Oregon Department of Education

Oregon Extended Assessment (ORExt)

Technical Documentation

Annual Technical Report

for the

2025–2026 Administration

Behavioral Research and Teaching

College of Education

University of Oregon

Brock Rowley

Gerald Tindal

Sevrina Tindal

Aaron Glasgow

Daniel Anderson

July 31, 2026

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

Dr. Andrea Lockard
Assistant Superintendent Office of Research, Assessment, Data, Accountability, and Reporting

Assessment Team Contacts		
TBD Director, Assessment & Student Reporting	*Holly Carter* Assessment Operations and Policy Analyst	*Crys Plattner* Administrative Specialist
Dr. Jon Wiens Director, Accountability Reporting	*Carla Martinez* Administrative Specialist	*Jennifer Bell* Mathematics Assessment Specialist
Tony Bertrand English Language Arts and Social Sciences Assessment Specialist	*Dr. Mariela Salas Bao* Science Assessment Specialist	*Sody Puth Fearn* K-2 Balanced Assessment Specialist
Dr. Audrey Lingley Survey Assessment Specialist	*Ben Wolcott* English Language Proficiency Assessment Specialist	

Table of contents

Overview

This document, prepared by Behavioral Research and Teaching (BRT) for the Oregon Department of Education (ODE), provides updated technical documentation for the Oregon Extended Assessment (ORExt), Oregon's alternate assessment based on alternate academic achievement standards (AA-AAAS). It describes the assessment's design, development, administration, scoring, technical characteristics, and use in reporting student achievement relative to grade-level academic standards under the Every Student Succeeds Act of 2015 (ESSA).

The ORExt measures achievement relative to Oregon's Alternate Academic Achievement Standards (AAAS). It was redesigned in 2014–15 to establish vertical scales for Grades 3–8 English Language Arts (ELA) and Mathematics. The redesign created a common developmental framework that supports longitudinal interpretation across adjacent grades.

Assessment content is derived from Essentialized Standards documented in the Essentialized Assessment Frameworks (EAFs). The EAFs document how Oregon's grade-level academic standards are systematically reduced in depth, breadth, and complexity while preserving the central academic constructs and providing Low, Medium, and High complexity expectations.

A comprehensive Standards Linkage and Item Alignment Study conducted in 2024–2025 evaluated the ORExt system against Oregon's current ELA, Mathematics, and Science standards. The study examined relationships among source standards, alternate standards, complexity parameters, and newly developed field-test items. Where revisions were needed, Essentialized Standards were updated through Oregon's SCORE process (Select, COde, Reduce, Essentialize) and validated through expert review. Findings supported the alignment and accessibility of the reviewed standards and items and identified limited evidence of potential bias.

The EAFs remain the foundation of the ORExt by providing accessible entry points to grade-level academic content. Achievement Level Descriptors (ALDs) and cut scores established in 2015 continue to define the four performance levels used for proficiency reporting.

Accessibility is supported through one-item-per-page presentation, read-aloud options, visual and object supports at lower complexity levels, and selected-response items with direct A/B/C response options. Additional accommodations and accessibility supports are permitted as described in the annual assessor training modules and applicable statewide accessibility guidance.

ORExt items are developed and reviewed by educators with expertise in special education and Oregon content standards. Each item is evaluated for alignment, accessibility, sensitivity, and potential bias before operational use.

Together, the 2025 Standards Linkage and Item Alignment Study, updated EAFs, revised alternate standards, and the technical evidence presented in this report have strengthened the basis for Oregon's continued use of a technically sound, accessible, equitable, and instructionally relevant alternate assessment system for students with the most significant cognitive disabilities.

1 Statewide Standards and Assessment System

1.1 Academic Content Standards

The Oregon State Board of Education adopts academic content standards that define statewide expectations for what students should know and be able to do. Oregon's statewide assessment system includes the ORExt, Oregon's AA-AAAS for eligible students with the most significant cognitive disabilities.

The ORExt measures achievement relative to Oregon's AAAS. The AAAS are linked to the state's grade-level academic content standards through the EAFs. The EAFs document how grade-level content is reduced in depth, breadth, and complexity while preserving the central academic knowledge and skills represented in Oregon's standards.

The relationship among Oregon's current academic content standards, the AAAS, the EAFs, and ORExt items was evaluated in the 2024–2025 Standards Linkage and Item Alignment Study. That study provides current evidence that Oregon's alternate assessment system remains linked to the state's adopted standards in ELA, Mathematics, and Science. Additional evidence regarding the linkage and alignment process is presented in Section 3.1.

1.1.1 ELA

Oregon adopted the Oregon English Language Arts and Literacy Standards in June 2019. These standards define grade-level expectations for reading, writing, speaking and listening, and language. The 2024–2025 Standards Linkage and Item Alignment Study evaluated the relationship among the ELA standards, the AAAS, the EAFs, and newly developed field-test items.

1.1.2 Mathematics

Oregon adopted the current Oregon Mathematics Standards in October 2021. These standards define grade-level expectations across the mathematical domains assessed within Oregon's statewide assessment system. The 2024–2025 Standards Linkage and Item Alignment Study evaluated the relationship among the current Mathematics Standards, the AAAS, the EAFs, and newly developed field-test items.

1.1.3 Science

Oregon adopted updated Oregon Science Standards in June 2022. These standards retain the three-dimensional structure of the Next Generation Science Standards (NGSS) while incorporating Oregon-specific revisions. The 2024–2025 Standards Linkage and Item Alignment Study evaluated the relationship among the updated Science Standards, the AAAS, the EAFs, and newly developed field-test items.

Together, these adopted content standards provide the grade-level foundation for the alternate academic achievement standards and assessment frameworks described in the following section.

1.2 EAFs

The EAFs define the academic content assessed by the ORExt. Through essentialization, Oregon's grade-level academic content standards are reduced in depth, breadth, and complexity while preserving the central academic constructs for students with the most significant cognitive disabilities. The EAFs provide the documented linkage among Oregon's grade-level content standards, the AAAS, assessment blueprints, and ORExt items.

During the 2024–2025 Standards Linkage and Item Alignment Study, the EAFs and AAAS were reviewed against Oregon's current ELA, Mathematics, and Science standards. The study examined whether the alternate standards and assessment content remained appropriately linked to the state's adopted grade-level standards. It also identified areas in which additional clarification, coverage, or alignment was needed and informed the development of new field-test items aligned to the proposed AAAS revisions. Recommendations were submitted to ODE for review and consideration in future development and implementation cycles.

A representative example of the relationship among grade-level standards, essentialized expectations, alternate achievement expectations, and anticipated item difficulty is presented below. Complete EAF documentation is maintained in the supporting program materials.

Table 1.1: Grade 5 Essentialized Assessment Framework Example

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 a story read to the student.	L - A sentence of 7 words or fewer contains 1 character, setting, or event read to the student. M - 2 short sentences contain 1 character, setting, or event read to the student. H - 2 medium-length sentences contain 1 character, setting, or event read to the 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 to its right and 1/10 of what it represents in the place to its left.	Use place value to compare numbers that are multiples of 10 and ones' versus tens' place and .5.	L - Identify multiples of 10: 10, 20, 30, 40, 50, and 60. M - Identify the relationship between the place values for the double-digit numbers 11, 22, 33, 44, and 55. H - Identify which number is in the tens place and which is in the ones place.

Table 1.1: Grade 5 Essentialized Assessment Framework Example *(continued)*

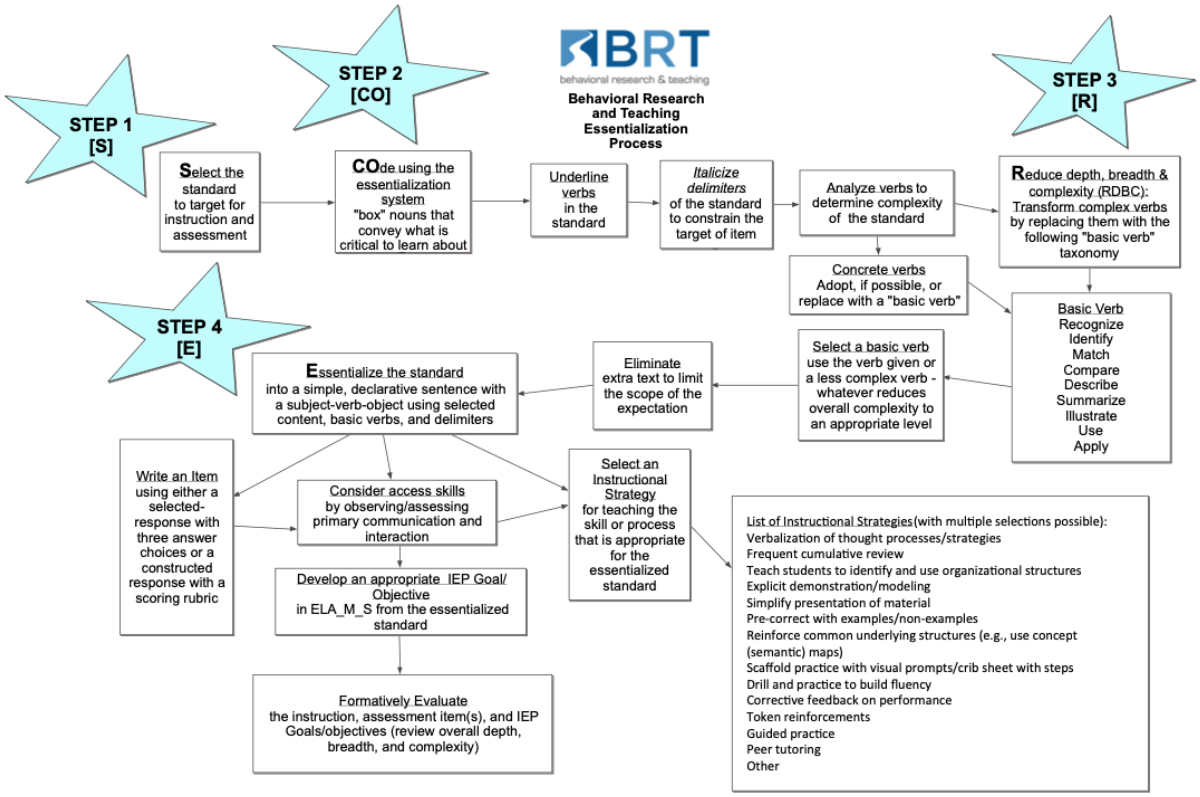
Area	Cluster	Standard	Sub-Standard	Essentialized Standard	L/M/H Descriptors
Science	Matter and Its Interactions	NGSS Standard: Measure and graph quantities to provide evidence that, regardless of the type of change that occurs when heating, cooling, or mixing substances, the total weight of matter is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

Note:
Science standards are linked to both the Next Generation Science Standards and Oregon Science Standards.

1.3 Essentialization and Instructional Use

The EAFs also support instruction by helping educators connect grade-level academic content to students' present levels of academic achievement and functional performance, instructional planning, and Individualized Education Program (IEP) goals. In this role, the EAFs provide a common content foundation for assessment and instruction while maintaining the expectation that students with the most significant cognitive disabilities receive access to grade-level academic content.

The essentialization process begins with Oregon's adopted grade-level academic content standards and systematically reduces their depth, breadth, and complexity. The resulting Essentialized Standards retain the central academic knowledge and skills, support development of the AAAS, and guide the creation of assessment blueprints and ORExt items. The process is summarized below.



1.4 Required Assessments

During the **2025–2026** school year, the ORExt was administered in ELA and Mathematics in Grades 3–8 and Grade 11. Science was administered in Grades 5, 8, and 11. This schedule corresponded to federal assessment requirements for annual assessment in ELA and Mathematics in Grades 3–8 and once in high school, and for assessment in Science once during each of the grade spans 3–5, 6–9, and 10–12.

Content Area	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
ELA	X	X	X	X	X	X	X
Mathematics	X	X	X	X	X	X	X
Science			X			X	X

1.5 Policies for Including All Students in Assessments

Oregon's statewide assessment policies establish the expectation that all students enrolled in the assessed grades participate in the Oregon Statewide Assessment System. Students with disabilities

participate in the general statewide assessment with appropriate accessibility supports unless the student's IEP team determines that the ORExt is the appropriate assessment. The decision to administer the ORExt is made by the student's IEP team, including the parent, using Oregon's alternate assessment participation criteria. A student may not be assigned to the ORExt solely on the basis of disability category, instructional setting, expected performance, or the need for accommodations.

Oregon law provides two distinct processes through which a student may not participate in a required statewide assessment. Under Oregon Revised Statutes (ORS) 329.479, a parent or adult student may annually opt out of the statewide summative assessments in ELA and Mathematics by submitting the state-issued opt-out form. This statutory opt-out process applies to both the general and alternate ELA and Mathematics assessments.

For other state-required assessments, including Science, a parent or adult student may request an exemption when participation conflicts with the student's disability or religious beliefs. These requests are governed by Oregon Administrative Rules (OAR) 581-021-0009 and require district review and approval. An exemption is not an IEP-team decision and may not be initiated by a school or district to exclude a student from assessment participation.

The distinction between selecting the ORExt and excusing a student from an assessment is important. The IEP team determines which statewide assessment is appropriate for the student. A parent or adult student determines whether to submit an allowable opt-out or exemption request. Oregon's participation policies therefore preserve access to the statewide assessment system for students with disabilities while recognizing the limited exceptions established in state law and rule.

Current participation, opt-out, exemption, accessibility, and administration requirements are communicated annually through the Test Administration Manual, Oregon's alternate assessment decision-making guidance, District Test Coordinator training, Qualified Test Coordinator training, and Qualified Assessor training. Together, these policies and procedures establish a statewide framework for including eligible students with the most significant cognitive disabilities in the ORExt and for documenting the limited circumstances in which a student does not participate.

English Learners

The ORExt is administered in English. Approved language supports may be provided when needed to help a student understand administration directions, provided that the support does not translate, simplify, explain, or otherwise alter the academic content, response options, scoring criteria, or construct measured by an item. Districts are responsible for ensuring that personnel who provide language support are appropriately qualified and follow the procedures established in the Oregon Accessibility Manual and the Test Administration Manual.

English learners, including English learners with disabilities, are included in Oregon's statewide assessment system. For a student who is also eligible for the ORExt, the IEP team determines the appropriate statewide assessment using the same participation criteria applied to all students considered for an alternate assessment. English learner status alone does not determine participation in the ORExt.

Additional statewide guidance regarding educational services for English learners is available through the Oregon Department of Education English Learner Program.

Native-Language Supports

When an assessment does not provide a translated or side-by-side version, approved administration directions may be interpreted in the student's language of origin by personnel whom the district has determined are qualified to provide educational interpretation. Any language support must be implemented in accordance with the Oregon Accessibility Manual and Test Administration Manual.

Native-language support is intended to provide access to administration directions, not to change the knowledge or skill required by an assessment item. Assessment content, response options, scoring criteria, and standardized administration procedures must remain consistent with the approved ORExt form. These requirements help include English learners in the statewide assessment system while preserving the validity and comparability of ORExt score interpretations.

Students Who Are Deaf or Hard of Hearing

Students who are Deaf or hard of hearing may use approved communication and accessibility supports, including American Sign Language (ASL), when the support is consistent with the student's documented needs and permitted by state policy. Personnel implementing these supports must follow the Oregon Accessibility Manual, Test Administration Manual, and applicable ORExt training requirements.

Communication supports are intended to provide access to assessment directions and content without translating, simplifying, explaining, or otherwise altering the academic construct measured by an item. Item content, response options, scoring criteria, and standardized administration procedures must remain consistent with the approved ORExt form.

Students Who Are Blind or Have Low Vision

Students who are blind or have low vision may use approved presentation formats and accessibility supports, including Braille and Large Print, when consistent with the student's documented needs and state policy. Accessible materials are developed from approved ORExt content and are intended to preserve the same academic knowledge, response expectations, and scoring criteria represented by the standard form.

Personnel administering accessible formats must follow the Oregon Accessibility Manual, Test Administration Manual, test-security requirements, and applicable ORExt training procedures. Additional technical evidence regarding accessible assessment formats is presented in Section 4.6.

1.6 Participation Data

Oregon monitors the participation of students with disabilities in statewide assessments as part of its federal reporting responsibilities under the Individuals with Disabilities Education Act (IDEA). Participation and performance for students with IEPs are reported through Indicator 3 of the State

Performance Plan and Annual Performance Report submitted to the U.S. Department of Education's Office of Special Education Programs.

Statewide assessment participation and achievement results are also reported through Oregon's public reporting systems. These data provide evidence regarding the inclusion of students in the assessed grades and support ongoing monitoring of participation requirements for all students, including students with disabilities and English learners. Additional statewide participation and performance information is available in the 2024–2025 Statewide Report Card.

The policies and participation evidence presented in this chapter establish the statewide standards and assessment context for the operational procedures described in Chapter 2.

2 Assessment Operations

2.1 Test Design and Development

The ORExt uses operational forms developed from Oregon's AAAS, EAFs, and content-specific assessment blueprints. Together, these sources define the academic content and reporting domains represented on each operational assessment.

Each 2025–2026 assessment contained 30 operational items used for scoring and six embedded field-test items used to support future item-bank development. Operational forms were constructed to satisfy blueprint requirements while representing an appropriate range of item difficulty for the students assessed. Candidate forms were reviewed for blueprint compliance, content coverage, item difficulty, field-test placement, accessibility, item order, and scoring accuracy before release.

For the vertically scaled ELA and Mathematics assessments in Grades 3–8, Test Characteristic Curves (TCCs) were used to evaluate the ordering and separation of expected scores across adjacent-grade forms. Test Information Functions (TIFs) were used concurrently to evaluate conditional measurement precision across the reporting scale and near the approved achievement-level cut scores. These prospective form-review procedures complemented the post-administration psychometric analyses presented in Chapters 3 and 4.

During 2025–2026, BRT continued to modernize the test development system. Improvements included integrated item management, blueprint monitoring, form assembly, item preview and editing, version control, and quality-assurance functions. These changes reduced reliance on manual file-based workflows and strengthened the consistency, documentation, and repeatability of annual test construction.

2.1.1 Score Reporting and Longitudinal Interpretation

ELA and Mathematics scores in Grades 3–8 are reported on vertical scales that support interpretation across adjacent grades.

Longitudinal interpretations presented in this report incorporated each student's conditional standard error of measurement (CSEM). Observed score changes were therefore evaluated in relation to measurement uncertainty rather than treated as error-free differences.

ORExt results are reported across four achievement levels. Levels 3 and 4 represent performance at or above Oregon's proficient standard for the AAAS, while Levels 1 and 2 represent performance below the proficient standard. The achievement levels, ALDs, and associated cut scores are described in Chapter 6.

2.1.2 Test Blueprint

The ORExt test blueprint specifies the required distribution of operational items across reporting domains and content strands for each assessment. The distribution reflects proportionate coverage of the standards represented within each content area. Together with the EAFs, the blueprint guides item selection, form assembly, and field-test placement while ensuring that each operational assessment appropriately represents Oregon's AAAS.

Blueprint requirements are verified throughout the form-development process. Candidate forms are evaluated for content representation, subdomain coverage, item sequence, accessibility, and psychometric characteristics before approval for operational use. The blueprint therefore serves as both a content specification and a quality-assurance tool supporting consistent assessment construction from year to year.

The 2024–2025 Standards Linkage and Item Alignment Study further examined the relationships among Oregon's current academic content standards, the EAFs, the AAAS, and newly developed field-test items. Findings from that study have informed revisions to the alternate standards and continue to guide item development, field testing, and future operational form construction.

2.1.3 Test Development

The ORExt is a fixed-form alternate assessment administered in ELA and Mathematics in Grades 3–8 and Grade 11 and in Science in Grades 5, 8, and 11. Assessment items are developed through a structured process that includes item writing, educator review, accessibility review, content revision, field testing, psychometric evaluation, and selection for operational use.

During the 2024–2025 Standards Linkage and Item Alignment Study, 393 newly developed field-test items were reviewed for standards alignment, cognitive complexity, accessibility, and potential sources of bias. Results from those reviews informed revisions to the EAFs and AAAS while identifying high-quality items for future operational use.

During 2025–2026, the test development environment was further modernized to support direct item import, integrated editing, blueprint monitoring, preview-based quality assurance, version control, and secure transfer of completed forms into the operational testing system. These enhancements reduced reliance on manual file-based workflows while improving the efficiency, consistency, and documentation of annual assessment development.

Before operational release, completed forms undergo multiple quality-assurance reviews evaluating blueprint compliance, content accuracy, accessibility, item order, scoring rules, and consistency between electronic and paper/pencil versions. ELA and Mathematics forms are also reviewed using TCCs and TIFs to confirm expected vertical ordering and measurement precision before administration.

Because the ORExt uses fixed operational forms rather than computer-adaptive testing, adaptive test assembly, item-selection algorithms, and online calibration procedures are not applicable.

2.1.4 Educator Participation and Preparation

ORExt item writers and reviewers are experienced Oregon educators with expertise in special education, general education, assessment, curriculum, or item development. Item writers are licensed educators with classroom experience, and preference is given to individuals with prior assessment development experience, advanced professional preparation, or demonstrated expertise serving students with the most significant cognitive disabilities.

Educators participate throughout standards review, item writing, accessibility review, field testing, and technical evaluation. Before participating, item writers and reviewers complete training addressing essentialization, item development guidelines, blueprint requirements, accessibility, cognitive complexity, universal design principles, and identification of potential sources of bias. These procedures promote consistent application of Oregon's assessment development standards across all content areas.

The 2024–2025 Standards Linkage and Item Alignment Study included 13 Oregon educators with direct experience serving students with the most significant cognitive disabilities and administering the ORExt. Most participants reported more than 10 years of special education experience, and nine had administered the ORExt for more than five years. Using standardized review criteria, participants evaluated newly developed field-test items for standards alignment, accessibility, cognitive complexity, and potential bias. Their recommendations contributed directly to revisions of the EAFs and AAAS and informed future item development.

2.2 Test Administration

The ORExt is administered under statewide policies and procedures established by ODE governing test security, standardized administration, accessibility, scoring, and reporting. These procedures are designed to promote consistent administration across districts while providing appropriate access for eligible students.

During the 2025–2026 administration, statewide implementation was supported by the Training and Proficiency (T&P) system, secure electronic test delivery, online practice tests, user account management, and Help Desk support. Together, these operational systems promoted standardized administration, timely technical assistance, and consistent implementation throughout the statewide testing window.

The following sections describe statewide administration procedures, training, electronic delivery, monitoring, security, and data integrity.

2.2.1 Administration

Qualified Assessors (QAs) complete annual T&P requirements before administering the ORExt. Training addresses standardized administration, appropriate implementation of accessibility supports, scoring procedures, minimum participation requirements, and specialized guidance for administering the assessment to students using ASL and other approved accessibility supports.

2.2.2 T&P Requirements

QAs, Qualified Test Coordinators (QTCs), and users assigned the Alternate Version of the Student Educational Equity Development Survey (Alt-SEED) Only User role complete annual online training and proficiency requirements through the OExt T&P system. The platform provides role-specific training, administration guidance, proficiency assessments, practice materials, and secure access to operational resources that support consistent statewide implementation of the assessment.

Only licensed and trained QAs may administer the OExt. QTCs coordinate local implementation, while District Test Coordinators (DTCs) provide district-level oversight and support. Each user must complete the training and proficiency requirements associated with the assigned role before receiving access to operational assessment materials. QAs must achieve a score of at least 80% on the required proficiency assessment before administering an operational assessment.

Beginning with the 2024–2025 administration, Oregon replaced the former Qualified Trainer model with the QTC role and transitioned fully to asynchronous online training. This approach provides consistent statewide training while allowing districts flexibility in assigning local coordination responsibilities.

The following table summarizes the distribution of recorded proficiency attempts for the 2025–2026 OExt T&P system. Results were based on the official annual proficiency summary prepared from the Learning Management System (LMS) and are reported by credential type.

Table 2.1: Recorded Attempts for Annual Proficiency Requirements

Number of Attempts	DTC	QTC	QA	Alt-SEED	ASL
1	134	114	624	12	104
2	16	8	28	2	16
3	2	11	27	0	2
4	1	0	2	0	0
5	0	0	1	0	0
6	0	0	0	0	0

Note. DTC = District Test Coordinator; QTC = Qualified Test Coordinator; QA = Qualified Assessor. ASL refers to the separate American Sign Language proficiency assessment administered as one component of the OExt T&P program.

Most users successfully completed the required proficiency assessment on their first attempt, with relatively few requiring additional attempts. These results provided evidence that the T&P program communicated the required knowledge and procedures effectively to most participating users. Proficiency results were interpreted together with direct observation, item-level scoring agreement, Help Desk activity, and other operational evidence when evaluating the consistency of statewide OExt administration.

Training effectiveness is evaluated using multiple operational evidence sources. Help Desk activity, frequently asked questions (FAQ) use, proficiency results, and user feedback are reviewed throughout and following each administration to identify recurring operational issues and opportunities to strengthen training, user documentation, and statewide support.

During the 2025–2026 administration, the Help Desk managed **216 support conversations** and responded to **361 messages**. The most common requests involved account access, credential

verification, student rostering, test administration, and assessment delivery. Together with FAQ analytics and annual T&P results, these operational data informed continued refinement of training materials, online resources, and statewide assessment administration.

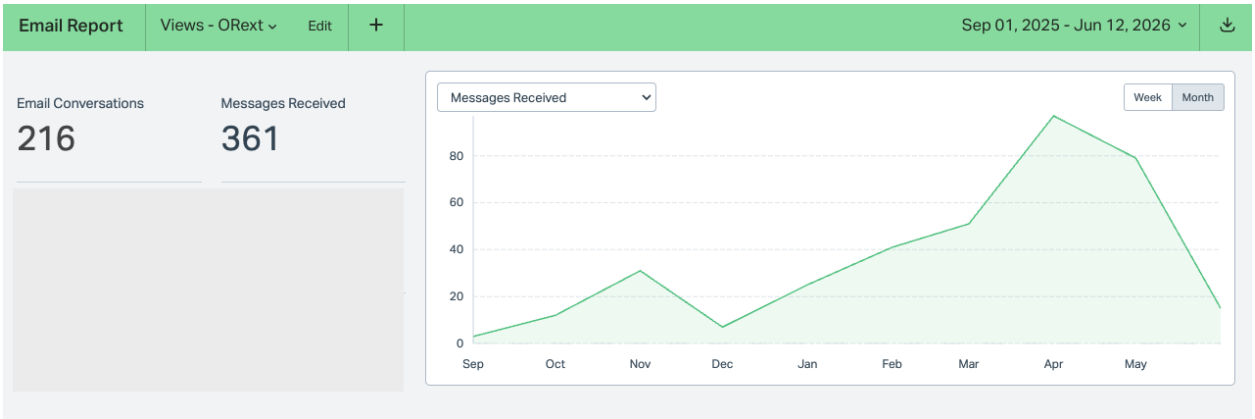
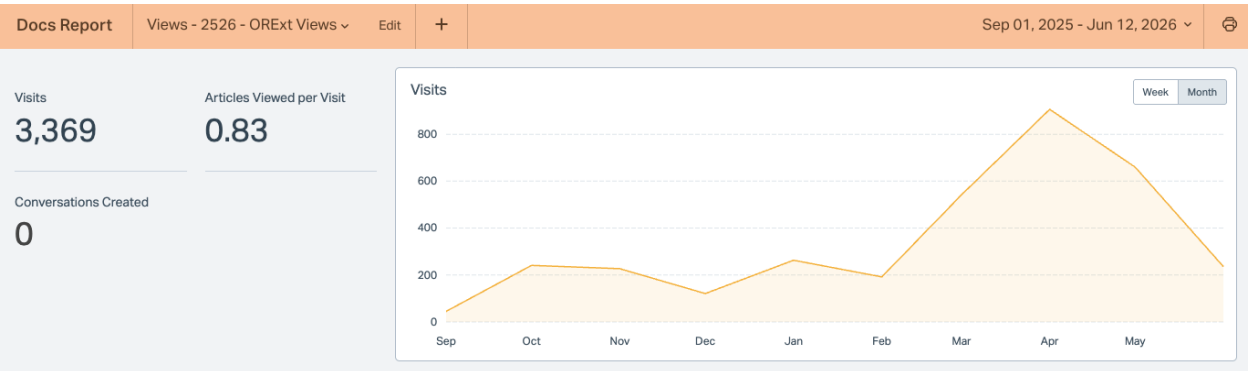


Figure 2.1: OREXt Help Desk Activity During the 2025–2026 Administration

The OREXt Help Desk FAQ supplements the T&P system by providing educators with immediate access to guidance on common assessment questions throughout the administration cycle. Frequently accessed topics include account access, student rostering, electronic test administration, instructional resources, data entry, credential verification, and assessment procedures.

During the 2025–2026 administration, the FAQ received **3,369 visits**, with an average of **0.83 articles viewed per visit**. No Help Desk conversations were initiated directly from the FAQ. This result indicated that users did not use the FAQ interface to escalate questions to the Help Desk, but it did not demonstrate whether users located all needed information. FAQ analytics were therefore interpreted together with Help Desk activity, proficiency results, and educator feedback when identifying recurring support needs and opportunities to improve self-service resources.

Help Desk FAQ analytics complement proficiency results, Help Desk activity, and educator feedback as part of Oregon's annual continuous improvement process. Collectively, these operational data help identify recurring information needs and guide updates to training materials, user documentation, online resources, and statewide assessment support.



2.2.3 Electronic Test Delivery

The OROExt is administered through a secure electronic test delivery system, with paper/pencil materials available when appropriate. Electronic and paper/pencil administrations follow the same standardized administration, accessibility, scoring, and reporting procedures to support comparable assessment experiences across delivery formats.

The electronic system supports secure user authentication, assessment delivery, practice tests, minimum-participation procedures, automated scoring where applicable, and centralized operational monitoring. During the 2025–2026 administration, modernization efforts focused on usability, accessibility, system performance, and consistency across assessment delivery while maintaining secure statewide administration. Operational feedback, Help Desk activity, user survey responses, and system analytics were reviewed to evaluate these changes and identify priorities for further improvement.

Browser and operating-system information automatically recorded by the electronic assessment system during the 2025–2026 administration was reviewed to inform compatibility testing, browser support, and continued system enhancement.

Chrome OS, Windows, and macOS accounted for the majority of assessment sessions, indicating that most OROExt sessions occurred within a limited number of widely used computing environments. Tablet access represented a smaller proportion of sessions, and Android devices accounted for very little operational activity.

Table 2.2: Operating Systems Used During the 2025–2026 OROExt Administration

OS	Sessions	Percent
Windows	4,147	39%
Chrome OS	2,995	28%
Mac OS X	2,183	21%
iOS	1,270	12%
Android	3	0%

Browser usage followed a similar pattern. Chrome accounted for the majority of assessment sessions, followed by Microsoft Edge, Safari Mobile, and Safari. These findings identify the platforms that should receive priority in compatibility testing and technical support.

Table 2.3: Browsers Used During the 2025–2026 OROExt Administration

Browser	Sessions	Percent
Chrome	8,357	79%
Safari Mobile	1,202	11%
Edge	503	5%
Safari	417	4%
Chrome Mobile	71	1%
Firefox	48	0%

2.3 Monitoring Test Administration

ODE and BRT monitor OExt administration through complementary operational and observational evidence. During the statewide testing window, credential completion, user access, student rostering, assessment delivery, scoring completion, Help Desk activity, reported irregularities, and other operational records are reviewed to identify issues requiring clarification, correction, or technical assistance. Following each administration, T&P results, FAQ analytics, educator feedback, and system-use data are reviewed to identify recurring patterns and guide revisions to training, documentation, and statewide support.

Direct evidence of administration fidelity is obtained through observations of QAs conducted by QTCs using a standardized protocol. The protocol addresses test security, assessment preparation, testing conditions, accessibility supports, level of support, minimum-participation procedures, scoring, and response recording. Results of the 2025–2026 observation study, including item-level scoring agreement, are presented in Section 4.1.1. Together, ongoing operational monitoring and direct observation provide evidence regarding implementation of standardized administration procedures and identify areas requiring follow-up in subsequent training and assessment cycles.

2.3.1 T&P Survey

BRT administered a statewide educator survey in Spring 2025 following implementation of the redesigned T&P system. The survey examined educators' perceptions of training quality, role-specific learning pathways, system navigation, credential verification, communication, and operational support.

Survey findings were interpreted together with proficiency results, Help Desk activity, FAQ analytics, and other operational evidence. This integrated review provided evidence regarding the effectiveness of the T&P program and identified opportunities for improvement before the 2025–2026 administration.

2.3.1.1 Survey Participants and Design

The voluntary Spring 2025 survey was distributed to Oregon educators who used the T&P system, including DTCs, QTCs, QAs, and Alt-SEED users. The survey combined Likert-scale items with open-ended questions addressing training effectiveness, site usability, credential verification, communication, and statewide operational support.

Respondents represented school districts throughout Oregon and reported substantial experience with both education and the OExt. More than 84% reported over 11 years of experience in education, and more than half reported at least five years of experience administering the OExt.

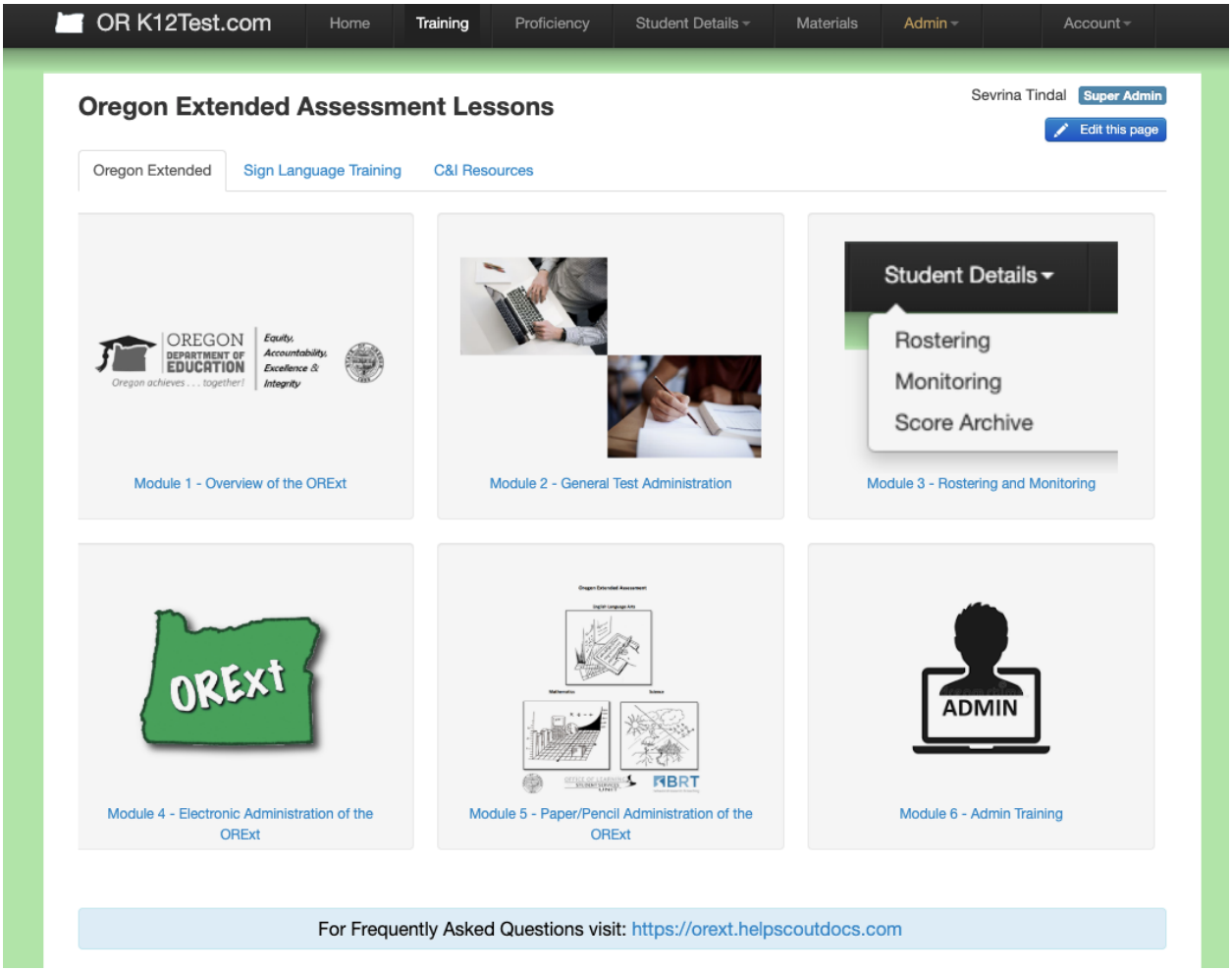
2.3.1.2 Findings and Program Improvements

Overall, educators reported that role-specific training aligned with their responsibilities, prepared them to administer the assessment, and was easy to navigate. Respondents also identified opportunities to strengthen onboarding for new users, improve guidance for specific roles, enhance

communication and credential verification, expand support for small and rural districts, and continue refining technical assistance.

Survey findings are reviewed annually alongside proficiency results, Help Desk activity, FAQ analytics, administration monitoring, and follow-up feedback from Oregon educators. Together, these sources guide revisions to training modules, onboarding resources, user documentation, credential guidance, and statewide operational support.

The T&P system serves as the central online environment for annual training, proficiency assessment, administration guidance, operational resources, and statewide support.



The T&P system also serves as a primary communication mechanism among ODE, BRT, and local assessment personnel. Annual updates to training materials, administration guidance, accessibility resources, and operational procedures are distributed through the system to promote consistent statewide implementation.

Official student, school, district, and statewide assessment reports are produced through ODE reporting systems. The T&P system complements these reporting systems by supporting assessment administration rather than serving as the official repository for state assessment results.

2.4 Test Security

Test security is governed by ODE policies and annual administration requirements designed to protect the validity of assessment results and the confidentiality of secure assessment materials. Before receiving access to operational materials, assessment personnel complete annual training addressing secure handling of materials, standardized administration, confidentiality, appropriate use of accessibility supports, and procedures for reporting testing irregularities.

DTCs oversee local administration and conduct the initial review of suspected testing irregularities. Reported incidents are reviewed by ODE in coordination with district personnel and, when appropriate, BRT. Confirmed incidents are addressed through established corrective procedures, which may include score corrections, administration corrections, additional training, or modification of user access privileges.

These statewide security procedures help ensure that assessment results are based on standardized administration conditions and that secure assessment materials remain protected throughout the testing cycle.

2.5 Data Integrity and Privacy

ORExt assessment materials and student data are maintained within secure systems throughout item development, administration, scoring, reporting, and archival storage. Access is controlled through authenticated user accounts, role-based permissions, credential verification, and annual security and confidentiality training.

Authorized and authenticated users receive access only to the student information and system functions required to perform their assigned responsibilities. Student-level data are restricted to personnel with a legitimate educational interest who have completed all required security, confidentiality, and T&P requirements.

Personally identifiable information (PII) is protected through ODE policies and secure data-management procedures that govern the collection, storage, transmission, reporting, and retention of assessment data. Public reporting applies minimum reporting thresholds and data-suppression rules to protect individual student privacy while supporting meaningful statewide, district, school, and subgroup reporting.

Together, Oregon's security, data-integrity, and privacy procedures help ensure that ORExt results are based on secure administration practices, protected throughout the assessment lifecycle, and reported in a manner that supports both appropriate educational use and student confidentiality.

3 Technical Quality: Validity

3.1 Validity Evidence Based on Test Content

Validity concerns the extent to which available evidence supports the intended interpretation and use of assessment scores. As described by Messick (1989), validity is evaluated by considering the relationship among the construct being measured, the assessment procedures used to measure that construct, and the intended uses of the resulting scores. A validity argument is therefore built from multiple complementary sources of evidence rather than from a single statistical result or study.

Consistent with the *Standards for Educational and Psychological Testing* (2014), this chapter organizes validity evidence into four categories directly aligned with the applicable Federal Assessment Peer Review Critical Elements: (a) evidence based on test content, (b) response processes, (c) internal structure, and (d) relations to other variables. Evidence regarding reliability, fairness, scoring, reporting, and the consequences and limitations of score use is addressed in the chapters that follow and contributes to the broader ORExt validity argument.

The validity argument presented throughout this report supports the intended use of the ORExt, Oregon's AA-AAAS, to measure student achievement relative to Oregon's grade-level academic content standards for students with the most significant cognitive disabilities. Evidence presented throughout this report includes standards linkage and alignment, assessment design, item development, accessibility, T&P, standardized administration, scoring procedures, operational monitoring, psychometric analyses, and longitudinal evaluation of the vertically scaled ELA and Mathematics assessments.

The ORExt is administered and scored using standardized statewide procedures, with approved accommodations implemented according to state policy. QAs receive annual training addressing administration, accessibility, scoring, and appropriate levels of support. During administration, QAs provide only the support permitted under Oregon's standardized administration procedures so that students can meaningfully access assessment content without altering the academic construct being measured. Collectively, the evidence presented in this chapter supports the intended interpretation and use of ORExt scores for statewide accountability, instructional planning, and educational decision making.

3.1A Alignment Between AA-AAAS and Academic Content Standards

Validity evidence based on test content begins with documentation describing the content measured by the assessment and the procedures used to develop operational forms. The ORExt assessment specifications (see OR Extended Assessment-Item Development Info) and the ORExt Test Blueprint define the assessment blueprint, reporting domains, test length, item formats, and the processes used to maintain alignment between Oregon's academic content standards and the ORExt. These documents provide the foundation for evaluating content validity.

All operational assessment items are linked to Oregon's grade-level academic content standards through the AAAS and the EAFs. Item development incorporates universal design principles together with established content-referenced item-writing practices. Particular attention is given to language complexity, accessibility, and the systematic reduction of the depth, breadth, and complexity (RDBC) of grade-level content while preserving the intended academic construct. The OREXt also employs a standardized scaffolded administration model that provides only the level of support necessary for students to access assessment content without altering the construct being measured.

Operational items are developed through an iterative process that includes item writing, educator review, accessibility review, technical review, field testing, and psychometric evaluation. The Item Writer Training documents the procedures used to prepare item writers and reviewers. BRT psychometricians, Oregon educators, and content specialists collaboratively review operational and field-test items for standards alignment, blueprint coverage, accessibility, cognitive complexity, statistical performance, and potential sources of bias. Items that do not perform as intended are revised, replaced, or removed while maintaining appropriate content representation, vertical scaling, and blueprint requirements. This annual review cycle supports continuous improvement while preserving the technical integrity of the assessment.

Evidence supporting alignment extends across multiple studies. Full documentation of the original 2014 linkage study and the independent alignment study completed in 2017 is available in the Oregon Extended Assessment Alignment Study. Most recently, the 2024–2025 Standards Linkage and Item Alignment Study examined the relationships among Oregon's updated academic content standards, the revised AAAS, the EAFs, and newly developed field-test items. Together, these studies provide a cumulative body of evidence supporting the continued content validity of the OREXt.

The independent 2017 alignment study found sufficient to strong alignment among the OREXt, Oregon's academic content standards, and the ALDs across ELA, Mathematics, and Science. Multiple rounds of independent professional review concluded that the standards, assessment content, and ALDs formed a coherent framework supporting the intended interpretation and use of OREXt scores. The study also identified opportunities for continued refinement that informed subsequent assessment development.

Annual field testing, psychometric review, and the 2024–2025 Standards Linkage and Item Alignment Study have continued to strengthen the alignment among Oregon's academic content standards, the AAAS, the EAFs, and operational assessment items. Field-test items are evaluated using psychometric, content-based, and standards-alignment evidence before operational use, allowing lower-performing items to be replaced while maintaining blueprint coverage, vertical scaling, and an appropriate distribution of item complexity consistent with Webb (2002).

Additional evidence supporting Oregon's AAAS is provided through the State's standard-setting process (see the OREXt Assessment Technical Report on Standard Setting). The Oregon State Board of Education adopted the AAAS, ALDs, and associated cut scores in June 2015. These standards continue to provide the interpretive framework for reporting and using OREXt results.

3.1B Operational Form Assembly and Blueprint Verification

Candidate operational forms are assembled within an integrated test development environment that supports systematic review of content representation and expected psychometric functioning before forms are approved for administration. The application links selected items to the ORExt item bank, identifies each item by grade, content area, and blueprint subdomain, and compares the number of selected items with the grade-specific minimum requirements established in the ORExt test blueprint.

The minimum subdomain requirements were retained when the ORExt transitioned from 48 total items to 36 total items. Retaining these requirements preserved the minimum expected representation of the assessed content domains while reducing overall test length. Each 2025–2026 form included 30 operational items used for scoring and 6 embedded field-test items used to support future assessment development.

The form-development environment also supports iterative review through complementary graphical and statistical diagnostics. Developers examine the distribution and sequence of calibrated Rasch item difficulties by grade, item position, and blueprint subdomain. Form-level summaries include the number of operational items, mean and median item difficulty, minimum and maximum difficulty, total difficulty range, and the number of items located below and above the Rasch scale origin.

For each candidate form, the application calculates TCCs by summing Rasch-model expected response probabilities across the selected operational items. TIFs are calculated by summing item information across the ability continuum and expressing the results on the ORExt Rasch Unit (RIT) reporting scale. Performance-level cut scores are included in the form-review displays so developers can evaluate expected score functioning and conditional measurement precision near the classification boundaries used for reporting. Raw-score-to-RIT conversion tables and the minimum raw and RIT scores associated with each performance level are reviewed as part of the same form-quality process.

Together, blueprint comparisons, item-difficulty diagnostics, TCCs, TIFs, and score-conversion tables provide a documented and repeatable quality-assurance process before operational administration. Following administration, point-measure correlations and outfit mean-square statistics provide complementary evidence regarding observed item functioning. These prospective and retrospective reviews connect operational form assembly directly to the validity and technical-quality evidence presented in Sections 3.3 and 4.1.

The approved minimum blueprint specifications used during 2025–2026 form construction are presented below. These tables report the minimum required representation of each subdomain. They do not represent the total number of items included on an operational form, nor do they by themselves demonstrate that each 2025–2026 operational form met every requirement.

Table 3.1: ELA 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 3.2: 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	–
Congruence	–	–	–	–	–	–	2
Right Triangles and Trigonometry	–	–	–	–	–	–	1
Expressing Geometric Properties with Equations	–	–	–	–	–	–	3
Geometric Measurement and Dimension	–	–	–	–	–	–	1
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 3.3: 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

3.1C AA-AAAS Linkage to General Content Standards

In Spring 2025, the ODE and BRT conducted a comprehensive Standards Linkage and Item Alignment Study to evaluate the relationship between Oregon's updated academic content standards and the OExt. The study examined the linkage of the Essentialized Standards (EsSt) to Oregon's grade-level standards in ELA (2019), Mathematics (2021), and Science (2022). It also evaluated the alignment of newly developed field-test items with the Essentialized Standards they were intended to measure.

A panel of 13 QTCs with extensive experience in special education, alternate assessment, and service to students with the most significant cognitive disabilities served as expert reviewers. Before beginning the study, reviewers completed standardized training in Oregon's SCORE framework, the study criteria, and BRT's Distributed Item Review (DIR) system.

The study included linkage reviews of 149 Essentialized Standards and alignment reviews of 393 field-test items, including 205 ELA items, 116 Mathematics items, and 72 Science items. Essentialized Standards receiving lower linkage ratings were revised through the SCORE process and reviewed again before the study was completed.

The study reported strong linkage between Oregon's current academic content standards and the Essentialized Standards across all three content areas. Field-test item ratings also supported alignment with the intended Essentialized Standards, and reviewers reported high levels of accessibility with limited evidence of potential bias. This technical report summarizes the study's conclusions; the review criteria, rating procedures, quantitative results, revision process, and item-level findings are documented in the complete study report linked below.

Collectively, the study provides current validity evidence supporting the following conclusions:

- Oregon's Essentialized Standards maintain strong linkage to the State's current grade-level academic content standards.
- ORExt field-test items align with the Essentialized Standards they are intended to measure.
- The reviewed items provide accessible representations of academic content for students with the most significant cognitive disabilities.
- Low, Medium, and High complexity parameters support variation in cognitive demand while maintaining alignment with the intended academic constructs.

Building on the 2014 internal linkage study and the independent alignment study completed in 2017, the 2025 Standards Linkage and Item Alignment Study extends the cumulative body of content-validity evidence supporting the ORExt. The findings demonstrate that Oregon's AAAS remain linked to the State's current academic content standards and that newly developed field-test items represent the intended essentialized academic expectations. The complete study is available in the ORExt Linkage and Alignment Study Report 2025.

Together, the assessment specifications, blueprint-verification procedures, alignment studies, educator reviews, and annual field-testing process provide evidence that ORExt content represents the intended academic constructs. The next section considers whether students' responses are elicited, supported, and scored in ways that are consistent with those intended constructs.

3.2 Validity Evidence Based on Response Processes

Validity evidence based on response processes concerns whether the knowledge, skills, judgments, and actions required by assessment items and administration procedures are consistent with the academic constructs the assessment is intended to measure. For the ORExt, this evidence is drawn from expert review of item cognitive complexity, standardized item presentation, scripted administration directions, defined levels of support, item-specific scoring guidance, annual T&P requirements, and monitoring of administration and scoring practices.

During the 2024–2025 Standards Linkage and Item Alignment Study, 13 Oregon educators with expertise in special education and alternate assessment reviewed 393 newly developed field-test items for standards alignment, cognitive complexity, accessibility, and potential sources of bias. Reviewers evaluated whether each item required an appropriate demonstration of the knowledge or skill represented by its Essentialized Standard. Results supported the use of Low, Medium, and High complexity parameters to vary cognitive demand while maintaining alignment with the

intended academic construct. Items receiving lower ratings were revised or identified for additional review before potential operational use.

QAs are trained to present items, provide only approved supports, recognize and record student responses, and apply scoring rules consistently. Administration procedures are designed to provide meaningful access for students with the most significant cognitive disabilities without changing the knowledge or skill required by an item. Direct operational evidence regarding implementation of these procedures, including standardized observations and item-level scoring-agreement results, is presented in Section 4.1.1.

Taken together, the expert-review and administration-fidelity evidence supports the conclusion that ORExt items are designed and administered to elicit responses consistent with the intended academic constructs. The current evidence is based primarily on expert judgment and observed administration practices rather than direct cognitive-laboratory research with students, identifying an area that may be strengthened through future response-process studies.

ORExt minimum-participation procedures require a student to be presented with at least the first five items before discontinuation is considered. These procedures help ensure that decisions to discontinue an assessment are based on the student's observed interaction with ORExt content rather than on assumptions about disability, communication mode, or expected performance.

During the 2025–2026 administration, the Oregon Observational Rating Assessment (ORora) remained available as an optional observational activity for students who could not meaningfully engage with ORExt academic content. ORora was not administered under the same standardized conditions as the ORExt and was not designed to measure the same academic constructs. ORora results are therefore not interpreted as response-process or validity evidence for ORExt scores. Detailed ORora score distributions, subscores, and correlations are not presented in the main body of this report.

Oregon is transitioning away from ORora as it develops the Pathways Communication Scale (PCS), which is intended to provide more useful communication-focused information for students who cannot meaningfully access ORExt academic content. Because the PCS has a purpose distinct from the ORExt, its future use will not alter the interpretation of ORExt scores as measures of academic achievement relative to Oregon's AAAS.

3.3 Validity Evidence Based on Internal Structure

Validity evidence based on internal structure examines whether relationships among assessment items, scores, and reporting components are consistent with the constructs the assessment is intended to measure. For the ORExt, this evidence includes blueprint representation, item-discrimination indices, Rasch item-fit statistics, expected score functioning, conditional measurement precision, performance-level distributions, and relationships among content-area and ELA subscore measures.

The EAFs and assessment blueprints define the intended organization of content within each grade and subject. The tables presented in Section 3.1B summarize the minimum distribution of Essentialized Standards across reporting domains for ELA, Mathematics, and Science. Those specifications establish the content structure that the psychometric analyses in this section evaluate.

Individual Essentialized Standards may be linked to more than one grade-level academic content standard. Therefore, the blueprint summaries describe the distribution of written Essentialized Standards and minimum subdomain requirements rather than the full range of content relationships documented in the EAFs. Evidence based on test content establishes what each assessment is intended to measure. Evidence based on internal structure evaluates whether student responses and resulting scores behave consistently with that intended design.

OExt scores are generated and reported primarily at the content-area test level. ELA also includes Reading and Writing subscores that provide supporting descriptive information. The analyses that follow evaluate whether operational items contribute positively to their intended measures, whether item responses conform adequately to Rasch-model expectations, whether vertically scaled forms demonstrate the expected ordering and measurement precision, and whether observed relationships among scores are consistent with the intended assessment structure.

The internal-structure evidence is interpreted together with the reliability, classification, fairness, and item-person targeting results presented in Chapter 4. Annual field testing and psychometric review use these findings to identify items or portions of the reporting scale that require additional review, replacement, or future item development.

3.3A Expected Score Functioning and Measurement Precision

Expected score functioning and conditional measurement precision were evaluated using TCCs and TIFs. TCCs summarize the expected total score at each location on the OExt reporting scale and provide evidence regarding the ordering and spacing of the vertically scaled Grades 3–8 ELA and Mathematics forms. TIFs describe how measurement precision varies across the reporting scale and in relation to the approved performance-level cut scores.

The detailed 2025–2026 TCC and TIF figures, calculation procedures, and content-specific interpretations are presented in Sections 4.1B through 4.1F. The TCCs provide evidence that adjacent-grade forms function in the intended order without inappropriate crossover. The TIFs show where each form provides stronger or weaker measurement precision and whether the approved classification points fall within adequately measured portions of the scale.

Together, these analyses support interpretation of the vertically scaled Grades 3–8 scores while identifying portions of the ability continuum where precision is comparatively lower. Findings from the annual review inform field testing, item development, and future form assembly. The next section examines whether individual operational items discriminate in the expected direction and contribute appropriately to the measured construct.

3.3B Item Discrimination and Alignment with the Measured Construct

Point-measure correlations were examined to determine whether individual operational items discriminated in the expected direction within the Rasch measurement framework. A point-measure correlation describes the relationship between a student's dichotomously scored response to an item and the student's estimated ability, or theta, on the measure containing that item.

Unlike a traditional point-biserial correlation based on total raw score, the point-measure correlation uses the Rasch ability estimate employed in calibration and score reporting. Positive values indicate that students with higher estimated ability were more likely to answer an item

correctly. Larger positive values provide stronger evidence that the item contributes as intended to the measured construct.

Point-measure correlations are interpreted together with outfit mean-square statistics, content coverage, accessibility evidence, and other item-review information. No single statistic determines whether an item is retained or replaced. Instead, the combined evidence informs annual decisions regarding continued operational use, revision, or replacement through the field-testing process.

The summaries below include only the 30 operational items used for scoring on each **2025–2026** assessment. Embedded field-test items are excluded. Tables report the mean, standard deviation (SD), median, minimum, and maximum. Point-measure correlations ranged from 0.02 to 0.73 in ELA, from -0.21 to 0.76 in Mathematics, and from -0.02 to 0.74 in Science. These correlations were calculated using students with nonmissing item responses and proficiency estimates. Positive values indicate that students with higher estimated proficiency were generally more likely to answer an item correctly. Values near zero or below zero identify items requiring targeted review rather than automatically demonstrating acceptable item functioning. The results are interpreted together with outfit statistics, blueprint representation, and other item-level evidence presented in this chapter.

Table 3.4: Point-Measure Correlation by Content and Grade

Grade	Mean	SD	Median	Min	Max
ELA					
Grade 3	0.50	0.13	0.52	0.23	0.72
Grade 4	0.53	0.10	0.54	0.27	0.69
Grade 5	0.48	0.10	0.49	0.20	0.65
Grade 6	0.52	0.11	0.53	0.24	0.67
Grade 7	0.46	0.12	0.47	0.13	0.69
Grade 8	0.44	0.15	0.45	0.02	0.63
High School	0.50	0.17	0.55	0.04	0.73
Math					
Grade 3	0.46	0.14	0.48	0.07	0.73
Grade 4	0.46	0.14	0.43	0.22	0.71
Grade 5	0.42	0.18	0.41	-0.08	0.76
Grade 6	0.31	0.21	0.37	-0.21	0.57
Grade 7	0.37	0.18	0.43	-0.15	0.60
Grade 8	0.34	0.26	0.38	-0.20	0.76
High School	0.41	0.15	0.43	0.09	0.64
Science					
Grade 5	0.47	0.16	0.47	-0.02	0.72
Grade 8	0.48	0.14	0.51	0.13	0.72
High School	0.49	0.14	0.52	0.21	0.74

The overall content-area results are supplemented by separate summaries for the Reading and Writing components of ELA. Because these components are nested within the broader ELA assessment, the results describe item functioning within each subscore model and should not be interpreted as evidence that Reading and Writing are independent operational assessments.

Table 3.5: Point-Measure Correlation of ELA Subscores by Grade

Grade	Mean	SD	Median	Min	Max
Reading					
Grade 3	0.50	0.11	0.52	0.34	0.70
Grade 4	0.54	0.08	0.56	0.31	0.63
Grade 5	0.50	0.10	0.53	0.27	0.63
Grade 6	0.51	0.12	0.54	0.24	0.67
Grade 7	0.46	0.10	0.49	0.29	0.66
Grade 8	0.44	0.10	0.47	0.22	0.59
High School	0.53	0.16	0.59	0.09	0.71
Writing					
Grade 3	0.67	0.14	0.69	0.42	0.83
Grade 4	0.60	0.11	0.61	0.40	0.72
Grade 5	0.62	0.07	0.58	0.55	0.71
Grade 6	0.63	0.07	0.60	0.55	0.74
Grade 7	0.68	0.04	0.68	0.64	0.74
Grade 8	0.63	0.09	0.62	0.50	0.79
High School	0.56	0.12	0.61	0.38	0.70

3.3C Item Fit to the Rasch Measurement Model

Outfit mean-square (OMS) statistics were examined to evaluate the extent to which observed item responses conformed to expectations under the Rasch measurement model. Values near 1.0 indicate that an item is functioning in a manner consistent with model expectations. Values substantially below 1.0 indicate overfit, meaning that responses are more predictable than expected. Values above 1.0 indicate underfit, meaning that responses are less predictable than expected.

For the annual operational review, item-model results with OMS values greater than 2.0 are flagged for closer examination. Exceeding this review criterion does not, by itself, indicate that an item should be removed. Item-retention decisions also consider content representation, response patterns, point-measure correlations, calibration evidence, accessibility, and the item's role within the operational form and item bank.

The analyses below first summarize OMS statistics for the operational content-area models and the Reading and Writing subscore models. A subsequent review identifies item-model results that exceeded the OMS criterion of 2.0. Together with the point-measure correlations presented in Section 3.3B, these results provide evidence regarding the extent to which operational items contributed to coherent Rasch measures of the intended academic constructs.

The content-area summaries describe the distribution of OMS values across the operational items contributing to each ELA, Mathematics, and Science score. Mean values near 1.0 indicate that item responses generally followed Rasch-model expectations, while the minimum and maximum values identify the range of item fit observed within each grade and content area.

Table 3.6: Mean-Square Outfit by Content and Grade

Grade	Mean	SD	Min	Max
ELA				
Grade 3	0.81	0.43	0.37	2.36
Grade 4	0.92	0.60	0.37	2.53
Grade 5	0.88	0.50	0.40	2.57
Grade 6	0.97	0.63	0.41	2.86
Grade 7	0.95	0.40	0.55	2.62
Grade 8	0.89	0.59	0.43	2.99
High School	0.89	0.34	0.37	1.93
Math				
Grade 3	0.79	0.31	0.42	1.56
Grade 4	0.81	0.38	0.40	2.04
Grade 5	0.79	0.33	0.46	1.98
Grade 6	0.84	0.38	0.52	2.44
Grade 7	0.85	0.41	0.33	2.09
Grade 8	0.83	0.66	0.31	2.77
High School	0.88	0.30	0.52	2.03
Science				
Grade 5	0.83	0.29	0.47	1.83
Grade 8	0.86	0.28	0.44	1.67
High School	0.86	0.46	0.20	1.76

The Reading and Writing summaries provide parallel evidence for the ELA subscore models. Because these subscores are derived from items contained within the overall ELA assessment, their results are interpreted as supporting information about item functioning within each subscore model rather than as evidence from separate operational assessments.

Table 3.7: Mean-Square Outfit for ELA Subscores by Grade

Grade	Mean	SD	Min	Max
Reading				
Grade 3	0.79	0.32	0.44	1.64
Grade 4	0.88	0.59	0.42	2.33
Grade 5	0.86	0.54	0.40	2.38
Grade 6	0.94	0.50	0.49	2.13
Grade 7	0.95	0.44	0.58	2.43
Grade 8	0.89	0.40	0.53	1.85
High School	0.88	0.35	0.50	1.81
Writing				
Grade 3	0.83	0.21	0.55	1.19
Grade 4	0.96	0.29	0.57	1.54
Grade 5	0.88	0.27	0.61	1.39
Grade 6	0.94	0.39	0.52	1.72

Table 3.7: Mean-Square Outfit for ELA Subscores by Grade *(continued)*

Grade	Mean	SD	Min	Max
Grade 7	1.01	0.15	0.81	1.18
Grade 8	1.04	0.35	0.55	1.47
High School	0.94	0.30	0.39	1.45

The following analysis identifies item-model results with OMS values greater than 2.0. Because the same ELA item may appear in both the overall ELA model and a Reading or Writing subscore model, the number of flagged item-model results is not necessarily the same as the number of distinct operational items requiring review.

The 2025–2026 review identified 26 item-model results with OMS values greater than 2.0. These results represented 22 unique item identifiers. Of the flagged results, 12 occurred in overall ELA scoring models, 8 occurred in Mathematics scoring models, and 6 occurred in Reading subscore models.

Because Reading items are also included in the overall ELA model, the same operational item may contribute to both the ELA and Reading counts. The 26 flagged item-model results therefore should not be interpreted as 26 distinct operational items. Flagged OMS values ranged from 2.01 to 2.99. No Science or Writing item-model results exceeded the review criterion.

One flagged identifier included a numeric suffix variant and was retained as a separate identifier in the summary. Accordingly, the reported number of unique identifiers should not be interpreted as the number of distinct item designs without additional review of the item-bank records.

Table 3.8: Summary of Operational Item-Model Results with Mean-Square Outfit Greater Than 2.0

Metric	Value
Flagged item-model results	26
Unique operational item identifiers flagged	22
Minimum flagged Outfit	2.01
Maximum flagged Outfit	2.99

Table 3.9: Flagged Item-Model Results by Scoring Model

Model	Flagged item-model results
ELA	12
Math	8
Reading	6

These flagged results identify item-model combinations requiring additional review rather than automatic removal. Final decisions consider the complete body of evidence for each item, including content coverage, point-measure correlation, response patterns, accessibility, calibration history, and whether the item was also flagged in another scoring model. The next section places these item-level findings within the broader distribution of student performance across achievement levels.

3.3D Performance-Level Distributions as Supporting Evidence

Performance-level distributions provide descriptive context regarding how students were classified across grades and content areas during the 2025–2026 administration. These distributions are not interpreted as primary evidence of internal structure because they are influenced by both the distribution of student performance and the locations of the approved achievement-level cut scores. Instead, they supplement the TCCs, TIFs, point-measure correlations, and Rasch item-fit results presented above.

Annual Measurable Objective (AMO) classifications were assigned by comparing student OREXt scores with Oregon's approved achievement-level cut scores. Students classified as **Meets** or **Exceeds** performed at or above the State's proficient standard for the AAAS. Across the reported grades and content areas, an average of 25% of students were classified as Meets or Exceeds, with a standard deviation of 10% across assessments. Detailed classification results by grade and content area are presented in the following table.

OREXt achievement results are not directly comparable with results from Oregon's general statewide assessments because the two systems assess different student populations against different achievement standards. A direct numerical comparison would therefore be misleading. Within the OREXt system, an average of 53% of students were classified in AMO 1 across the grade-by-content distributions, and approximately three-quarters were below the OREXt proficient standard.

These results describe the distribution of achievement during a single administration and should not be interpreted as evidence that one grade or content area is inherently stronger than another. Differences in classification rates may reflect variation in student performance, test difficulty, score distributions, and the locations of the approved cut scores. Longitudinal interpretations are addressed separately using vertically scaled scores and conditional measurement error. Together, achievement-level classifications and measurement-aware growth results provide complementary information about current performance and change over time.

Table 3.10: Performance-Level Distribution by Grade

Content and Grade	AMO 1 (Does Not Yet Meet)	AMO 2 (Nearly Meets)	AMO 3 (Meets)	AMO 4 (Exceeds)
ELA				
Grade 3	50 (51%)	26 (27%)	16 (16%)	6 (6%)
Grade 4	47 (54%)	15 (17%)	15 (17%)	10 (11%)
Grade 5	53 (50%)	34 (32%)	13 (12%)	7 (7%)
Grade 6	48 (50%)	23 (24%)	14 (15%)	11 (11%)
Grade 7	33 (43%)	25 (32%)	14 (18%)	5 (6%)
Grade 8	50 (66%)	16 (21%)	9 (12%)	1 (1%)
High School	20 (20%)	40 (40%)	24 (24%)	16 (16%)
Math				
Grade 3	56 (67%)	12 (14%)	13 (16%)	2 (2%)
Grade 4	44 (53%)	16 (19%)	20 (24%)	3 (4%)
Grade 5	46 (47%)	32 (33%)	16 (16%)	3 (3%)
Grade 6	47 (62%)	11 (14%)	18 (24%)	0 (0%)
Grade 7	50 (68%)	6 (8%)	14 (19%)	4 (5%)
Grade 8	51 (71%)	10 (14%)	11 (15%)	0 (0%)
High School	58 (62%)	18 (19%)	14 (15%)	3 (3%)
Science				
Grade 5	52 (54%)	22 (23%)	12 (12%)	10 (10%)
Grade 8	43 (51%)	19 (22%)	13 (15%)	10 (12%)
High School	23 (25%)	18 (20%)	35 (38%)	16 (17%)

The modal achievement level was AMO 1 (6 of 7 grades) and AMO 2 (1 of 7 grades) for ELA, AMO 1 (7 of 7 grades) for Mathematics, and AMO 1 (2 of 3 grades) and AMO 3 (1 of 3 grades) for Science. These summaries identify the most frequently observed classification within each content area across the reported grades. They do not indicate that the same students or the same proportion of students occupied the modal category at every grade.

Relatively few students were classified in AMO 4, **Exceeds**, particularly in several Mathematics grades. The classification tables describe this pattern but, by themselves, do not identify its cause. Annual review therefore considers the full score distributions, approved cut scores, TCCs, TIFs, classification statistics, and item-level psychometric results before determining whether a pattern reflects student performance or indicates a need for additional form, scale, or item-bank evaluation.

The Reading and Writing results presented next provide descriptive information about performance within the two reported ELA subscores. Because both subscores are derived from items included in the overall ELA assessment, they should be interpreted as components of the broader ELA construct rather than as separate operational assessments.

Table 3.11: Performance-Level Distribution by Grade: 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	49 (50%)	32 (33%)	13 (13%)	4 (4%)
Grade 4	45 (52%)	19 (22%)	9 (10%)	14 (16%)
Grade 5	49 (46%)	39 (36%)	8 (7%)	11 (10%)
Grade 6	43 (45%)	30 (31%)	13 (14%)	10 (10%)
Grade 7	33 (43%)	24 (31%)	15 (19%)	5 (6%)
Grade 8	49 (64%)	16 (21%)	6 (8%)	5 (7%)
High School	26 (26%)	44 (44%)	19 (19%)	11 (11%)
Writing				
Grade 3	51 (53%)	25 (26%)	20 (21%)	0 (0%)
Grade 4	46 (53%)	23 (26%)	12 (14%)	6 (7%)
Grade 5	57 (53%)	21 (20%)	16 (15%)	13 (12%)
Grade 6	48 (51%)	29 (31%)	5 (5%)	13 (14%)
Grade 7	40 (52%)	13 (17%)	14 (18%)	10 (13%)
Grade 8	26 (57%)	13 (28%)	3 (7%)	4 (9%)
High School	28 (28%)	31 (31%)	9 (9%)	32 (32%)

The Reading and Writing distributions provide descriptive context for ELA subscore classifications across grades. Because both subscores are derived from items included in the overall ELA assessment, they are not interpreted independently as evidence of dimensionality, measurement precision, or separate academic constructs.

Compared with Mathematics, Science, and overall ELA, the Reading and Writing subscores were distributed more broadly across the four achievement levels. Variation among grades is reviewed annually alongside item-level statistics, score distributions, and other psychometric evidence. These results inform ongoing evaluation of the ELA reporting structure and should be interpreted as supporting evidence rather than as a stand-alone basis for conclusions about internal structure.

3.3E Synthesis of Internal-Structure Evidence

Section 3.3A demonstrated the expected ordering of the vertically scaled forms and identified where measurement precision was stronger or weaker. Section 3.3B showed that most operational items contributed positively to their intended Rasch measures, while identifying items requiring targeted review. Section 3.3C indicated that most item responses were broadly consistent with Rasch-model expectations, although several item-model results exceeded the annual outfit review criterion. Section 3.3D showed that approximately half of the assessed population remained in AMO 1 across the grade-by-content distributions.

Taken together, the evidence is generally consistent with use of a dominant content-area score for each OREXt assessment. It does not constitute a direct dimensionality study using factor-analytic or comparable methods. Conclusions regarding internal structure are therefore limited to the observed coherence of item functioning, score relationships, measurement precision, and blueprint

representation within the current Rasch scoring framework. The findings also identify specific item-model results and portions of the ability continuum that require continued monitoring through annual psychometric review, field testing, and item development.

3.4 Validity Evidence Based on Relations to Other Variables

Validity evidence based on relations to other variables examines whether OREXt scores demonstrate patterns of association that are consistent with the assessment's intended interpretation and use. A foundational framework for this evaluation is the multitrait-multimethod approach described by Campbell and Fiske (1959). Within this framework, the term *trait* may be interpreted as the academic knowledge and skills represented by a content-area score.

Measures intended to represent related academic skills should generally demonstrate stronger positive relationships than measures intended to represent distinct content areas. At the same time, correlations should remain below unity, indicating that the assessments are not measuring a single undifferentiated ability. Evidence is therefore evaluated by considering the direction, magnitude, and pattern of relationships rather than by applying a single universal correlation threshold.

The current-year analyses examined relationships among ELA, Mathematics, Science, Reading, and Writing scores. The expected pattern is a stronger association between overall ELA and its Reading and Writing components than among distinct academic content areas. Because Reading and Writing items are included within the overall ELA measure, these relationships are interpreted as evidence regarding the coherence of the ELA reporting structure rather than as correlations among fully independent assessments. The analyses are also interpreted together with prior OREXt validation research reported by Tindal et al. (2003).

These within-OREXt relationships provide useful evidence about convergence among related scores and distinction among content areas. They do not, however, provide direct criterion-related evidence from an independent external academic measure administered to the same students. The findings are therefore interpreted as one component of the broader validity argument and are considered together with standards-linkage evidence, internal-structure analyses, administration-fidelity evidence, and longitudinal score patterns.

Descriptive performance comparisons across disability groups are presented in Chapter 4 as part of the assessment's fairness and comparability review. Group differences are not, by themselves, interpreted as validity evidence or evidence of bias. Instead, they provide context for determining whether score patterns require additional investigation alongside Differential Item Functioning (DIF), participation, accessibility, and other technical evidence.

Longitudinal analyses provide an additional source of evidence regarding adjacent-year score differences. Because the Grades 3–8 ELA and Mathematics assessments are vertically scaled, changes across school years should be interpreted in relation to the precision of both scores. Chapter 4 classifies adjacent-year differences as reliable increases, reliable decreases, or changes not distinguishable from expected measurement error using each student's CSEM. The cross-sectional relationships examined in this section and the measurement-aware longitudinal results in Chapter 4 address different components of the broader validity argument.

3.4A Convergent and Divergent Validity Documentation

The Oregon Observational Rating Assessment (ORora) remained available during 2025–2026 as an optional observational activity for students who could not meaningfully engage with ORExt academic content. ORora was not administered under the same standardized conditions as the ORExt and was not designed to measure the same academic constructs. ORora score distributions, density comparisons, and correlations with ORExt scores are therefore not interpreted as convergent or divergent validity evidence.

ORora reporting in this technical report is limited to participation context and its relationship to the ORExt minimum-participation procedures described in Section 3.2. Oregon is transitioning away from ORora as it develops the Pathways Communication Scale (PCS), which is intended to provide communication-focused information for students who cannot meaningfully access ORExt academic content.

Current evidence based on relations to other variables is drawn primarily from patterns of association among ORExt content-area scores, relationships between overall ELA and its Reading and Writing components, and measurement-aware classification of adjacent-year score differences on the vertically scaled ELA and Mathematics assessments. These analyses evaluate whether related scores converge as expected, whether content-area score patterns remain nonidentical, and whether observed longitudinal differences can be distinguished from expected measurement error.

Descriptive comparisons across student groups are not treated as convergent or divergent validity evidence. They are presented in Chapter 4 as part of the fairness and comparability review and are interpreted together with DIF, participation, accessibility, and other technical evidence. The absence of a current independent external academic criterion administered to the same students limits the strength of criterion-related conclusions and identifies an area for future validity research.

3.4B Analyses Within and Across Subject Areas

Correlational analyses were conducted to examine relationships among student performance in ELA, Mathematics, and Science and among overall ELA, Reading, and Writing scores. These analyses evaluated whether observed score relationships were generally consistent with the intended organization of the ORExt reporting system.

Positive correlations were expected among content areas because many of the same students participated in multiple assessments and academic performance was related across subjects. The magnitude of these relationships was interpreted descriptively. Correlations below unity are consistent with nonidentical score patterns across content areas, but they do not by themselves establish that the assessments measure fully distinct constructs.

Stronger associations were expected between overall ELA and its Reading and Writing components because the subscore items are included within the overall ELA measure. These relationships therefore provided evidence regarding the coherence of the ELA reporting structure. Less-than-perfect relationships between Reading and Writing may indicate that the components provide somewhat distinguishable descriptive information, although they are not interpreted as independent assessments or direct evidence of separate dimensions.

The cross-sectional analyses were complemented by measurement-aware analyses of the OREXt vertical scales. Because Grades 3–8 ELA and Mathematics scores are intended to support interpretation across adjacent grades, observed year-to-year differences were evaluated relative to the conditional measurement error associated with both scores. Together, the within-ELA, across-content, and adjacent-year analyses provide supporting evidence regarding whether OREXt score patterns are consistent with the intended assessment and reporting design.

3.4C Correlational Analysis Results

Pearson product-moment correlations were calculated by grade among available OREXt content-area and ELA subscore measures. The results are presented separately for grades with and without a Science assessment because Science is administered only in Grades 5, 8, and 11.

The interpretation focuses on the magnitude and pattern of the correlations rather than statistical significance alone. With statewide assessment samples, even modest correlations may be statistically significant. More relevant to the validity argument is whether relationships are positive and whether overall ELA is more strongly related to Reading and Writing than to different content-area scores. Correlations below unity are consistent with nonidentical score patterns, but they do not establish separate constructs.

The analyses used listwise complete-case deletion within each grade-specific correlation matrix. The resulting coefficients therefore describe students with nonmissing scores on every measure shown for that grade and should not be generalized automatically to every tested student. The tables and summaries that follow are interpreted together with content-alignment, internal-structure, and measurement-aware longitudinal evidence.

Table 3.12: Content Area Correlations (Grades Without a Science Assessment)

	ELA	Math	Reading	Writing
Grade 3				
ELA	1.00			
Math	0.82	1.00		
Reading	0.98	0.81	1.00	
Writing	0.95	0.80	0.89	1.00
Grade 4				
ELA	1.00			
Math	0.80	1.00		
Reading	0.97	0.76	1.00	
Writing	0.94	0.75	0.89	1.00
Grade 6				
ELA	1.00			
Math	0.64	1.00		
Reading	0.98	0.66	1.00	
Writing	0.93	0.55	0.89	1.00
Grade 7				
ELA	1.00			
Math	0.66	1.00		

Table 3.12: Content Area Correlations (Grades Without a Science Assessment) *(continued)*

	ELA	Math	Reading	Writing
Reading	0.99	0.67	1.00	
Writing	0.83	0.61	0.80	1.00

Table 3.13: Content Area Correlations (Grades With a Science Assessment)

	ELA	Math	Reading	Science
Grade 5				
ELA	1.00			
Math	0.75	1.00		
Reading	0.96	0.68	1.00	
Science	0.81	0.69	0.76	1.00
Writing	0.92	0.78	0.84	0.77
Grade 8				
ELA	1.00			
Math	0.19	1.00		
Reading	0.87	0.26	1.00	
Science	0.65	0.41	0.68	1.00
Writing	0.82	0.10	0.52	0.37
High School				
ELA	1.00			
Math	0.76	1.00		
Reading	0.95	0.76	1.00	
Science	0.90	0.73	0.86	1.00
Writing	0.92	0.67	0.79	0.82

Results of the Pearson product-moment correlation analyses across content areas ranged from:

- ELA and Mathematics: 0.19 to 0.82
- ELA and Science: 0.65 to 0.90
- Mathematics and Science: 0.41 to 0.73

The correlations were positive across content-area pairs, indicating that student performance was related across ELA, Mathematics, and Science, as expected. Because these correlations varied substantially across grades and remained below unity, the results are consistent with related but nonidentical score patterns across content areas. Correlations alone do not establish that the assessments measure fully distinct constructs, so these findings are interpreted together with content-alignment, item-functioning, and internal-structure evidence.

For ELA and its Reading and Writing components, correlations ranged from:

- ELA and Reading: 0.87 to 0.99
- ELA and Writing: 0.82 to 0.95
- Reading and Writing: 0.52 to 0.89

The stronger relationships between overall ELA performance and the Reading and Writing components are expected because both subscores are derived from items included in the overall ELA measure. The less-than-perfect relationship between Reading and Writing suggests that the two components provide somewhat distinguishable descriptive information within the broader ELA score, but it does not by itself establish separate dimensions or independent constructs.

Taken together, the pattern of stronger relationships within ELA and generally more moderate relationships across content areas is consistent with the assessment's intended reporting structure. These cross-sectional results are interpreted as supporting evidence rather than as a stand-alone demonstration of construct distinction. They are complemented by the longitudinal analyses presented in Chapter 4, which evaluate whether vertically scaled ELA and Mathematics scores demonstrate interpretable patterns across adjacent grades and school years.

4 Technical Quality

This chapter evaluates the technical quality of the 2025–2026 OExt through evidence concerning score reliability, conditional measurement precision, classification accuracy and consistency, fairness, accessibility, coverage of the performance continuum, scoring, and ongoing technical maintenance. The chapter emphasizes the operational meaning of the results: whether scores are sufficiently precise for their intended uses, whether administration and scoring procedures are implemented consistently, and whether the assessment functions comparably across relevant student groups.

The analyses are presented as a connected technical-quality argument rather than as isolated statistical exercises. Reliability and measurement-precision evidence is considered first, followed by fairness and accessibility evidence, coverage of the full performance continuum, and the procedures used to sustain technical quality over time.

4.1 Reliability, Precision, and Classification Quality

Reliability refers to the consistency and precision with which an assessment measures student performance across administrations, contexts, and scoring conditions. For the 2025–2026 OExt, the primary current-year reliability evidence is model-based and includes marginal reliability, TIFs, overall standard errors and CSEM values, classification accuracy and consistency, and agreement evidence from administration and scoring observations.

Earlier OExt technical reports documented classical internal-consistency indices, including Cronbach's alpha and split-half estimates. Those historical statistics provide continuity with prior reporting, but they are not presented here as current-year results. Because the OExt is calibrated and scored using Rasch measurement models, the information-based and classification-focused indices reported in this chapter provide the most direct evidence regarding conditional score precision and the dependability of operational performance-level classifications.

The longitudinal analyses extended the practical application of score-precision evidence. Rather than interpreting observed year-to-year score differences as error-free, the measurement-aware analyses incorporated each student's CSEM. This approach reduced the risk of treating routine score variation as meaningful change. The resulting classifications were interpreted as evidence about the dependability of observed score differences, not as direct evidence about the causes of student growth or decline.

4.1.1 Consistency of Administration and Scoring

4.1.1.1 Background

Inter-rater reliability (IRR) studies provide evidence regarding the consistency of judgments made by different observers or scorers when applying standardized administration and scoring procedures (Hallgren 2012). For the OExt, IRR evidence evaluates whether QAs administer and score the assessment consistently across students, schools, and districts. This evidence supports the Federal Assessment Peer Review requirements related to reliability (Critical Element 4.1).

The 2025–2026 IRR study included a sample of Oregon QAs administering the paper/pencil version of the OExt. Observations focused on adherence to standardized administration procedures, the accuracy of item scoring and response recording, and the appropriate implementation of allowable accessibility supports.

4.1.1.2 Methods

IRR observations are conducted as part of Oregon's ongoing quality-assurance process. During 2025–2026, QTCs observed QAs administering the OExt using a standardized observation protocol. The protocol evaluates adherence to administration procedures, scoring accuracy, appropriate implementation of accessibility supports, and accurate recording of student responses.

The observation protocol is based on the statewide administration procedures described in the OExt Assessment Technical Report on Standard Setting and supports ongoing monitoring of administration quality throughout the state. Observations were conducted across participating Oregon school districts representing a range of geographic regions, district sizes, and instructional settings. The study was based on observed paper/pencil administrations in participating districts. It did not constitute a probability sample of all statewide administrations and did not directly evaluate comparability with electronic administration.

4.1.1.3 Observation Domains

The standardized observation protocol evaluates the following aspects of assessment administration:

1. Test security
2. Assessment materials and preparation
3. Administration environment
4. Accessibility supports
5. Appropriate level of support
6. Appropriate use of praise and motivation
7. Score interpretation
8. Minimum participation procedures
9. Administration accuracy
10. Scoring and data entry accuracy

4.1.1.4 Agreement Results

QTCs (25) from school districts across Oregon participated in the IRR study by observing paper/pencil administrations of the OrExt. Observations represented ELA, Mathematics, and Science (ELA (40%), Math (40%), Science (20%)) and were conducted under typical operational testing conditions within participating schools.

QTC ratings indicated that administration procedures were implemented consistently across the 10 domains included in the observation protocol. For nearly every domain, at least 96% of observations were rated either *Exemplary* or *Appropriate*. The strongest ratings were observed for level of support, motivation, praise, printed materials, and distraction-free testing conditions. Score interpretation showed comparatively greater variability, with 76% of observations rated *Exemplary* or *Appropriate*, 16% rated *Somewhat Appropriate*, and 8% rated *Inappropriate*. These results identify score interpretation as the primary area for continued training and technical assistance.

Table 4.1: Inter-rater Observation Ratings by Administration Domain

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%

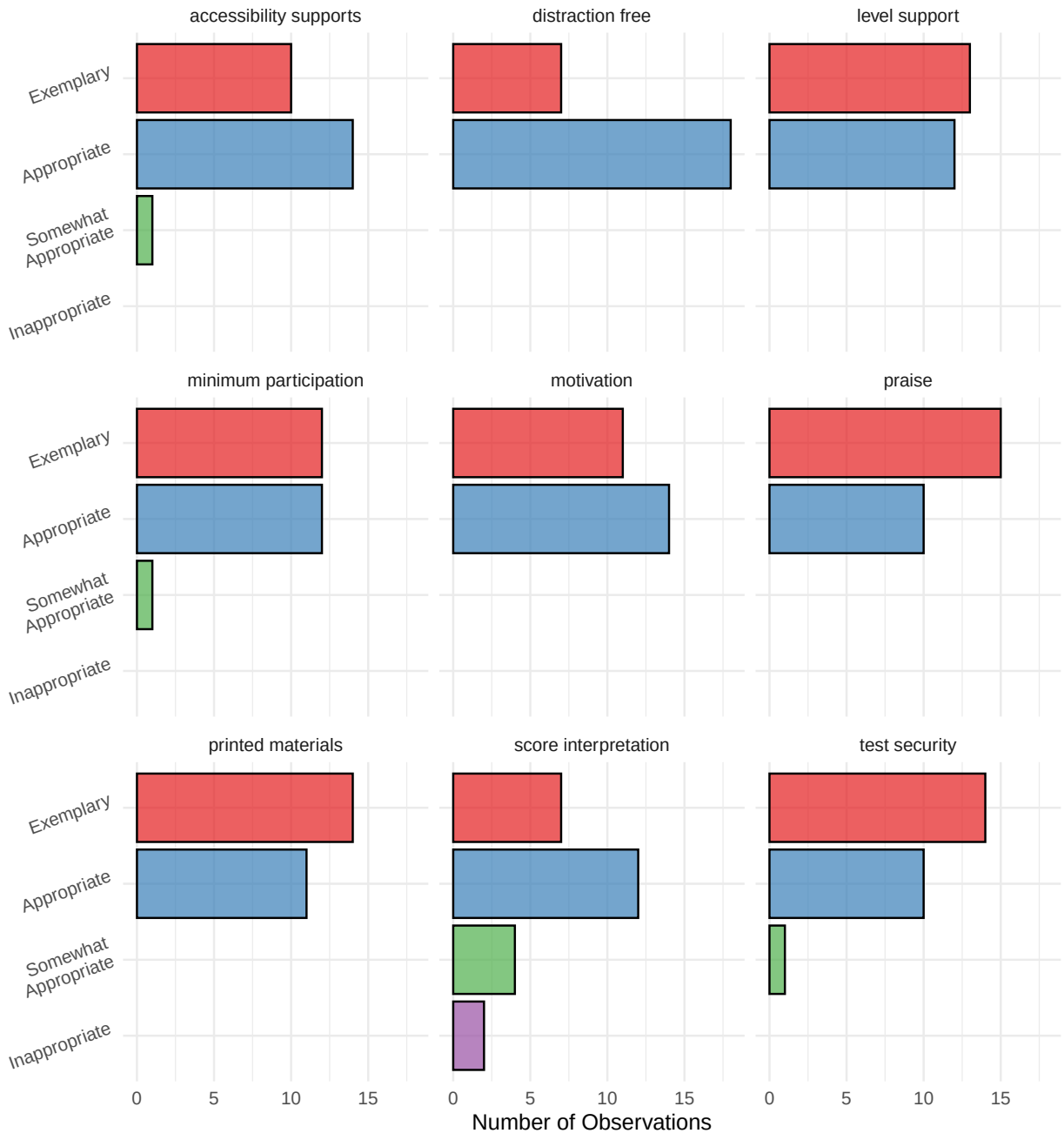
Agreement between QAs and QTCs was also examined at the individual item level. Among administered items, the two raters agreed on 99.9% of scoring decisions. Only 1 scoring disagreement was observed. Items that were not administered were reported separately and were not treated as scoring disagreements. The result provides strong evidence that the scoring rules were applied consistently within the observed paper/pencil administrations. The finding should not be generalized automatically to unobserved administrations, other delivery formats, or the statewide population of assessors.

Table 4.2: Agreement Between QAs and QTCs

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

The distribution of ratings across the administration domains is displayed in the following figure. The figure reinforces the tabled results: ratings were concentrated in the *Exemplary* and *Appropriate* categories, with comparatively few observations in the lower two categories. The

greater spread of ratings for score interpretation is also visible and supports continued emphasis on this topic in annual assessor training.



4.1.1.5 Results

The observed administrations showed a high level of agreement in the application of ORExt scoring rules and generally favorable ratings across the administration domains. This evidence supports the consistency of administration and scoring within the participating paper/pencil observations.

Score interpretation showed greater variability than the other observed domains. This finding identifies interpretation of achievement levels and cut scores as a priority for subsequent training, technical assistance, and monitoring. Because the study included 25 participating paper/pencil observations rather than a statewide probability sample, the results are interpreted as direct but limited evidence of implementation fidelity.

4.1A Internal Consistency and Score Reliability

Marginal reliability estimates the proportion of observed score variance attributable to differences in students' underlying achievement rather than measurement error. Estimates for each operational ORExt administered during the **2025–2026** school year are presented below.

Across the reported content-area scores and ELA subscores, marginal reliability coefficients ranged from 0.57 to 0.95. Most total content-area estimates were comparatively strong. Several Reading and Writing subscore estimates were lower, indicating that these subscores should be used as supporting descriptive information rather than as independent bases for high-stakes decisions. Estimates were based on the operational responses used to calculate reported scores; embedded field-test items were excluded from reported-score interpretations.

Marginal reliability summarizes score consistency at the total-test or subscore level. Interpretation of an individual score or year-to-year difference also requires information about conditional measurement precision at the student's score location. The TIFs and CSEM values presented in the following sections therefore provide essential complementary evidence.

Table 4.3: Marginal Reliability by Content and Grade

	ELA	Math	Science	Reading	Writing
High School	0.95	0.92	0.93	0.90	0.89
Grade 3	0.89	0.89	–	0.85	0.73
Grade 4	0.89	0.88	–	0.82	0.73
Grade 5	0.89	0.87	0.88	0.84	0.72
Grade 6	0.89	0.84	–	0.82	0.70
Grade 7	0.88	0.83	–	0.86	0.57
Grade 8	0.84	0.72	0.91	0.75	0.64

4.1B Conditional Measurement Precision

Prospective form-quality review complements the post-administration reliability and item-functioning analyses presented throughout this chapter. Before operational use, candidate assessment forms are evaluated using calibrated Rasch item parameters, expected score functions, TIFs, and conditional measurement precision.

The TIFs presented below show how information varies across the ORExt reporting scale. Each figure is calculated from the 30 operational items contributing to reported scores. The six embedded field-test items are excluded. Vertical reference lines identify the approved achievement-level cut scores. Shaded regions provide approximate marginal-reliability reference bands and should not be treated as universal pass-fail thresholds.

Across forms, precision is generally greatest in the central portion of the reporting scale and declines where operational item coverage becomes less dense. Precision at the achievement-level boundaries varies by grade, content area, and cut score. The figures are therefore interpreted together with the cut-score CSEM values in Section 4.1G, the marginal-reliability estimates, and the item-person maps in Section 4.3.

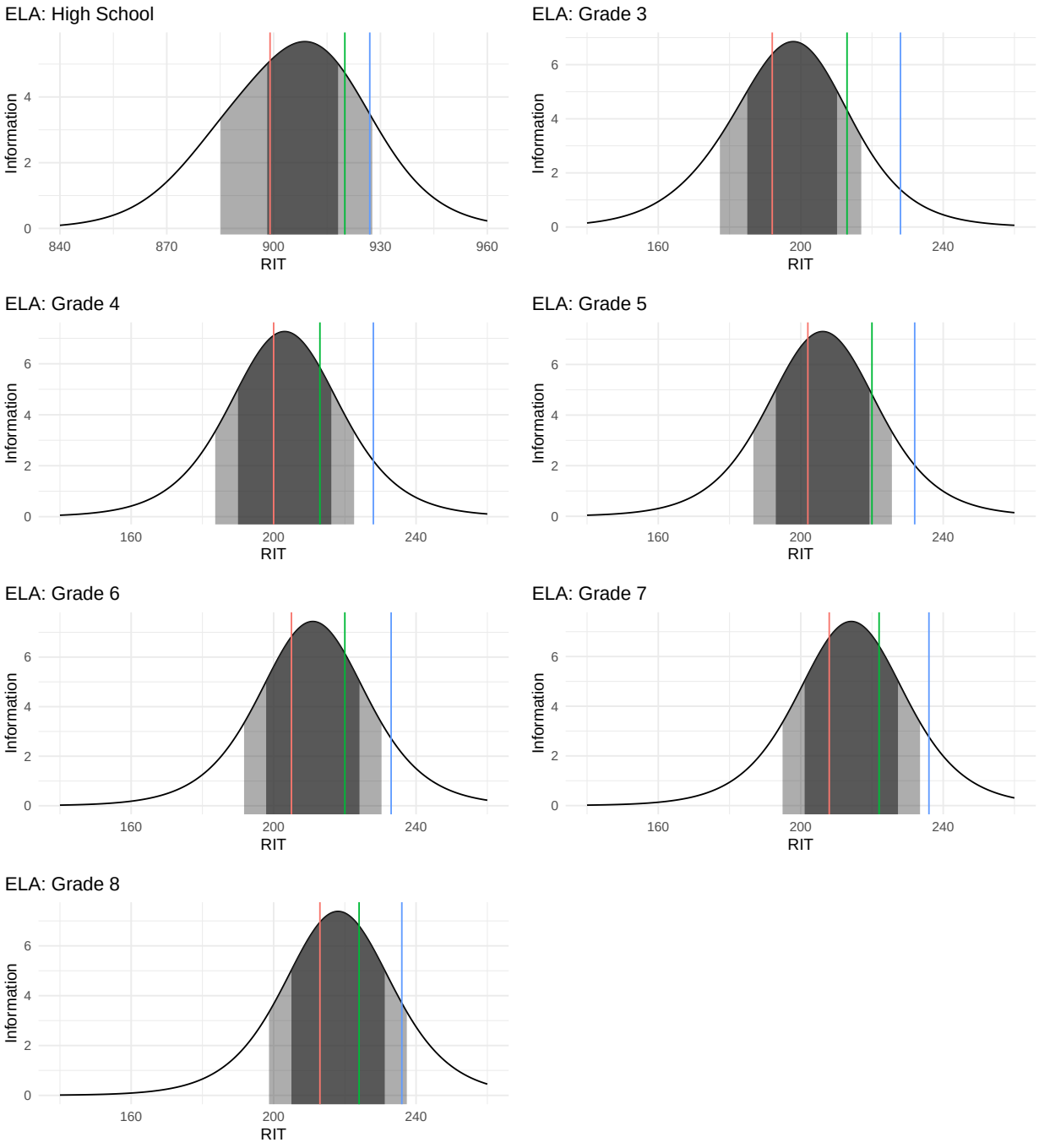
Because precision varies across the scale, longitudinal interpretations use the conditional standard error associated with each student's score rather than a single common error estimate.

4.1C ELA Measurement Precision

The ELA TIFs show the locations on each reporting scale where the operational forms provide more or less information. Precision is strongest where the operational item difficulties are most concentrated and decreases toward portions of the score continuum with fewer targeted items. The amount of information near individual achievement-level boundaries varies across grades.

These results support interpretation of ELA scores within the portions of the scale that are measured most precisely while identifying locations requiring continued monitoring and targeted item development. Exact CSEM values at the achievement-level cut scores are presented in Section 4.1G.

Colored vertical lines indicate performance-level cut scores. Gray shading identifies regions where marginal reliability is approximately .70 and .80 or greater.

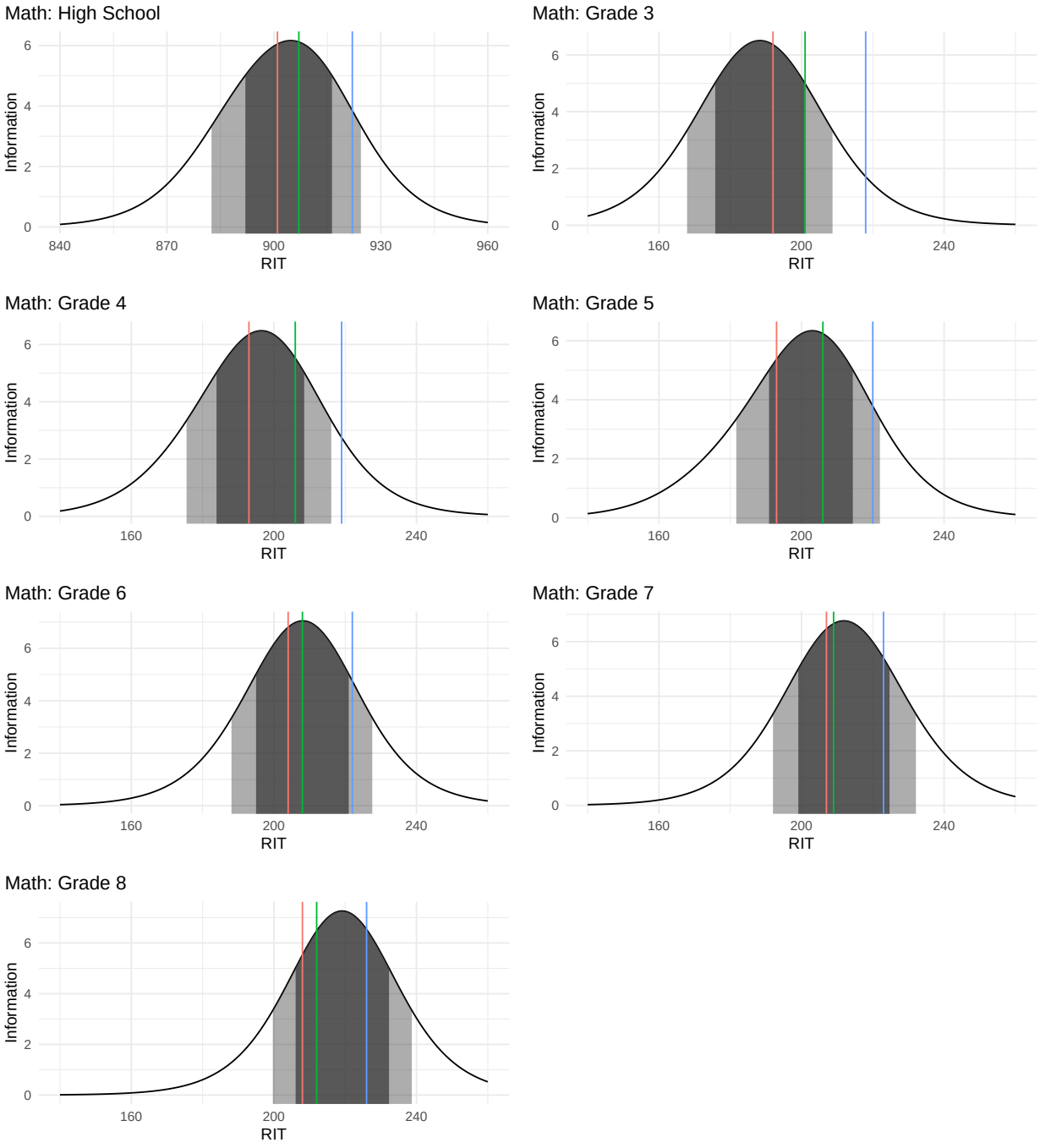


4.1D Mathematics Measurement Precision

The Mathematics TIFs show that measurement information is concentrated across the central portion of each operational reporting scale. Precision near the achievement-level cut scores and at the lower and upper tails varies by grade. The figures should therefore be interpreted as conditional evidence rather than as a single statement that all score locations are measured equally well.

The results identify where the current forms support more precise interpretation and where future

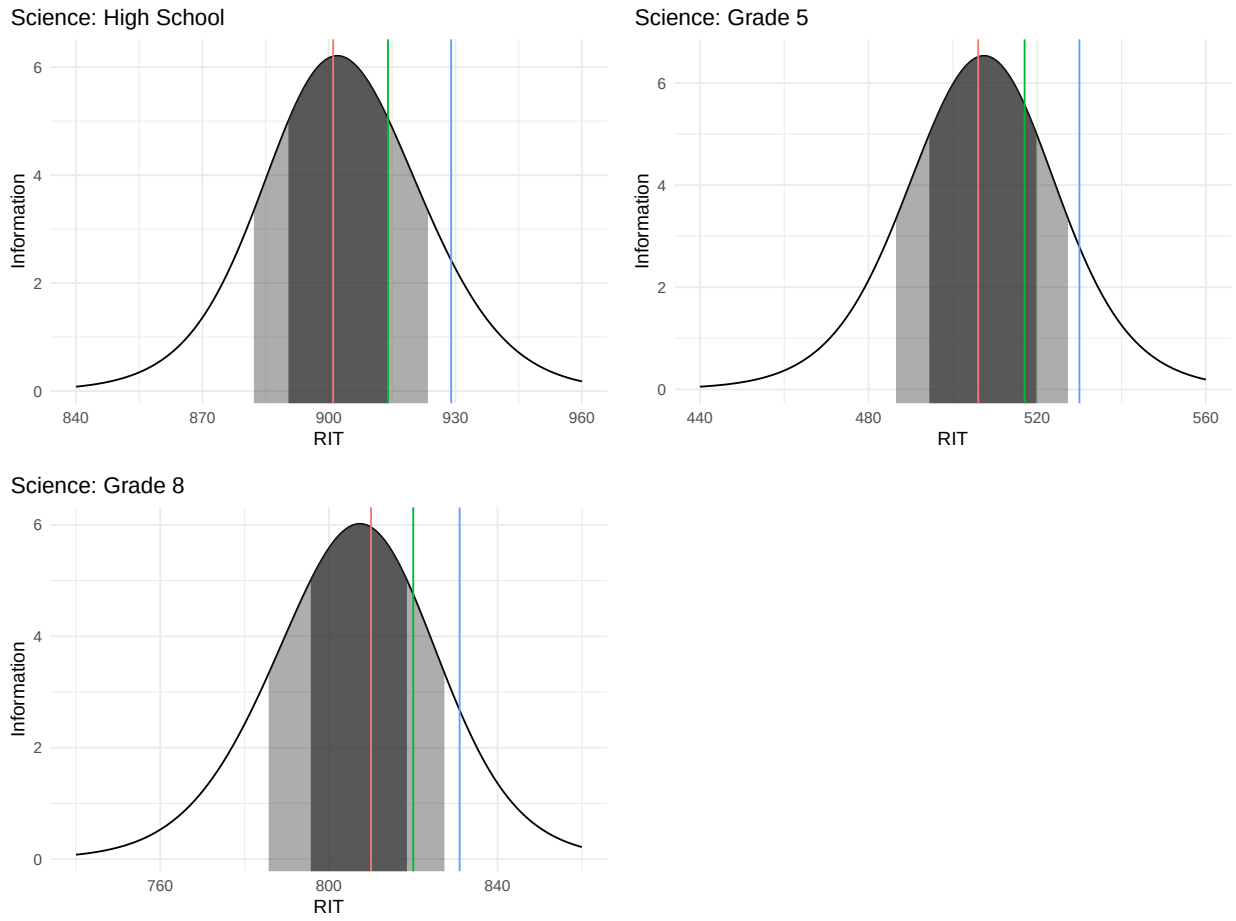
field testing and item development may improve coverage. Exact CSEM values at the achievement-level cut scores are presented in Section 4.1G.



4.1E Science Measurement Precision

The Science TIFs indicate that the operational forms provide their greatest precision in the middle portion of the grade-specific reporting scales. Precision decreases toward score locations where fewer operational items are available, including portions of the upper continuum on some forms.

The figures support score interpretation within the best-measured portions of the scale and identify a continuing need for operational items targeted to less densely measured regions. Exact CSEM values at the achievement-level cut scores are presented in Section 4.1G.

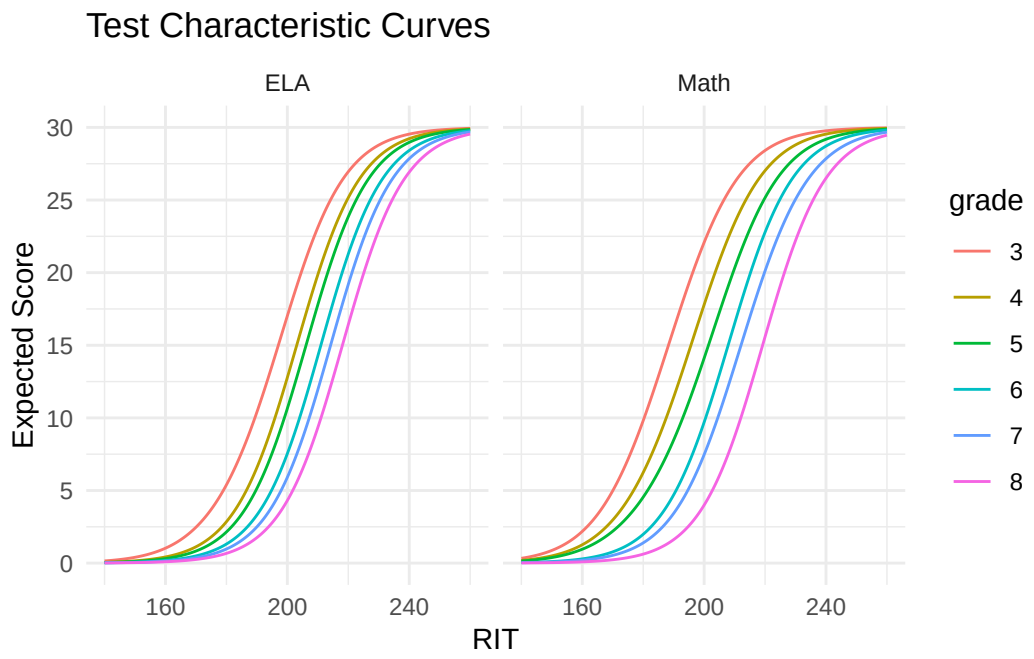


4.1F Expected Score Progression Across Vertically Scaled Forms

TCCs summarize the expected operational score at each location on the OREXt reporting scale by summing Rasch expected-response probabilities across the 30 scored items in each form. TCCs are presented for the vertically scaled Grades 3–8 ELA and Mathematics assessments. Grade 11 and Science use separate reporting scales and are not included in this comparison.

All Rasch analyses and TCC visualizations were conducted using R (R Core Team 2026), the **exirt** package (Anderson and Loan 2022), and the **TAM** package (Robitzsch et al. 2022). The curves provide evidence about the expected ordering, spacing, and relative difficulty of the operational forms on the shared vertical scale.

TCCs are one component of the vertical-scale review. They do not, by themselves, demonstrate empirical longitudinal stability, year-to-year comparability, or student growth. Those interpretations also require evidence from linking procedures, score distributions, conditional measurement precision, and linked-student longitudinal analyses.



The TCC and TIF figures are presented in this chapter because they provide complementary evidence about expected score functioning and conditional precision. Their implications for validity and internal structure are synthesized in Section 3.3.

4.1G Overall and Conditional Standard Errors of Measurement

AMO performance levels define Oregon's statewide achievement classifications for ELA, Mathematics, and Science. CSEM values quantify the uncertainty associated with scores at particular locations on the reporting scale and are especially important near achievement-level cut scores.

Unlike a single overall reliability estimate, CSEM varies with score location. The **2025–2026** CSEM values at each approved AMO cut score are presented below. The values are rounded to two decimal places.

The AMO cut scores define the boundaries between adjacent performance levels:

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

The table shows that measurement uncertainty is not constant across grades, content areas, or classification boundaries. Individual score interpretations and adjacent-year comparisons should therefore use location-specific measurement error rather than treating all observed differences as equally precise.

The following figure applies the CSEM evidence to adjacent-year ORExt scores. Observed differences are classified as reliable score increases, reliable score decreases, or differences not distinguishable from expected measurement error. These classifications are an application of

Table 4.4: Cut Score and Conditional Standard Error of Measurement by Content and Grade

	AMO 2 Cutscore (CSEM)	AMO 3 Cutscore (CSEM)	AMO 4 Cutscore (CSEM)
ELA			
Grade 3	193 (3.93)	214 (4.98)	232 (10.21)
Grade 4	202 (3.72)	215 (4.37)	230 (7.35)
Grade 5	203 (3.73)	223 (4.93)	233 (7.34)
Grade 6	207 (3.74)	221 (4.14)	238 (7.33)
Grade 7	209 (3.81)	223 (4.00)	241 (7.33)
Grade 8	214 (3.76)	225 (3.90)	237 (5.39)
High School	901 (3.08)	921 (3.32)	929 (4.01)
Math			
Grade 3	193 (3.99)	202 (4.56)	225 (10.29)
Grade 4	194 (3.95)	207 (4.35)	220 (6.23)
Grade 5	195 (4.21)	208 (4.06)	222 (5.54)
Grade 6	205 (3.81)	209 (3.77)	225 (4.97)
Grade 7	208 (3.91)	211 (3.85)	225 (4.52)
Grade 8	209 (4.18)	214 (3.84)	228 (4.04)
High School	902 (2.86)	908 (2.88)	925 (3.93)
Science			
Grade 5	507 (3.91)	518 (4.34)	531 (6.23)
Grade 8	811 (4.12)	821 (4.65)	832 (6.33)
High School	903 (2.84)	915 (3.22)	934 (5.44)

Note: AMO = Annual Measurable Objective; CSEM = conditional standard error of measurement at the achievement-level cut score.

score-precision evidence. They are not themselves reliability coefficients and do not establish why a student's score changed.

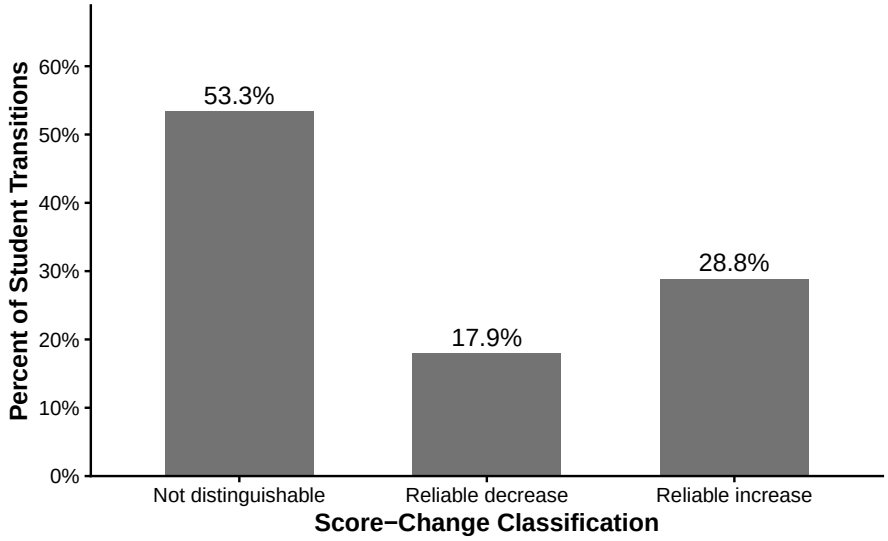


Figure 4.1: Measurement-Aware Classification of Adjacent-Year Score Change. Adjacent-year score differences are classified as reliable increases, reliable decreases, or differences not distinguishable from expected measurement error after accounting for the conditional standard error associated with both scores. The classifications describe the dependability of observed score differences and do not establish the causes of change.

4.1H Classification Accuracy and Consistency

Classification accuracy and classification consistency were evaluated for the **2025–2026** OrExt administration using Rudner's model-based classification indices (Rudner 2005). Classification accuracy estimates the probability that an observed performance level corresponds to the modeled underlying classification. Classification consistency estimates the probability of receiving the same classification on a modeled comparable administration.

The analyses used operational Rasch item difficulties, student ability estimates, and associated standard errors from a single administration. The resulting indices should therefore be interpreted under the assumptions of the Rasch and Rudner models rather than as empirical test-retest results.

Across content areas and grades, the model-based estimates were generally high. They provide supporting evidence for the dependability of achievement-level decisions when interpreted together with marginal reliability, TIFs, CSEM, and the observed administration and scoring evidence.

Table 4.5: Classification Accuracy and Consistency by Content and Grade

	AMO 2 (Acc | Con)	AMO 3 (Acc | Con)	AMO 4 (Acc | Con)
ELA			
Grade 3	0.96 | 0.94	0.94 | 0.92	0.97 | 0.96
Grade 4	0.95 | 0.93	0.96 | 0.93	0.95 | 0.93
Grade 5	0.97 | 0.95	0.93 | 0.91	0.97 | 0.95
Grade 6	0.93 | 0.90	0.95 | 0.93	0.96 | 0.95
Grade 7	0.94 | 0.90	0.93 | 0.90	0.97 | 0.96
Grade 8	0.93 | 0.90	0.97 | 0.95	0.98 | 0.97
High School	0.95 | 0.94	0.92 | 0.90	0.93 | 0.91
Math			
Grade 3	0.94 | 0.92	0.96 | 0.94	0.99 | 0.98
Grade 4	0.95 | 0.94	0.92 | 0.90	0.97 | 0.96
Grade 5	0.94 | 0.91	0.94 | 0.91	0.98 | 0.98
Grade 6	0.92 | 0.89	0.92 | 0.89	0.99 | 0.99
Grade 7	0.92 | 0.88	0.93 | 0.89	0.97 | 0.97
Grade 8	0.91 | 0.88	0.94 | 0.91	0.99 | 0.98
High School	0.92 | 0.89	0.96 | 0.94	0.98 | 0.98
Science			
Grade 5	0.92 | 0.89	0.95 | 0.92	0.95 | 0.94
Grade 8	0.90 | 0.87	0.95 | 0.92	0.94 | 0.92
High School	0.98 | 0.97	0.93 | 0.91	0.94 | 0.93

Note: AMO 2 = Nearly Meets; AMO 3 = Meets; AMO 4 = Exceeds; Acc = classification accuracy; Con = classification consistency.

The OrExt uses fixed operational test forms rather than computer-adaptive testing. Consequently, adaptive test precision indices are not applicable to this assessment.

4.2 Fairness, Accessibility, and Comparability

ODE and BRT have incorporated multiple procedures to support fairness and accessibility throughout the development and administration of the OrExt. Assessment items undergo review by Oregon educators for standards alignment, accessibility, sensitivity, and potential sources of bias (see the Oregon Extended Assessment Alignment Study). Most recently, the 2024–2025 Standards

Linkage and Item Alignment Study extended this work through systematic evaluations of standards alignment, accessibility, cognitive complexity, and item bias for newly developed field-test items.

In addition to content reviews, psychometric analyses are conducted annually to evaluate item functioning across relevant student populations. Collectively, these procedures support the development of assessment items that are aligned with Oregon's academic content standards while minimizing construct-irrelevant sources of variance and promoting equitable access for students with the most significant cognitive disabilities.

4.2A DIF Screening

DIF analyses were conducted to identify operational items that may function differently for demographic groups after matching students on overall performance. DIF is a statistical screening result, not a determination that an item is biased. Large effect-size results and adjusted statistical flags require psychometric and content review before an operational decision is made.

Analyses used the Mantel-Haenszel procedure with iterative purification (Holland and Thayer 1988; Kamata and Vaughn 2004), as implemented in the **difR** package (Magis et al. 2010). Two comparisons were examined where sample sizes permitted: Female/Male and White/Non-White. Missing item responses were retained as missing rather than recoded as incorrect. Embedded field-test items were excluded. Benjamini-Hochberg (BH) adjustments were applied within each grade-by-content analysis and are reported separately from the Educational Testing Service (ETS) effect-size categories.

The pooled Non-White group combines students from several race and ethnicity categories and is therefore heterogeneous. Results from this comparison should be interpreted as a broad screening analysis rather than evidence about any specific racial or ethnic group.

Mantel-Haenszel DIF Screening Summary

The tables report two complementary results. ETS A, B, and C categories describe Mantel-Haenszel effect-size magnitude, while the BH-flagged column reports the number of item comparisons with BH-adjusted p values below .05. Reporting these results separately prevents a nonsignificant result from being treated as evidence of negligible effect size.

- **Category A:** Negligible ETS delta magnitude, $|\Delta_{MH}| < 1.0$
- **Category B:** Moderate ETS delta magnitude, $1.0 \leq |\Delta_{MH}| < 1.5$
- **Category C:** Large ETS delta magnitude, $|\Delta_{MH}| \geq 1.5$
- **BH Flagged:** BH-adjusted p value below .05
- **NE:** The item comparison was not estimable from the available responses

Table 4.6: DIF: Female and Male

	Negligible (A)	Moderate (B)	Large (C)	BH Flagged	Not Estimable
ELA					
Grade 3	29	1	0	0	0
Grade 4	28	0	2	0	0
Grade 5	26	2	2	0	0
Grade 6	27	0	3	0	0

Table 4.6: DIF: Female and Male *(continued)*

	Negligible (A)	Moderate (B)	Large (C)	BH Flagged	Not Estimable
Grade 7	20	6	4	0	0
Grade 8	25	2	3	0	0
High School	20	6	4	0	0
Math					
Grade 3	29	0	1	0	0
Grade 4	26	3	1	0	0
Grade 5	28	1	1	0	0
Grade 6	27	0	3	0	0
Grade 7	27	3	0	0	0
Grade 8	30	0	0	0	0
High School	22	4	4	0	0
Science					
Grade 5	27	3	0	0	0
Grade 8	23	5	2	0	0
High School	20	7	1	2	2

Across all grade-by-content Female/Male analyses, 31 item comparisons demonstrated a large ETS delta magnitude, 2 remained statistically flagged after the BH adjustment, and 2 were not estimable. Effect-size magnitude and adjusted statistical detection are reviewed together at the item level; neither result alone establishes item bias.

Table 4.7: DIF: Non-White and White

	Negligible (A)	Moderate (B)	Large (C)	BH Flagged	Not Estimable
ELA					
Grade 3	28	1	1	0	0
Grade 4	26	1	3	0	0
Grade 5	27	1	2	0	0
Grade 6	30	0	0	0	0
Grade 7	24	2	4	0	0
Grade 8	27	0	3	0	0
High School	19	5	6	0	0
Math					
Grade 3	24	2	4	0	0
Grade 4	26	3	1	0	0
Grade 5	27	3	0	0	0
Grade 6	28	1	1	0	0
Grade 7	26	2	2	0	0
Grade 8	27	0	3	0	0
High School	25	3	2	0	0
Science					
Grade 5	26	2	2	0	0

Table 4.7: DIF: Non-White and White *(continued)*

	Negligible (A)	Moderate (B)	Large (C)	BH Flagged	Not Estimable
Grade 8	22	4	4	0	0
High School	15	8	6	1	1

Across all grade-by-content White/Non-White analyses, 44 item comparisons demonstrated a large ETS delta magnitude, 1 remained statistically flagged after the BH adjustment, and 1 was not estimable. The pooled comparison is used only as a broad screening procedure, and item-level effect sizes and adjusted statistical results are reviewed together.

The Category C counts did not show a consistent grade-by-grade increase across content areas. High School ELA and Science had the largest Category C counts within their respective content areas. Plausible contributors include grade-specific item sets, differences in subgroup size and composition, and the heterogeneity of the pooled Non-White group. The available analyses do not distinguish among these explanations. The pattern is therefore treated as a screening signal for item-level review rather than as evidence of developmental change or systematic bias.

All Category C results are reviewed jointly by psychometric and content specialists. Review considers item content, accessibility, response patterns, sample composition, item-fit evidence, and prior calibration history before an item is retained, revised, replaced, or removed.

4.2B Participation by Race and Ethnicity

Demographic characteristics of students participating in the OExt are summarized below by race/ethnicity and disability category. Student race/ethnicity was reported using seven categories: (a) American Indian or Alaska Native, (b) Asian, (c) Black or African American, (d) Hispanic, (e) Multi-Racial, (f) Native Hawaiian or Other Pacific Islander, and (g) White.

White students represented the largest racial/ethnic group across grades (45% to 54%), reflecting the overall demographic composition of students participating in the OExt. Additional statewide demographic information is available in the State Annual Report Card.

The demographic summaries presented below provide descriptive information about the assessment population and establish context for the DIF analyses presented in Section 4.2A. These summaries are not intended to evaluate subgroup performance or fairness outcomes; rather, they describe the student populations represented in the statewide analyses.

Table 4.8: Racial/Ethnic Demographic Composition of the Sample

	White	Hispanic	Multiracial	Asian
Grade 3	50 (45%)	29 (26%)	16 (14%)	–
Grade 4	57 (53%)	25 (23%)	11 (10%)	–
Grade 5	67 (51%)	30 (23%)	11 (8%)	10 (8%)
Grade 6	68 (54%)	36 (29%)	–	–
Grade 7	47 (46%)	28 (27%)	–	–
Grade 8	49 (45%)	41 (38%)	10 (9%)	–
High School	63 (48%)	46 (35%)	–	–

Note: Cells with fewer than 10 students are suppressed to protect student privacy.

4.2C Participation by Disability Category

Student disability classifications are reported using Oregon's special education eligibility categories. The following eligibility codes describe the primary disability classifications represented in the statewide ORExt participation summaries and subsequent subgroup analyses.

Eligibility Codes

- 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)

Autism Spectrum Disorder (IDEA Code 82) was the most common primary eligibility category in Grades 3–6 and 8. Intellectual Disability (IDEA Code 10) was most common in High School. Codes 10 and 82 were tied in Grade 7.

Table 4.9: Distribution of Primary and Secondary IDEA Eligibility Categories by Grade. Primary and secondary IDEA eligibility codes are summarized together to provide a complete description of the disability classifications represented within each tested grade. Categories with fewer than 10 students are suppressed to protect student confidentiality.

Grade	Primary IDEA Codes	Secondary IDEA Codes
Grade 3	82 (68.8%); 80 (9.8%); 10 (8.9%)	0 (74.1%)
Grade 4	82 (48.1%); 10 (32.4%); 80 (12.0%)	0 (69.4%)
Grade 5	82 (58.0%); 10 (19.1%); 80 (16.0%)	0 (76.3%)
Grade 6	82 (46.4%); 10 (28.8%); 80 (16.0%)	0 (69.6%); 50 (12.0%); 10 (9.6%)
Grade 7	10 (43.7%); 82 (43.7%)	0 (65.0%); 50 (15.5%); 80 (10.7%)
Grade 8	82 (45.0%); 10 (33.0%); 80 (14.7%)	0 (73.4%)
Grade 11	10 (47.7%); 82 (33.3%); 80 (7.6%)	0 (70.5%); 80 (8.3%)

Note: Cells with fewer than 10 students are suppressed to protect student privacy.

4.2D Descriptive Performance Patterns Across Student Groups

The following tables summarize observed score means and SDs by content area and grade. ELA and Mathematics scores for Grades 3–8 are reported on the vertically scaled OREXt reporting scale centered on 200, while Grade 11 scores are reported on a separate reporting scale centered on 900. Science scores are reported on grade-specific scales centered on 500 for Grade 5 and 800 for Grade 8. These reporting scales distinguish assessment programs while supporting consistent interpretation of OREXt results within their intended reporting frameworks.

Observed score locations differed across grades, but the cross-sectional means and medians were not strictly monotonic. Each grade represented a different student cohort, so these descriptive comparisons were not direct evidence of individual growth or empirical vertical-scale stability. The summaries provided context for the demographic comparisons that follow. Longitudinal interpretations rely instead on linked-student adjacent-year analyses together with the TCC, TIF, and CSEM evidence presented elsewhere in this report.

Table 4.10: Observed RIT Score: Measures of Central Tendency by Grade and Content

	RIT Median	RIT Mean (SD)
ELA		
Grade 3	191	187.88 (27.17)
Grade 4	199	196.59 (25.65)
Grade 5	203	200.79 (23.32)
Grade 6	206	204.83 (23.61)
Grade 7	213	209.10 (19.77)
Grade 8	206	204.41 (17.77)
High School	913	910.01 (20.15)
Math		
Grade 3	184	178.90 (25.09)
Grade 4	188	184.24 (24.86)
Grade 5	197	186.34 (23.83)

Table 4.10: Observed RIT Score: Measures of Central Tendency by Grade and Content *(continued)*

	RIT Median	RIT Mean (SD)
Grade 6	198	194.68 (16.25)
Grade 7	201	197.09 (17.22)
Grade 8	197	195.65 (16.79)
High School	899	893.99 (17.40)
Science		
Grade 5	505	498.70 (24.33)
Grade 8	810	805.51 (23.54)
High School	915	912.39 (24.99)

Performance by Gender

The following table summarizes average performance by gender for each grade and content area. Welch's two-sample t-tests were treated as exploratory and were adjusted for multiple comparisons using the BH procedure. No grade-by-content comparison remained statistically significant after the BH correction. Mean differences are descriptive and are not, by themselves, evidence of assessment bias or comparability.

Table 4.11: Mean (SD) RIT Scores by Gender, Grade, and Content

	Female RIT Mean (RIT SD)	Male RIT Mean (RIT SD)
ELA		
Grade 3	186.69 (28.03)	188.38 (26.99)
Grade 4	188.56 (23.77)	199.82 (25.86)
Grade 5	196.47 (21.78)	202.79 (23.89)
Grade 6	205.51 (26.07)	204.44 (22.30)
Grade 7	213.00 (17.42)	207.00 (20.80)
Grade 8	206.21 (19.01)	203.35 (17.13)
High School	909.61 (19.84)	910.26 (20.49)
Math		
Grade 3	178.96 (24.22)	178.88 (25.69)
Grade 4	178.08 (26.67)	186.90 (23.79)
Grade 5	177.82 (23.45)	190.94 (22.92)
Grade 6	193.65 (17.93)	195.22 (15.47)
Grade 7	199.12 (16.79)	196.12 (17.51)
Grade 8	191.34 (16.16)	198.56 (16.77)
High School	890.56 (15.70)	895.97 (18.14)
Science		
Grade 5	497.55 (21.43)	499.30 (25.86)
Grade 8	803.03 (22.86)	806.93 (24.02)
High School	912.29 (25.09)	912.46 (25.16)

Performance by Race and Ethnicity

The following integrated table summarizes mean RIT scores and SDs by content area, grade, and the race/ethnicity groups with sufficient sample sizes for stable public reporting. Smaller groups remain in the source data and are available for internal review but are omitted from the displayed summary to protect privacy and avoid unstable descriptive estimates.

Table 4.12: Mean (SD) RIT Scores by Content, Grade, and Race/Ethnicity

Grade	White	Hispanic	Multiracial
ELA			
Grade 3	195.59 (26.66)	182.08 (25.38)	196.75 (29.28)
Grade 4	204.56 (25.64)	188.64 (23.55)	–
Grade 5	203.32 (21.94)	198.25 (24.99)	–
Grade 6	208.13 (23.36)	196.67 (23.80)	–
Grade 7	206.83 (19.61)	209.29 (19.42)	–
Grade 8	207.91 (18.63)	201.19 (15.38)	–
HS	915.42 (14.20)	903.63 (22.08)	–
Math			
Grade 3	186.49 (19.58)	171.54 (26.80)	187.18 (27.13)
Grade 4	190.74 (24.87)	178.62 (23.52)	–
Grade 5	189.04 (21.23)	183.48 (28.90)	–
Grade 6	193.22 (17.54)	195.68 (15.37)	–
Grade 7	194.65 (17.51)	196.55 (16.77)	–
Grade 8	196.88 (16.69)	196.29 (16.18)	–
HS	897.29 (17.01)	889.47 (17.07)	–
Science			
Grade 5	501.90 (23.05)	499.76 (25.82)	–
Grade 8	806.58 (22.07)	806.66 (23.94)	–
HS	918.12 (18.57)	903.76 (29.62)	–

Note: Cells with fewer than 10 students are suppressed.

Performance by Disability Category

The following table summarizes mean RIT scores and SDs by primary IDEA eligibility code, grade, and content area.

Table 4.13: Mean (SD) RIT Scores by Primary IDEA Eligibility Code

Primary IDEA Code	ELA	Math	Science
Grade 3			
82	186.13 (25.36)	179.31 (23.24)	–
Grade 4			
10	201.76 (25.30)	190.46 (26.57)	–
80	174.25 (20.54)	161.08 (21.19)	–
82	200.62 (23.48)	188.34 (20.11)	–
Grade 5			

Table 4.13: Mean (SD) RIT Scores by Primary IDEA Eligibility Code *(continued)*

Primary IDEA Code	ELA	Math	Science
10	205.18 (19.92)	189.05 (22.47)	511.37 (17.93)
80	199.20 (22.86)	182.62 (22.28)	493.67 (26.47)
82	200.65 (23.18)	185.02 (24.47)	495.69 (22.74)
Grade 6			
10	209.59 (25.15)	196.12 (18.30)	–
80	204.94 (24.72)	197.91 (15.14)	–
82	201.45 (21.76)	194.00 (14.47)	–
Grade 7			
10	212.36 (18.29)	197.82 (16.56)	–
82	206.94 (19.74)	196.13 (18.71)	–
Grade 8			
10	207.74 (13.04)	193.74 (16.23)	815.03 (13.54)
80	–	196.36 (13.82)	794.27 (26.29)
82	204.44 (18.34)	197.45 (17.94)	801.82 (25.42)
High School			
10	908.00 (18.64)	889.37 (16.72)	910.00 (24.31)
82	911.53 (22.83)	899.53 (15.98)	916.12 (24.68)

Note:
Cells with fewer than 10 students are suppressed to protect student privacy.

Taken together, item-review procedures, DIF screening, participation summaries, and descriptive subgroup comparisons provide complementary evidence for the annual fairness review. Descriptive group differences are not evidence of bias, and DIF flags are not bias determinations. Defensible conclusions require documented content review and disposition of flagged items.

4.3 Coverage of the Full Performance Continuum

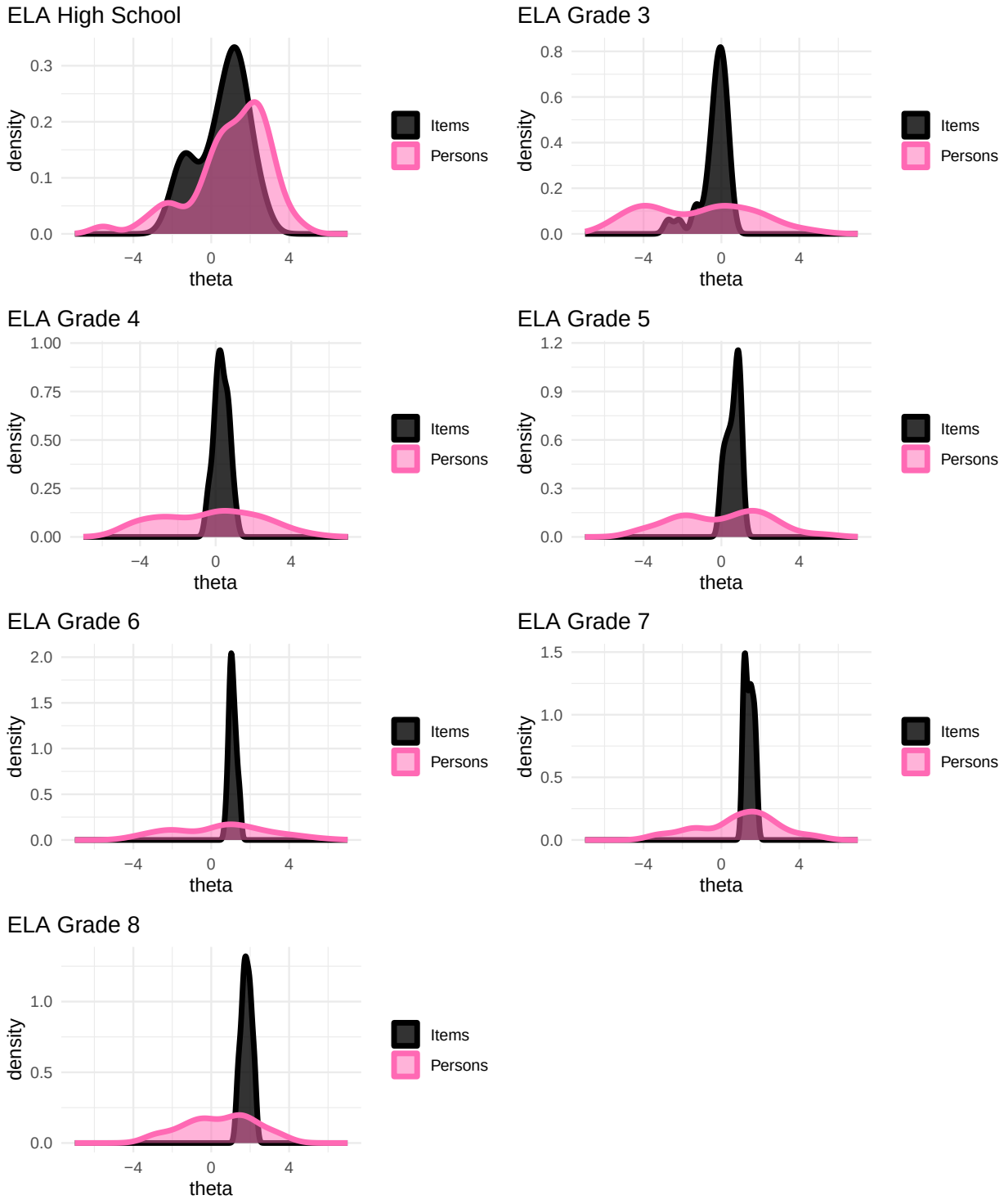
Critical Element 4.3 concerns whether each assessment provides adequate precision across the performance continuum, including for students performing at the lower and upper ends. The evidence considered here includes operational TIFs, CSEM at the achievement-level cut scores, score and performance-level distributions, and item-person maps based on the 30 scored operational items.

The item-person maps compare the distribution of operational item difficulties with the distribution of student ability estimates. Precision is strongest where the distributions overlap and weaker where students fall beyond densely represented item difficulties. The maps do not show equal precision across the entire observed range and should be interpreted together with the TIF and CSEM evidence.

ORora is not used as evidence that the ORExt covers the academic performance continuum. ORora has a different purpose and construct and cannot compensate for limited academic measurement precision at the lower end of an ORExt form.

4.3A ELA Item-Person Maps

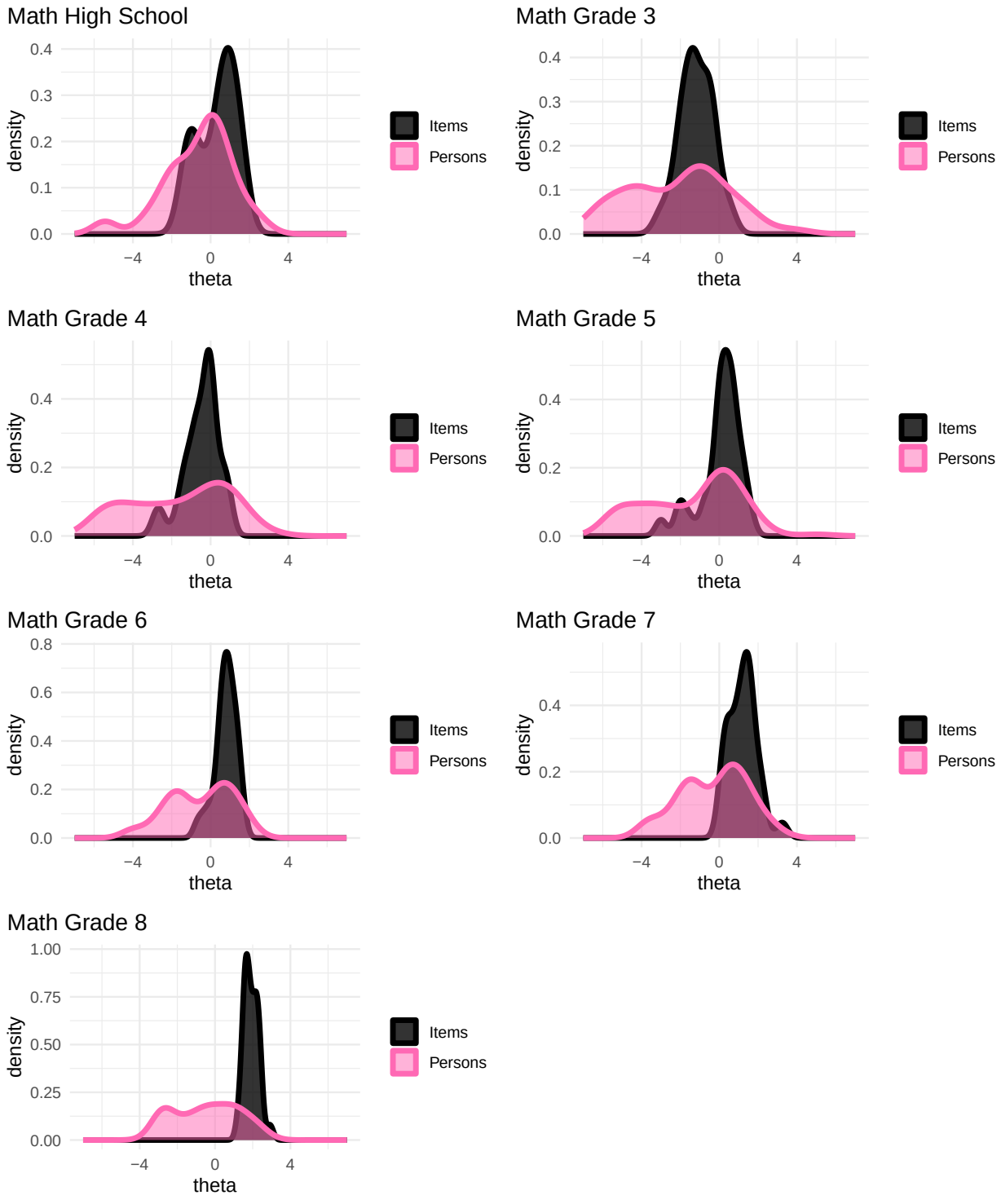
The ELA item-person maps show overlap between operational item difficulties and the central portion of the student ability distributions. Several forms have comparatively narrow concentrations of item difficulty, and coverage is less dense for students at the lower and upper tails. The maps therefore support useful measurement across much of the observed distribution while identifying locations where additional targeted operational items may improve precision.



4.3B Mathematics Item-Person Maps

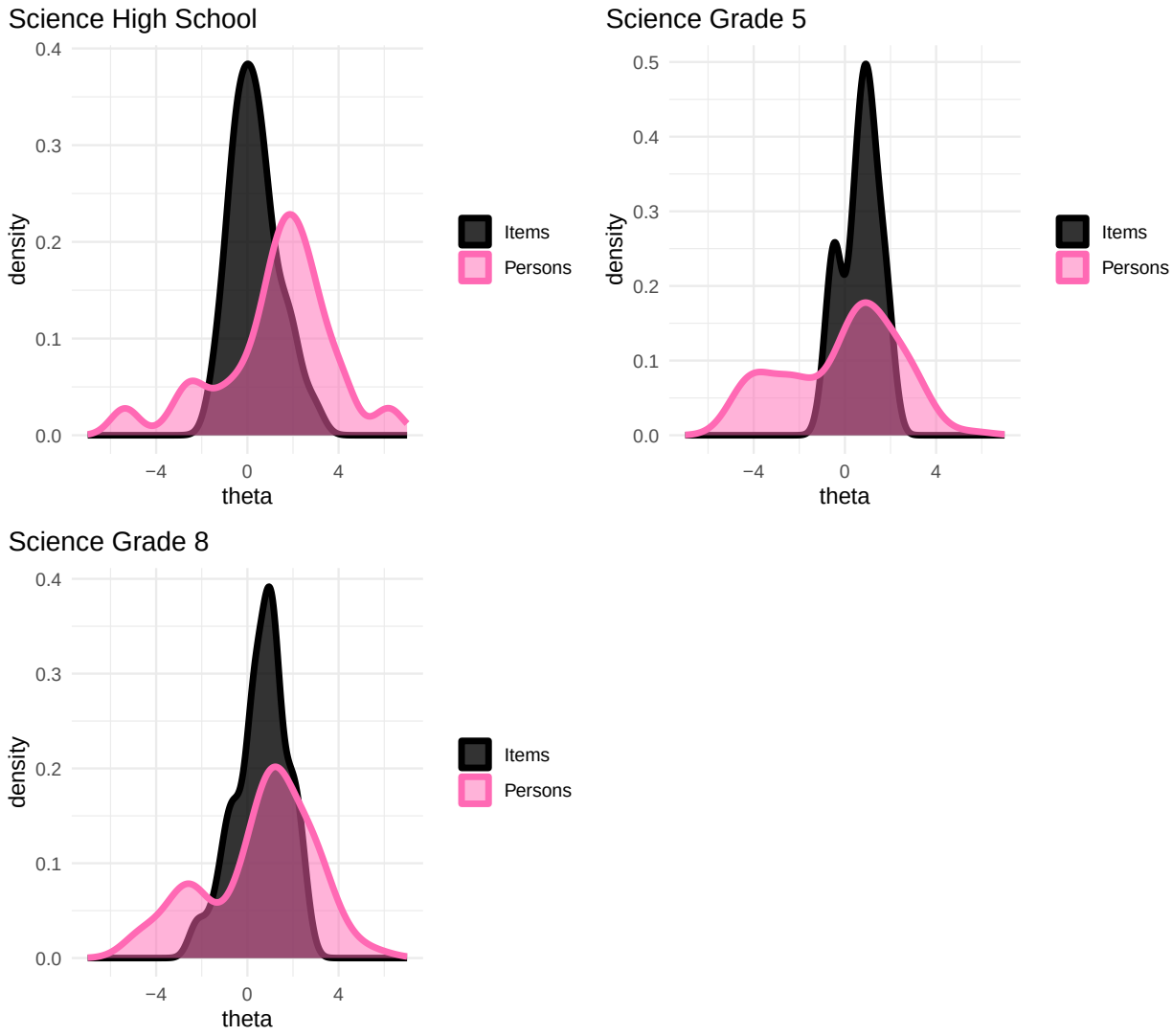
The Mathematics item-person maps show central overlap between operational item difficulty and student ability, but the amount of coverage varies across forms. Several maps include students below the range where operational items are most densely concentrated, and upper-tail coverage

also differs by grade. These patterns identify specific portions of the continuum for targeted item development rather than demonstrating uniform precision across all score locations.



4.3C Science Item-Person Maps

The Science item difficulties are concentrated primarily in the middle portion of the grade-specific ability distributions. Operational coverage is less dense at both tails, particularly for students whose estimated performance falls outside the central cluster of item difficulties. The maps support useful measurement within the central range while identifying a need for additional items targeted to the lower and upper portions of the continuum.



Across content areas, the evidence supports comparatively stronger precision where operational item coverage is concentrated and reduced precision at less densely measured score locations. Annual field testing and form development should prioritize the lower and upper regions identified by the TIF and item-person evidence.

4.4 Scoring and Score Interpretation

Scoring procedures for the ORExt are documented in the T&P system and the annual Test Administration Manual. Operational items are scored dichotomously, with one point for a correct response and zero points for an incorrect response. The six embedded field-test items are excluded from reported raw scores, operational ability estimates, RIT scores, and achievement-level classifications.

Assessment records distinguish correct responses, incorrect responses, omitted responses, incomplete submissions, discontinued assessments, and reopened test sessions. Duplicate and anomalous records are reviewed through data-validation and operational-monitoring procedures before statewide score files are finalized. Minimum-participation rules are applied separately from the scoring rules for completed assessments.

Items requiring human judgment are scored using approved item-specific guidance and rubrics. QAs complete annual training and proficiency requirements before administering or scoring the assessment. Manual-scoring completion is monitored during the assessment window. The observed scoring agreement reported in Section 4.1.1 provides current-year implementation evidence for the participating paper/pencil administrations.

Reported operational responses are converted to Rasch ability estimates and transformed to the approved ORExt reporting scales. Achievement levels are assigned using Oregon's adopted cut scores. Reading and Writing subscores are reported as supporting components of ELA and should be interpreted in light of their lower reliability in some grades.

The Individual Student Report (ISR) communicates the student's RIT score, achievement level, applicable subscores, and interpretive guidance. Section 6.4C describes the report and its intended limitations.

4.5 Operational Form Design

During the **2025–2026** administration, the ORExt used one fixed operational form for each grade and content area. Each form contained 30 scored operational items and 6 embedded field-test items. Because only one form was used within each grade and content area during the current administration, within-year alternate-form comparability is not applicable.

Comparability across school years remains relevant because operational item combinations change. Grades 3–8 ELA and Mathematics use established vertical reporting scales and calibrated item parameters; Grade 11 and Science use separate reporting scales. Blueprint review, calibrated-item selection, TCCs, TIFs, score-conversion review, and field testing contribute to cross-year form maintenance. The current chapter does not present a dedicated year-to-year linking stability or item-parameter drift study.

4.6 Accessible Assessment Formats

The ORExt is available in standard, Large Print, and Braille formats. Accessible versions are developed from the approved operational content and are intended to preserve the academic

knowledge and skills measured by the standard form. Adaptations are limited to presentation and response changes judged necessary for access.

Large Print and Braille materials undergo content and accessibility review before operational use. When an item cannot be represented in Braille without changing the intended construct, the accessible-format procedures determine whether the item is excluded or replaced and how the resulting score is calculated. Those rules must be documented explicitly because changing the number or composition of scored items can affect score precision and interpretation.

The number of students using alternate formats is generally too small for stable large-sample comparability analyses. Current evidence is therefore primarily procedural and should be interpreted as limited.

4.7 Ongoing Technical Review and Maintenance

The ORExt annual technical review integrates reliability estimates, TIFs, TCCs, CSEM, item-fit results, DIF screening, participation summaries, scoring observations, accessible-format review, operational incidents, and longitudinal evidence. BRT and ODE use these sources to identify issues requiring follow-up in item development, field testing, form assembly, training, scoring, accessibility, reporting, or technical documentation.

A technical finding does not result automatically in item removal or a program change. Psychometric, content, accessibility, and operational evidence are considered together when determining whether an item or form should be retained, revised, replaced, or subjected to additional monitoring. Review decisions and subsequent actions are documented through the program's annual technical maintenance procedures.

When appropriate, findings are considered through ODE's broader assessment governance and technical-review processes. Public technical documentation and supporting program records together provide the evidence needed to evaluate the continued appropriateness of ORExt score interpretations.

4.8 Technical Quality Synthesis

The evidence in this chapter supports several components of the ORExt technical quality argument. Marginal reliability is comparatively strong for most content-area scores, while some Reading and Writing subscore estimates are lower. TIFs, CSEM, and item-person maps show useful precision across the central portions of the reporting scales and reduced precision at less densely measured score locations. Classification accuracy and consistency estimates are generally high under the applied Rasch and Rudner models.

Administration observations provide strong scoring-agreement evidence within the participating paper/pencil sample, while also identifying score interpretation as an area for continued training. DIF analyses function as screening procedures and require documented content review before conclusions about item bias or fairness are reached. Descriptive subgroup differences are not interpreted as evidence of bias.

Taken together, the available evidence supports continued use of OREXt scores for their stated purposes when the limitations described in this chapter are respected. The evidence does not support claims of equal precision at every score location, empirical equivalence across every delivery format, or cross-year comparability without the associated operational and linking artifacts. Continued technical maintenance should focus on score-range coverage, subscore precision, documented disposition of flagged items, accessible-format comparability, and stability of score interpretations across administrations.

5 Inclusion of All Students

This chapter synthesizes the policies, procedures, and operational evidence used to include students with disabilities and English learners in Oregon's statewide academic assessment system. Rather than repeating the detailed accessibility, administration, and monitoring information presented in Chapters 1, 2, and 4, the chapter focuses on the evidence relevant to Federal Assessment Peer Review Critical Elements 5.1 through 5.4.

5.1 Procedures for Including Students With Disabilities

Oregon's statewide assessment system includes all public elementary and secondary school students with disabilities. As described in Section 1.5, assessment decisions are made individually by the team responsible under the applicable disability law. An IEP team determines participation for a student served under IDEA. For a student served under Section 504 or Title II, the responsible district team determines the accessibility supports needed for participation in the general assessment.

Participation in the ORExt is limited to students whose IEP teams determine, case by case, that the student meets Oregon's alternate-assessment participation criteria. The parent participates as a member of the IEP team. The decision may not be based solely on disability category, instructional setting, anticipated low performance, English learner status, or the need for accommodations.

Clear Explanation of the Assessment Options

The Test Administration Manual, the Oregon Extended Assessment page, and Oregon's alternate-assessment decision-making guidance describe the available statewide assessment options. These resources distinguish among the general academic assessment without accommodations, the general assessment with designated supports or accommodations, and the ORExt.

ORExt results are not directly comparable with results from Oregon's general statewide assessments because the assessment systems serve different student populations and apply different achievement standards. IEP teams and parents should understand this distinction before an ORExt participation decision is finalized.

Individualized Eligibility Decisions

The IEP team applies Oregon's Alternate Assessment Decision-Making Guidance to determine whether the ORExt is appropriate for an individual student. The decision considers the student's cognitive functioning, adaptive behavior, instructional needs, and ability to participate in the general assessment when appropriate accommodations are provided.

Assessment Selection and Accessibility Supports

Students who do not meet the ORExt participation criteria take Oregon's general academic assessments with the universal tools, designated supports, and accommodations needed to address their individual access needs. Students who meet the criteria take the ORExt with the supports permitted for that assessment.

The Oregon Accessibility Manual provides the statewide framework for selecting supports. ORExt-specific administration and accessibility procedures are described in Sections 2.2 and 4.6 and reinforced through annual T&P requirements for QAs and QTCs.

Accessibility Within the ORExt Design

As described in Sections 1.2, 2.2.1, and 4.6, accessibility is incorporated through the EAFs, reduced depth, breadth, and complexity, controlled language, visual and object supports where appropriate, standardized item presentation, multiple response modes, and approved Large Print, Braille, and ASL supports. These features are intended to reduce construct-irrelevant barriers while preserving the academic knowledge or skill represented by each item.

Practice tests and sample items allow students and educators to become familiar with assessment presentation, response requirements, and administration procedures before operational testing.

Eligibility Is Not Based on Disability Category Alone

No disability category automatically qualifies or disqualifies a student from participation in the ORExt. Eligibility depends on the IEP team's documented application of Oregon's participation criteria to the individual student. Students served only under Section 504 or Title II take the general assessment with appropriate accessibility supports unless they are also eligible under IDEA and their IEP team determines that the ORExt criteria are met.

Oregon also monitors district participation under the federal 1.0% participation requirement. Districts projecting participation above the threshold submit justification and planning information to ODE. District projections, justification plans, ODE review records, and follow-up documentation provide evidence that the participation criteria are monitored in practice.

Parent Information and Diploma Considerations

Parents participate in the IEP-team decision and must be informed that the student's achievement will be measured against the AAAS. Information provided to parents should explain that ORExt results are not directly comparable with general-assessment results and should describe any potential effects of participation on diploma pathways or other educational decisions.

Participation in the ORExt does not preclude a student from attempting to complete the requirements for a regular high school diploma.

Access to the Grade-Level General Education Curriculum

The ORExt is linked to Oregon's grade-level academic content standards through the AAAS and EAFs. The EAFs reduce depth, breadth, and complexity while retaining the central academic knowledge and skills represented by the grade-level standards.

The 2025 Standards Linkage and Item Alignment Study, summarized in Section 3.1C, provides current evidence regarding the linkage among Oregon's academic content standards, the Essentialized Standards, and newly developed field-test items. That evidence supports the academic-content relationship underlying the ORExt; it does not establish that every student's daily instruction or IEP is linked to grade-level content.

Oregon supports implementation through standards-based IEP guidance, instructional resources, professional-development materials, and state and district monitoring activities intended to promote access to and progress in the grade-level general education curriculum.

5.2 Procedures for Including English Learners

English learners, including English learners with disabilities, are included in Oregon's academic assessment system. English learner status alone does not determine participation in the ORExt. For an English learner with an IEP, the IEP team first applies Oregon's alternate-assessment participation criteria. Linguistic supports are then selected according to the student's individual language and accessibility needs.

Oregon's English Learner resources, Test Administration Manual, and Oregon Accessibility Manual provide the policy framework for selecting linguistic supports. Decisions should consider the student's proficiency in English and the home language, prior use of supports during instruction, assessment demands, and the support most likely to yield accurate information about the student's academic knowledge and skills.

The ORExt is administered in English. Approved language support may be used for administration directions when permitted by state policy, but it may not translate, simplify, explain, or otherwise alter item content, response options, scoring criteria, or the construct measured. Personnel providing oral interpretation must meet state and district qualification, training, fidelity, and test-security requirements.

Alternate English Language Proficiency Assessment (Alt-ELPA) eligibility and domain-exemption procedures are documented separately and are not evaluated through ORExt score evidence.

5.3 Accommodations

The Oregon Accessibility Manual provides the statewide framework for universal tools, designated supports, and accommodations. The manual applies to students with disabilities, students served under Section 504 or Title II, English learners, and English learners with disabilities. The Test Administration Manual and T&P materials provide ORExt-specific implementation requirements.

Accommodation decisions are made individually. A support should address a documented access barrier, be consistent with the student's instructional or practice experience when appropriate, and

be implemented without changing the academic construct measured by the assessment. Allowable accommodations do not change the ORExt reporting scale, achievement standards, or type of score report provided to the student.

Accommodations for Students With Disabilities

Students with disabilities may use accommodations identified by their IEP team, Section 504 placement team, or other district team authorized under Title II, provided that the supports are permitted for the assessment and administered according to state procedures. Available ORExt formats and supports include standard, Large Print, Braille, ASL, and other approved presentation or response supports described in the state manuals.

When assistive technology is used, applicable state or ORExt-specific technical guidance should document compatibility, configuration, security, and administration requirements. This report does not establish interoperability with every assistive-technology device or configuration.

Linguistic Accommodations for English Learners

English learners may receive approved linguistic supports based on individual need. Language support is limited by the requirement to preserve the academic content and scoring expectations of each ORExt item. Interpretation of allowable directions does not authorize translation or explanation of secure item content unless a specific state policy expressly permits that support.

These procedures are documented through the Oregon Accessibility Manual, Test Administration Manual, training materials for personnel providing language support, and district accommodation records.

Evidence of Appropriateness and Effectiveness

Evidence concerning ORExt accessibility and accommodations is drawn from educator accessibility reviews during item development, the 2025 Standards Linkage and Item Alignment Study, review of accessible formats, available accommodation-use records, Help Desk and operational records, and the administration observations presented in Section 4.1.1.

These sources provide evidence that accessibility is considered during development and that allowable supports can be implemented under standardized procedures. Accommodation frequencies and favorable expert reviews do not, by themselves, demonstrate that every accommodation is effective or that scores from accommodated and non-accommodated administrations are fully comparable. Direct empirical comparability evidence remains limited, particularly for low-incidence formats and supports.

Requests for Additional or Unique Accommodations

Oregon maintains a process through which educators and district personnel may request an accommodation not routinely listed for the assessment. A request should document the student's access need, the proposed support, why routinely available supports are insufficient, and how the proposed support will preserve the assessed construct.

ODE maintains the request form, submission instructions, decision criteria, review records, approval or denial communications, and any conditions placed on use as supporting program records.

5.4 Monitoring Test Administration for Special Populations

The statewide monitoring framework described in Section 2.3 includes ESSA monitoring procedures, district and school responsibilities, annual T&P requirements, operational system records, reported irregularities, Help Desk activity, and direct observations of ORExt administration. ODE's resources include monitoring guidance, a rotation of districts, self-assessment materials, and monitoring rubrics.

Direct ORExt evidence is presented in Section 4.1.1. During the 2025–2026 administration, QTCs observed 25 paper/pencil administrations using a standardized protocol that addressed accessibility supports, level of support, minimum-participation procedures, scoring, and response recording. Ratings were concentrated in the Exemplary and Appropriate categories for most domains, and item-level scoring agreement was 99.9%. Score interpretation showed greater variability and was identified as the primary area for continued training and technical assistance.

Because the study involved a limited sample of observed paper/pencil administrations, the findings should not be generalized to every district, delivery mode, student, or accommodation. The study also did not document every assessment-selection decision or accommodation assignment; state policy and implementation records provide additional evidence regarding those processes.

Together, the policy framework, monitoring activities, direct observations, and documented follow-up provide evidence for evaluating whether students with disabilities and English learners were assigned to the appropriate assessment, received appropriate supports, and were included in assessment administration with fidelity.

6 Standards and Reporting

Oregon's achievement standards and reporting system translates ORExt results into interpretations of student achievement relative to Oregon's AAAS. Achievement standards establish the categories used to describe performance, while reporting communicates those results to students, families, educators, districts, and the public.

For ELA and Mathematics in Grades 3–8, the ORExt also uses vertically scaled scores. The longitudinal analyses in Chapters 3 and 4 evaluated adjacent-year score differences in relation to measurement error and provide technical context for interpreting change over time.

6.1 State Adoption of Alternate Academic Achievement Standards for Students With the Most Significant Cognitive Disabilities

The ORExt is Oregon's statewide AA-AAAS for eligible students with the most significant cognitive disabilities. It is administered in ELA and Mathematics in Grades 3–8 and 11 and in Science in Grades 5, 8, and 11. Participation is determined through the student's IEP team using Oregon's alternate-assessment participation criteria.

The ORExt AAAS are linked to Oregon's grade-level academic content standards through the EAFs. The EAFs document how academic expectations are reduced in depth, breadth, and complexity while preserving the central academic knowledge and skills represented in Oregon's standards. Chapters 1 and 3 present the content-linkage and assessment-alignment evidence supporting this relationship.

6.1A State Formally Adopted Alternate Academic Achievement Standards

The Oregon State Board of Education formally adopted the ORExt AAAS, including the four achievement levels, ALDs, and associated cut scores, on June 25, 2015. The adopted standards establish the statewide performance expectations used to classify and report ORExt results in ELA, Mathematics, and Science. Oregon's current alternate-assessment resources are available through the Alternate Academic Achievement Standards.

6.1B State Applies AAAS to Eligible Students in the Tested Grades

Oregon applies the adopted AAAS statewide to eligible public school students with the most significant cognitive disabilities who participate in the ORExt in the required grades and content areas. The standards do not apply on the basis of disability category alone. The student's IEP team determines whether the ORExt is the appropriate statewide assessment and documents that decision in the student's IEP.

6.1C State's AAAS Include Achievement Levels, ALDs, and Cut Scores

Oregon's AAAS include four achievement levels. Levels 3 and 4 represent performance at or above Oregon's proficient standard for the AAAS, while Levels 1 and 2 represent performance below the proficient standard. Three adopted cut scores define the boundaries among the four levels for each tested grade and content area.

The achievement levels provide categorical interpretations for accountability and reporting. Vertically scaled ELA and Mathematics scores in Grades 3–8 provide complementary continuous-score information. Interpretations of change across adjacent grades should be made in conjunction with the measurement-precision evidence presented in Chapters 3 and 4.

6.2 Achievement Standard Setting

Oregon established the ORExt achievement standards through a formal standard-setting study conducted at the University of Oregon in June 2015. Separate panels were convened for Science, Mathematics, and ELA. Panelists included Oregon educators with expertise in academic content, assessment, special education, and instruction for students with the most significant cognitive disabilities.

Panelists completed standardized training and applied a structured bookmarking methodology consistent with established psychometric practice (Hambleton and Pitoniak 2006; Yen et al. 2006; Cizek 2012). The process resulted in recommendations for cut scores and ALDs. The Oregon State Board of Education subsequently adopted the resulting achievement standards.

The ORExt Assessment Technical Report on Standard Setting is the primary technical record of the method, panel composition, training, judgment process, recommended cut scores, impact information, and independent audit. This annual technical report summarizes that earlier work and documents how the adopted standards continue to be used. No new standard-setting study was conducted during 2025–2026.

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 vertically scaled ELA and Mathematics reporting scales use a common reporting scale centered at 200 across Grades 3–8. Grade 11 assessments are reported on independent scales centered on 900 and are not intended for longitudinal comparisons. Science assessments are reported on independent scales centered on 500 for Grade 5 and 800 for Grade 8.

Oregon adopted updated academic content standards after the 2015 standard-setting study. The 2024–2025 Standards Linkage and Item Alignment Study examined the linkage among those standards, the Essentialized Standards, and newly developed field-test items. That study provides current content-alignment evidence, but it did not re-establish the achievement-level cut scores or repeat the 2015 standard-setting process.

6.3 Challenging and Aligned Academic Achievement Standards

Oregon's AAAS are intended to set challenging expectations for students with the most significant cognitive disabilities while remaining linked to the grade-level academic content standards in which each student is enrolled. The EAFs document the relationship among Oregon's grade-level standards, the Essentialized Standards, the assessment blueprints, and ORExt items.

The 2015 standard-setting process contributes evidence that the achievement levels reflect professional judgment regarding the performance expectations appropriate for the assessed population. The 2017 independent alignment study and the 2024–2025 Standards Linkage and Item Alignment Study provide separate content-based evidence regarding the relationship among Oregon's grade-level standards, Essentialized Standards, and ORExt content. The more recent study examined 149 Essentialized Standards and 393 newly developed field-test items across ELA, Mathematics, and Science.

Summary results are presented below. Complete documentation of the earlier alignment study is available in the Oregon Extended Assessment Alignment Study, and the current standards-linkage findings are documented in the ORExt Linkage and Alignment Study Report 2025.

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.

Taken together, the EAFs, standard-setting documentation, alignment studies, assessment blueprints, and item-review procedures support the conclusion that ORExt achievement

expectations are linked to Oregon's grade-level academic content standards and are intended to promote access to the general education curriculum. The IEP team designates the ORExt for each student to whom the AAAS apply.

The current annual technical evidence addresses standards linkage, assessment alignment, professional judgment, score reliability, and classification quality. It does not directly evaluate postsecondary education or competitive integrated employment outcomes for students meeting the ORExt proficient standard.

6.4 Reporting

Oregon's reporting system communicates ORExt results at the student, school, district, and state levels. Chapters 3 and 4 address the technical quality of the reported scores; this section describes how achievement results are summarized and communicated. ORExt results are reported relative to Oregon's AAAS and are not directly comparable with results from Oregon's general statewide assessments.

The 2025–2026 technical analyses included scaled scores, achievement-level classifications, standard errors of measurement, conditional measurement precision, and measurement-aware longitudinal summaries. Public and individual reports present the reporting elements appropriate to their intended audiences and purposes.

6.4A Public Reporting

Oregon publicly reports assessment participation and achievement results through state reporting systems. Results are reported for all students and required student groups at the school, district, and state levels, subject to Oregon's minimum-reporting and suppression rules for protecting student confidentiality. The State Annual Report Card provides the broader statewide reporting context.

This annual technical report does not reproduce the complete public reporting system. Instead, it documents the design, scoring, reliability, validity, fairness, and measurement-precision evidence underlying the ORExt results reported through Oregon's public systems.

6.4B State Reports Interpretable Results

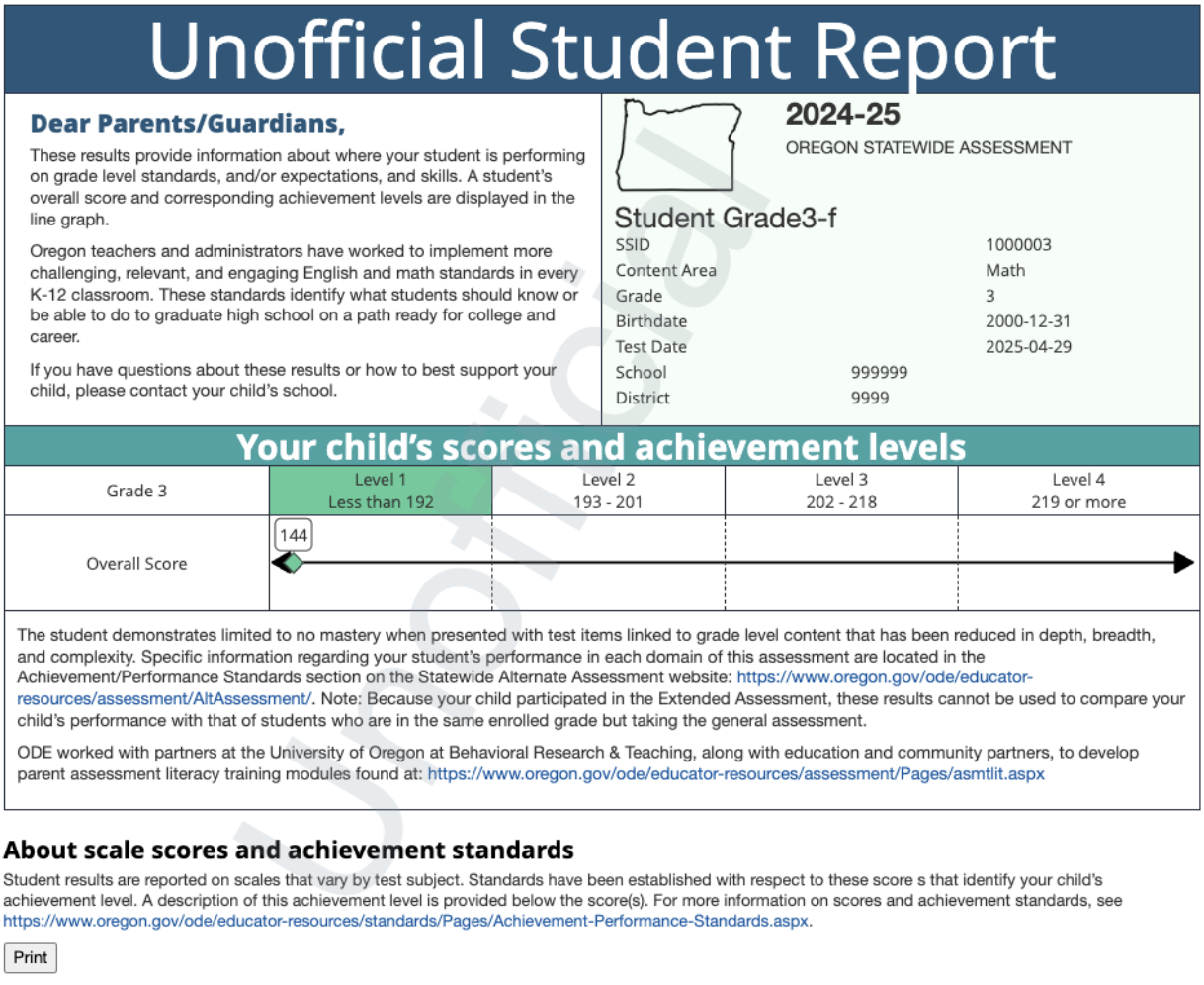
ORExt reports communicate student performance using the reporting scale and achievement-level framework adopted by the state. The illustrative report identifies the student's content area and grade, displays the overall score in relation to the four achievement-level score ranges, and provides a concise performance interpretation. These elements help educators and families connect the reported score with Oregon's AAAS.

Standard errors of measurement and the measurement-aware longitudinal analyses presented in Chapter 4 provide technical context for interpreting scores and year-to-year differences. These analyses supplement the information presented on the ISR and support appropriate interpretation of reported results.

6.4C State Provides Individual Student Reports

The ISR is intended to provide educators and families with a concise summary of student performance relative to Oregon's AAAS. The illustrative report shown below includes student and administration information, the assessed content area and grade, the overall score, the score ranges defining the four achievement levels, and a brief interpretation of the student's performance.

The report can support conversations about current achievement, access to grade-level academic content, instructional priorities, and the need to consider additional classroom evidence. It is not designed to diagnose a disability, prescribe a specific instructional program, serve as the sole basis for an IEP or placement decision, or establish meaningful growth from a small score change.



Unofficial 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.

2024-25
OREGON STATEWIDE ASSESSMENT

Student Grade3-f

SSID	1000003
Content Area	Math
Grade	3
Birthdate	2000-12-31
Test Date	2025-04-29
School	999999
District	9999

Your child's scores and achievement levels

Grade 3	Level 1 Less than 192	Level 2 193 - 201	Level 3 202 - 218	Level 4 219 or more
Overall Score	144			

The student demonstrates limited to no mastery when presented with test items linked to grade level 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.

Print

The displayed example is labeled **Unofficial Student Report** and reflects a 2024–2025 illustrative report for a fictitious student. It demonstrates the basic reporting layout and the types of information communicated through the ISR.

Guidance regarding interpretation of scaled scores, ALDs, and student performance is available in Decision Making Related to Scaled Scores.

6.4D Timely Delivery and Distribution of Student Reports

Following completion of annual scoring, data-quality review, and final reporting procedures, ISRs are made available to districts for distribution to schools, educators, students, and families. District and school personnel are responsible for providing the reports to families and may supplement them with translated explanations or other locally appropriate interpretive support.

Collectively, Oregon's adopted achievement standards, standards-linkage and alignment evidence, score-reporting framework, and technical documentation provide a defensible basis for communicating ORExt achievement results. The reporting system is most appropriately interpreted when scaled scores and achievement levels are considered together with measurement precision, instructional opportunity, classroom evidence, and the limitations described in this chapter.

7 Overall Conclusions

This annual technical report documents the design, administration, scoring, and technical quality of the ORExt for the 2025–2026 administration. Taken together, the evidence supports continued use of the ORExt for statewide accountability and for reporting student achievement relative to Oregon's AAAS, provided that scores are interpreted within the purposes, procedures, and limitations described throughout the report.

The content foundation of the assessment is supported by Oregon's academic content standards, AAAS, EAFs, assessment blueprints, and cumulative standards-linkage and item-alignment studies. The 2024–2025 Standards Linkage and Item Alignment Study provides current evidence that the Essentialized Standards remain linked to Oregon's adopted academic content standards and that newly developed field-test items generally represent the intended essentialized expectations. Operational form assembly, blueprint review, educator participation, field testing, and annual psychometric review provide complementary procedures for maintaining content representation over time.

Operational evidence indicates that the assessment is administered and scored through established statewide procedures supported by annual T&P requirements, standardized administration guidance, secure test delivery, operational monitoring, and technical assistance. The 2025–2026 observation study found very high item-level scoring agreement within the observed paper/pencil administrations while also identifying score interpretation as an area requiring continued training and support. Because the observation sample was limited, these findings provide evidence regarding the consistency of the observed administrations while supporting ongoing statewide monitoring of administration quality.

The psychometric evidence is generally supportive while also identifying specific areas for continued review. Marginal reliability, TIFs, CSEM, classification accuracy and consistency, TCCs, and item-person maps provide complementary evidence regarding score precision, classification quality, expected grade-level ordering, and item-person targeting. Precision is strongest across the portions of the reporting scale where most participating students perform and is comparatively lower at portions of the upper and lower achievement continuum. Some Reading and Writing subscore reliability estimates were lower than the corresponding overall ELA estimates, reflecting the smaller number of items contributing to those subscores.

Item-level analyses also identified results requiring follow-up. Several operational items produced point-measure correlations near zero or below zero, and some item-model results exceeded the annual outfit review criterion. These findings do not invalidate an assessment automatically, but they require documented psychometric and content review before items are retained, revised, replaced, or removed. DIF analyses similarly identify items requiring additional investigation rather than establishing item bias by themselves. Descriptive subgroup score differences are not interpreted as direct evidence of fairness or bias.

Accessibility is incorporated through essentialized academic content, standardized presentation, approved supports, multiple response modes, annual assessor training, and accessible assessment formats. Current comparability evidence for Large Print and Braille relies primarily on common content, controlled adaptation procedures, accessibility review, standardized scoring, and operational monitoring because the number of students using these formats is generally too small to support stable large-sample statistical comparisons. Accessibility conclusions are supported by those procedures and reviews.

The Grades 3–8 ELA and Mathematics vertical scales support interpretation across adjacent grades when observed differences are considered together with the measurement precision of both scores. TCCs provide evidence of expected grade-level ordering, and the measurement-aware longitudinal analysis distinguishes reliable increases, reliable decreases, and differences that are not distinguishable from expected measurement error. These results do not support treating every observed year-to-year difference as meaningful growth. Grade 11 and Science scores are reported on separate scales and are not intended for the same adjacent-grade longitudinal interpretations.

Important limitations remain within the current validity evidence. Response-process evidence is based primarily on expert review, standardized procedures, and observed administration practices rather than direct cognitive-laboratory research with students. Relations-to-other-variables evidence is drawn largely from relationships among OExt scores and does not include a current independent external academic criterion administered to the same students. The internal-structure evidence does not include a direct factor-analytic or comparable dimensionality study. These limitations do not negate the available evidence, but they define the boundaries of the conclusions that can be drawn and identify priorities for future research.

Continuous improvement remains central to the OExt program. Annual priorities include reviewing flagged item results, strengthening measurement at less precise portions of the achievement continuum, monitoring DIF and accessibility, documenting operational form and item dispositions, evaluating the vertical scale, improving score interpretation guidance, and expanding response-process and external validity evidence. These activities connect the findings of the annual technical review to subsequent cycles of item development, field testing, administration, scoring, reporting, and training.

Taken as a whole, the 2025–2026 evidence supports the intended use of OExt scores for statewide accountability and for describing student achievement relative to Oregon's AAAS. The evidence is strongest when the assessment is interpreted as a coordinated system of standards linkage, accessible test design, standardized administration, Rasch-based scoring, conditional measurement precision, classification, and annual technical review. Appropriate use requires that individual scores, achievement levels, subgroup results, and longitudinal differences be interpreted together with their documented measurement uncertainty and the limitations identified in this report.

Appendix A: Abbreviations

The following abbreviations are used in this report. Plural forms are created by adding *s* where appropriate.

Abbreviation	Full Term
AA-AAAS	Alternate Assessment Based on Alternate Academic Achievement Standards
AAAS	Alternate Academic Achievement Standards
ALD	Achievement Level Descriptor
Alt-ELPA	Alternate English Language Proficiency Assessment
Alt-SEED	Alternate Version of the Student Educational Equity Development Survey
AMO	Annual Measurable Objective
ASL	American Sign Language
BH	Benjamini-Hochberg
BRT	Behavioral Research and Teaching
CSEM	Conditional Standard Error of Measurement
DIF	Differential Item Functioning
DIR	Distributed Item Review
DTC	District Test Coordinator
EAF	Essentialized Assessment Framework
ELA	English Language Arts
EsSt	Essentialized Standard
ESSA	Every Student Succeeds Act of 2015
ETS	Educational Testing Service
FAQ	Frequently Asked Questions
HS	High School
IDEA	Individuals with Disabilities Education Act
IEP	Individualized Education Program
IRR	Inter-Rater Reliability
ISR	Individual Student Report
LMS	Learning Management System
NE	Not Estimable
NGSS	Next Generation Science Standards
OAR	Oregon Administrative Rules
ODE	Oregon Department of Education

Abbreviation	Full Term
OMS	Outfit Mean-Square
OREXt	Oregon Extended Assessment
ORora	Oregon Observational Rating Assessment
ORS	Oregon Revised Statutes
OS	Operating System
PCS	Pathways Communication Scale
PII	Personally Identifiable Information
QA	Qualified Assessor
QTC	Qualified Test Coordinator
RDBC	Reduction in Depth, Breadth, and Complexity
RIT	Rasch Unit
SCORE	Select, COde, Reduce, Essentialize
SD	Standard Deviation
SEED	Student Educational Equity Development Survey
T&P	Training and Proficiency
TAM	Test Administration Manual
TCC	Test Characteristic Curve
TIF	Test Information Function

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 2014. *Standards for Educational and Psychological Testing*. American Educational Research Association.

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

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

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.

Magis, David, Sébastien Béland, Francis Tuerlinckx, and Paul De Boeck. 2010. "A General Framework and an r Package for the Detection of Dichotomous Differential Item Functioning." *Behavior Research Methods* 42 (3): 847–62. https://doi.org/10.3758/BRM.42.3.847.

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

R Core Team. 2026. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

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.

Tindal, Gerald, Marilee McDonald, Marick Tedesco, et al. 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.