Oregon Extended Assessment Technical Documentation

Brock Rowley Sevrina Tindal Christopher M. Loan
Gerald Tindal

8/31/23

OREGON DEPARTMENT OF EDUCATION
Oregon achieves . . . together!

Equity, Accountability, Excellence & Integrity

It is the policy of the State Board of Education and a priority of the Oregon Department of Education that there will be no discrimination or harassment on the grounds of race, color, religion, sex, sexual orientation, national origin, age or disability in any educational programs, activities or employment. Persons having questions about equal opportunity and nondiscrimination should contact the Deputy Superintendent of Public Instruction with the Oregon Department of Education.

OREGON DEPARTMENT OF EDUCATION
255 Capitol Street NE
Salem, OR 97310
(503) 947-5600
http://www.oregon.gov/ode

Dr. Charlene Williams
Director of the Oregon Department of Education

Dan Farley
Assistant Superintendent Office of Research, Assessment, Data, Accountability, and Reporting

Test Administration Manual Contacts		
Andrea Lockard Director, Assessment & Student Reporting	*Holly Carter* Assessment Operations and Policy Analyst	*Crys Plattner* Administrative Specialist
Jon Wiens Director, Accountability Reporting	*Mason Rivers* Special Education Assessment Specialist	*Mohammed Zaidan* Administrative Specialist
Tony Bertrand English Language Arts and Social Sciences Assessment Specialist	*Mariela Salas Bao* Science Assessment Specialist	*Sody Puth Fearn* K-2 Balanced Assessment Specialist
Andrew Byerley Mathematics Assessment Specialist	*Beth LaDuca* NAEP State Coordinator	*Ben Wolcott* English Language Proficiency Assessment Specialist

Table of contents

Introduction

It is the policy of the State Board of Education and a priority of the Oregon Department of Education that there will be no discrimination or harassment on the grounds of race, color, religion, sex, sexual orientation, national origin, age or disability in any educational programs, activities or employment. Persons having questions about equal opportunity and nondiscrimination should contact the Deputy Superintendent of Public Instruction with the Oregon Department of Education.

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.

Peer Review Critical Elements Reference Tables

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

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: Test reliability of the State's assessments estimated for its student population; Overall and conditional standard error of measurement of the State's assessments; 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 assessments and their uses, 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 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 Essentialized Standards (EsSt) that are part of comprehensive Essentialized Assessment Frameworks (EAFs) that were written at three levels of complexity (low, medium, and high). The EsSt have been linked to grade level content and expectations, but systematically reduced in terms of depth, breadth, and complexity (RDBC).

A statewide sample of 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. In addition, Achievement Level Descriptors (ALDs) were also reviewed for alignment to the EsSt. See Sections 1.1, 1.2, 6.1, and 6.3 for additional information related to the comprehensive grade level standards to EsSt linkage, as well as alignment of items to the EsSt.

The ORExt 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 recorded exactly as given (A,B,or C) and not manually scored by assessors.

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 Alternate Academic Achievement Standards (AAAAS) 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 Measurable Objective (AMO) reports were finalized on July 10, 2015.

Though an alignment study was conducted in the fall of 2014 as described above, Non-Regulatory Guidance from the U.S. Department of Education, published on September 25, 2015, included an expectation that all alignment studies must be independent (see Critical Element 3.1). An independent contractor, Dr. Dianna Carrizales, was therefore hired to perform an additional alignment study in the spring of 2017.

A two year pilot tablet study was conducted in the 2015-2016 and 2016-17 school years. Over the two year study, 26 students were administered all subject areas of the ORExt in tablet format in grades 5, 8, and 11. The 2017-18 school year marked the first year the ORExt was available in tablet/online format for all grades in all subject areas.

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. 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. 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 newly adopted academic content standards were then reduced in depth, breadth, and complexity through a process called essentialization. The new Essentialized Assessment Frameworks (EAFs) were then used for item writing for the ORExt. The tables below provide examples of essentialized standards in grades 5, 8, & 11 in the subject areas of English language arts (ELA), mathematics, and science. In the right column are designations for estimated difficulty of an item: L (low), M (medium), and H (high). More information on the essentialization process can be found in section 1.2. See the EAF User Guide for a User Guide that explains the development process and intended uses for the EAFs. A selection (grade level sample) of the guide below.

Table 1.1: Grade 5

Area	Cluster	Standard	Sub-Standard	Essentialized Standard	L/M/H Descriptors
Reading Standards for Literature K-5	Key Ideas and Details	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 story read to student	L- Sentence of 7 words or less 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	Understand the place value system	Recognize that in a multi-digit number, a digit in one place represents 10 times as much as it represents in the place ot its right and 1/10 of what is 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.
Science	Matter and Its Interactions	NGSS Standard: Measure and graph quantities to provie evidence that regardless of the type of change that occurs when heating, cooling, or mising substances , the total weight of matter is conserved	OR Science Standards: 5.35.1. 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 ovservations 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 common objects in various phases of matter using graphs and data

Note:
The science esssentialized 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.

Table 1.2: Grade 8

Area	Cluster	Standard	Sub-Standard	Essentialized Standard	L/M/H Descriptors
Reading Standards for Literature K-5	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 the story read to student	L - 3 sentences that contain 2 charaters 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

Table 1.2: Grade 8 *(continued)*

Area	Cluster	Standard	Sub-Standard	Essentialized Standard	L/M/H Descriptors
Math	Statistics & Probability	1. Investigate patterns of associations 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 I ntercept. For example, in a linear model for a biology experiment, interpret a slope of 1.5 cm/hr as a 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 investigation to determin the relationship 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	ORScience 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 an dcold (e.g., objects, outside); M - Recognize that hot and cold are related to measures of temperature, including changes in temperature; H - identify exmples of heat transfer, and how such transfer might be minimized/maximized (e.g., wearing a coat to stay warm)

Table 1.3: Grade 11

Area	Cluster	Standard	Sub-Standard	Essentialized Standard	L/M/H Descriptors
Reading Standards for Literature K-5	2. Craft and Structure	4. Determine the meaning of words and phrases as they are used in 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 (including 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.

Table 1.3: Grade 11 *(continued)*

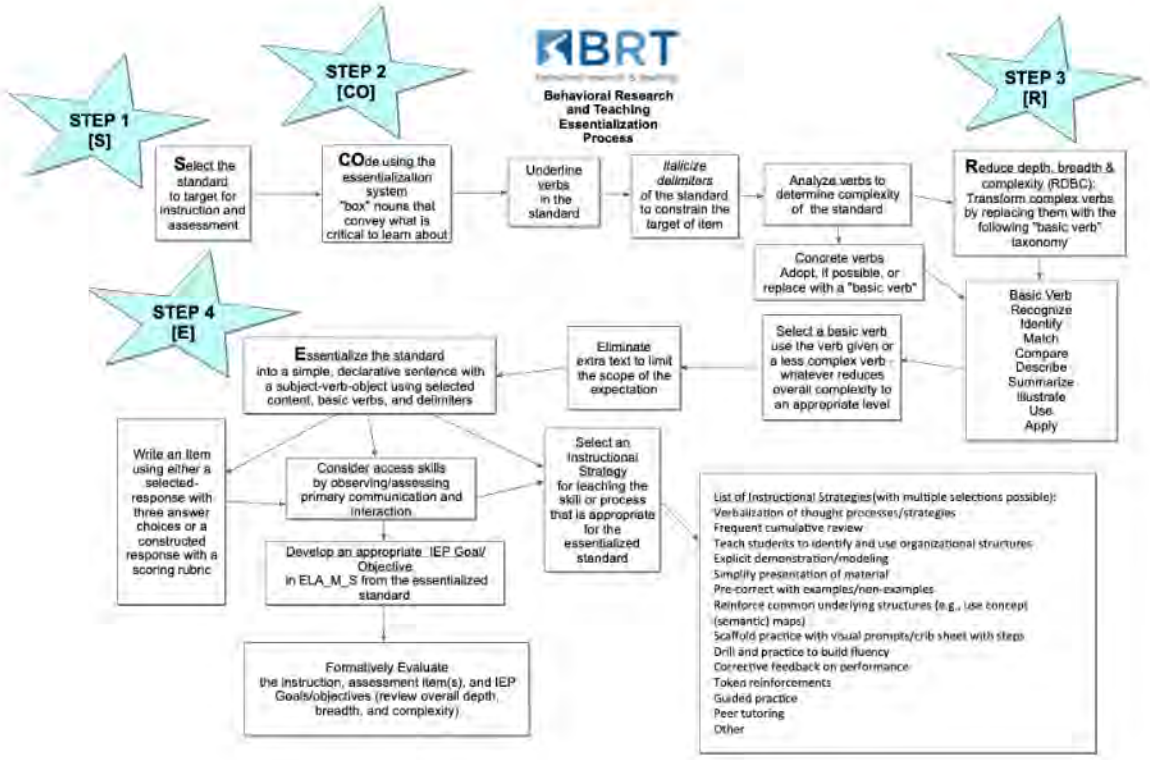
Area	Cluster	Standard	Sub-Standard	Essentialized Standard	L/M/H Descriptors
Math	Expressing Geometric Properties with Equations	2. Use coordinates to prove simple geometric theorems algebraically	7. Use coordinates to compute perimeters of polygons and areas of triangles and rectangles, e.g., using the distance formula	Identify the perimeter of triangles, squares, rectangels, and petagons	L - Identify the 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)
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	H1E.2 Describe the structure and composition of Earth's atmosphere, geosphere, and hydrosphere. H2E.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. H2E.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	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/creaters - which is an example of volcanism? - pictures of volcano, river, rain). H - Extend M-level questions by linking features to the geoscience process (e.g., wwhich type of erosion pocess likely led to the canyon? - river, rain, wind; which feature is associated with 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 (see Carmichael et al. (2010)). 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.

The new Essentialized Assessment Frameworks (EAFs) are publicly available. A User Guide is provided to instruct educators regarding the intended uses of the Essentialized Standards (EsSt), 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 2022-23 school year in ELA and Math in grades 3-8 and grade 11; Science was 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 7	Grade 8	Grade 11
English Language Arts	X	X	X	X	X	X
Mathematics	X	X	X	X	X	X

Content Area	Grade 3	Grade 4	Grade 5	Grade 7	Grade 8	Grade 11
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 Executive Numbered Memo 003-2015-16 - Exemption from Statewide Summative Tests Update.

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..." Test Administration Manual.

1.4A English Learners

English learners are included as appropriate in Oregon's statewide assessment system, see Oregon Department of Education English Learner Program Guide. 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 Oregon Secretary of State Administrative Rules. A bilingual test administrator who is trained and endorsed by a district in Spanish or the students' language of origin should provide any language translation support. see Oregon Accessibility Manual.

1.4B Native Language Assessments

For all assessments that do not have a side-by-side version, such as the ORExt, directions may be interpreted by the personnel designated as competent by their district to make language interpretations for educational purposes. Translations must be conducted by a person whom the district has determined is qualified to administer such translation. Students who need additional support understanding the test directions may benefit from this resource. A bilingual test administrator who is trained and endorsed by a district in Spanish or the students' language of origin should provide any language translation support.

1.5 Participation Data

Oregon's participation data indicate that most students in the tested grade levels are included in our assessment system. 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 2021-22, see Statewide Report Card.

2 Assessment 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 the OR Extended Assessment-Item Development Info. 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. Only Grade 7 Math field test items were developed in 2017-18 which were in accordance with the 2014-15 test specifications, and are the most current available. A brief historical review of the ORExt is provided at the following link: Historical Review ORExt.

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 the Decision Making Related To Scaled Scores.

2.1B ORExt Test Blueprint

The ORExt Test Blueprint includes the entire test blueprint for the ORExt, as conveyed by the balance of representation across content areas and domains. Field-testing is conducted 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 the ORExt Item Development Process. 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 "ORExt Item Development Process" articulates the process used to generate the test administrator Scoring Protocol with comma separated value files used to create item templates that feed into Adobe InDesign© through a data merge. Student Materials are generated through the electronic test application, and then merged to pdf for a 1:1 correspondence for electronic and paper/pencil Student Materials. Final test packages are reviewed for accuracy and content and then disseminated to Oregon Qualified Assessors and Qualified Trainers through the secure password protected electronic test application platform, and paper/pencil materials are available for download on the secure password protected Training and Proficiency site.

2.1D Computer-Adaptive Considerations

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

2.2 Item Development

Item writers were recruited by ODE staff using an existing Qualified Assessor/Qualified Trainer listserv.

Needs	Content Area	Grade Level(s)
2 teachers (2 SPED)	ELA	Elementary (G 3-5)
2 teachers (1 GEN-ED; 1 SPED)	ELA	Middle (G 6-8)
1 teachers (1 GEN-ED)	ELA	High (G 11)
	Total Number Needed ELA	**5**
1 teachers (1 SPED)	Math	Elementary (G 3-5)
1 teachers (1 SPED)	Math	Middle (G 6-8)
2 teachers (1 SPED; 1 GEN-ED)	Math	High (G 11)
	Total Number Needed Math	**4**
3 teachers (2 SPED; 1 GEN-ED)	Science	G 5, 8, & 11
Total Number Needed Science		**3**
Total Oregon Teacher Item Writers Needed		**12**

2.2A Project Description:

Behavioral Research and Teaching at the University of Oregon recruited Oregon teachers to participate in item development for a new alternate assessment prior to the 2014-2015 school year. Selected teachers were asked to develop 360 items in English Language Arts, Mathematics, or Science over the course of the summer, from mid-June through end of August. The Project Director worked with lead item developers to provide training, ongoing review and feedback, and quality assurance. All participants were expected to provide documentation of their qualifications and sign test security agreements. In addition, all item developers were expected to participate in a half-day item development training based upon the following schedule: ELA - Tuesday, from 8 AM to 12 PM; Math - Wednesday, from 8 AM to 12 PM; Science - Thursday, from 8 AM to 12 PM.

2.2B Minimum Qualifications:

All licensed Oregon public school teachers with at least three years of teaching in a life skills/severe needs program (SPED) or a general education classroom (GEN-ED), respectively, were encouraged to apply. Preference was given for item writing experience, additional years of teaching experience, and higher education degree status.

2.2C Compensation:

Teachers who participated in this process were compensated at a rate of $20/hr via professional service contracts. It was anticipated that teachers would produce 4 ELA

items/hr, 6 Science items/hr, and 8 Math items/hr. As such, the maximum contract amount for ELA was $1,800, for Science $1,440, and for Math $900. Item development focused primarily on writing the stem and 3 options, with no need to produce graphics (rather use labels for a BRT graphic designer to produce).

2.2D Contact:

Because the timeline required work over the summer, Oregon teacher recruitment was challenging. BRT researchers thus performed an additional on-campus recruitment within the College of Education using the same information. The final pool of item writers included 18 item writers: seven Oregon teachers (all with MA degrees), five PhD candidates within the COE, and six BRT researchers (four PhD candidates, one PhD, and one with an MA). Item writers averaged 11.5 years of teaching experience. The teachers recruited all had prior experience developing items for the ORExt, as did all of the BRT researchers. The five PhD candidates within the COE had no prior item development experience. All item development was reviewed by BRT researchers and the Project Manager.

The item development process followed is elaborated in the Item Writer Training 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 complexity 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 ORExt Assement Vertical Scaling Project). Sample items are provided in the ORExt Electronic Practice Tests 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 Summative Testing Administration Manual. Important updates to the testing process are distributed via the Assessment and Accountability Updates listserve, as well. 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 in the Oregon Accessibility Manual, and the Oregon Accommodations Manual. 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 in the Oregon Accessibility Manual. 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. A bilingual test administrator who is trained and endorsed by a district in Spanish or the students' language of origin should provide any language translation support.

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.**(Here might be a spot to include some ELL #)** Accommodations information was collected this year as an option for data entry. Entering accommodations information will be required each year. Analyses of the impact of accommodation provision on the ORExt is feasible after each years 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 2022-23: How to Select, Administer, and Evaluate Accommodations for Oregon's Statewide Assessments in the Oregon

Accommodations Manual. The Oregon Extended assessments are also designed according to universal design principles and utilize a simplified language approach (see Reducing the Depth, Breadth, and Complexity of Items).

In the 2021-2022 school year, the state redesigned a training and proficiency program for sign language interpretation of its assessments and has significantly updated the site during this time. The Sign Language Training process included videos of interpreters administering items to students, materials that support appropriate administration (i.e., transcripts, closed captioning and PowerPoint slides that supplement the video administrations and the current ODE accessibility manual), and proficiency testing to support standardized interpretation for Oregon's assessments, including the ORExt. An 11-item proficiency test was administered, with an 80% required for passing (9/11 items correct). In 2022-23, the site was used to train 56 participants. All participants except one passed the assessment within 2 attempts. The overall average scores on the proficiency test were between 94%-96%.

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 Reducing the Depth, Breadth, and Complexity of Items). Any given student's communication system may include home signs, school signs, English words, and Spanish words, for example. With the exception of items that require independent reading, a bilingual test administrator who is trained and endorsed by a district in Spanish or the students' language of origin should provide any language translation support, including American Sign 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 the following QT Training Video. Through an online distribution and assessment system, QA/QT Training and Proficiency is determined annually. 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 returning QTs in Zoom regional trainings. The regional trainings were provided at two separate time slots on November 30th and December 2nd for a total of two QT training opportunities (see QT Training Video link above). 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

Table 2.1: Passing Scores (as percentage correct) by User Type, Assessment, and Number of Attempts.

	Count	Number of Attempt	Mean (SD)	Range of Passing Scores
Qualified Assessor				
new assessor	230	1	0.92 (0.07)	[80%, 100%]
new assessor	28	2	0.89 (0.07)	[80%, 100%]
new assessor	5	3	0.94 (0.09)	[80%, 100%]
refresher	479	1	0.92 (0.07)	[80%, 100%]
refresher	53	2	0.87 (0.05)	[80%, 100%]
refresher	9	3	0.93 (0.07)	[85%, 100%]
refresher	2	4	0.9 (0.14)	[80%, 100%]
Qualified Trainer				
new assessor	46	1	0.9 (0.06)	[80%, 100%]
new assessor	3	2	0.89 (0.07)	[83%, 97%]
refresher	48	1	0.94 (0.06)	[80%, 100%]
refresher	2	2	0.85 (0.07)	[80%, 90%]

Note:
The New Assessor Training Exam has 30 items; the Refresher has 20 items

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 andProficiency Website. All assessors need to complete training each year to retain their status for administering the Extended Assessments.

All assessors, new assessors and returning assessors needed further training in 2022-23 and were required to pass a proficiency with a score of 80% or higher. The proficiency assessment covered areas in Administration, English Language Arts (ELA), Mathematics, and Science. Returning QAs or QTs for the 2022-23 school year needed to pass the same proficiency assessment, again with a score of 80% or higher. The tables below contain data on the number of assessors (participants) in proficiency tests. 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.

The table below outlines registered users from the Oregon Extended Assessment Training and Proficiency Website.

A higher number of assessors completed the Refresher Proficiency test than the New Assessor Proficiency tests reflecting a greater number of returning assessors compared to new assessors. Data showed the majority of those who took either the Refresher or New Assessor

Proficiency tests passed the first attempt. A limited number of participants had to take a second attempt, and very few to none had to take a third and fourth attempt.

OREXt Trainer Training Confidence Scale Percentages

Following this training of the OREXt system, I feel confident:	Confidence 1	2	3	4
1. In my understanding of the administration (i.e., paper-pencil & tablet), scoring, and data entry of the OREXt.		**3**	**37**	**60**
2. In my understanding of the administration, scoring, and data entry of the Oregon Observational Rating Assessment (ORora).		**2**	**43**	**55**
3. In my understanding of the qualification process for Qualified Trainers and Qualified Assessors.			**32**	**68**
4. Making statewide assessment decisions (as part of an IEP team) for students with significant cognitive disabilities (SWSCD).			**24**	**76**
5. Training others in the administration, scoring, and secure test/data entry of the OREXt system			**39**	**61**
6. In my use of the OREXt online training and proficiency website.			**34**	**66**

KEY: 1 = strongly disagree 2 = disagree 3 = agree 4 = strongly agree

Figure 2.1: Trainer Confidence Scale Percentages

All technical assistance questions that we received from the field as part of our HelpDesk are tagged and reviewed through HelpScout. The most common inquiries for the 2022-2023 test administration window involved status upgrades, missing students,adding additional schools and districts to user accounts, credential verification, and rostering. Some other common inquiries included student registration, access to monitoring for DTC's, and technical issues with individual tablets.

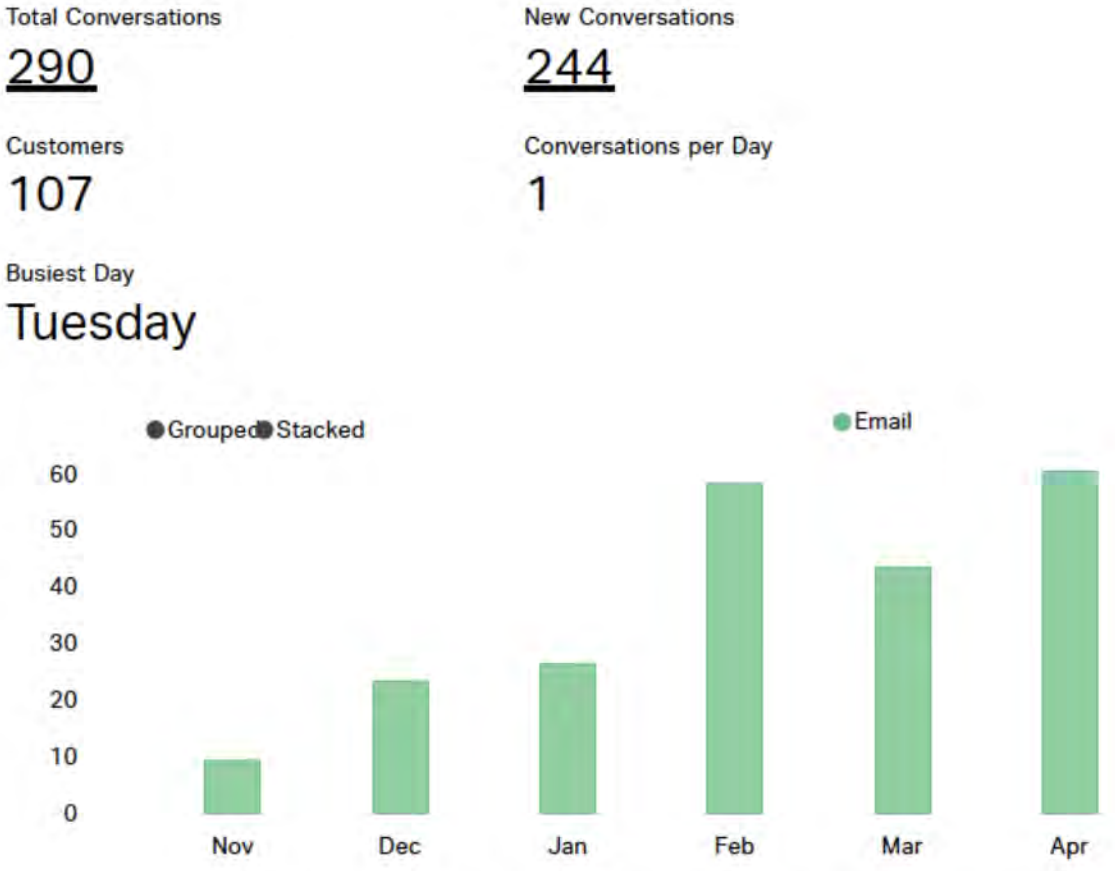


Figure 2.2: Help Desk Email Report

A HelpScout FAQ website was also created to assist in answering questions from the field. Some of the frequently viewed pages included Logging into the OREXt Electronic app, Rostering, Supporting Electronic Test Administration, Curricular and Instructional Materials, Data Entry, and Credential Verification. All help desk inquiries will be taken into consideration for the 2022-2023 QA/QT training, and as the training site and FAQ site are updated prior to opening the 2022-2023 testing window.

A summary of visits to the FAQ page on OREXt HelpScout is below.

Figure 2.3: FAQ Conversations Created

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. Different decision rules for unique items versus horizontally- or vertically-scaled anchor items were imployed. 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%).

A consequential validity study was implemented in 2018-19 that surveyed QAs and QTs regarding the academic and social consequences of the ORExt, both intended and unintended. The Consequential Validity report is published in the Consequential Validity Survey Results. ODE and BRT staff reviewed the results of the survey to determine what program improvements were 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 were framed based upon current consequential validity approaches for alternate assessments in the literature, as well as issues that were of specific value in Oregon. The survey included 121 respondents. This was 11% of the solicited respondents, who were all Qualified Assessors (QAs) and Qualified Trainers (QTs) in the or.k12test.com database. The sample was 83% 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 demonstrated that QAs and QTs continued to feel that the ORExt test items were easy to administer and score (64.2% Strongly Agree) and felt confident in their ability to interpret scaled scores and Achievement Level Descriptors for the ORExt (69.8% Strongly Agree and Agree). They also felt that the items were accessible for students who participated (78% Strongly Agree and Agree) and that the ORExt reflected the academic content that SWSCD should be learning (68.4% Strongly Agree and Agree). QAs and QTs felt marginally positive about the educational impacts of the ORExt and marginally negative about its social impacts. The results again demonstrated that the ORExt content area assessments generally required up to one hour to administer.

The qualitative results revealed two 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); and, 2) overall item and test design (i.e., one item per page, visual supports, scoring protocol and student materials design, accessibility of test questions). Teachers recommended the following areas of improvement, not all of which are actionable: 1) Option to administer the assessment electronically was beneficial, 2) A functional skills assessment should be added, 3) New items for very low functioning students should be developed, and 4) request for a math assessment composed of more practical/life skills problems involving time and money. Complete results, including anticipated responses, from the survey can be found in the Consequential Validity Survey Results.

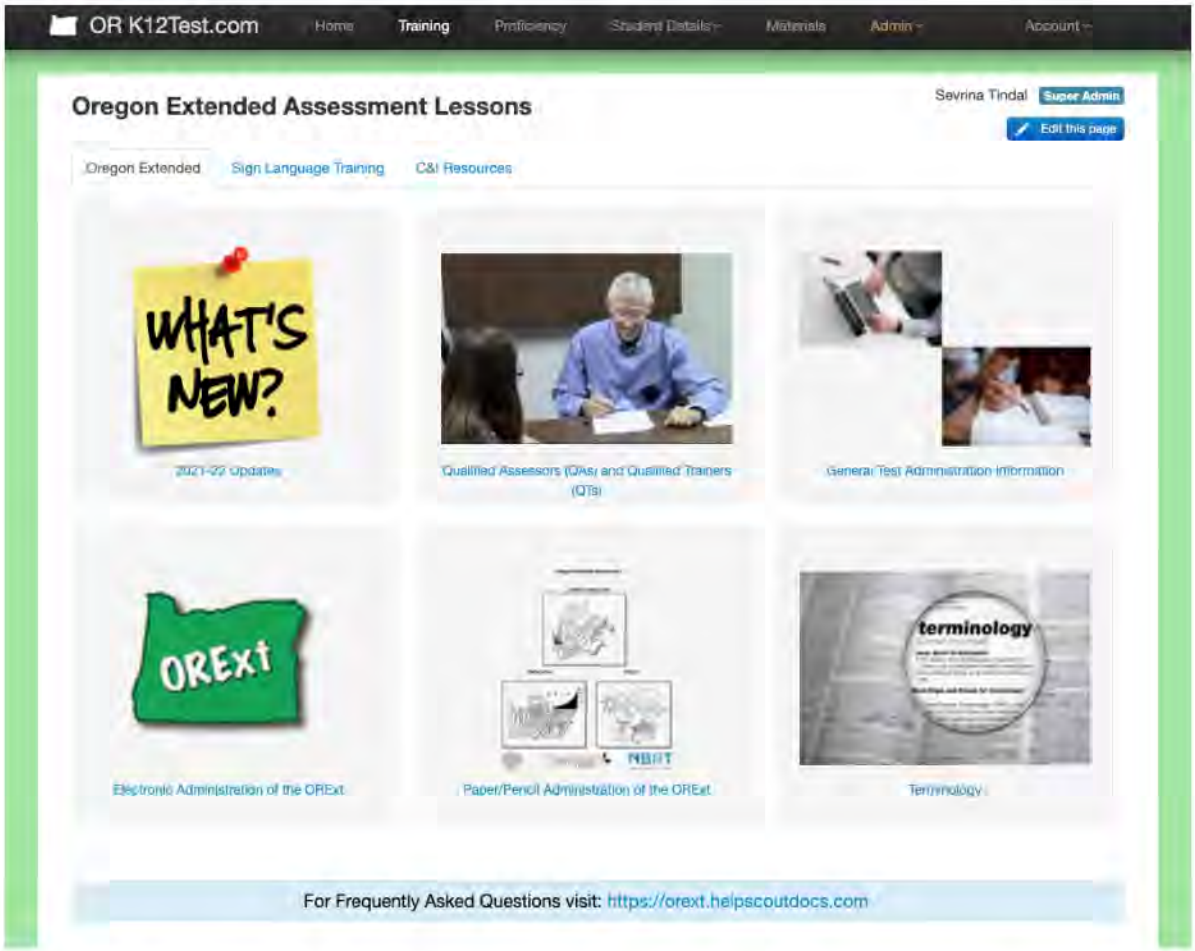
2.3C Technology-based Assessments

The ORExt was implemented using a technology-based platform as Phase 3 of the ORExt Tablet Administration. The 2017-18 testing window was the first year all grade level and subject area assessments were available on a tablet application/web-based platform (see ORExt Pilot Tablet Study Report). Administration of the tablet application mirrors

paper/pencil administration with each item read aloud to the student, and the student asked to select one of three answer choices. Tablet functionality includes optional discontinuation if the student misses 5 out of the first 10 items, directing the assessor to administer the ORora. To support understanding of the system by both teachers and students, a separate practice test tablet application is available. Helpdesk inquiries and feedback from the field indicated much preference of the tablet administration versus paper/pencil. Qualified Trainers and Qualified Assessors reported their students' were more focused during tablet administration, and because the tablet application scores automatically it was much more efficient for assessors. Improvements will be made to the electronic test based on technology improvements and feedback from the field. Data entry for all platforms is now maintained and monitored by secure BRT servers.

2.4 Monitoring Test Administration

ODE maintains a rigorous training system to support standardized test administration for the secure Oregon K12 website see screenshot 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 (with video examples), supporting student access to items without violating the test construct, content area trainings that demonstrate how to administer items in ELA, Math, and Science (with video examples and 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):

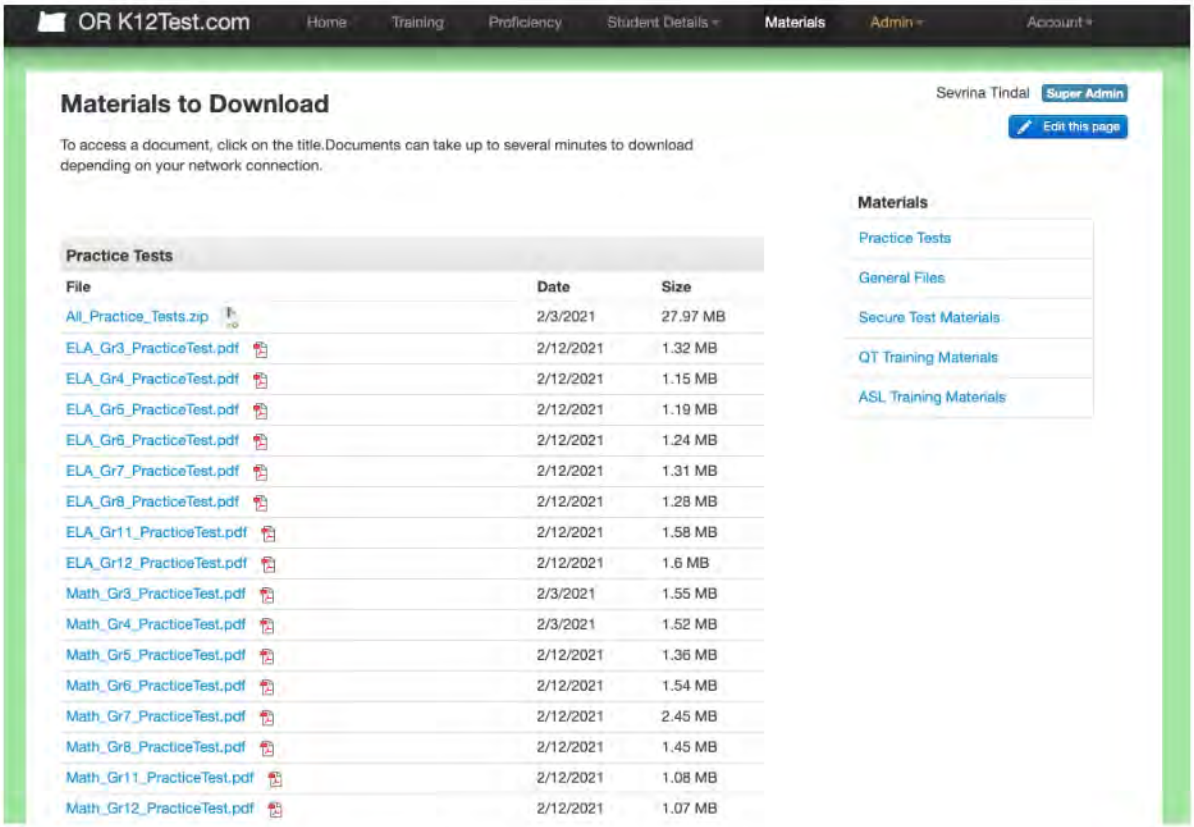


Figure 2.4: Materials Page

In addition, monitoring and unofficial reports related to test administration for the ORExt can be found in the Student Details tab of the training and proficiency website. Official reports are 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 Test Administration Manual. 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 Test Administration Manual).

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 Test Administration Manual).

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 Test Administration Manual). 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 OExt 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 OExt. Item reviews necessary to provide alignment, bias, and sensitivity information are conducted online using the secure Distributed Item Review (DIR) platform (secure website, but see DIR Overview for a system overview).

For the 2022-2023 school year, all paper/pencil secure test distribution and data entry was hosted by BRT through the secure training site.

The secure tablet application and web-based platform distribution and data entry were hosted by BRT servers. All technology based secure administration and data entry was password-protected. Download of the tablet app was dependent on the type of device, all instructions and download links are available in the Oregon K12 website secure platform. Additional information regarding test security can be located in the general assessment system Peer Review evidence submission.

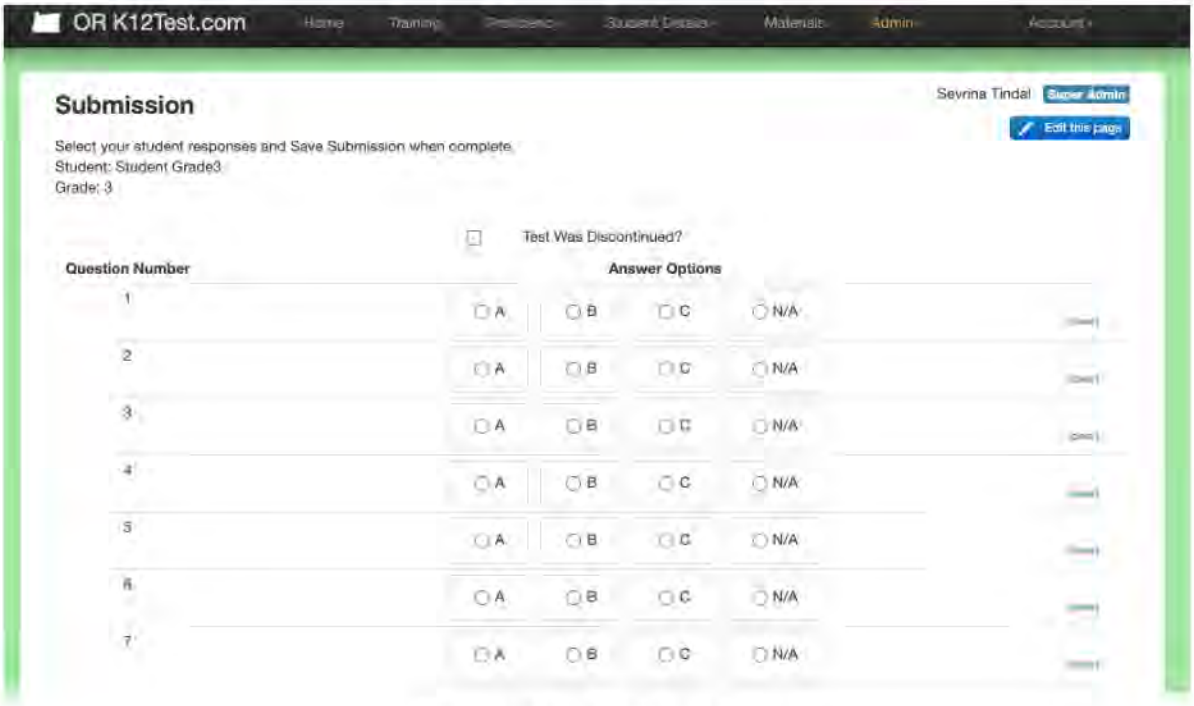


Figure 2.5: Data Entry Page

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 Oregon K12 website. Only trained users with a vested educational interest who have signed test security agreements are authorized to access to online data entry systems.

2.6C Protecting Personally Identifiable Information

All confidential, personally identifiable student information is protected by policy and supported by training (see Test Administration Manual). 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 State Annual Report Card). Similar rules are applied to student subgroups, including students with disabilities, English

learners, and students from diverse racial/ethnic backgrounds (see State Annual Report Card).

3 Technical Quality: Validity

3.1 Overall Validity, Including Validity Based on Content

As elaborated by Messick (1989), 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, answers for two important questions are addressed. Regardless of whether the students tested have disabilities, the questions are identical: (1) How valid is the 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. The Association et al. (2018) is used 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. 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 the ORExt Electronic Practice Tests.

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 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 the QT Training Video.

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

The foundation of validity evidence from content coverage for the ORExt comes in the form of test specifications (see OR Extended Assessment-Item Development Info) and the ORExt Test Blueprint. Among other things, the Association et al. (2018) suggest specifications should "define the content of the test, the proposed test length, the item formats...".

All items are linked to grade level standards and a prototype was developed using principles of universal design with traditional, content-referenced multiple-choice item writing techniques. 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. One test form for the ORExt was developed 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.

The ORExt tests were developed iteratively by developing items. The Item Writer Training conveys the item writer training materials, piloted, reviewed, and edited in successive drafts. Existing panels of veteran teachers were used 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.

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 (2002).

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 initial 2014 linkage study and a new, independent alignment study conducted in spring, 2017 is provided in the Oregon Extended Assessment Alignment Study. Based on student performance from the 2016-2017 testing year, new and Grade 7 Math field test items were written in fall 2017.

The summary section of the independent alignment study report states that, "Oregon's Extended Assessments (ORExt) in English Language Arts, Mathematics, and Science were evaluated in a low-complexity alignment study conducted in Spring of 2017. Averages of reviewer professional judgments over five separate evaluations were gathered, reviewed, and interpreted in the pages that follow. In the three evaluations that involved determining the relationship between standards and items, reviewers identified sufficient to strong relationships among assessment components in all grades and all subject areas. In the two evaluations involving Achievement Level Descriptors, reviewers identified thirty instances of sufficient to strong relationships out of thirty-four possible relationship opportunities resulting in an overall affirmed relationship with areas for refinements identified."

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, see ORExt Assessment Technical Report on Standard Setting. 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

Results of the analysis of the linkage of the new Essentialized Assessment Frameworks, (EAF), composed of Essentialized Standards (EsSt), 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 EsSt are sufficiently linked to grade level standards, while the

OExt items are aligned to the EsSt. In addition to presenting linkage information between grade level content standards and the EsSt, the linkage study presents alignment information related to the items on the new OExt in comparison to the EsSt. 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 as part of two comprehensive linkage studies, one performed in 2014 and an independent alignment study performed in 2017 (see Section 3.1A). The professional reviewers in an internal study in 2014 and an independent study in spring 2017 included both special and general education experts, with content knowledge and experience in addition to special education expertise.

According to the independent linkage study report, the spring 2017 review was conducted by expert reviewers with professional backgrounds in either Special Education (the population), Assessment, or in Oregon's adopted content standards. Reviewers were assigned to review grade-level items relative to their experience and expertise. In all, 39 reviewers participated. Thirty-four (34) participated in all 5 evaluations: thirteen (13), for the English Language Arts review, fifteen (15) for the Mathematics review, and six (6) for the Science review. All participants were assigned to at least one specific content area as shown in Table 1. Note: Four individuals were assigned to two areas of review. The thirty-nine individuals who participated in the study had a robust legacy of experience in the field and in the state. Participants represented 25 unique school districts across the state representing both urban and rural perspectives. All 39 of the individuals participating in the study held current teaching licenses. Two individuals also held administrative licenses. Years of experience in their area ranged from 3 - 30 years of experience with an average of 17 years of experience. (Mode = 11 years, Median = 16 years). One individual indicated 50 years of experience in the field. Three of the 39 individuals held a Bachelor's degree only. Thirty-six held a Bachelor's degree and at least one Master's degree. Two held a Bachelor's degree, at least one Master's degree, and a doctoral degree. Fourteen (36%) of the individuals identified as experts in a specific Content area and 25 (64%) of the individuals identified Special education as their primary area of expertise.

These skilled 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 and on Excel spreadsheets shared with the researcher electronically, (see DIR Overview for a system 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). When averaged across reviewers, 1.00-1.29 was considered in the low range, 1.30

- 1.69 was sufficient, and 1.70 - 2.0 was strong. Additional comment was requested for any essentialized standard or item whose linkage was rated 0.

Overall, the 2017 independent alignment study concluded that: "First, reviewers were asked to conduct an affirmational review of the rationale used by test developers to omit certain content standards." This finding was used to infer that the final standards selected for inclusion or omission in Oregon's Extended Assessment were chosen rationally and that the final scope of content standards can be considered justifiable for the population for the subject area.

Conclusion: This review, with a lowest average rate of .82 (on a scale of 1), permited the inference: the scope of the standards selected for translation to Essentialized Standards were rationally selected. None of the standards de-selected (for inaccessibility or for being covered elsewhere) were strongly identified for re- inclusion, nor were identified as a critical hole for this population of students.

Second, reviewers were asked to identify the strength of the link between the source standard and the Essentialized Standard. This finding was used to infer that the process undertaken to essentialize a given Source Standard did not fundamentally or critically alter the knowledge or skill set intended by the source standard for this population of students (further confirming that the content selected for the assessment was comparable).

Conclusion: This review, with a range of 1.5 - 1.9 (on a scale of 2) permited the inference: the Essentialized Standards were found to link sufficiently to the source standards on average beyond the "sufficient" average of 1.0.

Third, reviewers were asked to identify the strength of the alignment between the Essentialized Standards and the items and to review the items developed using the Essentialized Standards for bias, and accessibility. The finding from this review was used to infer that the items written for this grade and subject area (using these Essentialized Standards) were adequately linked to the Essentialized Standards, were free from bias, and were accessible to students with significant cognitive disabilities.

Conclusion: The alignment review (1.32 - 1.89), accessibility review (.67 - 1.0), and freedom from bias review (.65 - 1.0) all permited the inference that the test items indicated a relationship with the source standards, the test items were not overly biased towards or against any particular group of individuals, and the test items were written such that the content and intent could be accessed by students with the most significant cognitive disabilities. (**Note: this range was skewed by feedback from one reviewer –ELA-Grade 3 - whose comments were noted in this study. Removing that individual's comments would result in a range of .90 - 1.0 accessibility range and .89 - 1.0 freedom from bias range respectively.)

Fourth, reviewers were asked to review the statements used to describe student achievement on the test (the Achievement Level Descriptors) and their alignment to the Essentialized Standards that the students were tested on. The findings from this review was used to infer

that the skills and achievements described by the Achievement Level Descriptors for each subject and grade level are aligned with the content standard being measured.

Conclusion: The reviews ranging from .68* - 1.0 permited the inference that the descriptions made regarding student skillset were an accurate reflection of the standards from which the assessment was developed at all three levels evaluated. (*One outlier for ELA-Grade 4 provided a review of a .52 average).

Fifth, and finally, reviewers were asked to review the alignment of the Achievement Level Descriptors to the items. The finding from this review was used to infer that each item in the developed assessment(s) was appropriately aligned to its associated Achievement Level Descriptor (further confirming that decisions made using this test were aligned with the intent of the source standard).

Conclusion: Fourteen of the seventeen grade-level reviews resulted in an average reviewer range of .67 - 1.0 indicating an appropriate alignment between ALDs and the items as written. This review permited the inference that, overall, the Achievement Level Descriptors are accurate reflections of the items. In three instances (Mathematics-Grades 3 and 4, and ELA-Grade 8) the average alignment by reviewer was .5 (indicating that one of the two individuals in that category did not agree that the items and ALDs were aligned)."

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" Association et al. (2018). 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).

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. The ORExt 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 Oregon Observational Rating Assessment (ORora) was first implemented in 2015-16. The ORora 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 2021-22 is provided:

Table 3.1: ORora Participation numbers (percent of total grade)

	Total ORExt N	ORora Subsample n (% of total)
3	396	140 (35.35%)
4	434	152 (35.02%)
5	401	139 (34.66%)
6	447	152 (34%)
7	432	148 (34.26%)
8	425	145 (34.12%)
11	336	117 (34.82%)
Total	2871	993 (34.59%)

Table 3.2: Total ORora Score Descriptives by Grade

Grade	Mean (SD) Score	Score Range
Grade 3	50 (14.98)	[20, 80]
Grade 4	51.21 (15.89)	[6, 80]
Grade 5	53.04 (15.69)	[13, 80]
Grade 6	53.67 (16.2)	[20, 80]
Grade 7	54.85 (17.01)	[6, 80]
Grade 8	56.45 (16.97)	[20, 80]
High School	53.73 (17.02)	[9, 80]
Total Average	53.28 (16.37)	[6, 80]

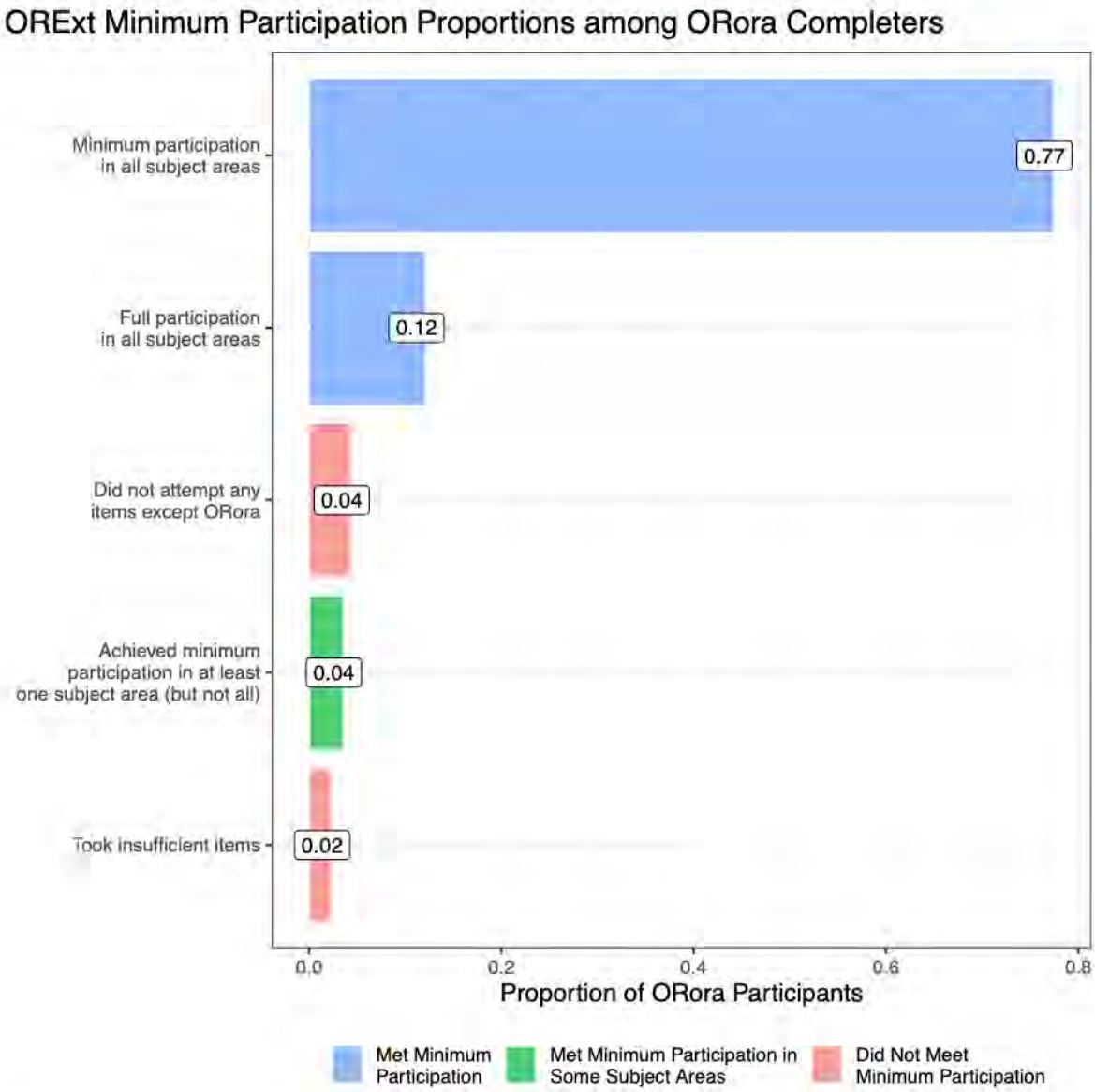
Table 3.3: ORora Subscore Descriptives by Grade

Grade	Mean (SD) Score	Score Range
Attn		
Grade 3	12.5 (3.57)	[5, 20]
Grade 4	12.64 (3.95)	[5, 20]
Grade 5	13.25 (3.67)	[5, 20]
Grade 6	13.39 (4.12)	[5, 20]
Grade 7	14.02 (4.14)	[5, 20]
Grade 8	14.24 (4.24)	[5, 20]
High School	13.76 (4.25)	[5, 20]

Table 3.3: ORora Subscore Descriptives by Grade *(continued)*

Grade	Mean (SD) Score	Score Range
Exp		
Grade 3	11.8 (4.55)	[5, 20]
Grade 4	12.32 (4.74)	[5, 20]
Grade 5	12.69 (4.88)	[5, 20]
Grade 6	12.84 (4.77)	[5, 20]
Grade 7	13.04 (5.16)	[5, 20]
Grade 8	13.49 (5.15)	[5, 20]
High School	12.81 (5.24)	[5, 20]
Math		
Grade 3	12.12 (4.17)	[5, 20]
Grade 4	12.65 (4.27)	[5, 20]
Grade 5	13.24 (4.34)	[5, 20]
Grade 6	13.32 (4.69)	[5, 20]
Grade 7	13.57 (4.6)	[5, 20]
Grade 8	14.15 (4.71)	[5, 20]
High School	13.02 (4.62)	[5, 20]
Recp		
Grade 3	13.58 (4.48)	[5, 20]
Grade 4	13.81 (4.39)	[5, 20]
Grade 5	14.02 (4.49)	[5, 20]
Grade 6	14.12 (4.5)	[5, 20]
Grade 7	14.51 (4.66)	[3, 20]
Grade 8	14.65 (4.71)	[5, 20]
High School	14.52 (4.72)	[5, 20]

Below is a breakdown of minimum participation on the OExt for those who took the ORora. Minimum participation is defined as having attempted at least 5 items. The vast majority (77%) of ORora participants achieved minimum participation on all subject areas (e.g., Math and ELA for grade 3; Math, ELA, and Science for grade 8) or full participation on all subject areas (12%); a small proportion (4%) met minimum participation in only 1 subject area but not in the other(s) (e.g., Math but not ELA for grade 4; Science and Math but not ELA for grade 11). A total of 7% of ORora participants did not meet minimum participation, with `r paste0(round(proportion_took_insufficient_items, 2)*100, '%')` of total being those who took insufficient items and `r paste0(round(proportion_did_not_attempt_any_items, 2)*100, '%')` of total being those who did not attempt any items besides ORora at all.



Fifty-one reviewers analyzed all OExt 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.

Substantive evidence that has been documented suggests that the OExt 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.

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 (EsSt) out of the total of 16 (4/16 = 25%).

The testblue 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 EsSt, as some EsSt link to multiple grade level content standards. However, the blueprints below reflect only the written EsSt and are thus an underrepresentation of the breadth of grade level content addressed by the OREXt.

Table 3.4: English Language Arts Subdomains by Grade

Subdomain	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
Reading Standards: Foundational Skills	2	2	2	–	–	–	–
Reading Standards: Informational Text	4	4	4	5	5	5	5
Reading Standards: Literature	4	4	4	5	5	5	5
Writing	4	4	4	4	4	4	4
Language	2	2	2	2	2	2	2
Total	16	16	16	16	16	16	16

Table 3.5: Mathematics Subdomains by Grade

Subdomain	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
Operations and Algebraic Thinking	7	4	3	–	–	–	–
Numbers and Operations in Base Ten	2	6	8	–	–	–	–
Numbers and Operations – Fractions	3	8	6	–	–	–	–
Measurement and Data	8	5	4	–	–	–	–
Geometry	2	3	2	3	3	4	7
Ratio and Proportional Relationships	–	–	–	3	2	–	–
The Number System	–	–	–	9	7	2	–
Expressions and Equations	–	–	–	6	2	6	–
Statistics and Probability	–	–	–	5	6	3	5
Functions	–	–	–	–	–	4	7
Numbers and Quantities	–	–	–	–	–	–	2
Algebra	–	–	–	–	–	–	2
Total	22	26	23	26	20	19	23

Table 3.6: Science Subdomains by Grade

Subdomain	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
Life Science Standards	–	–	4	–	–	9	8
Physical Sciences	–	–	4	–	–	7	9
Earth and Space Science	–	–	4	–	–	6	6
Engineering, Technology, and Applications	–	–	2	–	–	2	–
Total	0	0	14	0	0	24	23

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).

3.3A Point Measure Correlations

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; as such, point measure correlation is interpreted as a correlation coefficient. All items included in the 2022-2023 operational assessment are represented. Point measure correlations ranged from 0.3 to 0.77 in ELA, 0.17 to 0.77 in Math, to 0.31 to 0.79 in Science.

Table 3.7: Point Measure Correlation by Content and Grade

Grade	Mean	Standard Deviation	Median	Min	Max
ELA					
Grade 3	0.56	0.11	0.59	0.33	0.74
Grade 4	0.56	0.10	0.57	0.30	0.72
Grade 5	0.59	0.05	0.60	0.51	0.68
Grade 6	0.61	0.05	0.60	0.49	0.69
Grade 7	0.62	0.06	0.63	0.43	0.71
Grade 8	0.62	0.05	0.63	0.51	0.69
High School	0.68	0.07	0.68	0.54	0.77
Math					
Grade 3	0.54	0.12	0.59	0.17	0.68
Grade 4	0.52	0.13	0.55	0.26	0.72
Grade 5	0.51	0.11	0.53	0.22	0.67
Grade 6	0.54	0.12	0.57	0.24	0.76
Grade 7	0.52	0.11	0.55	0.28	0.72

Table 3.7: Point Measure Correlation by Content and Grade *(continued)*

Grade	Mean	Standard Deviation	Median	Min	Max
Grade 8	0.43	0.09	0.42	0.25	0.60
High School	0.52	0.11	0.52	0.23	0.69
Science					
Grade 5	0.63	0.08	0.66	0.31	0.74
Grade 8	0.64	0.08	0.67	0.40	0.74
High School	0.70	0.05	0.71	0.61	0.79

Table 3.8: Point Measure Correlation of ELA Subscores by Grade

Grade	Mean	Standard Deviation	Median	Min	Max
Reading					
Grade 3	0.58	0.09	0.61	0.37	0.70
Grade 4	0.56	0.07	0.57	0.43	0.65
Grade 5	0.58	0.03	0.59	0.52	0.65
Grade 6	0.62	0.05	0.62	0.51	0.69
Grade 7	0.64	0.05	0.65	0.55	0.73
Grade 8	0.64	0.04	0.65	0.53	0.72
High School	0.68	0.06	0.69	0.57	0.76
Writing					
Grade 3	0.65	0.10	0.62	0.49	0.78
Grade 4	0.68	0.09	0.68	0.53	0.79
Grade 5	0.68	0.06	0.68	0.59	0.77
Grade 6	0.65	0.04	0.63	0.58	0.72
Grade 7	0.65	0.09	0.67	0.45	0.76
Grade 8	0.68	0.05	0.68	0.57	0.74
High School	0.72	0.08	0.73	0.59	0.82

3.3.0.1 Outfit Mean Square Distributions

Outfit mean square (OMS) values below 1.0 demonstrate that values are too predictable and perhaps redundant, while values above 1.0 indicate unpredictability. Another way to think about OMS is that values closer to 1.0 denote minimal distortion of the measurement system. Items above 2.0 are deemed insufficient for measurement purposes and flagged for replacement.

Table 3.9: Mean Square Outfit by Content and Grade

Grade	Mean	Standard Deviation	Min	Max
ELA				
Grade 3	1.02	0.34	0.58	2.33
Grade 4	0.92	0.21	0.51	1.48
Grade 5	1.03	0.28	0.73	1.83
Grade 6	0.93	0.20	0.64	1.43
Grade 7	0.94	0.28	0.52	2.12
Grade 8	0.94	0.26	0.60	1.78
High School	0.77	0.24	0.42	1.32
Math				
Grade 3	0.95	0.23	0.64	1.46
Grade 4	0.94	0.20	0.61	1.52
Grade 5	0.95	0.21	0.71	1.74
Grade 6	0.89	0.20	0.50	1.45
Grade 7	0.88	0.21	0.63	1.45
Grade 8	0.91	0.18	0.70	1.39
High School	0.84	0.18	0.56	1.32
Science				
Grade 5	0.91	0.27	0.53	1.59
Grade 8	0.88	0.25	0.52	1.54
High School	0.89	0.34	0.38	2.05

Table 3.10: Mean Square Outfit by Content and Grade

Grade	Mean	Standard Deviation	Min	Max
Reading				
Grade 3	1.12	0.49	0.58	2.75
Grade 4	1.02	0.15	0.72	1.27
Grade 5	1.11	0.27	0.80	1.73
Grade 6	0.94	0.21	0.66	1.47
Grade 7	0.99	0.36	0.58	2.30
Grade 8	0.96	0.22	0.63	1.67
High School	0.74	0.21	0.35	1.08
Writing				
Grade 3	0.93	0.40	0.45	1.61
Grade 4	0.74	0.31	0.40	1.27
Grade 5	0.89	0.30	0.53	1.39

Table 3.10: Mean Square Outfit by Content and Grade *(continued)*

Grade	Mean	Standard Deviation	Min	Max
Grade 6	0.93	0.20	0.62	1.21
Grade 7	0.95	0.37	0.56	1.89
Grade 8	0.90	0.32	0.55	1.57
High School	0.76	0.49	0.30	2.11

While most OMS values in ELA were between 0.5 and 1.5, 8 items across 3 grades (High School, Grade 3, Grade 7) and 4 contents and/or subdomains (Writing, Science, Reading, ELA) were above 2. The exact OMS values above 2 can be seen in the table below, arranged by test and grade.

Table 3.11: Mean Square Outfit of Items above 2 by Grade and content

Grade	Outfit
Writing	
High School	2.11
High School	2.11
Science	
High School	2.05
High School	2.05
Reading	
Grade 3	2.75
Grade 7	2.30
ELA	
Grade 3	2.33
Grade 7	2.12

3.3B 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.

Two categories are used to designate meeting AMOs among participants: “Meets” and “Exceeds”. Across content areas and grades an average of 32% (SD = 6%) were in these categories. The table below shows exact AMO breakdowns by grade and content area.

Table 3.12: Annual Measureable Objectives Frequencies and Percentages

Content and Grade	AMO 1 (Does Not Yet Meet)	AMO 2 (Nearly Meets)	AMO 3 (Meets)	AMO 4 (Exceeds)
ELA				
Grade 3	106 (27%)	175 (44%)	94 (24%)	19 (5%)
Grade 4	146 (34%)	150 (35%)	107 (25%)	27 (6%)
Grade 5	130 (33%)	153 (39%)	74 (19%)	37 (9%)
Grade 6	163 (37%)	131 (30%)	103 (23%)	46 (10%)
Grade 7	159 (37%)	131 (31%)	80 (19%)	58 (14%)
Grade 8	194 (46%)	91 (22%)	83 (20%)	51 (12%)
High School	113 (35%)	94 (29%)	43 (13%)	74 (23%)
Math				
Grade 3	191 (49%)	92 (24%)	100 (26%)	7 (2%)
Grade 4	155 (36%)	187 (44%)	72 (17%)	14 (3%)
Grade 5	135 (34%)	163 (42%)	78 (20%)	16 (4%)
Grade 6	240 (55%)	50 (11%)	128 (29%)	18 (4%)
Grade 7	229 (54%)	14 (3%)	167 (39%)	18 (4%)
Grade 8	209 (50%)	68 (16%)	136 (33%)	2 (0%)
High School	167 (53%)	66 (21%)	78 (25%)	7 (2%)
Science				
Grade 5	175 (44%)	93 (23%)	82 (21%)	47 (12%)
Grade 8	183 (45%)	78 (19%)	74 (18%)	76 (18%)
High School	113 (35%)	61 (19%)	70 (22%)	78 (24%)

Across all years, the most common AMOs were AMO 1 (4 of 7 grades) and AMO 2 (3 of 7 grades) for ELA, AMO 1 (5 of 7 grades) and AMO 2 (2 of 7 grades) for math, and AMO 1 (3 of 3 grades) for science.

Across subjects there are often few students in AMO 4 compared to the other 3. Considering this is the highest AMO, this is unsurprising; however, ELA and science have much higher rates of AMO 4 than math for most grades.

In some cases, a very small range of scaled scores exist because of the small range of observed scores. The smallest is Math grade 7, which only exists between scaled scores 207 and 209. Math grades 6 and 8 are also very small in terms of scaled scores, each existing between 4 scaled score points.

For comparison, the smallest AMO range for other contents areas is 7, which is high school ELA, followed by science at 10. In these cases, error can make a greater difference; theoretically, this could lead to lower test-retest consistency. This may be why there are higher percentages in AMO 3 for math in grades 7 and 8, compared to other years of math.

One to two more low-complexity items to relevant mathematic tests may help address this concern, as well.

Table 3.13: Annual Measureable Objectives Frequencies and Percentages: ELA Subscores (Reading and Writing)

Content and Grade	AMO 1 (Does Not Yet Meet)	AMO 2 (Nearly Meets)	AMO 3 (Meets)	AMO 4 (Exceeds)
Reading				
Grade 3	115 (29%)	155 (39%)	100 (25%)	23 (6%)
Grade 4	124 (29%)	163 (38%)	111 (26%)	32 (7%)
Grade 5	120 (30%)	164 (42%)	63 (16%)	47 (12%)
Grade 6	162 (37%)	136 (31%)	77 (17%)	67 (15%)
Grade 7	164 (38%)	118 (28%)	74 (17%)	72 (17%)
Grade 8	189 (45%)	91 (22%)	66 (16%)	70 (17%)
High School	112 (35%)	105 (32%)	25 (8%)	82 (25%)
Writing				
Grade 3	128 (32%)	185 (47%)	36 (9%)	45 (11%)
Grade 4	177 (41%)	110 (26%)	96 (22%)	47 (11%)
Grade 5	164 (42%)	131 (33%)	35 (9%)	64 (16%)
Grade 6	184 (42%)	105 (24%)	93 (21%)	61 (14%)
Grade 7	150 (35%)	118 (28%)	95 (22%)	65 (15%)
Grade 8	195 (47%)	108 (26%)	39 (9%)	77 (18%)
High School	117 (36%)	93 (29%)	31 (10%)	83 (26%)

For subscores, the most common AMOs across years were AMO 1 (6 of 7 grades) and AMO 2 (1 of 7 grades) for writing, AMO 1 (4 of 7 grades) and AMO 2 (3 of 7 grades) for reading.

These subscores—compared to math, science, and overall ELA—display broader coverage of ELA categories across grades, on average. Compared to other grades, grades 7 and 8 writing have relatively higher AMO 1 groups. For these grades, a better balance may be seen if existing difficult items are replaced with an easier ones.

3.4 Validity Based on Relations to Other Variables

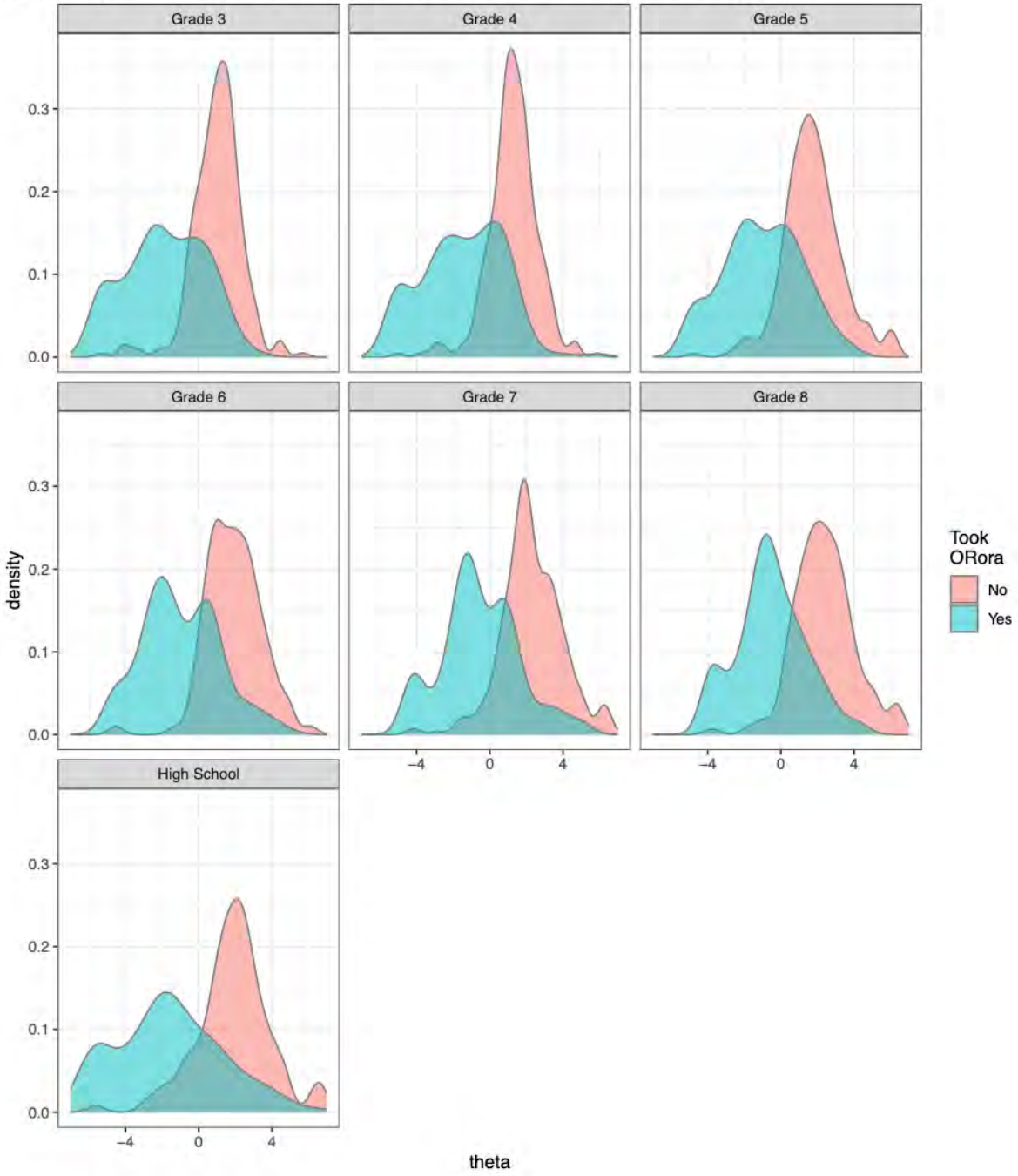
Perhaps the best model for understanding criterion-related evidence comes from Campbell and Fiske (1959), in their description of the multi-trait, multi-method analysis [translate the term 'trait' to mean 'skill']. In this process (several) different traits are measured using (several) different methods 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. Data is presented that quite consistently reflects higher relations among items within an academic subject than between academic subjects. Data is also present which performance on items is totaled within categories of disability, expecting relations that would reflect appropriate differences Tindal et al. (2003).

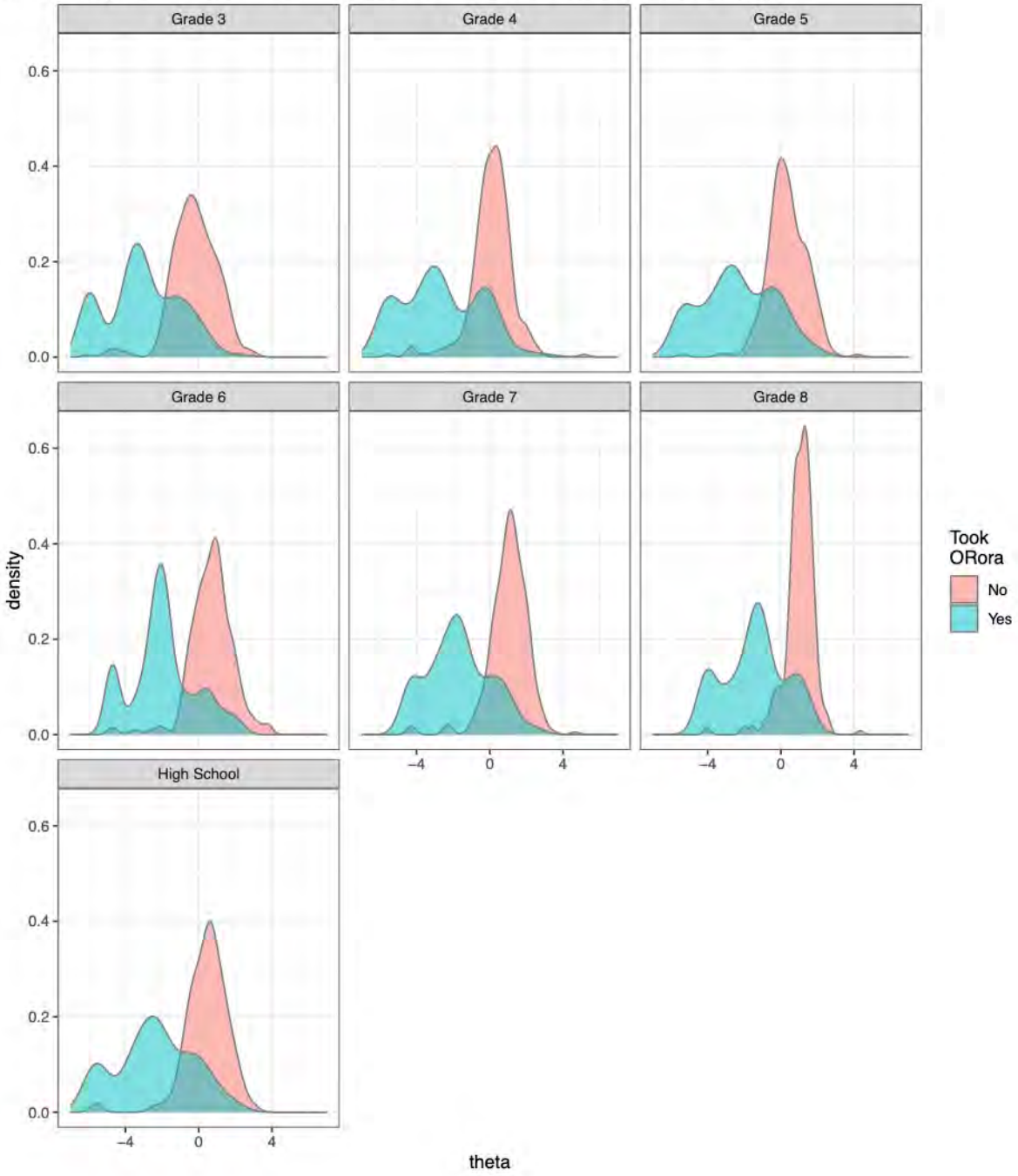
3.4A Convergent and Divergent Validity Documentation

Criterion validity information is difficult to document with AA-AAAS, as most SWSCD do not participate in any standardized assessment outside of the ORExt and/or ORora in Oregon. Divergent validity evidence is garnered via comparisons of ORExt results to ORora outcomes and shows that students whose ORExt assessments are discontinued exhibit serious limitations in attention, basic math skills, and receptive and expressive communication skills. Density distributions show that there are very different measures of central tendency (i.e., means and medians) for all grades and contents, when comparing those who did and did not take ORora.

ELA ability estimates (theta) for those who did vs. not not take the ORora, by grade



Math ability estimates (theta) for those who did vs. not not take the ORora, by grade



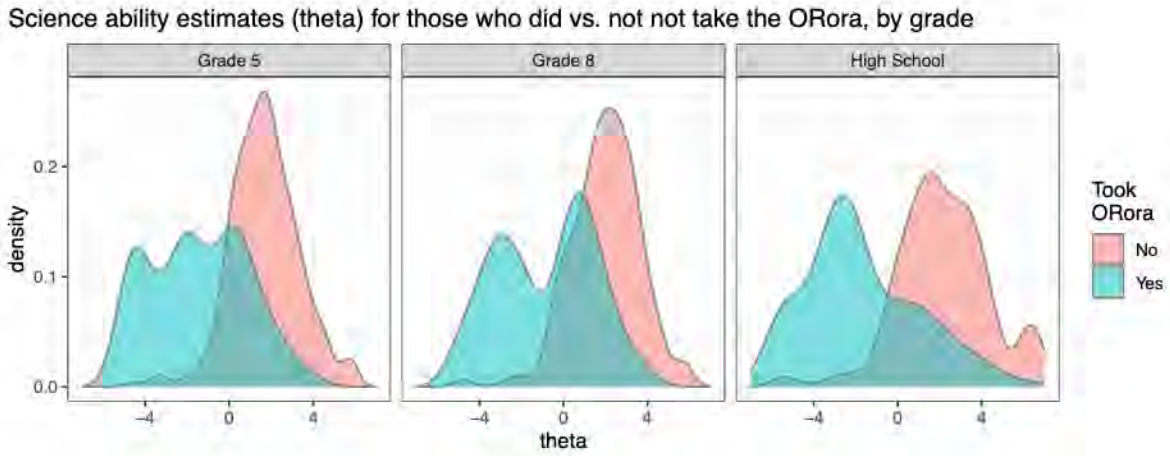


Table 3.14: Pearson correlation of ORora and Content areas, for those who took ORora and at least one ORExt test

	ORora	ELA	Math	Science
ORora	1	–	–	–
ELA	0.369	1	–	–
Math	0.318	0.626	1	–
Science	0.41	0.694	0.587	1

Pearson correlations between the total raw scores on the ORExt and the total raw score on the ORora were conducted to address the relationship between total performance on each assessment. The correlation between ELA and ORora scores was 0.369, between Math and ORora scores was 0.318, and between Science and ORora scores was 0.41. As expected, the ORora results provide divergent validity evidence for the ORExt. A strong relationship is not expected between the scores, as students whose OREXt testing is discontinued are generally unable to access the academic content on the OREXt, even with the requisite reductions in depth, breadth, and complexity.

Furthermore, content area correlations other than ORora (i.e., ELA ~ Math, ELA ~ Science, Math ~ Science) are different than these correlation for those who did not take the ORora (See content area correlations below).

Convergent evidence that the OREXt is assessing appropriate academic content is provided by QA and QT responses to the consequential validity survey. Respondents to the survey generally agree that, "The items in the Oregon Extended Assessment accurately reflect the academic content (what the student should know) that my students with significant cognitive disabilities should be learning, as defined by grade level content standards (CCSS/NGSS) and the Essentialized Assessment Frameworks" (85% Strongly Agree or Agree). In addition, they also agreed with the statement that, "The items in the Oregon Extended Assessment,

which primarily ask students to match, identify, or recognize academic content, are appropriate behaviors to review to determine what my students with significant cognitive disabilities are able to do" (85% Strongly Agree or Agree). The consequential validity results demonstrate that the OREXt is sampling academic domains that the field of QAs and QTs deem appropriate in the area of academics. See the Consequential Validity Survey Results for complete consequential validity study results.

3.4B Analyses Within and Across Subject Areas

Correlational analyses was conducted to further explore the validity of the OREXt. The purpose of the analysis was described, as well as our anticipated results. Then observed results were discuss before concluding with an overall evaluative judgment of the validity of the test.

Correlational analysis were explore 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. However, tests would correlate quite strongly given that the same students were assessed multiple times. Therefore, moderately strong correlations (e.g., .70 - .90) would be expected simply because of the within-subject design. Idiosyncratic variance associated with the individual student is thus captured.

3.4C Correlational Analyses Results

Full results of the Pearson's product-moment correlation analysis by content area and grade level are reported below. The results are significant, yet the overall correlations across content areas suggest that different, though strongly related, constructs are being measured.

Table 3.15: Content Area Correlations (years without science test)

	ELA	Math	Reading	Writing
Grade 3				
ELA	1			
Math	0.86	1		
Reading	0.97	0.81	1	
Writing	0.93	0.82	0.85	1

Table 3.15: Content Area Correlations (years without science test) *(continued)*

	ELA	Math	Reading	Writing
Grade 4				
ELA	1			
Math	0.79	1		
Reading	0.97	0.77	1	
Writing	0.92	0.71	0.83	1
Grade 6				
ELA	1			
Math	0.79	1		
Reading	0.97	0.78	1	
Writing	0.95	0.76	0.87	1
Grade 7				
ELA	1			
Math	0.82	1		
Reading	0.97	0.78	1	
Writing	0.94	0.8	0.85	1

Table 3.16: Content Area Correlations (years with science test)

	ELA	Math	Reading	Science	Writing
Grade 5					
ELA	1				
Math	0.81	1			
Reading	0.97	0.78	1		
Science	0.79	0.78	0.75	1	
Writing	0.93	0.76	0.84	0.77	1
Grade 8					
ELA	1				
Math	0.73	1			
Reading	0.97	0.69	1		
Science	0.81	0.78	0.78	1	
Writing	0.93	0.72	0.85	0.79	1
High School					
ELA	1				
Math	0.79	1			
Reading	0.97	0.77	1		
Science	0.84	0.82	0.84	1	

Table 3.16: Content Area Correlations (years with science test) *(continued)*

	ELA	Math	Reading	Science	Writing
Writing	0.96	0.77	0.89	0.81	1

Results of the Pearson's product-moment correlation analysis across domains (i.e., ELA, Science, and Math) ranged from:

- ELA and math: 0.733 to 0.859
- ELA and science: 0.789 to 0.836
- Math and science: 0.779 to 0.82

Across domains, higher scores are certainly correlated, with those scoring higher on any test being likely to score highly on another. However, these correlations are low enough to support that different cognitive domains are being measured.

For ELA and it's subdomains (i.e., ELA:Reading:Writing), correlations of:

- ELA and reading: 0.966 to 0.975
- ELA and writing: 0.919 to 0.96
- Reading and writing: 0.826 to 0.887.

Within subdomains of ELA, very high correlations are observed. ELA and reading may be so correlated that they are measuring nearly the same information. Reading and writing display lower correlation with one another, though, supporting the assumption that they are measuring unique constructs.

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. Typical strategies for addressing reliability include documentation of internal consistency, split-half reliability, and test-retest reliability. The implementation plan for the ORExt included initial documentation of internal consistency (Cronbach's alpha). The 2015-16 technical report included internal consistency estimates, split-half reliability analyses, as well as a small test-retest assessment of reliability comparisons by means of the pilot tablet administration study. There is only one test form for the ORExt, so test form comparisons are not possible.

4.1.1 Inter-Rater-Reliability

4.1.1.1 Background

Pursuant to Hallgren (2012) the assessment of IRR may be necessary to demonstrate consistency among observational ratings provided by multiple assessors. The results of the study will be used to address the requirements within the USED's Peer Review process (Critical Element 4.1). A sample of Oregon's Qualified Assessors (QAs) who administer the paper/pencil version of the Oregon Extended Assessment (ORExt) were observed to determine reliability of administration and scoring. A tablet administration was not included in the Oregon Observational Rating.

4.1.1.2 Methods

QTs in districts across the state observe a sample of their respective QAs using the observation protocol (see Oregon Extended Assessment Technical Report on Standard Setting) and enter their data online. The QA reads the item stem and the student selects from three possible answer choices (A, B, or C) then, the QA records the answer choice. QTs (observer) records the students answer choice, then records the answer choice recorded by the QA for agreement. Only the English Language Arts Writing porting of the ORExt requires additional analysis by the assessor to determine if the written response (answer) meets (1) or doesn't meet (0) provided critera. Districts from across the state of Oregon participated in

the study, matching the state's student population demographics, including large, medium, and small districts, across all regions. The observation protocol was completed for the identified QA, but the student(s) and content area(s) observed were selected by the QT or QA. BRT researchers contacted district-level QTs at the beginning of the test window, which runs from February 15 - April 26, 2018, to arrange observations that could hopefully be completed within one school day. In addition to addressing inter-rater reliability, the study also evaluated test administration procedures. The methods, results, and interpretation are provided here, in addition to recommended next steps. The observation was composed of three sections:

- First, QT's reviewed ORExt paper/pencil test preparation and administration using the rubric (see Oregon Extended Assessment Technical Report on Standard Setting). Test preparation/administration domains were rated on a four-point scale from Inappropriate (I) to Exemplary (E):
  - Inappropriate (I) denotes a level of concern that could clearly affect the accuracy of the test results gathered from the test administration. Ratings at this level require substantive retraining of the QA involved.
  - Somewhat Appropriate (SA) rating denotes a level that includes some minor aspects that could be improved, but the accuracy of the test results are likely not compromised.
  - Appropriate (A) denotes a level that is consistent with all test administration requirements.
  - Exemplary (E) level performance suggests that the QA incorporated approaches to test administration that could become models for best practice.
- Second, QT's scored the student alongside the QA using the scoring sheet. QT's compared results after this observation to ensure that the QA entered accurate data.
- Finally, QT's observed the QA completing the data entry process to ensure that no errors are made during data entry and document the number of errors (see Oregon Extended Assessment Technical Report on Standard Setting). #### Domain Definitions

1. Test Security – The QA utilized a system to ensure that all test materials were stored in a secure location,. The QA also had a district Assurance of Test Security form on file.
2. Printed Materials – the QA had all materials required to administer the ORExt ready for test administration.
3. Distraction-Free Environment – the QA arranged to provide the ORExt in a one-on-one test administration in a location that ensured that the student focused attention on the assessment.
4. Accessibility Supports – the QA provided all necessary accessibility supports for the student and ensured that all support systems were functional prior to testing.
5. Level of Support – The QA provided an appropriate level of support throughout testing that did not compromise the validity of the score.

6. Praise – The QA utilized praise appropriately to support student involvement without leading the student to the correct answer.
7. Motivation – The QA appropriately maintained the student's motivation during the assessment using relevant strategies, such as token systems.
8. Score Interpretation – The QA demonstrated an appropriate understanding of how to use the cut scores and achievement level descriptors to interpret scores (i.e., ask the QA to describe how they interpret scores for parents).
9. Minimum Participation Rule - The QA demonstrated an appropriate understanding of the minimum participation rule (i.e., ask the QA to define the rule if it is not used).
10. Qualified Assessor Testing Preparation and Administration Rubric - Participants are told to record an "X" in the cell that corresponds to their rating. An example of a filled out form is shown below (example made by most common response by item).

Table 4.1: Example Responses

Domain	Exemplary	Appropriate	Somewhat Appropriate	Inappropriate
Accessibility Supports		X		
Distraction Free		X		
Level Support	X			
Minimum Participation	X	X		
Motivation		X		
Praise	X			
Printed Materials	X			
Score Interpretation		X		
Test Security	X			

Note:
There was a tie on Minimum Participation between "Exemplary" and "Appropriate"

4.1.1.3 Inter-rater Agreement Results

Qualified Trainers (n = 25) from around Oregon participated in the Inter-Rater-Reliability study by doing at least one observation on the Oregon Extended Assessment via paper/pencil administration. Not all subjects were equally represented ELA (40%), Math (40%), Science (20%). Observations were done at individual student's typical testing location.

The following two tables display the percentage of responses in the nine different domains and percentage of agreement between assessors and observers.

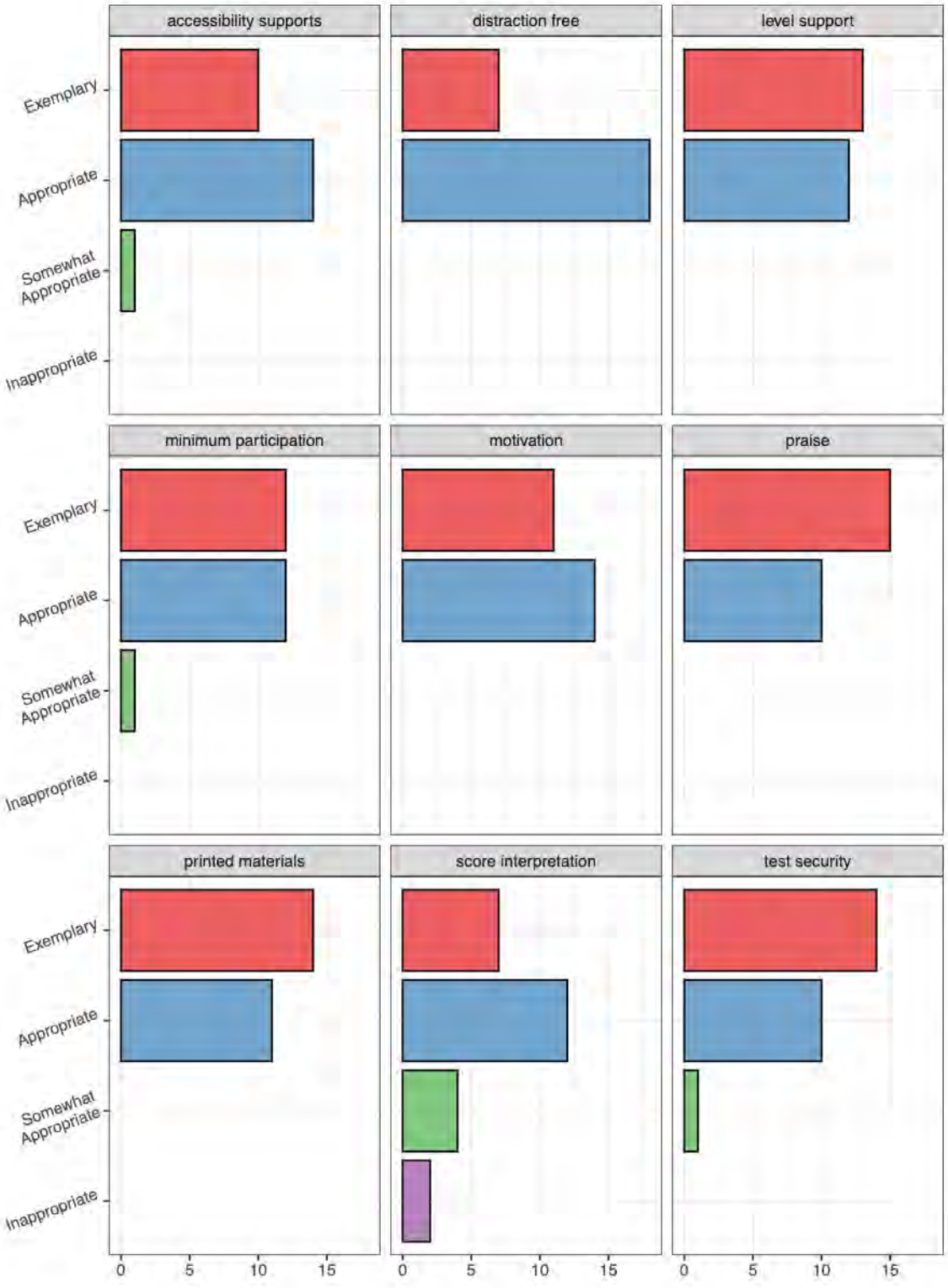
Table 4.2: Percentage for responses

Domain	Exemplary	Appropriate	Somewhat Appropriate	Inappropriate
Accessibility Supports	40%	56%	4%	0%
Distraction Free	28%	72%	0%	0%
Level Support	52%	48%	0%	0%
Minimum Participation	48%	48%	4%	0%
Motivation	44%	56%	0%	0%
Praise	60%	40%	0%	0%
Printed Materials	56%	44%	0%	0%
Score Interpretation	28%	48%	16%	8%
Test Security	56%	40%	4%	0%

Table 4.3: Student Answers and Agreement between QA and QT

	Responses
Student Answer Correct (QA and QT Agreed)	645 (53.75%)
Student Answer Incorrect (QA and QT Agreed)	310 (25.83%)
Not Administered	244 (20.33%)
QA said Student Answer Correct; QT Disagreed	1 (0.08%)

The following plots provides a visual display of the responses from the nine different domains observed.


accessibility supports
distraction free
level support
minimum participation
motivation
praise
printed materials
score interpretation
test security
Exemplary
Appropriate
Somewhat Appropriate
Inappropriate
0
5
10
15

4.1.1.4 Results:

OExt's Selected response format provides for a high percentage of inter-rater reliability. One response out of the 1200 observed where observes disagreed with raters was in the ELA Writing scoring. 'Score Interpretation' appears to be a domain in need of additional training. Qualified Trainers indicated that 16% of observed Qualified Assessors were Somewhat Appropriate and 8% were Inappropriate in their understanding of how to use cut scores and achievement level descriptors to interpret scores.

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 0.8 to 0.94. Each table below provides the content area, grade, and the marginal reliabilities. All test forms were composed of 36 operational items; marginal reliability was only calculated with operational items.

Table 4.4: Marginal Reliability by Content and Grade

	ELA	Math	Science	Reading	Writing
Grade 3	0.92	0.9	–	0.88	0.82
Grade 4	0.92	0.9	–	0.88	0.83
Grade 5	0.92	0.9	0.91	0.88	0.82
Grade 6	0.92	0.9	–	0.88	0.82
Grade 7	0.92	0.89	–	0.87	0.82
Grade 8	0.91	0.87	0.92	0.86	0.8
High School	0.94	0.92	0.94	0.91	0.91

4.1B Test Information Functions

Test information functions shown below indicate cutpoints as vertical lines and acceptable marginal reliabilities (specifically dark gray = 0.8 and light gray = 0.7). All tests have the first two thresholds better than 0.8; some grades and contents include the final cut above 0.8.

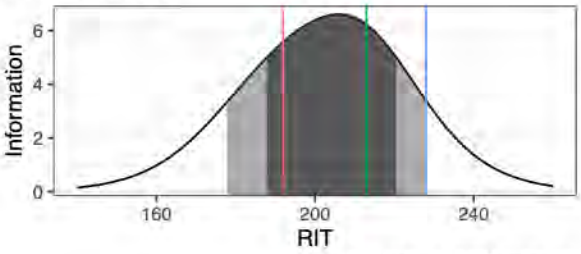
In a few cases, information at the final cut is not quite 0.8, but is still above 0.7. These can be seen as the tests with blue lines in the lighter gray region.

This indicates that at the more difficult ends of the test, math items are providing more true variance, relative to the error variance. If other subjects' items can be more greatly aligned with the constructs, these test information functions can be improved. Alternatively, more difficult item may be added to tests to improve the ratio of true variance to error variance in these tests.

4.1C English Language Arts TIFs

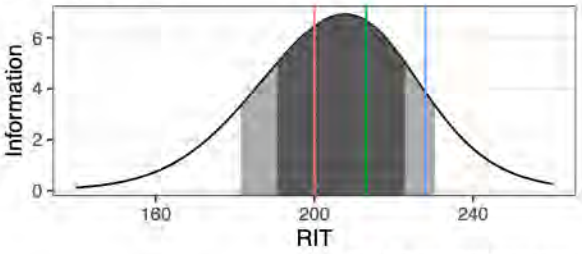
All tests except grade 7 have the final cutpoint in the 0.7-0.8 marginal reliability range. To continue using this cutpoint, future iterations should replace some simpler items with more difficult items. Seventh grade's final cutpoint has only adequate marginal reliability.

ELA: Grade 3



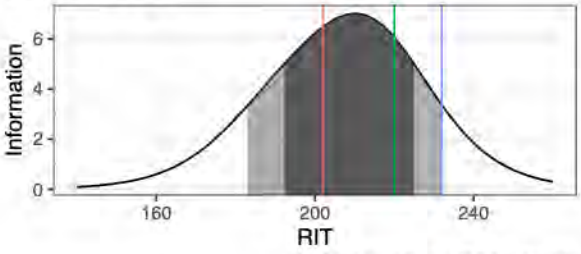
Lines indicate AMO cutpoints
marginal reliability 0.7 and 0.8 shaded

ELA: Grade 4



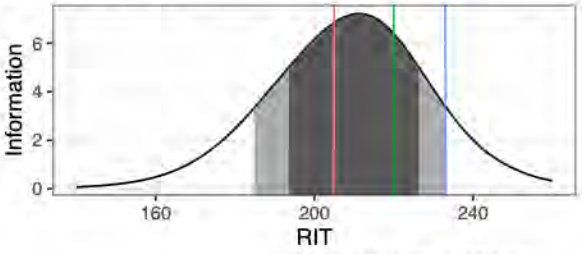
Lines indicate AMO cutpoints
marginal reliability 0.7 and 0.8 shaded

ELA: Grade 5



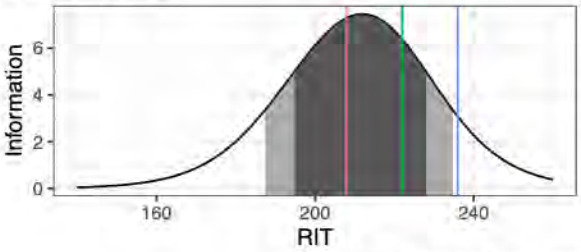
Lines indicate AMO cutpoints
marginal reliability 0.7 and 0.8 shaded

ELA: Grade 6



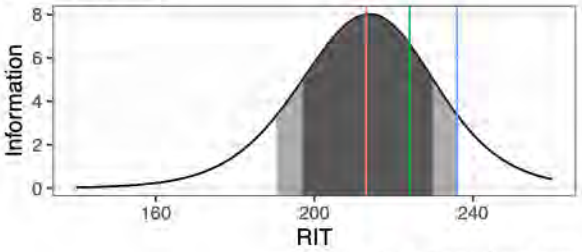
Lines indicate AMO cutpoints
marginal reliability 0.7 and 0.8 shaded

ELA: Grade 7



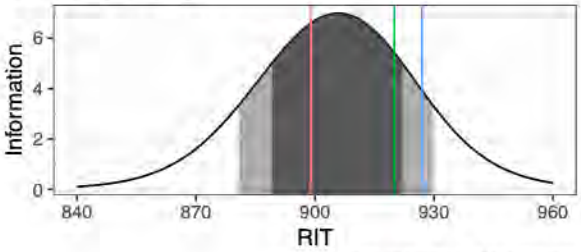
Lines indicate AMO cutpoints
marginal reliability 0.7 and 0.8 shaded

ELA: Grade 8



Lines indicate AMO cutpoints
marginal reliability 0.7 and 0.8 shaded

ELA: High School

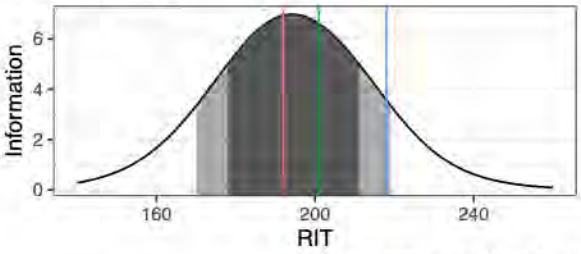


Lines indicate AMO cutpoints
marginal reliability 0.7 and 0.8 shaded

4.1D Mathematics TIFs

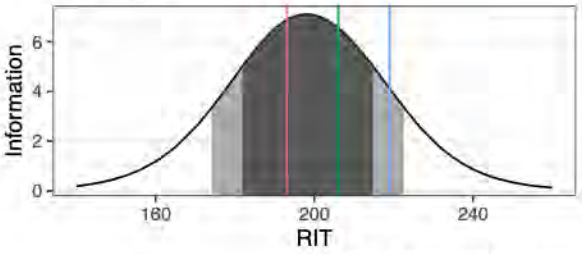
Math grades 7 and 8 have upper cutpoints at or above a marginal reliability of 0.8, and all others are between 0.7 and 0.8. Future math forms should include harder item with particular emphasis on grade 3, grade 4 and high school.

Math: Grade 3



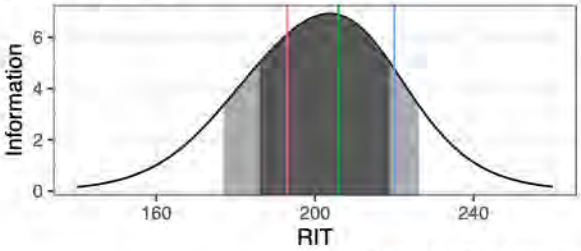
Lines indicate AMO cutpoints
marginal reliability 0.7 and 0.8 shaded

Math: Grade 4



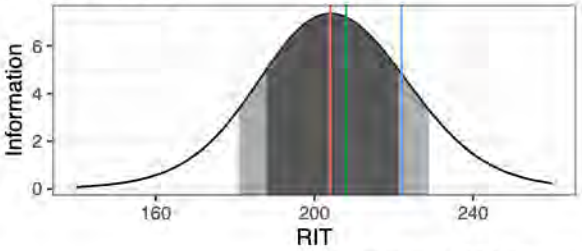
Lines indicate AMO cutpoints
marginal reliability 0.7 and 0.8 shaded

Math: Grade 5



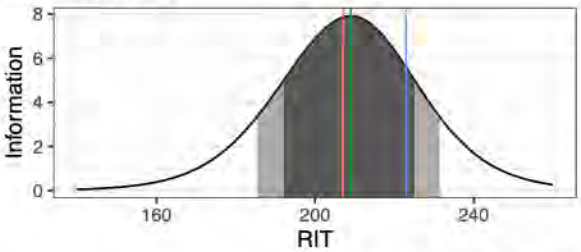
Lines indicate AMO cutpoints
marginal reliability 0.7 and 0.8 shaded

Math: Grade 6



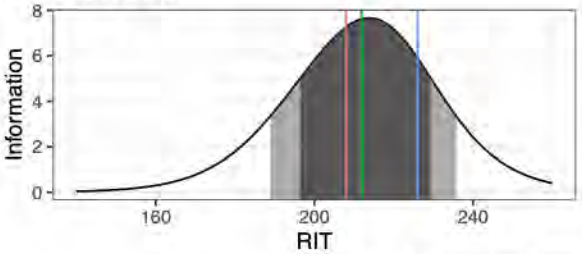
Lines indicate AMO cutpoints
marginal reliability 0.7 and 0.8 shaded

Math: Grade 7



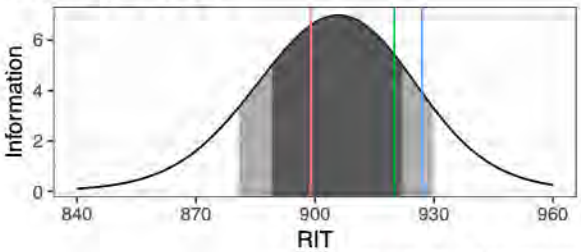
Lines indicate AMO cutpoints
marginal reliability 0.7 and 0.8 shaded

Math: Grade 8



Lines indicate AMO cutpoints
marginal reliability 0.7 and 0.8 shaded

Math: High School

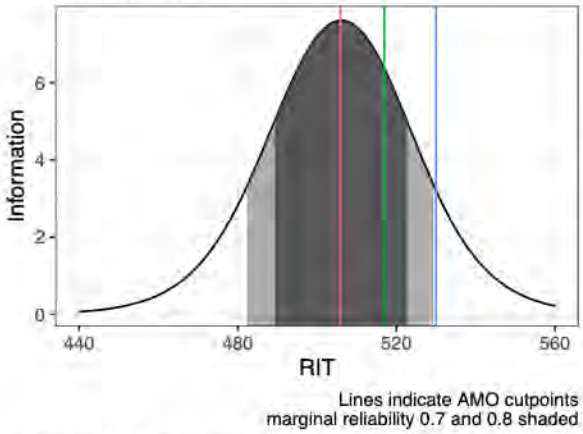


Lines indicate AMO cutpoints
marginal reliability 0.7 and 0.8 shaded

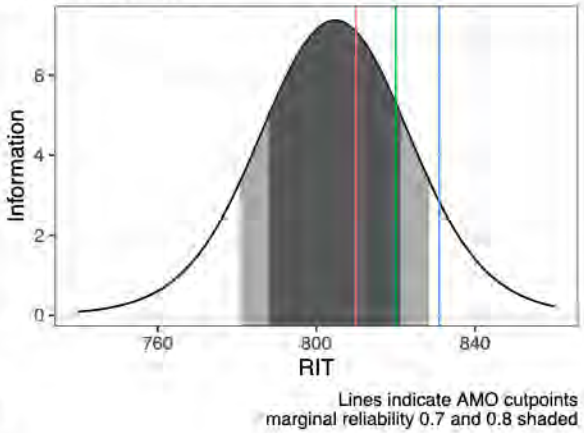
4.1E Science TIFs

For all science grades, the final cutpoint is just short of the desired 0.7 mark. Future tests should exchange some simpler items with more difficult items, especially in grade 8. Importantly, cuts other than "exceeds" cut are well within desired ranges.

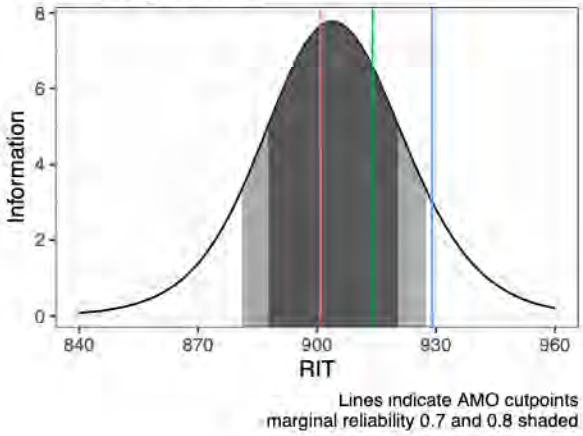
Science: Grade 5



Science: Grade 8



Science: High School



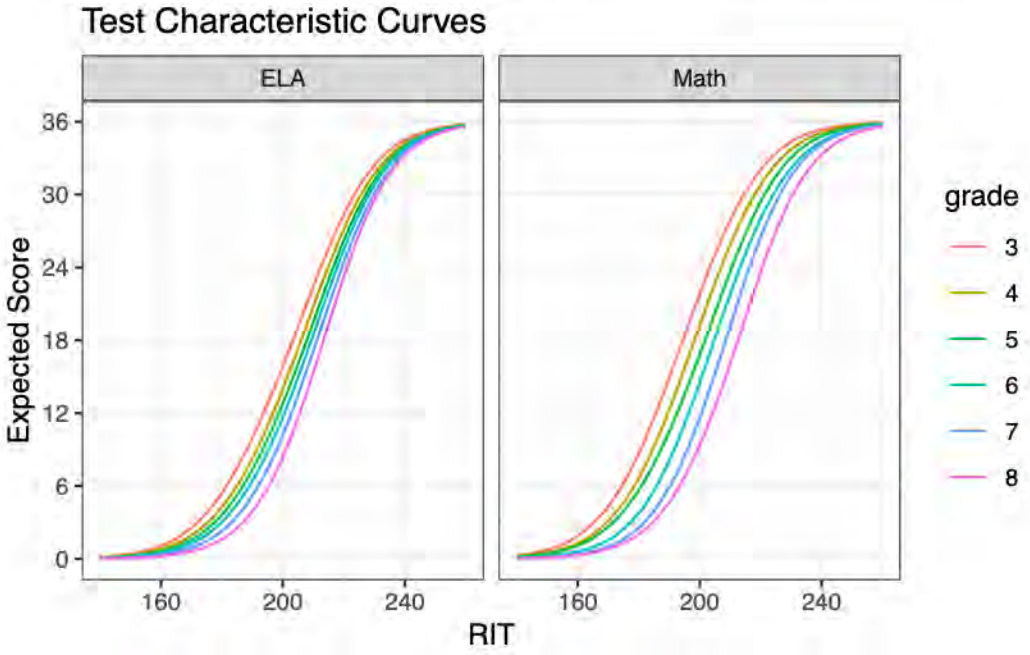
4.1F Validation of OrExt Vertical Scales

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, Grade 11 and science tests are not vertically scaled; TCCs are thus not presented for Grade 11 or science. All Rasch model scaling, as well as the data visualizations for the TCCs were conducted in the R software version 4.2.3 environment (R Core Team, 2021) using the exirt package (Anderson and Loan (2022)), built upon the TAM package for Rasch modeling

(Robitzsch, Kiefer, and Wu (2022)). This package framework estimates analogously to Winsteps, but with additional functionality and better integration with important visualization and analytic ecosystems.

Test characteristic curves demonstrate a very clear vertical scale. Improving on last year's assessment which displayed one small cross-over, no tests cross over. This is important for demonstrating the consistency with the vertical scale. Additionally, the spacing of curves looks roughly even across within tests. If one area were modified preferably, future tests should focus on the upper end of ELA, particularly in higher grades. This is where the greatest compression of either scale is seen.

The magnitude of spacing for the math assessment is slightly preferable to ELA whereby greater distinctions exist between test forms.



It is clear that the overall testing framework is vertically integrated and able to distinguish progress through their respective constructs.

Taking this information alongside the TIFs, systematic progress can be most easily achieved by making all test forms slightly more difficult. For the most part, difficulty should be added to test-forms in parallel, with slighltly more focus on upper-end ELA tests.

4.1G Overall and Conditional Standard Errors of Measurement (SEM)

Annual Measurable Objectives (AMO) are yearly learning targets set by the state in ELA, Math, and Science. Standard Error of Measurement (SEM) estimates how repeated measures

of a person on the same instrument tend to be distributed around his or her "true" score. The average SEM associated with each cut score for 2022-23 student data are presented in the table below. See Section 4.2 below for means and standard deviations by grade and subject area. The numbers below, AMO is on the left and the SEM associated with the cut score is in parentheses, rounded to two decimals.

The AMO 2 cutscore represents the threshold between AMO 1 and AMO 2; similarly AMO 3 cutscore represents the threshold between AMO 2 and 3; AMO 4 represents the cutscore between AMO 3 and 4.

AMO levels correspond to the following descriptions:

- AMO Level 1 = Does Not Yet Meet
- AMO Level 2 = Nearly Meets
- AMO Level 3 = Meets
- AMO Level 4 = Exceeds

Table 4.5: Cutscore (Conditional Standard Error of Measurement) by Content and Grade

	AMO 2 Cutscore (SEM)	AMO 3 Cutscore (SEM)	AMO 4 Cutscore (SEM)
ELA			
Grade 3	194 (4.16)	214 (4.05)	229 (5.56)
Grade 4	202 (3.89)	215 (3.93)	231 (5.55)
Grade 5	203 (3.88)	221 (4.15)	236 (6.21)
Grade 6	206 (3.78)	221 (4.01)	237 (6.19)
Grade 7	210 (3.67)	224 (4.11)	238 (6.19)
Grade 8	215 (3.53)	225 (3.99)	239 (6.12)
High School	901 (2.72)	921 (3.09)	928 (3.62)
Math			
Grade 3	193 (3.79)	202 (3.92)	220 (5.59)
Grade 4	194 (3.79)	207 (3.95)	221 (5.14)
Grade 5	194 (3.99)	208 (3.84)	221 (4.73)
Grade 6	205 (3.69)	209 (3.75)	224 (4.8)
Grade 7	209 (3.56)	210 (3.56)	224 (4.37)
Grade 8	209 (3.7)	213 (3.62)	227 (4.18)
High School	902 (2.6)	908 (2.63)	924 (3.53)
Science			
Grade 5	507 (3.63)	519 (4.09)	533 (6.18)
Grade 8	812 (3.81)	822 (4.5)	832 (6.25)
High School	902 (2.55)	915 (2.82)	932 (4.44)

Table 4.5: Cutscore (Conditional Standard Error of Measurement) by Content and Grade *(continued)*

	AMO 2 Cutscore (SEM)	AMO 3 Cutscore (SEM)	AMO 4 Cutscore (SEM)

Note:
AMO = Annual Measureable Objective
SEM = Standard Error of Measurement associated with the cut score

4.1H Classification Accuracy & Consistency

Results from the 2022-23 OrExt test administration were analyzed 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 (consistency). The calculation utilizes item difficulty and theta value distributions, as well as related standard errors of measurement, to generate probabilistic estimates based on one test administration. Complete results, generated from the cacIRT package in R (Lathrop (2015)), are provided below. Results denote very high levels of classification accuracy and consistency.

Table 4.6: Classification Accuracy and Consistency, by content and grade

	AMO 2 (Nearly Meets)	AMO 3 (Meets)	AMO 4 (Exceeds)
ELA			
Grade 3	Acc: 0.96 | Con: 0.94	Acc: 0.92 | Con: 0.88	Acc: 0.97 | Con: 0.96
Grade 4	Acc: 0.94 | Con: 0.92	Acc: 0.92 | Con: 0.88	Acc: 0.97 | Con: 0.96
Grade 5	Acc: 0.94 | Con: 0.92	Acc: 0.93 | Con: 0.91	Acc: 0.96 | Con: 0.95
Grade 6	Acc: 0.93 | Con: 0.9	Acc: 0.94 | Con: 0.91	Acc: 0.95 | Con: 0.94
Grade 7	Acc: 0.94 | Con: 0.92	Acc: 0.94 | Con: 0.92	Acc: 0.95 | Con: 0.93
Grade 8	Acc: 0.94 | Con: 0.92	Acc: 0.94 | Con: 0.91	Acc: 0.95 | Con: 0.94
High School	Acc: 0.98 | Con: 0.97	Acc: 0.95 | Con: 0.93	Acc: 0.95 | Con: 0.93

Table 4.6: Classification Accuracy and Consistency, by content and grade *(continued)*

	AMO 2 (Nearly Meets)	AMO 3 (Meets)	AMO 4 (Exceeds)
Math			
Grade 3	Acc: 0.92 | Con: 0.89	Acc: 0.93 | Con: 0.9	Acc: 0.98 | Con: 0.97
Grade 4	Acc: 0.93 | Con: 0.9	Acc: 0.92 | Con: 0.88	Acc: 0.98 | Con: 0.98
Grade 5	Acc: 0.92 | Con: 0.89	Acc: 0.93 | Con: 0.9	Acc: 0.97 | Con: 0.96
Grade 6	Acc: 0.92 | Con: 0.88	Acc: 0.91 | Con: 0.88	Acc: 0.97 | Con: 0.96
Grade 7	Acc: 0.91 | Con: 0.87	Acc: 0.9 | Con: 0.87	Acc: 0.97 | Con: 0.95
Grade 8	Acc: 0.88 | Con: 0.83	Acc: 0.87 | Con: 0.82	Acc: 0.99 | Con: 0.99
High School	Acc: 0.94 | Con: 0.92	Acc: 0.94 | Con: 0.92	Acc: 0.99 | Con: 0.98
Science			
Grade 5	Acc: 0.94 | Con: 0.92	Acc: 0.93 | Con: 0.9	Acc: 0.95 | Con: 0.93
Grade 8	Acc: 0.94 | Con: 0.91	Acc: 0.92 | Con: 0.89	Acc: 0.92 | Con: 0.9
High School	Acc: 0.98 | Con: 0.97	Acc: 0.96 | Con: 0.94	Acc: 0.95 | Con: 0.93

For ELA accuracies are seen of 0.92 to 0.98 and consistencies of 0.88 to 0.97; For Math accuracies are seen of 0.87 to 0.99 and consistencies of 0.82 to 0.99; For Science accuracies are seen of 0.92 to 0.98 and consistencies of 0.89 to 0.97.

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 Oregon Extended Assessment Alignment Study). In addition, test functioning was reviewed 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.

4.2A Differential Item Functioning Analyses

To investigate Differential Item Functioning (DIF), the Mantel-Haenszel test using a purification process was conducted (Holland and Thayer 1988; Kamata and Vaughn 2004) with the R software using the difR package (Magis et al., 2013). When using the Mantel-Haenszel test to investigate DIF, contingency tables are constructed, and the resulting odds for the focal group answering the item correctly are compared to the odds for the reference group. Given n-size limitations (Scott et al. (2009)), we were able to conduct two analyses: a) White/Non-White and b) Male/Female. Whites and Males were the focal groups and Non-Whites and Females were the reference groups, respectively. The contingency table summarizes correct and incorrect responses to each item by respondents' total raw score by subgroup (Kamata and Vaughn (2004)). If there is no difference in performance for the two groups, the odds ratio of the focal group performance to reference group performance will equal one. An odds ratio greater than one means the focal group is performing better than the reference group, with the opposite being true for odds ratios less than one.

The difR package contains a built in algorithm to conduct purification automatically, of interest was how this algorithm functioned relative to the iterations conducted manually using SPSS. Criteria was used as outlined by the Educational Testing Service (ETS) for DIF Classification (Holland and Thayer (1988)) to determine whether or not items exhibited DIF. The Holland and Thayer criteria were used for all Mantel-Haenszel analyses. As the difR package reports delta values by default, defined as

$$\Delta_{MH} = -2.35 * ln(\alpha_{MH})$$

.

It is common for negligible differences to exist between groups in data, for that reason, the focus of the DIF analysis is to remove items which display larger DIF. Below, a 3 category magnitude of effect for DIF is reported, with increasing magnitudes of the difference: A, B, and C. Refer to these as "Negligible", "Small" and "Substantial", respectively.

i DIF Grades

- A: $0 > \delta <= 1$
- B: $1 > \delta <= 1.5$
- C: $1.5 > \delta$

Items that were flagged as "C" were reviewed by BRT researchers for potential biases. If biases are identified, the item is removed from the item pool. Items categorized as "A" or "B" were considered smaller differences and do not rise to the level of exclusion in this wave of test administration. Continued examination of these items over time will ensure the differences remain below the threshold of $\delta < 1.5$.

Table 4.7: Differential Item Functioning Female vs. Male

	Negligible Differences (A)	Small Differences (B)	Substantial Differences (C)
ELA			
Grade 3	25	8	3
Grade 4	28	7	1
Grade 5	29	3	4
Grade 6	27	8	1
Grade 7	30	2	4
Grade 8	28	8	0
High School	28	7	1
Math			
Grade 3	27	7	2
Grade 4	31	3	2
Grade 5	32	4	0
Grade 6	32	1	3
Grade 7	30	5	1
Grade 8	32	2	2
High School	30	3	3
Science			
Grade 5	30	4	2
Grade 8	26	5	5
High School	27	4	5

In terms of the Male/Female analyses, a total of 39 items were flagged as "C". This means there were substantial differences in the performance of those items by coded student sex, at the same ability level. Of these ELA favored 5 items for girls, Math favored 9 items for girls, Science favored 7 items for girls. That means that ELA favored 9 items for boys, Math favored 4 items for boys, Science favored 5 items for boys. This suggests that the items display bias towards female more often than male by a total of 3 items. Items will be modify, such that items do not favor one group over the other; if that is not possible, achieving balance between number of items which favor these groups is important.

Table 4.8: Differential Item Functioning Non-White vs. White

	Negligible Differences (A)	Small Differences (B)	Substantial Differences (C)
ELA			
Grade 3	34	2	0
Grade 4	30	4	2
Grade 5	34	1	1
Grade 6	26	8	2
Grade 7	33	3	0
Grade 8	31	5	0
High School	31	5	0
Math			
Grade 3	30	6	0
Grade 4	33	2	1
Grade 5	32	2	2
Grade 6	29	4	3
Grade 7	33	3	0
Grade 8	33	2	1
High School	29	4	3
Science			
Grade 5	32	1	3
Grade 8	32	2	2
High School	28	6	2

In terms of the White/non-White analyses, a total of 22 items were flagged as "C". Of these ELA favored 4 items for white participants, Math favored 6 items for white participants, Science favored 4 items for white participants. That means that ELA favored 1 items for non-white participants, Math favored 4 items for non-white participants, Science favored 3 items for non-white participants. This suggests that the items display bias towards White more often than non-White by a total of 6 items. A balance of these items will be achieved in the next test administration. Items will be modify such that items do not favor one group over the other; if that is not possible, achieving balance between number of items which favor these groups is important.

4.2B Race - Ethnicity Percentages and Totals by Content Area and Grade Level

Table 4.9: Racial Ethnic Demographic Percentages of Sample

	White	Hispanic	Multi-ethnic	Asian	Black	Native American or Alaskan
Grade 3	208 (53%)	110 (28%)	26 (7%)	22 (6%)	13 (3%)	–
Grade 4	210 (48%)	140 (32%)	21 (5%)	27 (6%)	19 (4%)	12 (3%)
Grade 5	204 (51%)	127 (32%)	32 (8%)	22 (5%)	–	–
Grade 6	243 (54%)	127 (28%)	27 (6%)	18 (4%)	20 (4%)	–
Grade 7	224 (52%)	144 (33%)	27 (6%)	12 (3%)	17 (4%)	–
Grade 8	218 (51%)	128 (30%)	34 (8%)	15 (4%)	20 (5%)	–
High School	186 (55%)	97 (29%)	21 (6%)	17 (5%)	–	–

Note:
Samples Less than 10 Omitted for Privacy
Pacific Islander n < 10 at all grades

The full ethnic and disability demographics for students taking the ORExt are reported below. Students ethnicity/race was reported in seven categories: (a) American Indian/Alaskan Native, (b) Asian, (c) Black or African-American, (d) Multi-ethnic, (e) Native Hawaiian or Other Pacific Islander, (f) Hispanic, or (g) White. Across grades, the majority of students were reported as White (48% to 55%). 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 State Annual Report Card).

4.2C Exceptionality Percentages By Content Area and Grade Level

Autism Spectrum Disorder was the most common in Grade 3 (48.23%); Autism Spectrum Disorder was the most common in Grade 4 (42.17%); Autism Spectrum Disorder was the most common in Grade 5 (44.14%); Autism Spectrum Disorder was the most common in Grade 6 (41.83%); Intellectual Disability was the most common in Grade 7 (39.35%);

Intellectual Disability was the most common in Grade 8 (40.71%); Intellectual Disability was the most common in Grade 11 (51.04%).

Table 4.10: Distribution of Primary IDEA Codes by Grade

IDEA Code	Amount of Sample
Grade 3	
Autism Spectrum Disorder	191 (48.23%)
Intellectual Disability	93 (23.48%)
Other Health Impairment	48 (12.12%)
Developmental Delay (3-10)	26 (6.57%)
Orthopedic Impairment	14 (3.54%)
Grade 4	
Autism Spectrum Disorder	183 (42.17%)
Intellectual Disability	135 (31.11%)
Other Health Impairment	62 (14.29%)
Communication Disorder	19 (4.38%)
Orthopedic Impairment	11 (2.53%)
Grade 5	
Autism Spectrum Disorder	177 (44.14%)
Intellectual Disability	125 (31.17%)
Other Health Impairment	63 (15.71%)
Grade 6	
Autism Spectrum Disorder	187 (41.83%)
Intellectual Disability	163 (36.47%)
Other Health Impairment	51 (11.41%)
Specific Learning Disability	19 (4.25%)
Grade 7	
Intellectual Disability	170 (39.35%)
Autism Spectrum Disorder	144 (33.33%)
Other Health Impairment	66 (15.28%)
Specific Learning Disability	15 (3.47%)
Communication Disorder	13 (3.01%)
Grade 8	
Intellectual Disability	173 (40.71%)
Autism Spectrum Disorder	141 (33.18%)
Other Health Impairment	58 (13.65%)
Orthopedic Impairment	16 (3.76%)
Specific Learning Disability	12 (2.82%)
Grade 11	
Intellectual Disability	172 (51.04%)

Table 4.10: Distribution of Primary IDEA Codes by Grade *(continued)*

IDEA Code	Amount of Sample
Autism Spectrum Disorder	111 (32.94%)
Other Health Impairment	30 (8.9%)

Note:
Samples Less than 10 Omitted for Privacy

Table 4.11: Distribution of Secondary IDEA Codes by Grade

Secondary IDEA Code	Amount of Sample
Grade 3	
Not Applicable	298 (75.25%)
Communication Disorder	39 (9.85%)
Other Health Impairment	21 (5.3%)
Intellectual Disability	11 (2.78%)
Grade 4	
Not Applicable	317 (73.04%)
Communication Disorder	54 (12.44%)
Other Health Impairment	22 (5.07%)
Intellectual Disability	15 (3.46%)
Grade 5	
Not Applicable	285 (71.07%)
Communication Disorder	40 (9.98%)
Other Health Impairment	30 (7.48%)
Intellectual Disability	22 (5.49%)
Grade 6	
Not Applicable	341 (76.29%)
Other Health Impairment	31 (6.94%)
Communication Disorder	29 (6.49%)
Intellectual Disability	19 (4.25%)
Grade 7	
Not Applicable	288 (66.67%)
Communication Disorder	46 (10.65%)
Other Health Impairment	38 (8.8%)
Intellectual Disability	23 (5.32%)
Grade 8	
Not Applicable	284 (66.82%)
Communication Disorder	38 (8.94%)

Table 4.11: Distribution of Secondary IDEA Codes by Grade *(continued)*

Secondary IDEA Code	Amount of Sample
Other Health Impairment	31 (7.29%)
Intellectual Disability	30 (7.06%)
Autism Spectrum Disorder	15 (3.53%)
Grade 11	
Not Applicable	223 (66.17%)
Communication Disorder	26 (7.72%)
Other Health Impairment	25 (7.42%)
Intellectual Disability	24 (7.12%)

Note:
Samples Less than 10 Omitted for Privacy

4.2D 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. 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, ORExt, ELPA, KA).

The vertically scaled scores generally convey incremental gains in achievement across grade levels, which is seen here by both measures of central tendency (i.e., mean and median) for all tests.

Table 4.12: Observed RIT Score: Measures of Central Tendency by Grade and Content

	RIT Median	RIT Mean (SD)
ELA		
Grade 3	205	200.05 (21.53)
Grade 4	209	203.25 (21.3)
Grade 5	211	208.06 (22.31)
Grade 6	211	209.47 (21.98)
Grade 7	217	213.28 (22.74)
Grade 8	216	213.68 (22.22)
High School	912	906.71 (28.43)
Math		
Grade 3	193	187.81 (20.53)

Table 4.12: Observed RIT Score: Measures of Central Tendency by Grade and Content *(continued)*

	RIT Median	RIT Mean (SD)
Grade 4	198	191.66 (20.13)
Grade 5	199	194.25 (19.35)
Grade 6	202	197.82 (18.92)
Grade 7	207	201.03 (18.5)
Grade 8	207	201.72 (16.67)
High School	900	894.49 (20.59)
Science		
Grade 5	510	506.16 (23.95)
Grade 8	815	810.3 (23.18)
High School	912	908.35 (29.96)

4.2.0.1 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 OREXt. Welch's two sample t-tests demonstrate that ELA Grade 6 favors girls significantly ($p = 0.011$); Math Grade 11 favors boys significantly ($p = 0.001$); Science Grade 11 favors boys significantly ($p = 0.015$).

Table 4.13: Mean (Standard Devaition) of RIT Scores by Gender, Grade, and Content

	Female RIT Mean (RIT SD)	Male RIT Mean (RIT SD)
ELA		
Grade 3	199.76 (21.24)	200.18 (21.69)
Grade 4	203.7 (21.47)	203.07 (21.26)
Grade 5	208.08 (23.51)	208.06 (21.77)
Grade 6	213.23 (20.68)	207.7 (22.39)
Grade 7	216.24 (23.64)	211.99 (22.25)
Grade 8	216.18 (21.97)	212.37 (22.31)
High School	903.77 (29.19)	908.1 (27.71)
Math		
Grade 3	187.74 (19.72)	187.84 (20.92)
Grade 4	190.27 (19.22)	192.22 (20.5)
Grade 5	191.46 (20.82)	195.57 (18.5)
Grade 6	198.17 (17.64)	197.66 (19.53)
Grade 7	201.19 (17.75)	200.96 (18.85)

Table 4.13: Mean (Standard Devaition) of RIT Scores by Gender, Grade, and Content *(continued)*

	Female RIT Mean (RIT SD)	Male RIT Mean (RIT SD)
Grade 8	201.45 (16.92)	201.81 (16.6)
High School	889.03 (21.78)	897.47 (19.24)
Science		
Grade 5	504.72 (23.92)	506.83 (23.97)
Grade 8	811.04 (23.7)	809.87 (22.98)
High School	902.99 (29.81)	911.42 (29.74)

4.2.0.2 Observed Means Reported by Race

The following table provides information regarding average student performance by grade level and race/ethnicity in each of the content areas assessed on the OREXt.

Table 4.14: Mean (Standard Deviation) for RIT Scores by Racial-Ethnic Group, Grade, and Content

	White	Hispanic	Multi-ethnic	Asian	Black	P	Native American or Alaskan
ELA							
Grade 3	201.5 (21.78)	198.94 (20.55)	195.96 (27.65)	196.73 (20.12)	199.5 (11.61)	195.5 (23.93)	–
Grade 4	205.5 (21.25)	200.94 (21.14)	201.14 (19.74)	202.77 (19.95)	204.47 (20.53)	–	200.58 (24.24)
Grade 5	209.18 (22.64)	205.58 (21.36)	209.32 (21.86)	209 (21.29)	–	–	–
Grade 6	210.78 (22.93)	208.99 (20.2)	201.74 (23.4)	214.17 (16.77)	206.65 (19.21)	–	–
Grade 7	214.54 (23.37)	212.74 (22.02)	210.73 (22.23)	203 (22.26)	211.59 (21.29)	–	–
Grade 8	215.94 (23.45)	211.41 (20.92)	212.29 (20.87)	210.93 (17.49)	208.79 (16.58)	–	–

Table 4.14: Mean (Standard Deviation) for RIT Scores by Racial-Ethnic Group, Grade, and Content *(continued)*

	White	Hispanic	Multi-ethnic	Asian	Black	P	Native American or Alaskan
High School	908.62 (29.07)	906.26 (26.43)	905.15 (25.86)	892.24 (27.23)	–	–	–
Math							
Grade 3	189.47 (20.43)	185.19 (21.22)	186.88 (23.21)	188.67 (17.79)	186.83 (12.1)	179.5 (21.53)	–
Grade 4	193.9 (19.46)	190.09 (20.25)	189.67 (17.5)	187.04 (21.44)	191.68 (22.69)	–	188.67 (25.12)
Grade 5	194.82 (18.95)	192.37 (19.91)	196.6 (20.85)	194.86 (17.68)	–	–	–
Grade 6	198.01 (18.94)	198.63 (18.87)	194.04 (17.01)	201.94 (20.83)	195.72 (19.57)	–	–
Grade 7	201.56 (18.58)	200.15 (18.22)	202.96 (18.46)	193.75 (22.09)	200.65 (20.31)	–	–
Grade 8	202.36 (16.5)	203.42 (15.92)	198.91 (17.62)	194.93 (19.1)	191.59 (18.29)	–	–
High School	895.69 (20.89)	893.56 (20.97)	896.89 (15.19)	890.12 (18.47)	–	–	–
Science							
Grade 5	507.86 (24.48)	503.42 (23.85)	507.44 (23.37)	503.73 (21.17)	–	–	–
Grade 8	811.89 (24.26)	809.33 (22.15)	808.5 (21.78)	805.27 (18.79)	806.18 (23.1)	–	–
High School	909.47 (31.96)	908.68 (26.51)	919.56 (21.49)	894 (26.65)	–	–	–

Note:

Samples Less than 10 Omitted for Privacy

Pacific Islander n < 10 at all grades

4.2.0.3 Observed Means Reported by Exceptionality Status

The following table is a number key for **Eligibility Codes:**

4.2.0.3.1 Eligibility Codes List

- 0 Not Applicable
- 10 Intellectual Disability
- 20 Hearing Impairment
- 40 Vision Impairment
- 43 Deafblindness
- 50 Communication Disorder
- 60 Emotional Disturbance
- 70 Orthopedic Impairment
- 74 Traumatic Brain Injury
- 80 Other Health Impairment
- 82 Autism Spectrum Disorder
- 90 Specific Learning Disability
- 98 Developmental Delay (3-10)

Table 4.15: Mean (Standard Deviation) of RIT scores by primary IDEA eligibility code

Primary IDEA Code	ELA	Math	Science
Grade 3			
Ages 3-10 Developmental Delay (98)	204.24 (20.28)	191.56 (23.38)	–
Autism Spectrum Disorder (82)	196.67 (22.28)	184.99 (20.38)	–
Intellectual Disability (10)	205.81 (19.44)	192.64 (18.76)	–
Orthopedic Impairment (70)	184.57 (22.67)	174.93 (18.96)	–
Other Health Impairment (80)	199.83 (19.83)	185.19 (21.2)	–
Grade 4			
Autism Spectrum Disorder (82)	200.8 (22.85)	188.67 (20.66)	–
Communication Disorder (50)	211.47 (11.76)	200.56 (15.15)	–
Intellectual Disability (10)	206.51 (18.52)	194.71 (17.45)	–
Orthopedic Impairment (70)	186.91 (22.67)	180.27 (22.24)	–
Other Health Impairment (80)	198.5 (22.79)	187.58 (21.55)	–
Grade 5			
Autism Spectrum Disorder (82)	205.46 (22.02)	191.89 (20.28)	499.56 (24.33)
Intellectual Disability (10)	212.55 (19.91)	198.18 (15.87)	513.23 (18.74)
Other Health Impairment (80)	205.85 (23.26)	192.05 (19.41)	506.46 (23.5)
Grade 6			
Autism Spectrum Disorder (82)	203.79 (21.78)	194.8 (19.85)	–

Table 4.15: Mean (Standard Deviation) of RIT scores by primary IDEA eligibility code *(continued)*

Primary IDEA Code	ELA	Math	Science
Intellectual Disability (10)	213.63 (20.07)	199.73 (17.22)	–
Other Health Impairment (80)	210.41 (22.71)	197.33 (17.57)	–
Specific Learning Disability (90)	228.32 (13.96)	214.28 (6.64)	–
Grade 7			
Autism Spectrum Disorder (82)	209.73 (24.48)	198.42 (20.01)	–
Communication Disorder (50)	225.92 (10.2)	213.23 (5.96)	–
Intellectual Disability (10)	215.92 (19.96)	201.95 (16.6)	–
Other Health Impairment (80)	211.98 (20.98)	202.05 (16.89)	–
Specific Learning Disability (90)	230.87 (13.59)	216.43 (10.2)	–
Grade 8			
Autism Spectrum Disorder (82)	209.68 (21.04)	199.84 (16.14)	806.33 (22.08)
Intellectual Disability (10)	215.67 (20.8)	201.78 (16.71)	811.38 (22.23)
Orthopedic Impairment (70)	188.33 (18.69)	188.07 (20.25)	785.33 (21.89)
Other Health Impairment (80)	215.39 (26.17)	203.32 (17.9)	812.96 (26.17)
Specific Learning Disability (90)	232.42 (14.93)	216.42 (4.68)	833.18 (7.78)
High School			
Autism Spectrum Disorder (82)	906.25 (30.44)	894.81 (19.4)	904.95 (30.73)
Intellectual Disability (10)	907.01 (24.93)	894.37 (19.16)	910.45 (26.98)
Other Health Impairment (80)	907.26 (35.69)	893.04 (26.97)	909 (36.63)

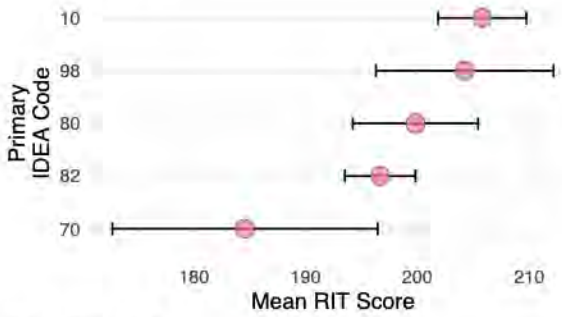
Note:

Samples Less than 10 Omitted for Privacy

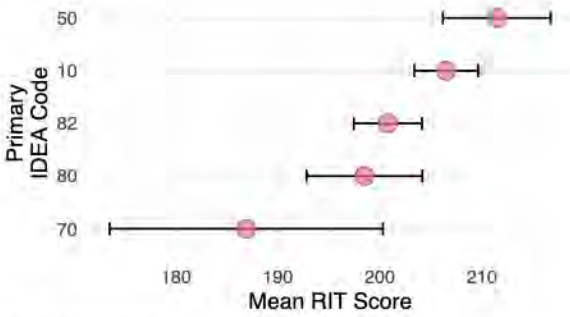
4.2.0.4 Graphs of Observed Means By Disability

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 are significantly different. In all cases, groups were not reported when smaller than 10 individuals for privacy.

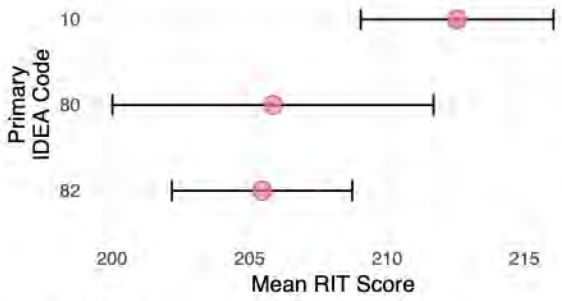
ELA Grade 3



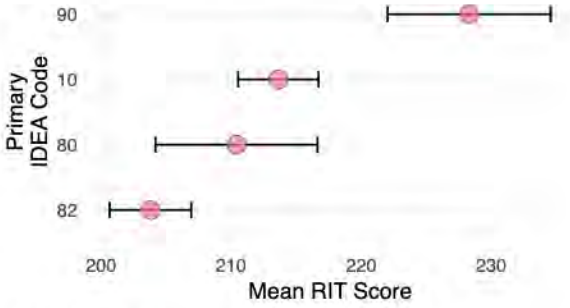
ELA Grade 4



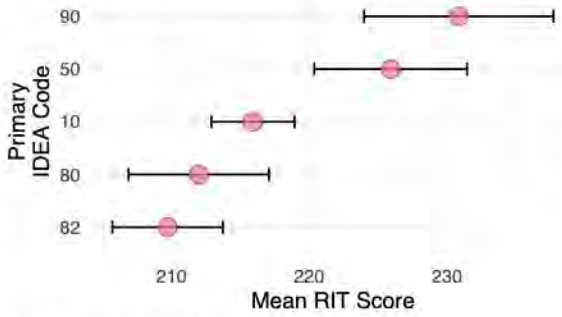
ELA Grade 5



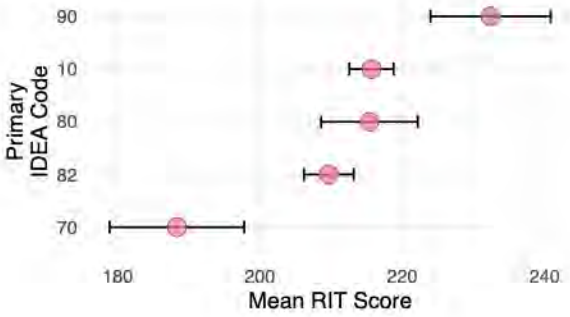
ELA Grade 6



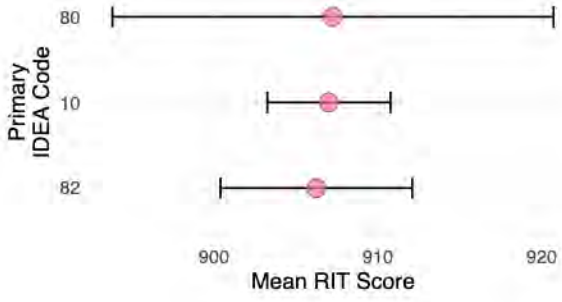
ELA Grade 7



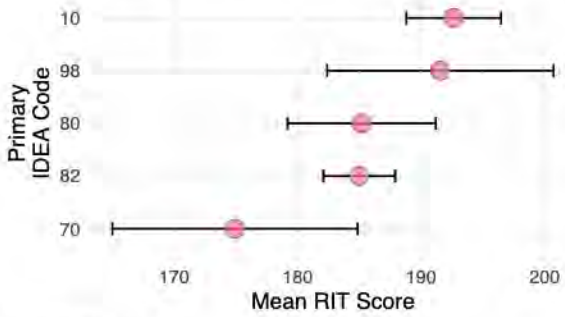
ELA Grade 8



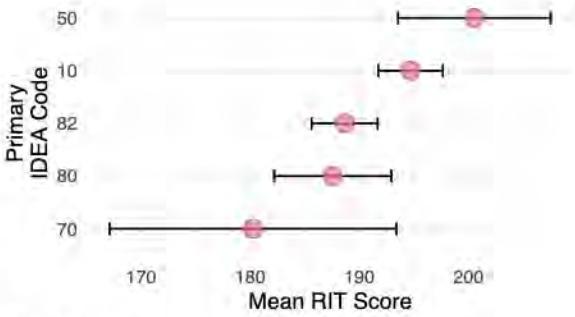
ELA High School



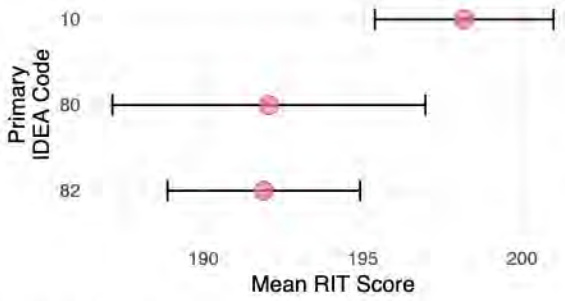
Math Grade 3



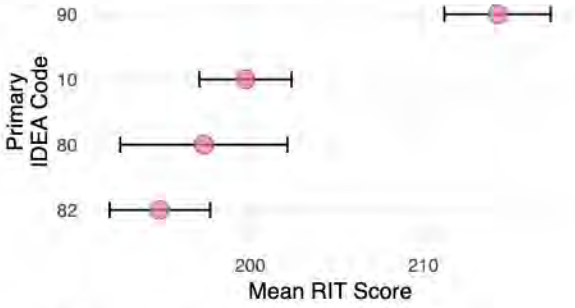
Math Grade 4



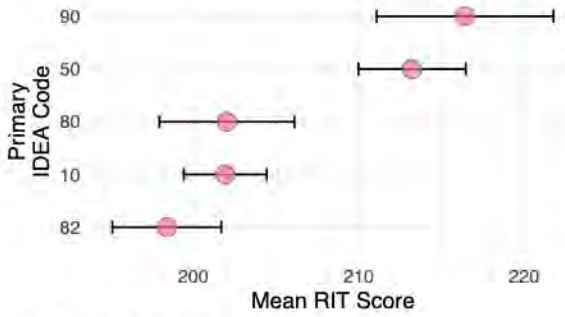
Math Grade 5



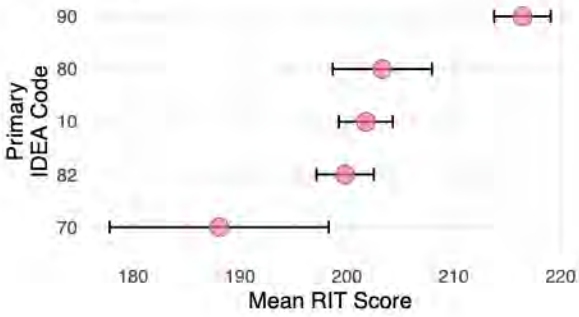
Math Grade 6



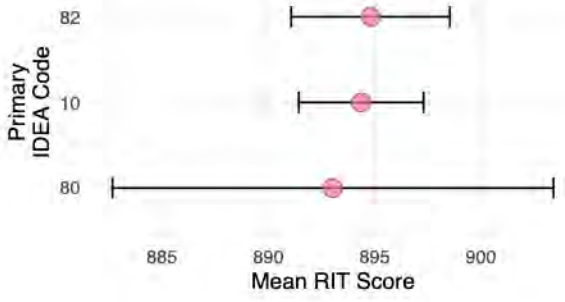
Math Grade 7

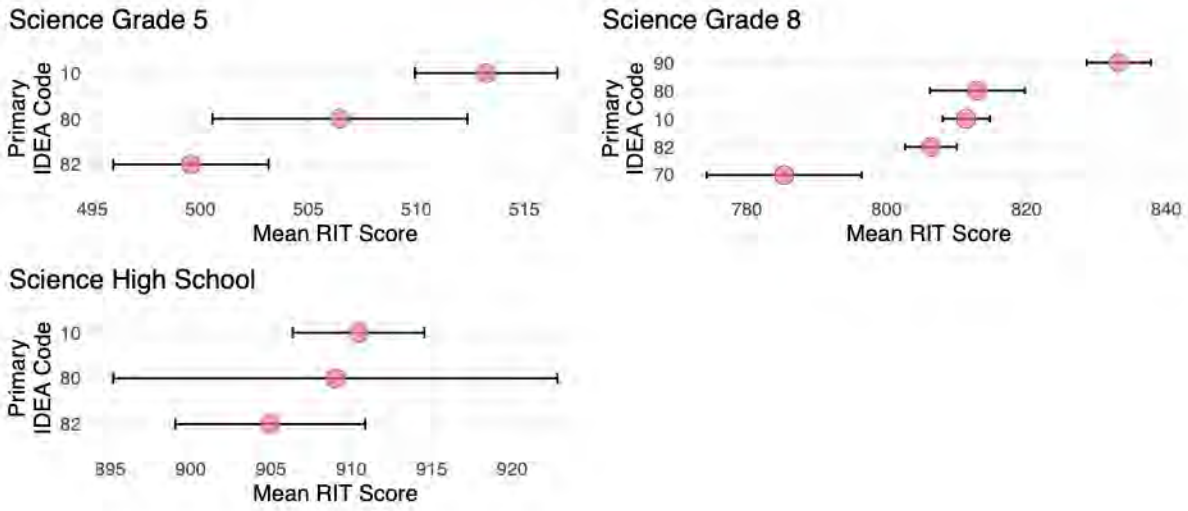


Math Grade 8



Math High School





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 purposeful, validated manner. The OREXt test blueprints convey the balance of representation exhibited by the assessment (see OREXt Test Blueprint). These test blueprints are supported by the OREXt Extended Assessment Frameworks, which define the assessable content on the OREXt that has been reduced in depth, breadth, and complexity (RDBC) using our defined process (see Reducing the Depth, Breadth, and Complexity of Items). 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 Oregon Extended Assessment Alignment Study).

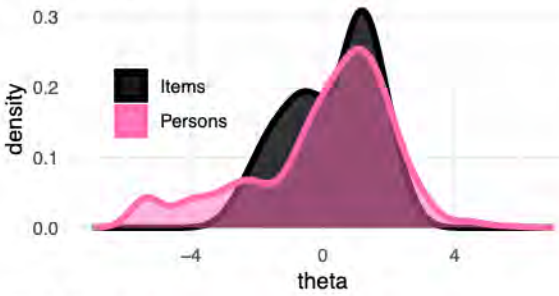
Though a simplified and standardized approach was taken to design items, and efficiency and access to the assessment increased for the majority of students (as evidenced by the decreased percentages of zero scores across all content areas), a small subgroup of students remains who cannot access an academic 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 targets communication (expressive and receptive) and basic skills (attention/joint attention and mathematics) and provides 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. The item person distributions provided below demonstrate that the ORExt is providing a performance continuum for students who participate.

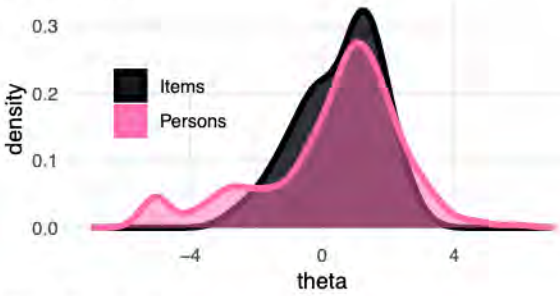
4.3A English Language Arts Person/Item Distributions

We see that every test has substantial overalp between the person and item distribution. It is clear, though, that there is over-representation of some areas of the test. From these plots, we can see that ELA tests should focus on covering a wider range of abilities (mostly at the upper end, but really in both directions), rather than overrepresenting moderate difficulty items.

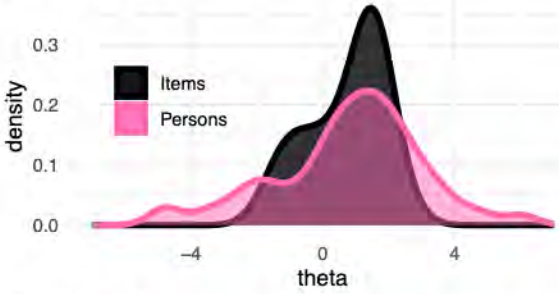
ELA Grade 3



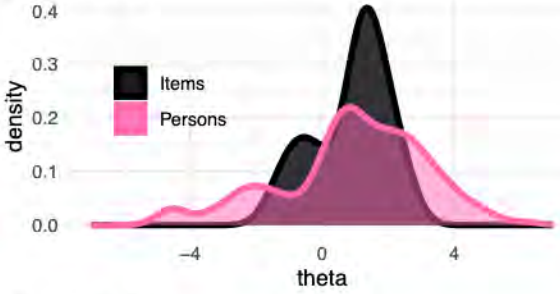
ELA Grade 4



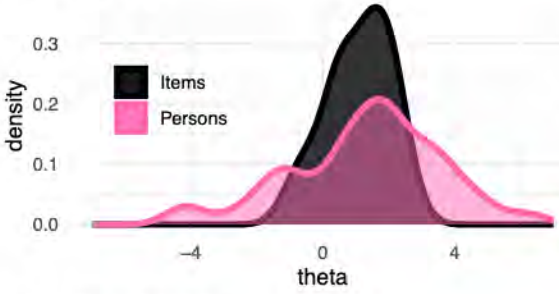
ELA Grade 5



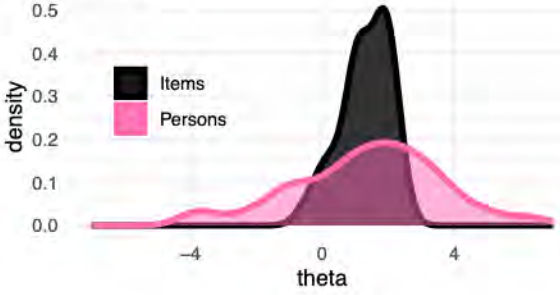
ELA Grade 6



ELA Grade 7



ELA Grade 8

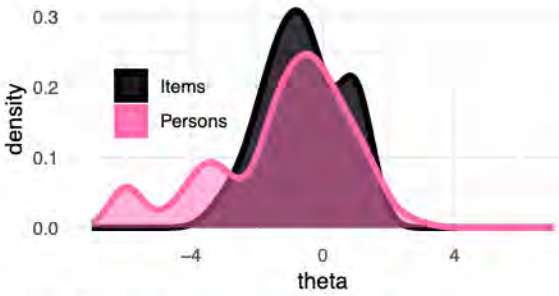


ELA High School

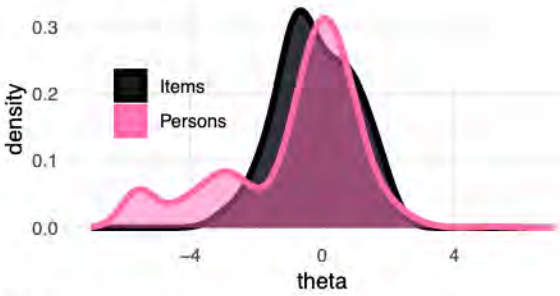
4.3B Mathematics Person/Item Distributions

From these plots, we can see that Math does a very good job of covering the range of abilities represented in the person distribution. Comparing this to the ELA plots, it is clear that math does a better job covering the upper end of the spectrum. However, there is a persistent lower tail to the distribution of students, which suggest testing modifications, inclusion criteria, or drop-out protocols could be re-evaluated. Considering the rest of the test statistics about math tests, it would be useful to consider the distribution without inclusion of ORORa-recommended students prior to large changes.

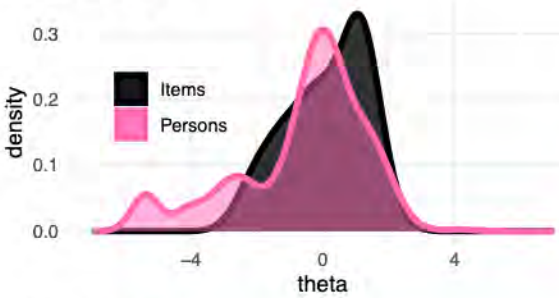
Math Grade 3



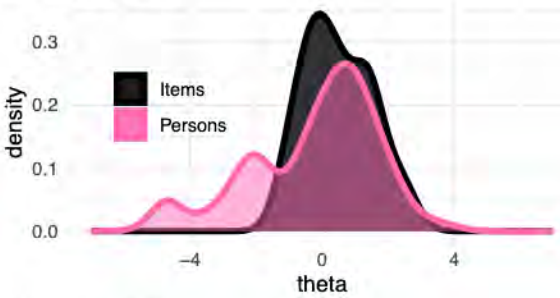
Math Grade 4



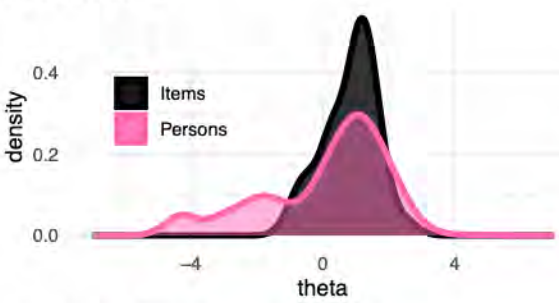
Math Grade 5



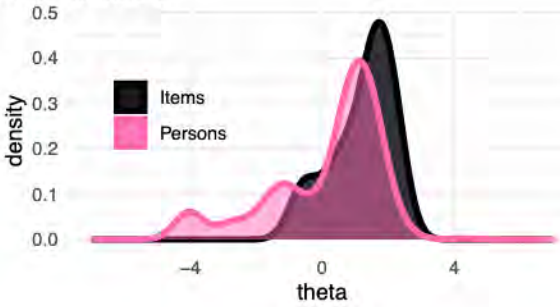
Math Grade 6



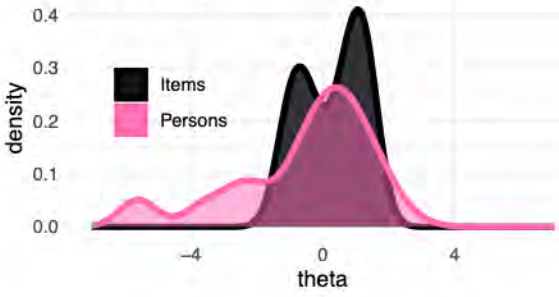
Math Grade 7



Math Grade 8

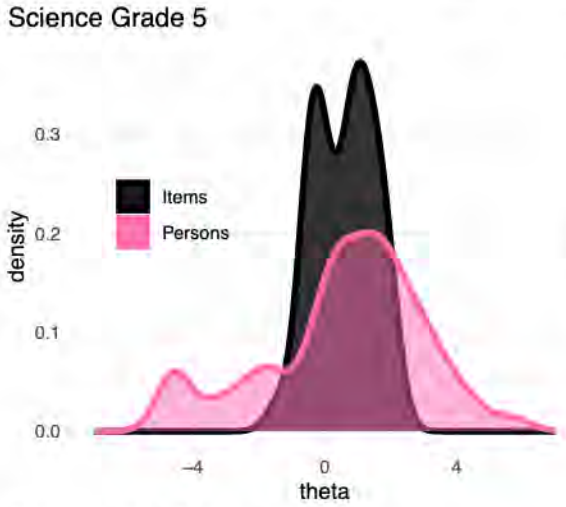


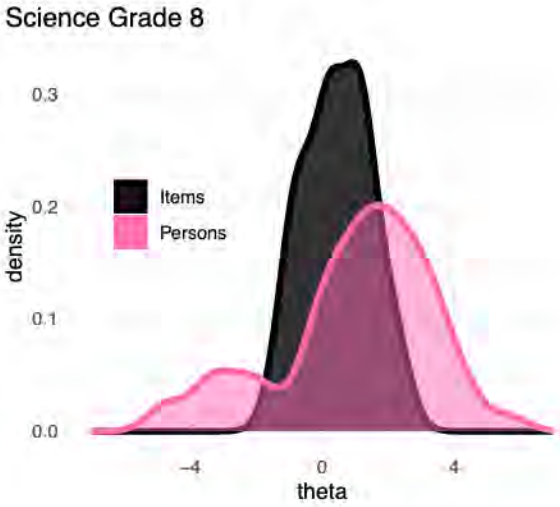
Math High School

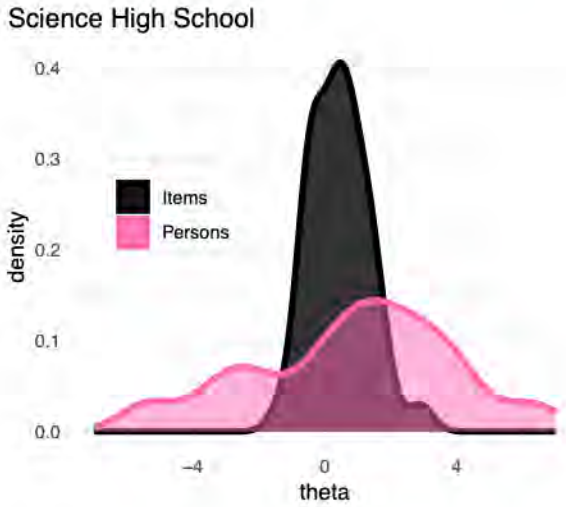


4.3C Science Person/Item Distributions

Similar to ELA, science items tend to be overly clustered in the moderate difficulty range. The test could be improved by better covering the range of person abilities, particularly at the upper end of the ability spectrum.







4.4 Scoring

All scoring expectations for the OREXt are established within the Test Administration Training. 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 Consequential Validity studies demonstrates that the assessment scoring

procedures are much more clear and easier to implement than prior scoring approaches (see Consequential Validity Survey Results). 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

Individual Student Report

Dear Parents/Guardians,

These results provide information about where your student is performing on grade level standards, and/or expectations, and skills. A student's overall score and corresponding achievement levels are displayed in the line graph.

Oregon teachers and administrators have worked to implement more challenging, relevant, and engaging English and math standards in every K-12 classroom. These standards identify what students should know or be able to do to graduate high school on a path ready for college and career.

If you have questions about these results or how to best support your child, please contact your child's school.

2015-16
OREGON STATEWIDE ASSESSMENT

Susan B. Anthony

SSID		5567
Grade		3
Birthdate		06/15/00
Test Date		06/12/16
Teacher		Janice Pembrook
School	1208	Super Star Elementary
District	31078	Amazing District
County	128	Dry Flats

Your child's scores and achievement levels

Grade 3	Level 1 Less than 192	Level 2 192 - 212	Level 3 213 - 227	Level 4 228 or more

The student demonstrates adept knowledge and skills when presented with test items linked to grade level content; i.e., content that has been reduced in depth, breadth, and complexity. Specific information regarding your student's performance in each domain of this assessment are located in the Achievement/Performance Standards section on the Statewide Alternate Assessment website: https://www.oregon.gov/ode/educator-resources/assessment/AltAssessment/. Note: Because your child participated in the Extended Assessment, these results cannot be used to compare your child's performance with that of students who are in the same enrolled grade but taking the general assessment.

ODE worked with partners at the University of Oregon at Behavioral Research & Teaching, along with education and community partners, to develop parent assessment literacy training modules found at: https://www.oregon.gov/ode/educator-resources/assessment/Pages/asmtlit.aspx

About scale scores and achievement standards

Student results are reported on scales that vary by test subject. Standards have been established with respect to these score s that identify your child's achievement level. A description of this achievement level is provided below the score(s). For more information on scores and achievement standards, see https://www.oregon.gov/ode/educator-resources/standards/Pages/Achievement-Performance-Standards.aspx.

for

annual performance, growth, and as part of Essential Skills requirements for very low performing students (see Decision Making Related To Scaled Scores).

4.5 Multiple Assessment Forms

The ORExt was administered in one form per subject area and grade level for the 2022-23 school year, with 36 operational items arranged in order of empirical difficulty.

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 review 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 Consequential Validity Survey Results).

Within our system of ongoing improvement is continuation of the development of additional curricular and instructional resources. This addresses an area of concern expressed by stakeholders. Training modules and templates continue to be developed to connect assessment results from the ORExt and ORora with curricular resources and instructional strategies aligned to the standards.

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 Test Administration Manual). 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 Test Administration Manual). 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 ORExt Eligibility Guidance).

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 the Oregon Accessibility Manual. Guidelines for making these determinations for SWSCD who participate in AA-AAAS are provided in the ORExt Eligibility Guidance.

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 Reducing the Depth, Breadth, and Complexity of Items). 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 the ORExt Electronic Practice Tests. All accessibility supports, designated supports, and accommodations for the ORExt are published in the Oregon Accessibility Manual. For students who have very limited to no communication and are unable to access even the most accessible items on the ORExt, an Oregon Observational Rating Assessment (ORora) was implemented in 2015-16. The ORora 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 for the ORora are included here (*Appendix A*).

5.1E Guidance Regarding Appropriate Accommodations

Guidance regarding appropriate accommodations is published in the Oregon Accessibility Manual. 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 ORExt Eligibility Guidance).

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 Test Administration Manual).

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 Oregon Extended Assessment Alignment Study). 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 the BRT Projects Website.

5.2 Procedures for Including ELs

In addition to the programmatic guidance provided in the Oregon Department of Education English Learner Program Guide related to EL program eligibility and services, ODE also provides guidance relevant to the inclusion of ELs in the statewide assessment program in the Test Administration Manual. Though the ORExt is currently published in English, A bilingual test administrator who is trained and endorsed by a district in Spanish or the students' language of origin should provide any language translation support, including American Sign Language. ODE has developed a training module to increase the standardization of ASL Training and Administration for its statewide assessments.

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 Oregon Accessibility Manual) 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 the Oregon Accessibility Manual. 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 in the Oregon Accessibility Manual.

5.3B Appropriate Accommodations are Available for ELs

A bilingual test administrator who is trained and endorsed by a district in Spanish or the students' language of origin should provide any language translation support. Appropriate accommodations for the ORExt are published in the Oregon Accessibility Manual. 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 in the Oregon Accessibility Manual.

5.3C Accommodations are Appropriate and Effective

In addition to the evidence gathered during the linkage study (see Oregon Extended Assessment Alignment Study) which suggests that the OREXt items were accessible and free of bias even before final editing, the appropriateness of the supports listed in the Oregon Accessibility Manual is reviewed annually by the Oregon Accommodations Panel. Practitioners may also request the addition of an accommodation through a formal process

see Appendix E in the Oregon Accessibility Manual. ODE has collected accommodations codes for the OrExt from Qualified Assessors who opt to enter this information in order to make performance comparisons feasible.

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 in the Oregon Accessibility Manual.

5.4 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.

6 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 Measurable 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 the Item Writer Training.

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 Adoption of Alternate Academic Achievement Standards). The ELA, Math, and Science AAAS, including both the ALDs and the requisite cut scores are included in the Alternate Academic Achievement Standards.

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 ORExt, 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 and Pitoniak (2006) as reported by Yen, Fitzpatrick, and Brennan (2006) (Educational Measurement, 4th Edition, pp. 433-470). Standard setting evaluation questions posed to participants were adapted from Cizek (2012), 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 the ORExt Assessment Technical Report on Standard Setting.

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 the Oregon Extended Assessment Alignment Study.

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 Test Administration Manual). 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 Consequential Validity Survey Results).

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 State Annual Report Card).

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 example ISR in 6.4C below.

6.4C 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).

Individual Student Report

Dear Parents/Guardians,

These results provide information about where your student is performing on grade level standards, and/or expectations, and skills. A student's overall score and corresponding achievement levels are displayed in the line graph.

Oregon teachers and administrators have worked to implement more challenging, relevant, and engaging English and math standards in every K-12 classroom. These standards identify what students should know or be able to do to graduate high school on a path ready for college and career.

If you have questions about these results or how to best support your child, please contact your child's school.

2015-16
OREGON STATEWIDE ASSESSMENT

Susan B. Anthony

SSID		5567
Grade		3
Birthdate		06/15/00
Test Date		06/12/16
Teacher		Janice Pembrook
School	1208	Super Star Elementary
District	31078	Amazing District
County	128	Dry Flats

Your child's scores and achievement levels

Grade 3	Level 1 Less than 192	Level 2 192 - 212	Level 3 213 - 227	Level 4 228 or more

The student demonstrates adept knowledge and skills when presented with test items linked to grade level content; i.e., content that has been reduced in depth, breadth, and complexity. Specific information regarding your student's performance in each domain of this assessment are located in the Achievement/Performance Standards section on the Statewide Alternate Assessment website: https://www.oregon.gov/ode/educator-resources/assessment/AltAssessment/. Note: Because your child participated in the Extended Assessment, these results cannot be used to compare your child's performance with that of students who are in the same enrolled grade but taking the general assessment.

ODE worked with partners at the University of Oregon at Behavioral Research & Teaching, along with education and community partners, to develop parent assessment literacy training modules found at: https://www.oregon.gov/ode/educator-resources/assessment/Pages/asmtlit.aspx

About scale scores and achievement standards

Student results are reported on scales that vary by test subject. Standards have been established with respect to these score s that identify your child's achievement level. A description of this achievement level is provided below the score(s). For more information on scores and achievement standards, see https://www.oregon.gov/ode/educator-resources/standards/Pages/Achievement-Performance-Standards.aspx.

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 the Decision Making Related To Scaled Scores.

6.5 Analytic Summary

6.5.1 Item-Level Information

Overwhelmingly, items appear appropriate based on point measure correlations and mean square outfit. Most items adequately contribute to the underlying constructs they are measuring for each grade and content area across these metrics, and those which do not will be flagged for review.

Two pairs of groups were assessed for differential item functioning: those coded as male vs. female and those coded as white vs. non-white. In both sets of analyses, substantial differential item functioning was observed on several items. There was some balance in who was favored for both sets of analyses, but in the end: (a) those coded as males were favored by 14 more items than those coded as females and (b) those coded as White were only favored by 1 more item than those coded as non-White.

There was good overlap in persons abilities and item difficulty.

Items were flagged for review if they had inadequate point measure correlations, unfavorable mean square outfit, or substantial differential item functioning.

6.5.2 Test-Level Information

For many grades' tests, one or more AMO level is not well represented. This is most often AMO 4 (Exceeds), but is occasionally AMO 2. In some cases, this is because the scaled score range which corresponds to a given AMO is very small (e.g., 2 scaled score points in Grade 7 math).

Conditional standard error of measurement (SEM) around the AMO cutscores indicate good separation between cutscores in most, but not all cases. Areas which display excessive overlap between SEMs should be considered if another standards setting occurs, as greater separation between these SEMs will lead to more accurate separation between AMOs.

Test reliabilities were good with (above 0.8) for all contents and grades at the total test level (between 0.8 and 0.94).

Test information functions overlaid with the thresholds of AMOs show the AMO 1-2 and AMO 2-3 thresholds are always in areas with acceptable marginal reliability (i.e., above 0.8). For several tests, the threshold of AMO 3-4 is between 0.8 and 0.7. For these tests, item composition should be explored to improve marginal reliability at this cut. This also relates to a general trend of excessive information at the lower end of the ability spectrum. If the test leveraged more difficult items, students could be assigned to AMO with greater confidence.

Test characteristic curves (TCCs) demonstrate a clear vertical scale across both ELA and Math for grades 3-8, with roughly even spacing across grades. The only exception to this is math grades 5 and 6 which overlap; modification of the item set for these grades can improve this cross over for the next year.

Across grades, content areas correlate with one another within ranges that demonstrate they are measuring similar but distinct concepts. Paired with validity from other aspects of items and tests, as well as expert opinion, indicates ORExt measures several related but distinct aspects of grade-appropriate ability. Correlation between ELA and its subscores (particularly reading) suggest there may be excessive representation of the subdomain; this is a place where item composition could be reconfigured or conceptualization of the construct is reconsidered.

6.5.3 Person-Level Information

Across years, the most common annual measurable objective (AMO) was most frequently:

- AMO 2 (Nearly Meets) then AMO 1 (Does Not Yet Meet) for ELA
- AMO 1 then AMO 2 for Math
- AMO 1 then AMO 3 (Meets) for Science

Rudner's (Rudner 2005) classification accuracy and consistency metrics show individuals were well classified into their AMO level across grades and tests.

There were differences in average RIT scores across primary IDEA code within a grade and content; occasionally these were significant. These differences, observable in text and systematic over- and underperformance by a specific IDEA code, should be reviewed for fairness and accessibility in test administration across these samples.

6.5.4 ORora

The Oregon Observational Rating Assessment (ORora) results demonstrate that approximately **(21-32%)** of the SWSCD who participated in the ORExt also took the ORora, depending upon grade level. The participants were primarily students with multiple, severe disabilities with very limited communication systems. Such students typically score

very low RIT scores, as ORora is administered either due to poor performance leading to discontinuation of the ORExt or by choice.

We have two pieces of evidence that differences in the population exist between those who take the ORora and those who do not: the comparison of ability estimates on the same test and the content area correlations.

Relatively small numbers of ORora takers had high scores on any subsection of the ORExt, which confirms that fewer individuals who took the ORora opted into it (compared to those who were redirected due to ORExt performance).

Content area correlations are different for those who take the ORora, compared to those who do not. For the most part, content areas did not correlate as highly among those who took the ORora, compared to those who did not. ORora scores also displayed very low correlations with ability estimates for all contents, confirming that ORora measures distinct constructs from those tests.

Among those who took the ORora, the vast majority (~76%) met minimum participation on the ORExt alongside completion of the ORora. There were, however, a non-negligible number of students that (a) did not meet minimum participation in all subject areas (~8%), (b) took insufficient items to meet minimum participation in all subject areas (~8%), or (c) did not attempt any items except the ORora (~8%). Future training should focus on ensuring QAs and QTs understand the minimum participation rules, especially in cases which result in students completing an ORora.

6.6 Conclusions

In sum, the rigor of the procedural development and statistical outcomes of the ORExt were substantive and support the assessments intended purpose. Procedural evidence includes essentialized standards development, item development, item content and bias reviews, an independent alignment study and item selection based upon item characteristics. Outcome-related evidence included measure reliability analyses, point measure correlations, outfit mean squares, item difficulty and person ability distributions, and convergent and divergent validity evidence. These sources of evidence were all quite good and provide important validity evidence.

The test development process adhered to procedural guidelines defined by the Association et al. (2018) AERA/APA/NCME Standards for Educational and Psychological Testing (2018), as well as incorporating procedures that are known in the field to be best practice. For example, an independent auditor evaluated alignment in 2016-17. Documentation collected in the alignment study report suggests that the ORExt assessment system is aligned based on five evaluation components: a) standard selection for essentialization, b) strength of linkage between essentialized standards and grade level content standards, c) alignment between items and essentialized standards, d) alignment between the essentialized standards

and the achievement level descriptors, and e) alignment between the achievement level descriptors and the ORExt test items. 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, as evidenced in our consequential validity study by teacher support of the academic content on the ORExt as well as the behaviors sampled during test administration.

The 2017-18 Oregon Consequential Validity study provided important information for future administrations of the ORExt. Results indicated historical concerns that were not possible to address, such as the ongoing tension between assessing life skills and academics, but also to some actionable steps with a focus toward continuous improvement. Respondents pointed to positive attributes of the ORExt, especially those involving test administration and design and felt somewhat positive regarding various educational impacts of the ORExt.

Feedback from the field and the number of students administered the tablet based ORExt indicated assessors preferred administration of the tablet/web-based assessment versus paper/pencil. Benefits expressed by the field indicated increased student engagement, improved standardization, ease of use by teachers, and resource protection (i.e., time, printing, expense). Practice tests were available to familiarize teachers and students to the tablet format prior to administration of the secure tests. Enhancements are in process to improve the tablet/web-based administration prior to the testing window.

Documenting evidence of validity remains an ongoing and continuous process. Our efforts to continue to improve the assessment system are outlined below, as well as in Sections 3 and 4 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.

6.7 Next Steps

Efforts to decrease the number of items with higher cognitive demand beyond the current minimum requirement of 36 items may necessitate a reassessment of standards alignment, item writing, and standards setting. To transition to matrix sampling or Computer Adaptive Testing (CAT), significant modifications would need to be made throughout the existing OrExt. The Alternate SEED Survey could undergo analysis to evaluate its utility for both the Oregon Department of Education (ODE) and end-users, including districts and teaching staff. Conducting Cognitive Labs would facilitate a better understanding of the effort involved, question quality, and alignment of items with students' experiences. These labs would focus on key concepts to provide valuable insights for ODE's decision-making process. Although the OrExt has had a vertical scale for the past eight years, little to no efforts have been made to report student growth in content areas across multiple testing years. Essentially, the vertical scale remained underutilized until this year. Test administrators now have access to a historical lookup of student performance, allowing for a simple line graph representation of score results across the years. Offering additional training on interpreting the vertical scale (line graph result) will enhance the understanding of score results and promote its utilization. Moreover, by analyzing the vertical scale data and utilizing new data visualization techniques, such as graphical representation, end-users can now be informed about the presence and usefulness of a growth model. Collaborating with ODE is crucial to ensure that the growth model serves its intended purpose and achieves desired outcomes. Currently, software is being developed to improve test construction through features like test information functions and test characteristic curves, among others. Enhancements to the Distributive Item Review platform will validate previous processes by automating scope and coverage verification.

References

Anderson, Daniel, and Christopher M. Loan. 2022. *Exirt: Analyze Data from the Oregon Extended Assessment.* https://github.com/datalorax/exirt.

Association, American Educational Research et al. 2018. *Standards for Educational and Psychological Testing.* American Educational Research Association.

Campbell, Donald T, and Donald W Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin* 56 (2): 81.

Carmichael, Sheila Byrd, Gabrielle Martino, Kathleen Porter-Magee, and W Stephen Wilson. 2010. "The State of State Standards–and the Common Core–in 2010." *Thomas B. Fordham Institute.*

Cizek, Gregory J. 2012. *Setting Performance Standards: Foundations, Methods, and Innovations.* Routledge.

Hallgren, Kevin A. 2012. "Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial." *Tutorials in Quantitative Methods for Psychology* 8 (1): 23.

Hambleton, Ronald K, and Mary J Pitoniak. 2006. "Setting Performance Standards." *Educational Measurement* 4 (1): 433–70.

Holland, PW, and DT Thayer. 1988. "Differential Item Performances and the Mantel-Haenszel Procedure." *Test Validity*, 129–45.

Kamata, Akihito, and Brandon K Vaughn. 2004. "An Introduction to Differential Item Functioning Analysis." *Learning Disabilities: A Contemporary Journal* 2 (2): 49–69.

Lathrop, Quinn N. 2015. *cacIRT: Classification Accuracy and Consistency Under Item Response Theory.* https://CRAN.R-project.org/package=cacIRT.

Messick, Samuel. 1989. "Meaning and Values in Test Validation: The Science and Ethics of Assessment." *Educational Researcher* 18 (2): 5–11.

Robitzsch, Alexander, Thomas Kiefer, and Margaret Wu. 2022. *TAM: Test Analysis Modules.* https://CRAN.R-project.org/package=TAM.

Rudner, Lawrence M. 2005. "Expected Classification Accuracy." *Practical Assessment, Research, and Evaluation* 10 (1): 13.

Scott, Neil W, Peter M Fayers, Neil K Aaronson, Andrew Bottomley, Alexander de Graeff, Mogens Groenvold, Chad Gundy, et al. 2009. "A Simulation Study Provided Sample Size Guidance for Differential Item Functioning (DIF) Studies Using Short Scales." *Journal of Clinical Epidemiology* 62 (3): 288–95.

Tindal, Gerald, Marilee McDonald, Marick Tedesco, Aaron Glasgow, Pat Almond, Lindy Crawford, and Keith Hollenbeck. 2003. "Alternate Assessments in Reading and Math: Development and Validation for Students with Significant Disabilities." *Exceptional Children* 69 (4): 481–94.

Webb, Norman L. 2002. "Depth-of-Knowledge Levels for Four Content Areas." *Language Arts* 28 (March).

Yen, Wendy M, Anne R Fitzpatrick, and RL Brennan. 2006. "Educational Measurement." Westport, CT: Praeger Publishers.

A Appendix A: ORora Admin Instructions

Oregon Observational Rating Assessment (ORora)
Administration Instructions

ORora Purpose

The ORora provides instructional and functional information for teachers and parents in four domains: attention, basic math concepts, and receptive and expressive communication. It is administered to students with significant cognitive disabilities (SWSCD) who are not able to access the academic demands of the Oregon Extended Assessment (ORExt), despite the provision of extensive supports and test design features founded in the concepts of universal design for assessment. Assessor(s) responsible for student's instruction should complete this rating scale. Qualified Assessors (QAs) are to use the following decision rule in determining whether or not to complete the ORora:

If testing for an ORExt content area assessment is discontinued in English language arts, Mathematics, or Science, QAs must complete the ORora (only one ORora per student must be completed).

Consequences of Discontinuing the ORExt

Students must complete 10 items on the ORExt to count for Annual Measurable Objective (AMO) participation. QAs should consider discontinuation of the ORExt administration if a student misses 10 items at any point within the administration of the first 15 items. If ORExt testing is discontinued, QAs must administer the ORora. However, teachers may elect to complete a full test administration in order to generate performance scores and still complete the ORora. Discontinuing the administration of the ORExt is a serious decision with many potential consequences; however, administering the ORExt when a valid score is not feasible is also an inefficient use of teacher and student time.

Two ORora Domains: LOI and Communication

This assessment includes both a level of independence (LOI) and a communication domain (COM), each with their own respective rating scales. The LOI scale helps stakeholders to define how much support a student needs from a teacher in order to become successful in specific areas. The COM scale helps to define the level of the student's functioning in terms of both understanding the intent of others as well as conveying their needs or wants to those around them.

Level of Independence (LOI)
In the LOI domain, the teacher rates how much assistance the student requires in order to bring them to success in a particular area, using a system of least prompts approach (Wolery, Ault, & Doyle, 1992), beginning with independent function, proceeding to the remaining levels of support only when needed, including verbal/gestural, partial physical, and/or full physical.

Level of Independence Rating Scale (LOI)

Level 1	Level 2	Level 3	Level 4
Full Physical Requires use of full physical supports from teacher (e.g., holding the elbow/hand) in order to attend to a task, as well as to complete the task.	**Partial Physical** Requires use of partial physical supports from teacher (e.g., touching the hand/shoulder) in order to attend to a task, as well as to complete the task.	**Verbal/Gestural** Requires use of verbal/gestural supports from teacher in order to attend to a task, as well as to complete the task.	**Independent** Able to complete task without direct support from teacher.

Clarifying Example
Here is an example of how a QA would work through a classroom activity using a system of least prompts. In a testing context, we are defining the level of support needed for different types of activities.
Level 4: Independent
Place preferred drink in front of student and wait 3-5 seconds to see if the student responds independently.
Level 3: Verbal/Gestural
If the student does not respond at Level 4 in 3-5 seconds, direct the child to the drink by pointing or providing a verbal prompt (*Indirect:* Are you thirsty? or *Direct:* Pick up your beverage so you can drink.)
Level 2: Partial Physical
If the student does not respond to Level 3 support in 3-5 seconds, use tactile physical assistance to prompt the student's hand, but do not use full physical assistance. Partial physical support can be paired with verbal prompting, as well.
Level 1: Full Physical
If the student does not respond to Level 2 support in 3-5 seconds, use full physical support (e.g., hand-over-hand) to fully assist the student to grab the beverage. Full physical support can be paired with verbal prompting, as well.

Communication (COM)

The **COM** rating is based on the following scale: 1 = Reactive, 2 = Proactive, 3 = Unconventional, 4 = Conventional. The COM rating captures communication behaviors below the pre-symbolic and symbolic levels assessed on the OREXt. The lowest functioning SWSCD likely have skills somewhere along this continuum—from staying awake and attending to functional and/or instructional objects in the classroom to beginning to work with objects and images. The COM rating scale is supported by a wide research base (Browder & Spooner, 2011; Browder, Wakeman, & Flowers, 2008; Browder, Wood, Thompson, & Ruboffo, 2011; McLean, Snyder-McLean, & Rowland, 1981; Rowland & Schweigert, 1990; Rowland, 2013).

Communication Rating Scale (COM)

Level 1	Level 2	Level 3	Level 4
Reactive Student's behavior is not purposeful, but may be reflective of the student's current status (e.g., level of comfort/energy, thirst, hunger). Teachers and parents are able to interpret the student's needs and wants by observing the behaviors (e.g., noises, facial expressions, moving body parts) and making inferences about what the student needs.	**Proactive** Student behaves purposefully, but does not realize that s/he can influence the behaviors of others by communicating needs at this level. Teachers and parents interpret the student's needs and wants by observing behaviors and making inferences.	**Unconventional** Student uses unconventional pre-symbolic communication. No use of symbols is included, nor does the student follow existing social communication norms. The student is attempting to interact with others to meet personal needs by making noises, facial expressions, and/or moving body parts.	**Conventional** Student uses conventional pre-symbolic behaviors to communicate with purpose. They are still below symbolic communication with abstract symbols (e.g., letters, numerals), but are communicating needs and wants in order to influence those around them in a socially accepted manner. Students may communicate by nodding, pointing, waving, hugging, looking toward a desired object, or using other socially appropriate gestures.

ORora Narrative Summary

In the open-ended narrative section, teachers can address or identify: (a) prerequisite skills that allow her/him to access instruction, (b) sensory support needs (hearing, vision, orthopedic, medical), (c) effective use of Assistive Technology (AT) (e.g., alternative communication devices), (d) relevant functional skills have developed over the past year, and, generally, (e) areas of growth that educators have noted in the prior year (e.g., comparing current to prior ORora scores, if available, or any context for determining the Present Levels of Academic and Functional Performance [PLAAFP] for SWSCDs).

Using Scores from the ORora

The ORora yields four sub-domain scores (**A.** Attention, **B.** Basic Math Concepts, **C.** Receptive Communication, and **D.** Expressive Communication), domain summary scores for the LOI and COM domains, and a summary score composed of both domain scores. These scores can be used for diagnostic purposes to represent student learning and change across time. Individualized Education Program (IEP) teams are encouraged to use the ORora results as one data source to develop appropriate and meaningful Present Levels of Academic and Functional Performance (PLAAFP) descriptions, as well as IEP goals and objectives. Here is an example of a student's ORora results reflected in a PLAAFP statement:

"Student achieved a total score of 70/80 on the ORora this year (87.5%), with a score of 19 in the Attention sub-domain, 18 in the Basic Math Concepts sub-domain, an 18 in the Receptive Communication sub-domain, and a 15 in the Expressive Communication sub-domain. These results reflect overall growth compared to last year's results, where s/he earned a 64/80 (80%). Student made impressive gains in communication, increasing by 4 points in the Expressive sub-domain and 2 points in the Receptive sub-domain."

IEP goals can also target overall improvement on the ORora, using other sources of data for assessment of objectives. Resources related to increasing student communication level will be published on BRT's curriculum and instruction website.

NOTE: For electronic and paper/pencil administration ORora scores are entered electronically either on the OREXt Training & Proficiency site in the Student Details, Monitoring tab OR electronic/tablet platform in the Data Entry tab.

Please contact Brad Lenhardt at ODE at brad.lenhardt@state.or.us with any questions.

References

Browder, D. M. & Spooner, F. (2011). *Teaching students with moderate and severe disabilities*. New York: Guilford Press.

Browder, D. M., Wood, L., Thompson, J., & Ribuffo, C. (2014). *Evidence-based practices for students with severe disabilities* (Document No. IC-3). Retrieved from University of Florida, Collaboration for Effective Educator, Development, Accountability, and Reform Center website: http://ceedar.education.ufl.edu/tools/innovation-configurations/

Browder, D. M., Flowers, C., & Wakeman, S. (2008). Facilitating participation in assessments and the general curriculum: Level of symbolic communication classification for students with significant cognitive disabilities. *Assessment in Education: Principles and Practice, 15*:2, 137-151. *doi* 10.1080/09695940802164176

McLean, J. E., Snyder-McLean, L., & Rowland, C. (1981). *Process-oriented educational programming for the severely/profoundly handicapped adolescent*. Parsons: University of Kansas, Bureau of Research.

Rowland, C. & Schweigert, P. (1990). *Tangible symbol systems: Symbolic communication for individuals with multisensory impairments.* Tucson, AZ: Communication Skill Builders. Retrieved from http://files.eric.ed.gov/fulltext/ED319154.pdf

Rowland, C. (2013). *Communication matrix for parents and professionals*. Oregon Health & Sciences University. Retrieved from http://communicationmatrix.org/uploads/pdfs/handbook.pdf

Wolery, M., Ault, M. J., & Doyle, P. M. (1992). *Teaching students with moderate to severe disabilities: Use of response prompting strategies*. New York: Longman.

B Appendix B: Inter Rater Observation Form

2017-18 Oregon Extended Assessment Rater Reliability

Observation Protocol

The Oregon Department of Education (ODE) plans to observe a sample of Oregon's Qualified Assessors (QAs) who administer the paper/pencil version of the Oregon Extended Assessment (OREXt) to determine reliability of administration and scoring. We do not include the tablet administration or the Oregon Observational Rating Assessment (ORora). You received this protocol because you were selected by ODE to participate as a Qualified Trainer (QT)/expert reviewer. The project will be conducted in two manners:

1) QTs in each district will observe a sample of their respective QAs using the observation protocol and enter their data online.
2) Expert reviewers from ODE and/or Behavioral Research & Teaching (BRT) will observe district-level QTs and those QAs who give the assessment in more than one school/district.

The observation protocol must be completed for the identified QA, but the student(s) and content area(s) observed will be selected by the QT or QA. BRT researchers will contact district-level QTs on day one of the test window, which runs from February 15 - April 26, 2018, to arrange multiple observations that can hopefully be completed within one school day. The observation is composed of three sections:

- **First**, you will be reviewing OREXt paper/pencil test preparation and administration using the rubric, see Page 2 for samples. Test preparation/administration domains are rated on a four-point scale from *Inappropriate (I)* to *Exemplary (E)*:
  - ***Inappropriate (I)*** denotes a level of concern that could clearly affect the accuracy of the test results gathered from the test administration. Ratings at this level require substantive retraining of the QA involved.
  - ***Somewhat Appropriate (SA)*** rating denotes a level that includes some minor aspects that could be improved, but the accuracy of the test results are likely not compromised.
  - ***Appropriate (A)*** denotes a level that is consistent with all test administration requirements,
  - ***Exemplary (E)*** level performance suggests that the QA incorporated approaches to test administration that could become models for best practice.
- **Second**, you will be scoring the student alongside the QA using the scoring sheet, see Page 3 for samples. You will compare results after this observation to ensure that the QA enters accurate data.
- **Finally**, you will observe the QA completing the data entry process to ensure that no errors are made during data entry and document the number of errors, see Page 4.

Qualified Assessor Testing Preparation and Administration Rubric

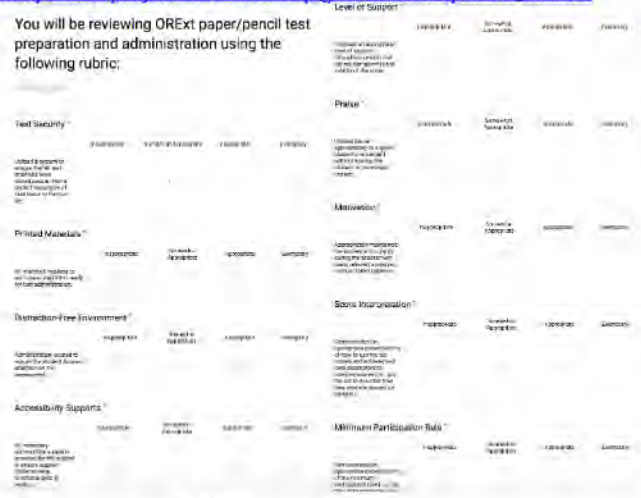
(Record an "X" in the cell that corresponds to your rating)

Domain Definitions

1. ***Test Security*** – The QA utilized a system to ensure that all test materials were stored in a secure location,. The QA also had a district Assurance of Test Security form on file.
2. ***Printed Materials*** – the QA had all materials required to administer the OREXt ready for test administration
3. ***Distraction-Free Environment*** – the QA arranged to provide the OREXt in a one-on-one test administration in a location that ensured that the student focused attention on the assessment.
4. ***Accessibility Supports*** – the QA provided all necessary accessibility supports for the student and ensured that all support systems were functional prior to testing.
5. ***Level of Support*** – The QA provided an appropriate level of support throughout testing that did not compromise the validity of the score.
6. ***Praise*** – The QA utilized praise appropriately to support student involvement without leading the student to the correct answer.
7. ***Motivation*** – The QA appropriately maintained the student's motivation during the assessment using relevant strategies, such as token systems.
8. ***Score Interpretation*** – The QA demonstrated an appropriate understanding of how to use the cut scores and achievement level descriptors to interpret scores (i.e., ask the QA to describe how they interpret scores for parents).
9. ***Minimum Participation Rule*** - The QA demonstrated an appropriate understanding of the minimum participation rule (i.e., ask the QA to define the rule if it is not used).

Domain #	Domain	I	SA	A	E
1.	Test Security				
2.	Printed Materials				
3.	Distraction-Free Environment				
4.	Accessibility Supports				
5.	Level of Support				
6.	Praise				
7.	Motivation				
8.	Score Interpretation				
9.	Minimum Participation Rule				

Online the form is found at the following link and will look like this:
https://docs.google.com/forms/d/e/1FAIpQLSdemN-sVqdmzNIWwanT4swSqUMM9YpncyzIt4AZ4TdeRDPSpQ/viewform?usp=form_confirm



2017-18 Oregon Extended Assessment – Rater Observation
Sample Scoring Sheet

QT/Expert Reviewer Name (First – Last) _____

Observed QA Name (First - Last):_____

Assessor completed required training on (date): _____

State Student ID: _____

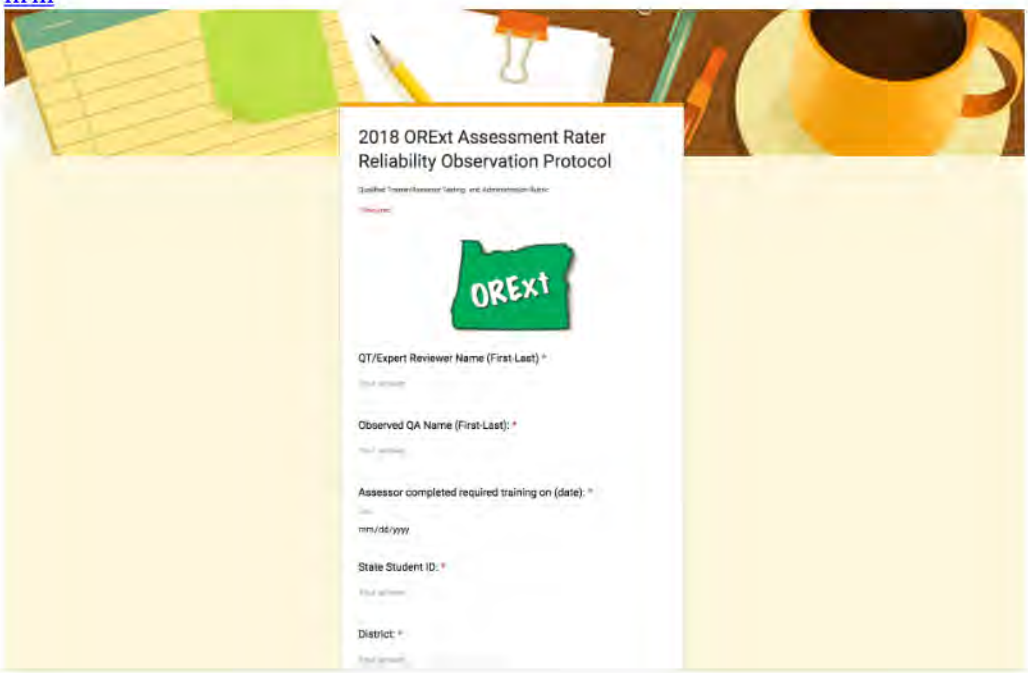
District: _____

School: _____

Student Grade: _____ Subject Area:_____

The online scoring sheet is found at the following link with a screen capture below.

https://docs.google.com/forms/d/e/1FAIpQLSdemN-sVqdmzNIWwanT4swSqUMM9YpncyzIt4AZ4TdeRDPSpQ/viewform?usp=form_confirm



Record all student responses for inter-rater reliability comparisons below (*Please circle all responses in which there was disagreement*).

1. ☐ 0 ☐ 1	17. ☐ 0 ☐ 1	33. ☐ 0 ☐ 1
2. ☐ 0 ☐ 1	18. ☐ 0 ☐ 1	34. ☐ 0 ☐ 1
3. ☐ 0 ☐ 1	19. ☐ 0 ☐ 1	35. ☐ 0 ☐ 1
4. ☐ 0 ☐ 1	20. ☐ 0 ☐ 1	36. ☐ 0 ☐ 1
5. ☐ 0 ☐ 1	21. ☐ 0 ☐ 1	37. ☐ 0 ☐ 1
6. ☐ 0 ☐ 1	22. ☐ 0 ☐ 1	38. ☐ 0 ☐ 1
7. ☐ 0 ☐ 1	23. ☐ 0 ☐ 1	39. ☐ 0 ☐ 1
8. ☐ 0 ☐ 1	24. ☐ 0 ☐ 1	40. ☐ 0 ☐ 1
9. ☐ 0 ☐ 1	25. ☐ 0 ☐ 1	41. ☐ 0 ☐ 1
10. ☐ 0 ☐ 1	26. ☐ 0 ☐ 1	42. ☐ 0 ☐ 1
11. ☐ 0 ☐ 1	27. ☐ 0 ☐ 1	43. ☐ 0 ☐ 1
12. ☐ 0 ☐ 1	28. ☐ 0 ☐ 1	44. ☐ 0 ☐ 1
13. ☐ 0 ☐ 1	29. ☐ 0 ☐ 1	45. ☐ 0 ☐ 1
14. ☐ 0 ☐ 1	30. ☐ 0 ☐ 1	46. ☐ 0 ☐ 1
15. ☐ 0 ☐ 1	31. ☐ 0 ☐ 1	47. ☐ 0 ☐ 1
16. ☐ 0 ☐ 1	32. ☐ 0 ☐ 1	48. ☐ 0 ☐ 1

Please enter all your observations at:
https://docs.google.com/forms/d/e/1FAIpQLSdemN-sVqdmzNIWwanT4swSqUMM9YpncyzIt4AZ4TdeRDPSpQ/viewform?usp=sf_link
for each Qualified Assessor whom you observe administering the ORExt.

FAX Oregon Extended Assessments
Behavioral Research & Teaching, University of Oregon
FAX: 541-346-5689

EMAIL orextended@gmail.com

If you want to share any anecdotal observations or explain sources of concern, please feel free to provide such on a separate email to the above email or fax. If you have any questions regarding the observation process, please contact Brock Rowley or Sevrina Tindal at the email address listed above or phone at (800) 838-3163. Thank you for your support of students with significant cognitive disabilities in Oregon.