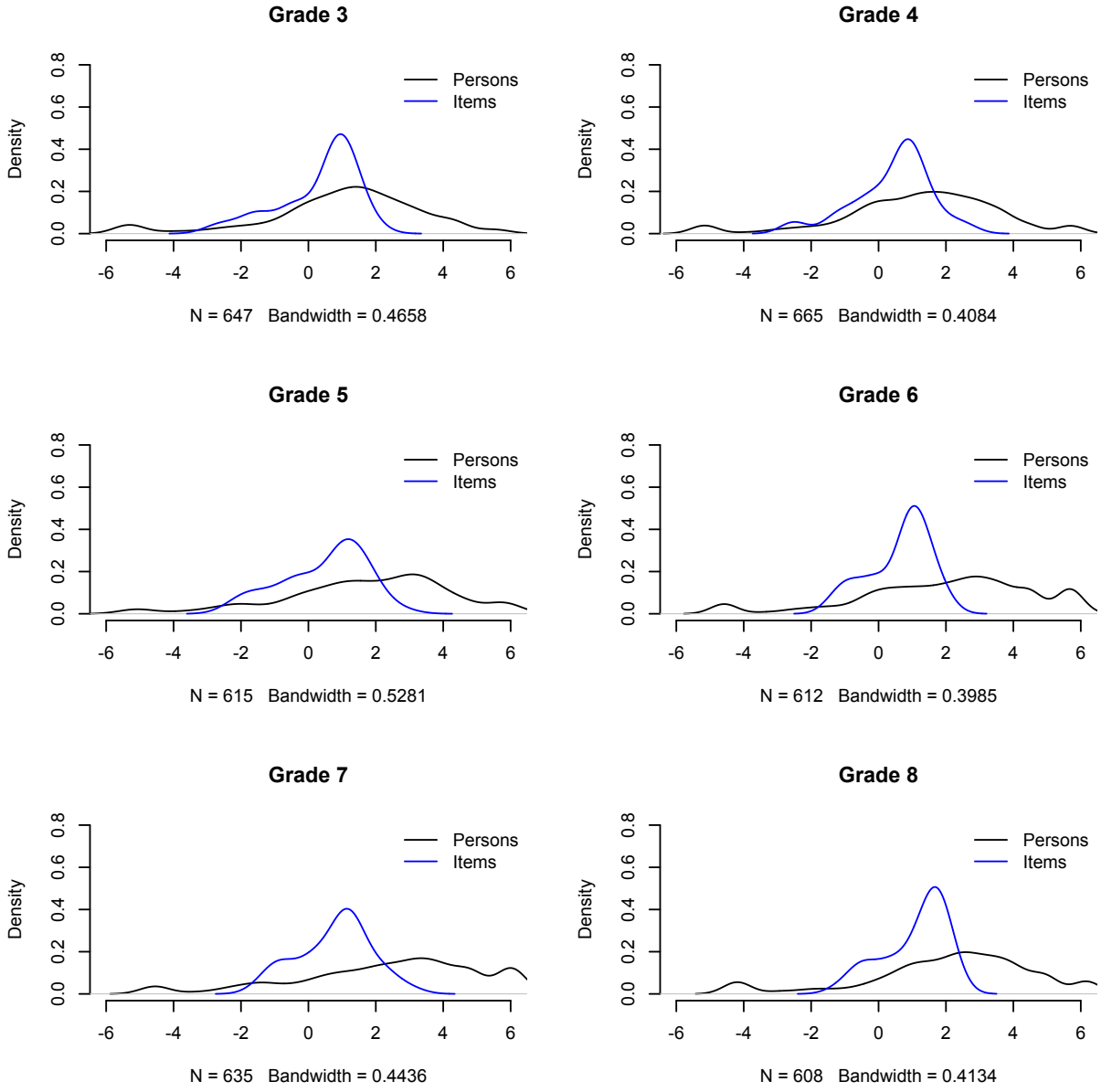
The OREXt is designed to sample the Common Core State Standards in English language arts (Reading, Writing, and Language) and mathematics, as well as the Oregon Science Standards and Next Generation Science Standards in science in a purposive, validated manner. The OREXt test blueprints convey the balance of representation exhibited by the assessment (see *Appendix 2.1B*). These test blueprints are supported by the OREXt Extended Assessment Frameworks (http://www.brtprojects.org/publications/training-modules), which define the assessable content on the OREXt that has been reduced in depth, breadth, and complexity (RDBC) using our defined process (see *Appendix 2.3A.3*). The decisions regarding which standards to target for essentialization, as well as the

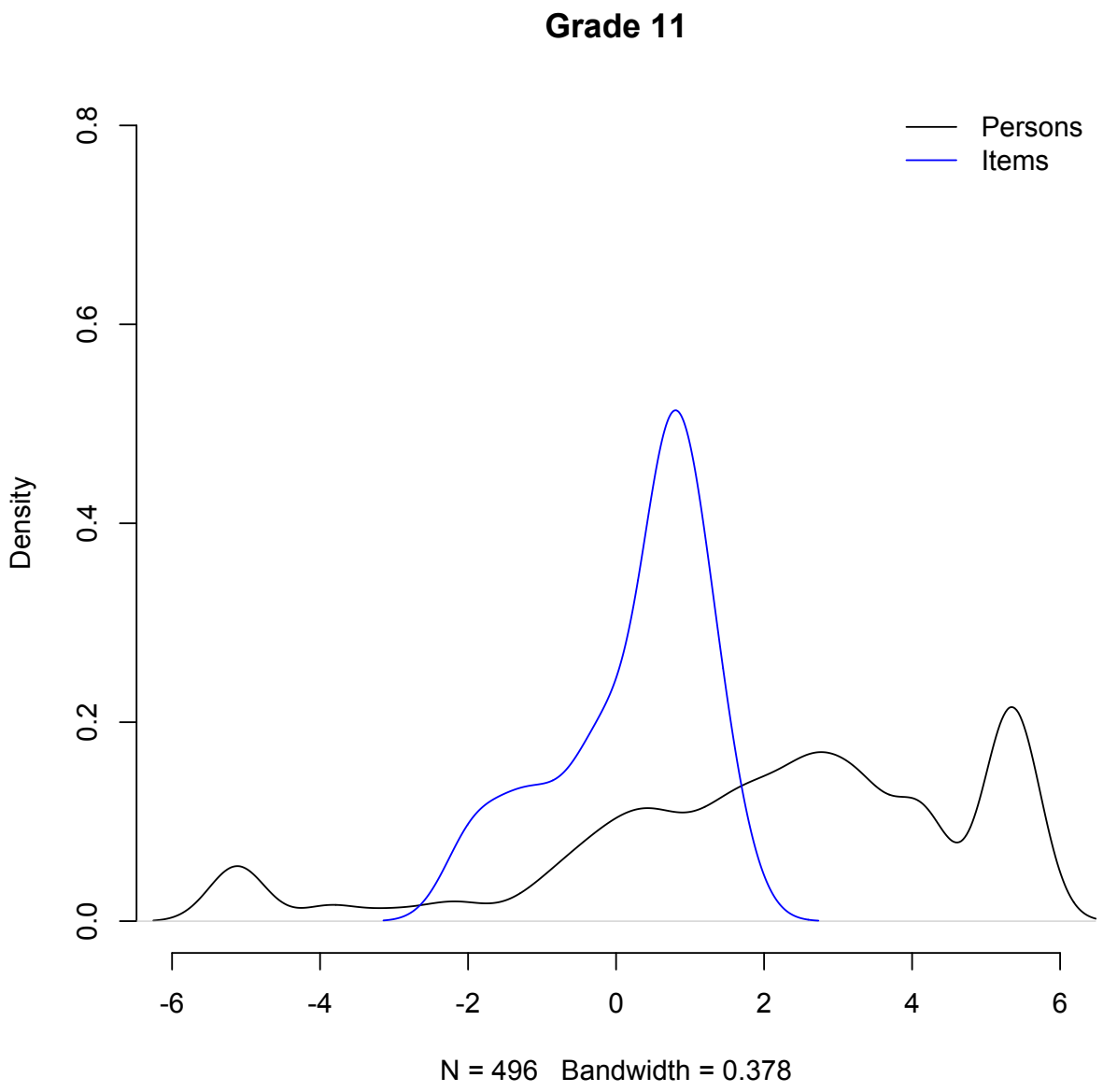
strength of linkage between the Essentialized Standards and the CCSS/ORSci/NGSS has been validated by Oregon teachers, as well (see *Appendix 3.1A*).

Though a simplified and standardized approach was taken to design item, increasing efficiency and access to the assessment for the majority of students (as evidenced by the decreased percentages of zero scores across all content areas), a small subgroup of students remains who cannot access an academic assessment. This is true even though items have been significantly RDBC at three levels of complexity (low-medium-high difficulty). As a response, ODE commissioned BRT to design and implement an observational rating scale for this group of very low-performing students, called the Oregon Observational Rating Assessment (ORora) for the spring 2016 administration. The ORora will target communication (expressive and receptive) and basic skills (attention/joint attention and mathematics) and provide documentation of student progress outside of our clearly defined academic domains.

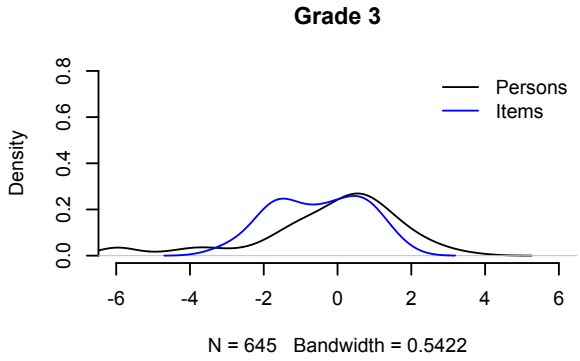
Items on all assessments were scored on a 2-point scale, with 1 point awarded for a correct response and 0 points awarded for an incorrect response. Plots are provided below for each content area and grade level, including the person ability and item difficulty distributions. In general, the descriptive statistics suggest that the test had an appropriate range of item difficulties represented, from easy to difficult, with item difficulties generally ranging from -4.0 to +4.0 on the Rasch scale. The assessments performed as expected across all grades and content areas with the exception of Grade 7 mathematics, as noted above. The item person distributions provided below demonstrate that the ORExt is providing a performance continuum for students who participate, though we plan to include more complex items for ELA and science in the 2016-17 ORExt administration.

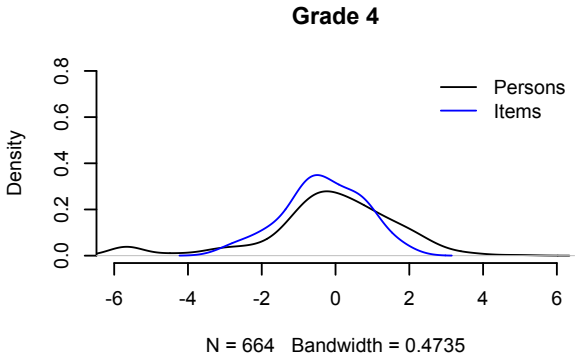
English Language Arts Person/Item Distributions

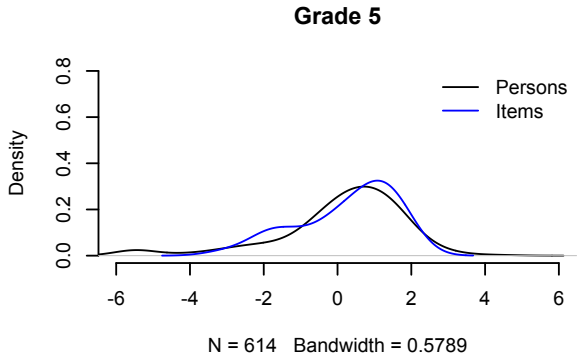


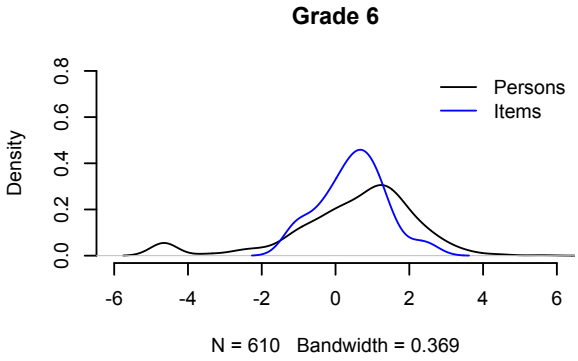

Grade 11
Persons
Items
Density
0.0
0.2
0.4
0.6
0.8
-6
-4
-2
0
2
4
6
N = 496 Bandwidth = 0.378

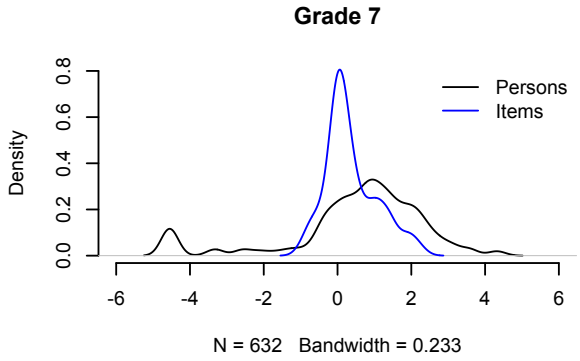
Mathematics Person/Item Distributions

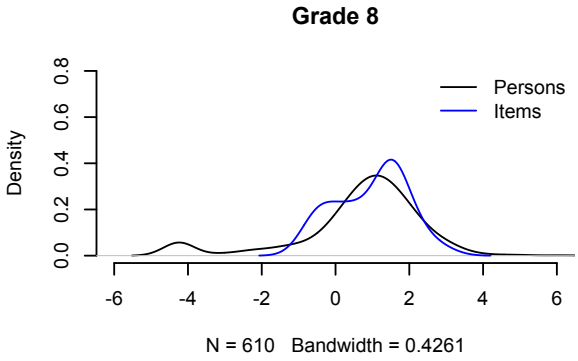


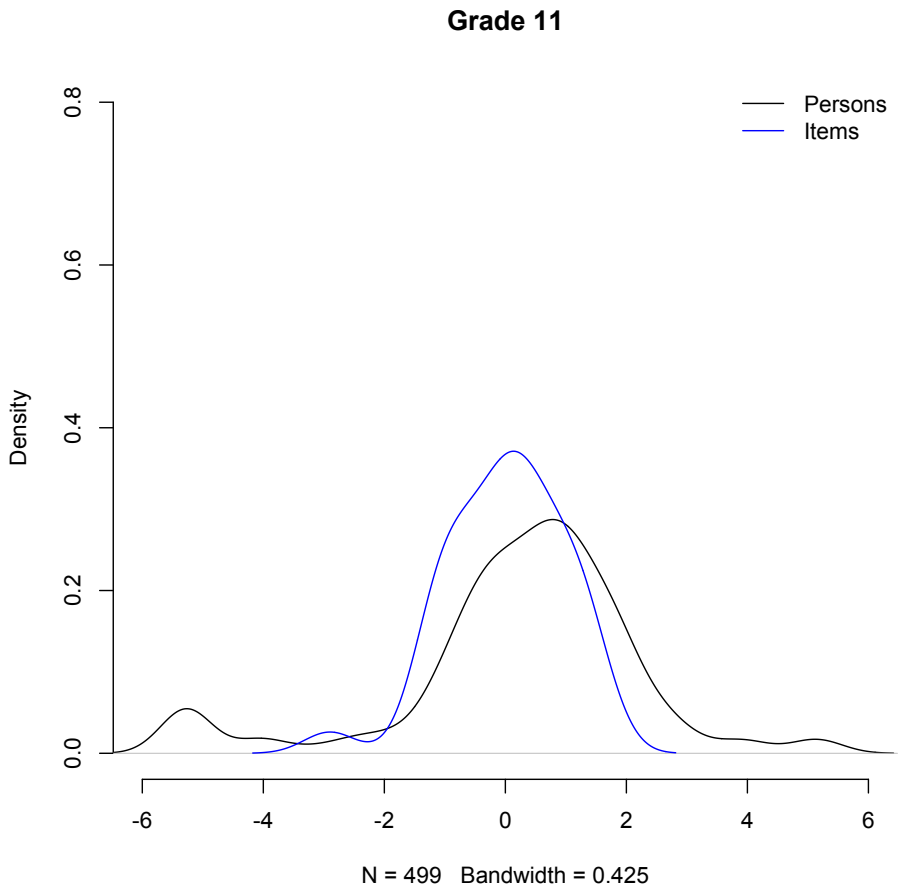




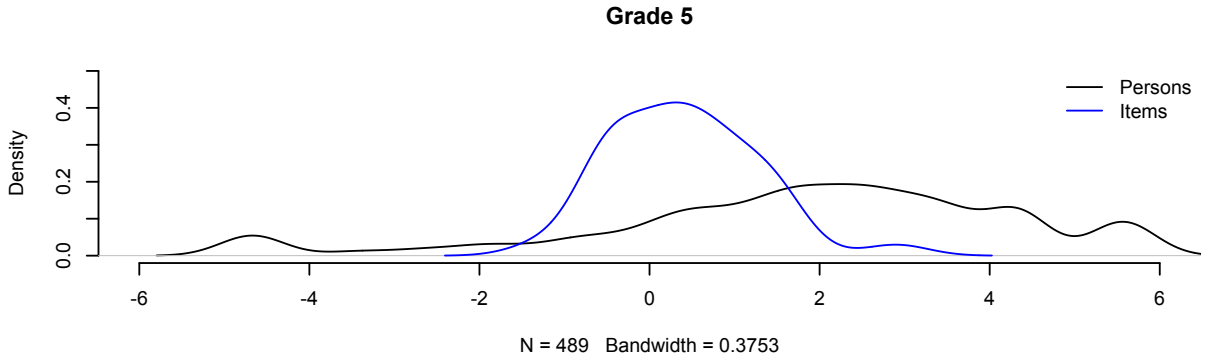


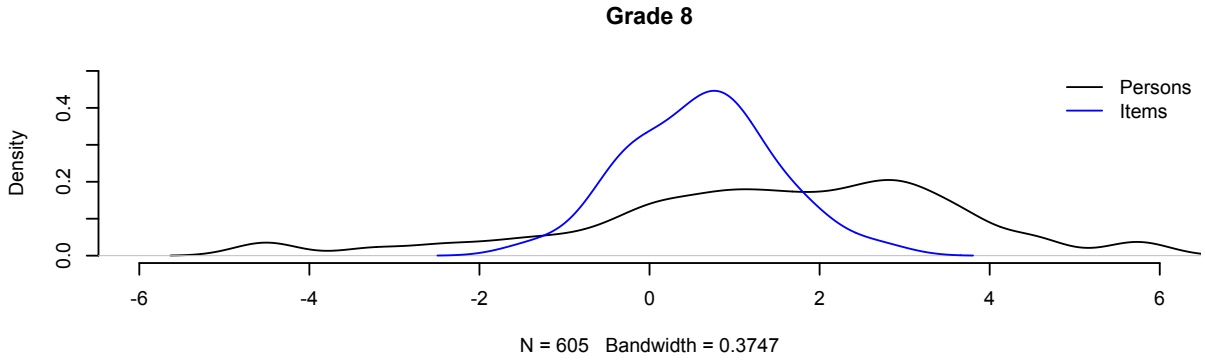


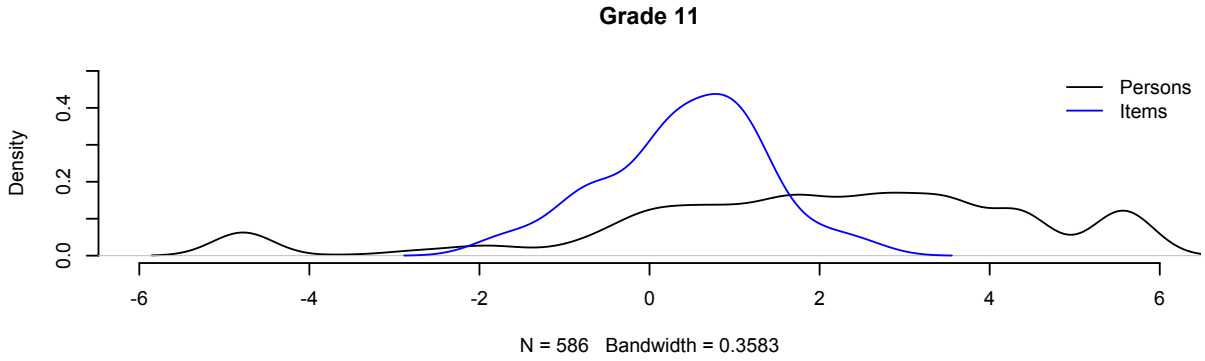



Grade 11
Persons
Items
Density
0.0
0.2
0.4
0.6
0.8
-6
-4
-2
0
2
4
6
N = 499 Bandwidth = 0.425

Science Person/Item Distributions







Person Ability and Item Difficulty Tables

English language arts

	Person Ability		Item Difficulty	
Grade	Mean	SD	Mean	SD
3	1.02	0.60	0.30	0.11
4	1.27	0.60	0.41	0.11
5	1.71	0.66	0.46	0.13
6	2.01	0.76	0.61	0.13
7	2.37	0.80	0.77	0.13
8	2.12	0.68	1.08	0.12
11	2.07	0.90	0.30	0.11

Mathematics

	Person Ability		Item Difficulty	
Grade	Mean	SD	Mean	SD
3	-0.23	0.53	-0.52	0.11
4	-0.18	0.51	-0.34	0.10
5	0.13	0.48	0.26	0.11
6	0.48	0.50	0.47	0.10
7	0.53	0.53	0.49	0.10
8	0.67	0.50	0.96	0.10
11	0.24	0.54	0.00	0.12

Science

	Person Ability		Item Difficulty	
Grade	Mean	SD	Mean	SD
5	1.41	0.60	0.61	0.11
8	1.90	0.76	0.42	0.13
11	1.80	0.74	0.40	0.14

4.4 Scoring

All scoring expectations for the ORExt are established within the Administration Manual (see *Appendix 2.3*, p. 15). The scoring procedures for the new ORExt have been simplified, with students receiving a 0 for an incorrect response or a 1 for a correct response. Input from the field gathered from review of our Consequential Validity study demonstrates that the assessment scoring procedures are much more clear and easier to implement than prior scoring approaches (see *Appendix 2.3B.10*). BRT was also commissioned to develop a scaled score interpretation guide, which describes specific strategies for interpreting student test scores and sub-test scores in Reading and Writing, and Achievement Level Descriptors (ALDs) published within the Individual Student Reports (see *Appendix 6.4C*) for annual performance, growth, and as part of Essential Skills requirements for very low performing students (see *Appendix 2.1A*).

4.5 Multiple Assessment Forms

The ORExt was administered in only one grade level form for the 2015-16 school year, with 36 operational items arranged in order of empirical difficulty and 12 embedded field test items.

4.6 Multiple Versions of An Assessment

The ORExt is provided in the standard format, but is also available in Large Print and Brailled formats. Test content is identical across all three versions, with an occasional item being eliminated on the Braille version due to inaccessibility. These items do not count for or against the student in reporting. Substantive test comparability analyses are not feasible, given the small *n*-sizes of the samples involved in the alternative versions.

4.7 Technical Analyses and Ongoing Maintenance

The ORExt technical analyses that document reliability and validity are included in this technical report (see *Sections 3* and *4*, respectively). ODE and BRT staff reviews these analyses annually. Necessary adjustments to the assessment are determined prior to implementation of the subsequent year's work plan, which elaborates the areas of improvement as well as aspects of the testing program that will be maintained. This decision-making is supported by input from the field gathered from the Consequential Validity study (see *Appendix 2.3B.10*).

One noteworthy example of the impact of our system of ongoing improvement this year is the development of the Oregon Observatioanl Rating Assessment (ORora), which addresses an area of concern expressed by stakeholders during the 2014-15 Consequential Validity study. The ORora was developed to address the inclusion of a small sub-group of very complex students in the assessment process, students whose disabilities are so severe that they cannot access even the reduced complexity demands of the ORExt. The ORora results and the test administration instructions are provided in *Appendix 5.1D*.

Critical Element 5 – Inclusion of All Students

5.1 Procedures for Including SWDs

The Oregon assessment system provides explicit guidance regarding the participation of all public school students in its statewide assessment program (see *Section 1.4*).

5.1A Clear Explanations of the Differences Between Assessments

The assessment options for all public school students in Oregon are elaborated in the Oregon Test Administration Manual (see *Appendix 1.4.2*, p. 7). These options include the Smarter Balanced Assessment in English language arts and mathematics in Grades 3-8 & 11, the Oregon Assessment of Knowledge and Skills in science in Grades 5, 8, & 11, and in the same content areas and grade levels for SWSCD who take the ORExt (see *Appendix 1.4.2*, p. 92-93). Social studies assessment is a district option within the OAKS portal, as well. In addition, expectations for the English Language Proficiency Assessment (ELPA) and the Kindergarten Assessment are provided.

5.1B Eligibility Decisions Made by IEP Teams

A student's IEP team determines how a student with disabilities will participate in the Oregon Statewide Assessment program. The IEP team must address the eligibility criteria for participation in the ORExt before determining that the assessment is the appropriate option (see *Appendix 5.1B*).

5.1C Guidelines for Assessment Selection

As noted earlier, IEP teams make decisions regarding how students with disabilities participate in the Oregon statewide assessment program. At present, students participate in one of three options: (a) student takes the general assessment with or without universal tools. (b) student takes the general assessment with designated supports and/or accommodations, or (c) student takes the ORExt. Guidelines for making universal support, designated support, and accommodations decisions for the general assessments are provided in *Appendix 2.3A.1*. Guidelines for making these determinations for SWSCD who participate in AA-AAAS are provided in *Appendix 5.1B*.

5.1D Information on Accessibility Options

Information regarding accessibility options for the general assessment can be found with the general assessment Peer Review evidence. For the ORExt, accessibility is treated holistically, with universal design for assessment concepts embedded in the item design and a wide variety of accommodations also available if needed. Items are crafted to be visually simple and clean. Graphic supports, which are always black/white line drawings, are embedded in all items at the low level of complexity but are phased out as items become more complex. Items are designed to incorporate simplified language unless specific academic vocabulary and concepts is what is being tested (see *Appendix 2.3A.3*). The items on the ORExt are all selected response, with three response options allowing for multiple modes of access (e.g., saying the answer, pointing to the answer, eye gaze, switch, etc.). All text presented to students is at least 18-pt font (larger, of course, in the large print version).

Sample items are presented in *Appendix 2.2.3*. All accessibility supports, designated supports, and accommodations for the ORExt are published in *Appendix 2.3A.1*, p. 37-44.

For students who have very limited to no communication and are unable to access even the most accessible items on the ORExt, an Observational Rating Assessment (ORA) was implemented in 2015-16. The ORA is completed by teachers and documents the student's level of communication complexity (expressive and receptive), as well as level of independence in the domains of attention/joint attention and mathematics. The administration instructions and 2015-16 results for the ORora are included in *Appendix 5.1D*.

5.1E Guidance Regarding Appropriate Accommodations

Guidance regarding appropriate accommodations is published in *Appendix 2.3A.1*. District and School Test Coordinators provide annual training on test security and administration. The ORExt approaches access as part of test design, as noted above in *Section 5.1D*. The complexity of SWSCD communication systems demands such an approach. In addition, comprehensive accommodations are allowed in order to decrease the chances that a disability may interfere with our ability to measure the student's knowledge and skills.

5.1F All SWDs Eligible for the ORExt

ODE's eligibility guidelines make it clear that all SWDs are eligible for the ORExt, regardless of disability category, and that specific disability category membership should not be a determining factor for considering participation (see *Appendix 5.1B*).

5.1G Parents Informed of AA-AAAS Consequences

The Parent FAQ section of the General Administration Manual makes it clear that parents must be informed of the potential consequences of having their child assessed against alternate achievement standards, including diploma options. Parents are also informed that alternate achievement standards are designed to reflect a significant reduction in depth, breadth, and complexity and are therefore not comparable to general academic achievement standards (see *Appendix 2.3*, p. 28-32).

5.1H State Ensures ORExt Promotes Access to the General Education Curriculum

The ORExt is strongly linked to the CCSS/ORSci/NGSS, as evidenced by our linkage study results (see *Appendix 3.1A*). The claim is based on the following warrants: (a) ORExt items are aligned to the Essentialized Standards; (b) the Essentialized Standards are strongly linked to the grade level content standards; therefore (c) the ORExt items are strongly linked to grade level content expectations. It is thus expected that the ORExt promotes access to the general education curriculum by assessing general education content that has been reduced in depth, breadth, and complexity yet maintains the highest possible standard for SWSCD.

In addition, ODE commissioned BRT to work with Oregon teachers of SWSCD in the 2015-16 school year to develop a variety of curricular and instructional resources that are aligned to the Essentialized Standards. These resources include: (a) curricular templates,

(b) video tutorials, and (c) supporting documents that provide specific guidance regarding how to develop lesson plans, Present Levels of Academic and Functional Performance (PLAAFP) statements, and Individualized Education Program (IEP) goals and objectives that are aligned with the Essentialized Standards. It is also expected that the essentialization process will generalize to many students who are performing off grade level, not merely to SWSCD. All resources are published on a BRT-sponsored website at http://lms.brtprojects.org.

5.2A – 5.2C Procedures for Including ELs

In addition to the programmatic guidance provided in *Appendix 1.4A.1* related to EL program eligibility and services, ODE also provides guidance relevant to the inclusion of ELs in the statewide assessment program in *Appendix 1.4.2*. Though the ORExt is currently published in English, an appropriately qualified interpreter can provide the assessment to any SWSCD from diverse language backgrounds, including American Sign Language. ODE has developed a training module to increase the standardization of ASL administration for its statewide assessments, available at http://lms.brtprojects.org.

Additional information regarding the inclusion of ELs in Oregon's general assessments is provided in the general assessment Peer Review evidence.

5.3 Accommodations

All statewide accommodation guidance is published in the Accessibility Manual (see *Appendix 2.3A.1*), outlining the universal tools and designated supports available to all students, and accommodations, available only to students with disabilities or students served by Section 504 Plans. In addition, the manual defines the supports as embedded, where they are provided by the online test engine (e.g., calculator, text-to-speech), or non-embedded, where they must be provided by a qualified assessor (e.g., read aloud, scribe). The manual also makes it clear that these supports are content-area specific, as a universal tool in one content area may be an accommodation in another.

5.3A Appropriate Accommodations are Available for SWD/ Section 504

Appropriate accommodations for the ORExt are published in *Appendix 2.3A.1*, p. 37-44. Additional accommodations for all statewide assessments are also published in this manual. The Oregon Accommodations Panel reviews the appropriateness of the supports listed annually. Practitioners may also request the addition of an accommodation through a formal process (see Appendix E: Approval Process for New Accessibility Supports within the manual, *Appendix 2.3A.1*, p. 101-103).

5.3B Appropriate Accommodations are Available for ELs

As noted in *Sections 5.2A-C*, the ORExt is accessible in any communication modality through the use of an interpreter. Appropriate accommodations for the ORExt are published in *Appendix 2.3A.1*, p. 37-44. Additional accommodations for all statewide assessments are also published in this manual. The Oregon Accommodations Panel reviews the appropriateness of the supports listed annually. Practitioners may also request the

addition of an accommodation through a formal process (see Appendix E: Approval Process for New Accessibility Supports within the manual, *Appendix 2.3A.1*, p. 101-103).

5.3C Accommodations are Appropriate and Effective

In addition to the evidence gathered during the linkage study (see *Appendix 3.1A*), which suggests that the ORExt items were accessible and free of bias even before final editing, the appropriateness of the supports listed in *Appendix 2.3A.1* is reviewed annually by the Oregon Accommodations Panel. Practitioners may also request the addition of an accommodation through a formal process (see Appendix E: Approval Process for New Accessibility Supports within the manual, *Appendix 2.3A.1*, p. 101-103). ODE is developing a process to collect specific accommodations codes for the ORExt in order to make performance comparisons feasible. It is hoped that this process will be in place in spring 2017, though the expected *n*-sizes may not be sufficient to generate reliable interpretations of such comparisons. The consequential validity study for 2017 will include questions regarding the appropriateness of the available accommodations, as well.

5.3D Accommodations are Appropriate and Effective

ODE has a formal process stakeholders can use to request accommodations that are not already published in the Accessibility Manual (see Appendix E: Approval Process for New Accessibility Supports within the manual, *Appendix 2.3A.1*, p. 101-103).

5.4A – 5.4E Monitoring Test Administration for Special Populations

ODE monitoring of test administration in its districts and schools is elaborated within the general assessment Peer Review evidence and is therefore not addressed here.

Critical Element 6 – Academic Achievement Standards and reporting

6.1 State Adoption of Alternate Academic Achievement Standards for SWSCD

The Oregon Extended assessment (OExt), Oregon's Alternate Assessment based on Alternate Academic Achievement Standards (AA-AAAS), is part of the Oregon Statewide Assessment System. The OExt is administered to Oregon students with the most significant cognitive disabilities (SWSCD) in English language arts and mathematics in Grades 3-8 and 11. The OExt is administered in science in Grades 5, 8, & 11. The OExt links to the CCSS in English language arts and mathematics. The new OExt is dually linked to Oregon's former science standards, as well as to the NGSS. Results from the English language arts and math administrations are included in calculations of participation and performance for Annual Measureable Objectives (AMO) – a provision of the No Child Left Behind Act (NCLB). Science participation is also included as part of the Title 1 Assessment System requirements, and is administered in grades 5, 8, & 11.

The revised OExt is built upon a vertical scale in order to support reliable determinations of annual academic growth in ELA and mathematics in Grades 3-8. The complete vertical scaling plan and operational item selection decision rules are located in *Appendix 2.2.1*.

6.1A State Formally Adopted Alternate Academic Achievement Standards

The State Board of Education formally adopted the AAAS and achievement level descriptors (ALDs) on June 25, 2015 (see *Appendix 6.1A.1*). The ELA, Math, and Science AAAS, including both the ALDs and the requisite cut scores are included in *Appendix 6.1.A.2*.

6.1B State Applies AAAS to All Public School SWSCD in Tested Grades

The state applies the AAAS to all public school-served SWSCD who participate in the OExt in Grades 3-8 & 11 in English language arts and mathematics, and in Grades 5, 8, & 11 in science.

6.1C State's AAAS Include At Least Three Levels, ALDs, and Cut Scores

The alternate academic achievement standards in Oregon are composed of four levels (though only three are required). In descending order, they are (a) Level 1, (b) Level 2, (c) Level 3, and (d) Level 4. Level 1 and Level 2 performances represent proficient achievement, while the bottom two levels represent achievement that is not yet proficient. The procedures followed to develop Oregon's alternate academic achievement standards were consistent with Title 1 assessment system requirements, including the establishment of cut scores, where relevant. In order to define four levels of proficiency, Oregon set three cut scores across all subject areas: (a) to separate Level 1 from Level 2, (b) to separate Level 2 from Level 3, and, (c) to separate Level 3 from Level 4.

The alternate academic achievement standards in English language arts, mathematics, and science for the OExt, including the achievement level descriptors (ALDs) and cut scores, were established during standard setting meetings held on June 15 (science), 16 (mathematics), and 17 (English language arts).

6.2 Achievement Standard Setting

Standard Setting meetings were held at the University of Oregon in Eugene, OR on June 15, 2015 (Science), June 16, 2015 (Mathematics), and June 17, 2015 (English language arts). A total of 53 standard setters were involved in the process: 11 in Science, and 21 in both English language arts and Mathematics. Panelists were assembled in grade level teams of three, where two members were special educators and one member was a content specialist.

The panelists were highly educated. Over 90% of the panel possessed a Master's degree or higher. Fifty-seven (57%) percent of the panelists had over 11 years of teaching experience. Seventy-six percent (76%) of the panelists had some experience working with students with significant cognitive disabilities with 64% licensed as Special Educators. The majority of panel members were female (87%), from the Northwest of the state (87%), and White (83%). No panel member self-identified with Oregon's major minority population (Hispanic).

In addition to the live training during standard setting meetings, panelists were asked to complete several training requirements prior to the standard setting meetings, which oriented them to the student population of students with significant cognitive disabilities (SWSCDs), the Oregon Extended Assessment test design and history, as well as the bookmarking standard setting method. Panelists were quite confident in their preparation and final judgments, as evidenced by responses to the questions: (a) " The training helped me understand the bookmark method and how to perform my role as a standard setter." (b) "I am confident about the defensibility and appropriateness of the final recommended cut scores." and, (c) "Overall, I am confident that the standard setting procedures allowed me to use my experience and expertise to recommend cut scores for the OrExt." The hearty majority of standard setters strongly agreed with these statements, while all participants agreed.

The nine-step process implemented for these standard setting meetings was based on Hambleton & Pitoniak (2006) as reported by R.L. Brennan (*Educational Measurement*, 4th Edition, pp. 433-470). Standard setting evaluation questions posed to participants were adapted from Cizek's *Setting Performance Standards* (2012). Standard setters set cut scores and recommended Achievement Level Descriptors (ALDs) for the Oregon State Board of Education to consider. The cut scores were articulated to reflect vertical development, or at least maintenance, of expectations across grades in a manner that respected standard setter judgments to the greatest possible degree. Six changes were made in ELA and Mathematics. Science is not built upon a vertical scale, so no cut score adjustments were necessary in Science. The cut scores are listed below.

English language arts (ELA)

Grade	Level 1	Level 2	Level 3	Level 4
3	191 or below	192 - 212	213 - 227	228 or above
4	199 or below	200 - 212	213 - 227	228 or above
5	201 or below	202 - 219	220 - 231	232 or above
6	204 or below	205 - 219	220 - 232	233 or above
7	207 or below	208 - 221	222 - 235	236 or above
8	212 or below	213 - 223	224 - 235	236 or above
11	898 or below	899 - 919	920 - 926	927 or above

Mathematics

Grade	Level 1	Level 2	Level 3	Level 4
3	191 or below	192 - 200	201 - 217	218 or above
4	192 or below	193 - 205	206 - 218	219 or above
5	192 or below	193 - 205	206 - 219	220 or above
6	203 or below	204 - 207	208 - 221	222 or above
7	206 or below	207 - 208	209 - 222	223 or above
8	207 or below	208 - 211	212 - 225	226 or above
11	900 or below	901 - 906	907 - 921	922 or above

Science

Grade	Level 1	Level 2	Level 3	Level 4
5	505 or below	506 - 516	517 - 529	530 or above
8	809 or below	810 - 819	820 - 830	831 or above
11	900 or below	901 - 913	914 - 928	929 or above

Note: The ELA and Math vertical scales for the OrExt are centered on 200 in grades 3-8 and can be used to document year-to-year growth. None of the other scales should be used for longitudinal comparisons. All Grade 11 scales are independent and centered on 900. The grade 5 Science scale is independent and centered on 500, while the Grade 8 Science scale is independent and centered on 800. An independent auditor evaluated the bookmarking standard setting process. The auditor's comprehensive report can be found in *Appendix 6.2.2.*

6.3 Challenging and Aligned Academic Achievement Standards

Oregon educators initially evaluated new Oregon *Essentialized Assessment Frameworks* in two respects. First, educators were asked to determine the appropriateness of the standards selected for inclusion and exclusion in the *Essentialized Standards* (yes/no). Second, the level of linkage between the *Essentialized Standards* and grade level content standard was evaluated (0 = no link, 1 = sufficient link, 2 = strong link). Summary results are provided in the tables below. A comprehensive essentialized standard to grade level

standard linkage study, as well as essentialized standard to item alignment study, is provided in *Appendix 3.1A.*

English language arts

Grade	# Essentialized Standards	# Raters	Ave. Linkage Rating (0-2)*	Ave. Agreement with Essentialization (0-6)*
3	27 (38)	6	1.74 (10)	5.68 (21)
4	30 (40)	6	1.78 (15)	5.77 (25)
5	28 (39)	6	1.73 (12)	5.79 (23)
6	25 (37)	6	1.80 (12)	5.76 (19)
7	24 (36)	6	1.77 (10)	5.79 (19)
8	25 (36)	6	1.79 (12)	5.80 (21)
11	24 (36)	6	1.82 (12)	5.79 (19)

Note. * Count of perfect ratings/agreement across all raters (in parenthetical) relative to number of essentialized standards.

Mathematics

Grade	# Essentialized Standards	# Raters	Ave. Linkage Rating (0-2)*	Ave. Agreement with Essentialization (0-3)*
3	22 (33)	3	2.00 (22)	2.77 (17)
4	26 (34)	3	1.99 (25)	2.81 (21)
5	23 (34)	3	1.99 (22)	2.78 (18)
6	27 (41)	3	1.98 (21)	2.68 (15)
7	20 (36)	3	1.95 (17)	2.90 (18)
8	19 (33)	3	1.96 (17)	2.37 (7)
11	23 (179)	3	2.00 (23)	2.52 (12)

Note. * Count of perfect ratings/agreement across all raters (in parenthetical) relative to number of essentialized standards.

Science

Grade	# Essentialized Standards	# Raters	Ave. Linkage Rating (0-2)*	Ave. Agreement with Essentialization (0-4)*
5	15 (16)	4	1.92 (10)	3.93 (14)
8	24 (59)	4	1.97 (21)	4.00 (24)
11	24 (71)	4	1.98 (22)	3.83 (20)

Note. * Count of perfect ratings/agreement across all raters (in parenthetical) relative to number of essentialized standards.

6.4 Reporting

Oregon's reporting system facilitates appropriate, credible, and defensible interpretation and use of its assessment data. With regard to the ORExt, the purpose is to provide the state technically adequate student performance data to ascertain proficiency on grade level state content standards for students with significant cognitive disabilities (see *Sections 3* and *4*). In addition, the state makes it clear that results from the Oregon Extended are not comparable to results from the SBA/OAKS (see *Appendix 2.3,* p. 28-32). Nevertheless, the test meets rigorous reliability expectations (see *Section 4.1*). Validity is considered here as an overarching summation of the Oregon Extended assessment system, as well as the mechanisms that Oregon uses to continuously improve the ORExt assessment (see *Appendix 2.3B.10*).

6.4A Public Reporting

Oregon reports participation and assessment results for all students and for each of the required subgroups in its reports at the school, district, and state levels. The state does not report subgroup results when these results would reveal personally identifiable information about an individual student. The calculation rule followed is that the number of students in the subgroup must meet the minimum cell size requirement for each AMO decision: participation, achievement in English language arts and math, attendance, and graduation, where appropriate (see *Appendix 2.6C*)

6.4B State Reports Interpretable Results

Oregon develops and disseminates individual student data upon final determination of accuracy. The state provides districts with individual student reports (ISRs) that meet most relevant requirements. The state incorporated the Standard Error of Measure (SEM) for each student score into the report templates. The SEM associated with each cut score is provided in *Section 4.1B*. Also, see the mock-up ISR in *Appendix 6.4C*.

6.4C1 – C5 State Provides Individual Student Reports

Oregon's student reports provide valid and reliable information regarding achievement on the assessments relative to the AAS. The reliability of the data is addressed in *Section 4.1.* Validity is considered here as an overarching summation of the Oregon Extended assessment system, as well as the mechanisms that Oregon uses to continuously improve the Oregon Extended assessment. The ISRs clearly demonstrate the students' scale score relative the AAAS that is relevant for that content area and grade level (see *Section 4.4* and *Appendix 6.4C*). The Oregon ISRs provide information for parents, teachers, and administrators to help them understand and address a student's academic needs. These reports are displayed in a simple format that is easy for stakeholders to understand. District representatives can translate results for parents as necessary. Scaled score interpretation guidance is published in *Appendix 2.1A*.

Conclusions and Next Steps

In sum, the rigor of the procedural development of the ORExt, including item development, item content and bias reviews, item selection based upon item characteristics, test form reliabilities, and correlation results were quite good and provide important validity evidence. The test development process adhered to procedural guidelines defined by the AERA/APA/NCME *Standards for Educational and Psychological Testing* (2014), as well as incorporating procedures that are known in the field to be best practice. For example, an independent auditor evaluated Standard Setting. In addition, the ORExt reflects what highly qualified Oregon educators believe represents the highest professional standards for the population of students with significant cognitive disabilities.

The test reliabilities for the ORExt were quite high, suggesting that the assessment items functioned consistently with the test as a whole. The correlations between students' content scores across subjects were not overly strong, implying that each test measures a distinct construct. The classification consistency analyses demonstrate that the ORExt is appropriately categorizing students into the proficient category, and capable of doing so in a consistent manner. The vertical scale developed in 2014-15 appears to be modeling incremental growth across Grades 3-8 in ELA and mathematics, as intended. The Grade 7 mathematics test demonstrated insufficient item difficulties across the range of low, medium, and high item complexity, however, and must be amended in the 2016-17 school year. The ELA and science assessments could benefit from the addition of more difficult items, as evidenced by comparisons of the average person abilities and item difficulties. Mathematics assessments appear to be functioning quite well in terms of person abilities and item difficulties. Now that the tested population has been more appropriately targeted, adjustments to item difficulties should have the desired impact of providing an appropriate level of challenge. As the administration of the ORExt continues, further documentation of validity will be gathered according to the plan presented below.

The Oregon Observational Rating Assessment (ORora) results demonstrate that approximately 20% of the SWSCD who participated in the ORExt also took the ORora. A total of 836 students were administered the ORora in the 2015-16 test administration. The participants were primarily students with multiple, severe disabilities with very limited communication systems. Analyses of missing data patterns for the ORExt demonstrate that QAs were able to adhere to the discontinuation rules. However, the percentages and input from stakeholders gathered during the administration suggest that assessors used the ORora in many cases where use was not required. They did this because the information was deemed valuable, in addition to that provided by the ORExt.

The 2015-16 Oregon Consequential Validity study provides important information for future administrations of the ORExt. The results demonstrate that the test continues to be easy to administer and score and is providing an accessible and appropriate representation of the knowledge and skills that should be required of SWSCD in Oregon. Study results demonstrate that approximately 29% of the SWSCD who participate in the ORExt come

from culturally and linguistically diverse backgrounds. Of this group, less than 5% are better able to use a language other than English for academic purposes. The survey results established that test administration for each content area takes up to one hour for most students. Results also suggest that students with Orthopedic Impairments (OI) are much more complex than the mere label suggests. Some sentiments shared regarding writing items suggests that training must address the various accommodations that are available on the ORExt, as a few QAs felt that the writing items were inaccessible because the student could not hold a pencil.

The 2015-16 Oregon Extended Assessment Pilot Tablet Administration demonstrates that Oregon teachers highly value the development of a tablet-based administration of the ORExt, for student engagement, teacher facility, and standardization. The results also suggest that the writing items require additional technological supports, such as the ability to scale text and erase input. In addition, the audio files must be adapted such that pace can be adapted (slower/faster). Focus Group members also recommended that practice items be developed in a tablet format so qualified assessors and students can practice with the tablet administration in preparation for the ORExt test window.

Documenting evidence of validity remains an ongoing and continuous process. Our efforts to continue to improve the assessment system are outlined in *Section 4.5* above. We also have studies planned over the course of the next three years that will help to solidify the evidence that is accumulating. All of the evidence we have at hand suggests that the ORExt is sufficient to its stated purpose of providing reliable determinations of student proficiency at the test level in order to support systems level analysis of district and state programs. The ORExt will hopefully continue to improve over time due to field-testing and constant monitoring and review, and additional validity evidence will be gathered.

As mentioned above in *Section 3.1A*, data are presented to support the claim that Oregon's AA-AAAS provides the state technically adequate student performance data to ascertain proficiency on grade level state content standards for students with significant cognitive disabilities – which is its defined purpose. In this technical report, we have provided content validity evidence related to the ORExt test development process (i.e., essentialization process, linkage study, distributed item review, test blueprint, item writer training and demographics, and item reviewer training and demographics), ORExt test reliability evidence, and ORExt consequential validity evidence. Further analyses over the coming years are planned to continue the development of technical documentation for overall construct validity of the ORExt. The technical documentation plan for the 2016 through 2019 school years is provided below:

Documentation Description	Anticipated Timeframe	Outcome
Follow-Up Independent Linkage/Alignment Study	Fall 2016	Independent contractor will provide comprehensive and representational documentation of linkage of Essentialized Standards (ES) to Grade Level standards and alignment of ORExt items to ES, as well as alignment of ALDs to ES.
ORExt Tablet Administration Pilot, Phase 2	Spring 2017	Focus on database communications between ODE and BRT to support the implementation of a successful statewide tablet administration in spring 2019.
DIF Analyses	Spring 2017	Provide documentation of item functioning that can be used to ensure that items are not biased toward student groups (i.e., based on gender and ethnicity).
Scoring Accuracy Study	Spring 2018	Provide inter-rater reliability documentation for the ORExt.
Accommodations Study	Spring 2019	Provide documentation related to the impact of accommodated test administration for the ORExt.
ORExt Eligibility Study	Fall 2019	Provide documentation of the consistency of IEP team decision making with the established ODE guidelines for ORExt eligibility.
ORExt Field Testing	*Ongoing*	Test and item characteristics are reviewed annually, with operational items that are not functioning as intended replaced by field test items that are functioning properly.

Appendix Table and Descriptions

Appendix Table

Topic	File Name
Essentialized Assessment Frameworks User Guide	App1.1_EAF_UserGuide
Independent review of rigor of CCSS standards	App1.2_FordhamCCSSReview
Memo from the Oregon Governor regarding parent opt-out of testing expectations	App1.4.1_ExecutiveMemo
ODE Test Administration Manual (TAM)	App1.4.2_TAM2015_16
ODE English Learner Program Guide	App1.4A.1_ODE_ELProgramGuide
Oregon regulation regarding language of assessment	App1.4A.2_OAR581_022_0102
Oregon's annual report to the state legislature regarding Oregon's public schools	App1.5_ORStateReportCard
ORExt Test Specifications	App2.1_ORExtTestSpecs
ORExt Scaled Score Interpretation Guide	App2.1A_ORExtScaleScoreInterp
ORExt Test Blueprint	App2.1B_ORExt_TestBlueprint_2016
ORExt Item Development Process Manual	App2.1C_ORExt_ItemDevt_Process
BRT Item Writer Training PPT Slides	App2.2.1_ItemWriter_Training
ORExt Vertical Scaling Plan	App2.2.2_ORExtVertScale
ORExt Sample Items	App2.2.3_ORExtSampleItems
ORExt Administration Manual	App2.3_ExAssessAdminMan2015_16
ODE Accessibility Manual	App2.3A.1_ODEAccessibilityMan2015_16
ODE Accommodations Guide	App2.3A.2_OSAAccomGuide
BRT Procedures for Reducing Depth, Breadth, and Complexity of Items	App2.3A.3_ORExtRDBC
ODE Qualified Trainer PPT Slides	App2.3B.1_QATraining2015_16
BRT Qualified Trainer PPT Slides	App2.3B.2_ORExtendQTTrng2015_16
2015-16 vs. 2014-15 OREXt Similarities and Differences Tables	App2.3B.3_ORExt_SameDif2015_16

Topic	File Name
ODE Statewide Assessment Test Schedule	App2.3B.4_TestSchd_2015_16
Qualified Trainer Suggested Agenda	App2.3B.5_QT_Training_Agenda2015_16
How to Access the or.k12test.com Website	App2.3B.6_ExtAssessAccessInstr2015_16
Qualified Trainer Responsibilities	App2.3B.7_Trainer Responsibilities2015_16
Frequently Asked Questions Document	App2.3B.8_ORExtFAQ2015_16
Help Desk Technical Assistance Log	App2.3B.9_HelpDeskLog2015_16
Consequential Validity Study Report	App2.3B.10_ORExtCVStudy2015_16
ORExt Pilot Tablet Study Report – Phase 1	App2.3C_ORExtPilotTabletStudyReport
Data Entry Guide	App2.6B_G3-8_11_DataEntryGuide2015_16
Report Card Rating Policy Technical Manual	App2.6C_ReportCardRating_PolicyTechManual
ORExt Item Linkage, Bias, and Sensitivity Study Report	App3.1A_DIRLinkageStudy
Overview of the Distributed Item Review website used for conducting alignment, bias, and sensitivity reviews	App3.1B_DIR_SystemIntro
History of the ORExt Document	App4.1B_HistoryORExt_2016
Report conveying the ethnic diversity of Oregon's students for the 2014-15 school year	App4.2_FallMembershipReport
ODE Eligibility Guidance for the ORExt	App5.1B_ORExtEligibilityGuidance
Results from the 2015-16 administration of the ORora and administration and scoring instructions.	App5.1D _OroraResults2015_16
Agenda and minutes from the State Board adoption of the AAAS for the ORExt	App6.1A.1StateBoardAAASAdoption
All AAAS for the ORExt, including ALDs and cutscores	App6.1A.2_ORExtAAS
Standard Setting Training PPT Slides	App6.2.1_ORExt_Standard_Setting
Independent Standard Setting Audit Report	App6.2.2_ORExtSSAuditRept
ORExt Individual Student Report Mock-Up	App6.4C_ORExt_ISR

Appendix Descriptions

Appendix 1.1

Appendix 1.1 explains the development process and intended uses for the Essentialized Assessment Frameworks (EAFs). The EAFs are the essentialized standards, which are linked to grade level content standards. The ORExt is aligned to the EAFs, as well. While the EAFs primarily guide item development, they are also intended to be used in the development of appropriate Present Levels of Functional and Academic Performance (PLAAFP) statements and Individualized Education Program (IEP) goals and objectives.

Appendix 1.2

Appendix 1.2 conveys the evaluation conducted by researchers at the Fordham Institute, which compared then-current state standards to the CCSS in terms of rigor. The findings generally show that the CCSS are as rigorous or more rigorous than state standards.

Appendix 1.4.1

Appendix 1.4.1 is the Executive Memo from the Governor of Oregon regarding parent opt-out expectations.

Appendix 1.4.2

Appendix 1.4.2 is the test administration manual (TAM) for all assessments in the Oregon statewide assessment system, including the SBA, OAKS, the ORExt, the Kindergarten Assessment, and the ELPA. The TAM elaborates all relevant test security and administration procedures.

Appendix 1.4A.1

Appendix 1.4A.1 is ODE's English Learner Program Guide, outlining English learner (EL) system requirements in the areas of student identification, services, reporting, and assessment for ELs in Oregon's public schools, including ELs who are SWD.

Appendix 1.4A.2

Appendix 1.4A.2 is Oregon’s regulations that require ODE to provide translated OAKS assessments for populations at or above 9% in grades K-12 within three years after the school year in which the language exceeds the threshold.

Appendix 1.5

Appendix 1.5 is Oregon’s annual report to the state legislature for the 2014-15 school year. The report includes student demographics and information on student groups, school funding and staff information, test results, graduation and drop out rates, charter school data and information on alternative education programs, early childhood data, and attendance and chronic absenteeism data.

Appendix 2.1

Appendix 2.1 is the test specifications document that describes our approach to assessment and test design for the ORExt. The document includes our approach to RDBC, an overview of the essentialization process and EAF documents, the anticipated operational test design for the ORExt, test development considerations, sample test items, item specifications, and universal tools/designated supports/accommodations.

Appendix 2.1A

Appendix 2.1A provides the field with comprehensive information related to scaled score interpretation for the ORExt. The guidance is published in three main areas: 1) Annual performance, 2) Annual growth, and 3) Performance for very low functioning students. Guidance regarding use and interpretation of reading and writing sub-scores is also provided.

Appendix 2.1B

Appendix 2.1B is the test blueprint for the ORExt, conveying the balance of representation of domains across the content areas and grade levels assessed. Operational items are selected to reflect the representation percentages included in the test blueprint.

Appendix 2.1C

Appendix 2.1C describes the eight-step item development process used to develop items for the ORExt, from standard selection to test booklet formation. The item development process is specific and explicit in order to increase transparency.

Appendix 2.2.1

Appendix 2.2.1 is the set of PPT slides that were used to train item writers for the ORExt. Item writers were also provided an orientation to the test specifications as part of training.

Appendix 2.2.2

Appendix 2.2.2 is a document that summarizes the balanced design vertical scaling plan employed for the ORExt in the 2014-15 administration. The document includes the domain sampling plan for all assessments, as well as the decision rules employed to remove items from the operational item pool prior to vertical scaling and standard setting procedures.

Appendix 2.2.3

Appendix 2.2.3 provides stakeholders with visual representation of the structure of the ORExt. Sample items are conveyed in English language arts, mathematics, and science, with the scoring protocol and student materials presented together. Stakeholders can see the structure of each item, as well as how the items are scored. They can also gather an idea about the types of formats that are used for answer choices that are included within the student materials documents.

Appendix 2.3

Appendix 2.3 is ODE's General Administration and Scoring Manual for 2015-16. The manual establishes ODE's expectations regarding the test window, utilizing the ORExt training and proficiency website, using the sign language interpreter training and proficiency website, and informing parents. It also provides the following information for stakeholders, including educators and parents:

- Overview of the Extended Assessments
- Assessing a Student
- Scoring
- Decision Making
- Information for Teachers.

The manual provides three appendices that provide guidance regarding the provision of supports, parent questions and answers, and a glossary.

Appendix 2.3A.1

Appendix 2.3A.1 is the 2015-16 accessibility options manual for all assessments in the Oregon statewide assessment system, including the SBA, OAKS, the ORExt, and the ELPA. Options include Universal Tools, Designated Supports, and Accommodations. The manual provides guidance regarding use of these options in instruction and assessment, as well as implementation strategies and use evaluation. Each accommodation is coded for use in data analysis related to assessment scores for the SBA and OAKS.

Appendix 2.3A.2

Appendix 2.3A.2 is ODE's How to Select, Administer, and Evaluate Accommodations on Oregon's Statewide Assessment manual for 2013-14. The manual trains users regarding how to implement and evaluate appropriate accommodations, from the student level to the systems level.

Appendix 2.3A.3

Appendix 2.3A.3 is a document that summarizes the procedures used during item development to reduce item depth, breadth, and complexity, in addition to the test specifications information found in Appendix 2.1. The document also provides more detail regarding how language complexity is addressed and reviewed in an effort to decrease the language load of items and make the test more accessible to all students. The document also discusses ways in which bias is addressed during test development.

Appendices 2.3B.1-2.3B.2

Appendices 2.3B.1 and 2.3B.2 are the PowerPoint (PPT) trainings that were used by ODE and BRT trainers to train new qualified assessors (QAs) and qualified trainers (QTs) in four regionally hosted webinar trainings in November 2015. QTs also used the package to train New Qualified Assessors for the 2015-16 school year. The training provides participants with the information needed to pass proficiency tests as part of the requirements to become a QA for the Oregon Extended Assessments and was delivered by QTs throughout the state. The training package addresses the following topics:

- What's new in 2015-16
- 2016 Test Window
- Eligibility – which students take AA-AAAS?
- Test administration
- Student Confidentiality & Test Security
- Test Administration (Physical & Logistic)
- Scoring & Data Entry
- Reports & Sharing Results with Parents
- Navigating the Training and Proficiency website
- Resources

Appendix 2.3B.3

Appendix 2.3B.3 is a document that provides a comparison of the prior year's Oregon Extended Assessment to the current Oregon Extended Assessment, using a table with what has remained the same and what has changed.

Appendix 2.3B.4

Appendix 2.3B.4 is the test calendar for the entire Oregon statewide assessment program, including the SBA, OAKS, the ORExt, the ELPA, the Kindergarten Assessment, and the NAEP.

Appendix 2.3B.5

Appendix 2.3B.5 is a sample agenda that ODE makes available to QTs around the state to train their respective new QAs as they implement the train-the-trainers model used by the Oregon Extended assessment.

Appendix 2.3B.6

Appendix 2.3B.6 is the list of instructions provided to new QAs and QTs regarding how to access the online training and proficiency website.

Appendix 2.3B.7

Appendix 2.3B.7 is the list of responsibilities associated with being a QT for the ORExt assessment.

Appendix 2.3B.8

Appendix 2.3B.8 is the document that contains the most commonly fielded questions and answers from stakeholders, including parents and teachers.

Appendix 2.3B.9

Appendix 2.3B.9 is the report that summarizes all of the technical assistance questions garnered from the field this year. Efforts are made to find any patterns that our team may use to improve training for the following year.

Appendix 2.3B.10

Appendix 2.3B.10 is the consequential validity report for the spring 2016 consequential validity study conducted by BRT. The report provides documentation of the perceptions in the field related to both intended and unintended academic and social consequences of the ORExt.

Appendix 2.3C

Appendix 2.3C is the ORExt Pilot Tablet Administration report for the spring 2016 tablet administration, Phase 1, study conducted by BRT. The report provides the research plan, summaries of results, and lessons learned regarding how to approach Phase 2 and Phase 3 of the planned study.

Appendices 2.6

Appendix 2.6 is the guidance that ODE has provided to assessors to walk them through the online data entry process for the ORExt on the secure ODE District Secure website.

Appendix 2.6C

Appendix 2.6C is the manual defining the state of Oregon's policies and procedures regarding how students are included in AMO reporting, including how achievement, growth, and graduation rates are reported for student groups and subgroups.

Appendix 3.1A

Appendix 3.1A is a document that summarizes the process and participants used to review the linkage between the Essentialized Standards and grade level content standards (CCSS in ELA and Math; ORSci and NGSS in Science), as well as the alignment between test items for the ORExt with those Essentialized Standards. In addition, reviewers rated the items for potential bias and access concerns. All data was gathered using the Distributed Item Review (DIR) website, supported by a webinar training and ongoing technical assistance.

Appendix 3.1B

Appendix 3.1B is a document that describes the Distributed Item Review (DIR) website used by Oregon teachers to evaluate the alignment between test items for the ORExt with Essentialized Standards. In addition, reviewers rated the items for potential bias and access concerns. All data was gathered using the DIR website, supported by a webinar training and ongoing technical assistance.

Appendix 4.1B

Appendix 4.1B conveys the historical development of the ORExt from 1999 to the present, including the grade levels/bands assessed, content areas assessed, and the targeted content standards.

Appendix 4.2

Appendix 4.2 includes the most current published state level data regarding Oregon's ethnic diversity.

Appendix 5.1B

Appendix 5.1B is the revised and rigorous guidance that ODE has provided to IEP teams to assist them in making appropriate assessment eligibility determinations for students with disabilities.

Appendix 5.1D

Appendix 5.1D includes a summary report of the statewide results and the administraiton and scoring instructions for the new Oregon Observational Rating Assessment (ORora). The ORora is adminsitered to all students whose ORExt testing was discontinued. It provides information regarding student progress in terms of functional skills in adaptive and communication domains for the small subgroup of students who are unable to meet the academic expectations in the ORExt.

Appendix 6.1A.1

Appendix 6.1A.1 is the agenda and minutes that document the hearing and adoption of the AAAS for the ORExt on June 25, 2015.

Appendix 6.1A.2

Appendix 6.1A.2 includes all of the achievement level descriptors (ALDs) and cutscores that define performance for the ORExt in qualitative and quantitative fashions, respectively. These Alternate Academic Achievement Standards (AAAS) describe what students should know and be able to do based upon their performance on the ORExt.

Appendix 6.2.1

Appendix 6.2.1 is the PPT slides used to train standard setters during the June 2015 standard setting meetings for ELA, math, and science.

Appendix 6.2.2

Appendix 6.2.2 is a standard setting report generated by an independent auditor. The report provides a comprehensive evaluation of the bookmark standard setting procedure employed for the ORExt on June 15-17, 2015.

Appendix 6.4C

Appendix 6.4C is a document that displays the individual student report (ISR) that ODE publishes for students who participate in the ORExt. The mock-up includes cut scores and achievement level descriptors (ALDs), as well as links to the ODE website for additional information.