

The Right Assessment for the Right Purpose

Guidance Document



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>

Colt Gill
 Deputy Superintendent of Public Instruction
Jennifer Patterson
 Assistant Superintendent
Candace Pelt
 Assistant Superintendent

ODE Contributors to The Right Assessment for the Right Purpose <i>Guidance Document</i>		
<i>Dan Farley</i> Director, Assessment	<i>Noelle Gorbett</i> Science Assessment Specialist	<i>Renée LeDoux</i> Administrative Specialist
<i>Jon Wiens</i> Director, Accountability Reporting	<i>Tim Boyd</i> Director, District & School Effectiveness	<i>Carla Wade</i> <i>Interim Director of Data, Operations, & Grants Management</i>
<i>Jamie Rumage</i> Interim Director, Standards & Instructional Support	<i>Laura Foley</i> Director, CTE/STEM/HSS	<i>Deborah Lange</i> Director, Federal Systems
<i>Holly Dalton</i> Early Literacy and Math Specialist	<i>Ben Wolcott</i> English Language Proficiency Assessment Specialist	<i>Carrie Thomas-Beck</i> Dyslexia Specialist
<i>Tony Bertrand</i> English Language Arts and Social Sciences Assessment Specialist	<i>Andrew Byerley</i> Mathematics Assessment Specialist	<i>Shawna Moran</i> Education Specialist
<i>Sarah Soltz</i> Education Specialist	<i>Jennifer Eklund-Smith</i> Education Specialist	<i>Lacey Rhoades</i> Office Support Specialist
<i>Lisa Darnold</i> Education Specialist	<i>Carla Martinez</i> Administrative Specialist	<i>Steve Slater</i> Psychometrician, Accountability Reporting
<i>Tom Tinkler</i> Psychometrician, Accountability Reporting	<i>Josh Rew</i> Psychometrician, Accountability Reporting	<i>Brad Lenhardt</i> Education Specialist
<i>Holly Carter</i> Operations and Policy Analyst	<i>Cristen McLean</i> Operations and Policy Analyst	

All or any part of this document may be photocopied for educational purposes without permission from the Oregon Department of Education and distributed for the cost of reproduction.

TABLE OF CONTENTS

<i>Executive Summary</i>	<i>i</i>
Balanced Assessment Systems	<i>i</i>
Balanced Assessment Includes Summative Assessment	<i>i</i>
Assessment Literacy is the Foundation	<i>iii</i>
<i>Background</i>	<i>1</i>
Toward a Balanced Assessment System	<i>1</i>
The Rationale for a Balanced Approach	<i>1</i>
<i>Components of a Balanced Assessment System</i>	<i>2</i>
Formative Assessment	<i>2</i>
Interim Assessment Systems	<i>2</i>
Summative Assessments	<i>3</i>
<i>ODE’s Plan to Support Districts</i>	<i>4</i>
Implementation Plan	<i>4</i>
<i>Problems of Practice</i>	<i>6</i>
Summative Assessment Parent Communication Resources	<i>6</i>
Assessment and Accountability	<i>7</i>
Appropriate Use of Time Spent Testing	<i>7</i>
Strategies	<i>7</i>
Inflection Point	<i>8</i>
<i>Assessment Literacy</i>	<i>9</i>
Appropriate Uses of Assessment Results	<i>9</i>
Misuses of Assessment Results	<i>9</i>
<i>Understanding Oregon’s Assessment System</i>	<i>11</i>
<i>Appendix A</i>	<i>12</i>
<i>Appendix B</i>	<i>17</i>

EXECUTIVE SUMMARY

BALANCED ASSESSMENT SYSTEMS

This guidance document was developed to assist Oregon educators — from educators in the classroom, to building- and system-level administrators — in effecting behaviors that leverage the **Right Assessment for the Right Purpose**.



The Oregon Department of Education's (ODE's) theory of action:

If Oregon educators increase appropriate uses and decrease inappropriate uses of assessments and assessment data by increasing assessment literacy,
then Oregon educators will make better instructional decisions that increase student learning.

After reading this document and included references, educators will be able to:

1. Describe the differences between formative, interim, and summative assessment practices
2. Describe the differences between foundational skills assessments and standards-based assessments
3. Describe the uses and limitations of each type of assessment at different levels in our education system (student, classroom, school, district, and state)

Assessment in Oregon's public education system must be student-centered, ultimately engaging students in the process of self-assessment. Teachers implement activities that lean on the student as the learner in this model, collecting evidence of learning along the way. When used appropriately, **assessment keeps educators on the cutting edge of student learning.**

BALANCED ASSESSMENT INCLUDES SUMMATIVE ASSESSMENT

Summative assessments are critical to increasing equity and excellence at the state, district, and school levels. ODE uses test results to support defensible and comparable evaluations of our state's educational system, helping community partners to determine whether our instructional programs are meeting the needs of all students, families, and communities. However, **summative assessment results cannot answer all educational questions** at every level within our system (student-classroom-school-district-state). Nor can they inform minute-to-minute, day-to-day, and week-to-week instructional decisions. Over the coming years, Oregon must develop a balanced system that provides all educators with evidence needed to inform their questions about student learning. A world-class educational system needs to emphasize **assessment for learning**, such as formative assessment practices and appropriate use of interim assessment systems, more than **assessment of learning**, such as statewide summative assessments.

ASSESSMENT LITERACY IS THE FOUNDATION

In order to build the foundation for this long-term plan, **communication must be founded in common definitions.** The importance of assessment literacy in our plan cannot be understated.

In the words of W. James Popham, a national expert in educational measurement,

“Now, after a lengthy career, I am convinced that the single *most cost-effective way* to improve our nation’s schools is to increase educator’s assessment literacy” (2018).

This guidance document is the ODE’s initial effort to develop assessment literacy, providing districts with clear communication about assessment and appropriate responses to assessment data as we build a more balanced approach to assessment in Oregon.



The document is a living document that will be informed by future consultation, community partner engagement, and research.

ODE developed this document in consultation with Oregon school districts, Education Service Districts (ESDs), Higher Education measurement experts, and the Oregon Education Association (OEA).

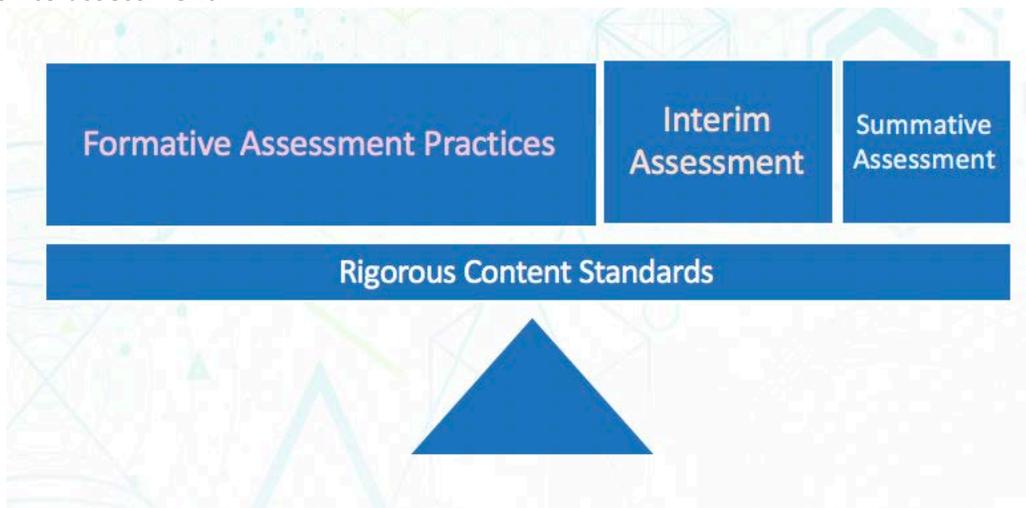
BACKGROUND

TOWARD A BALANCED ASSESSMENT SYSTEM

Over the past five years, community partners have consistently called for ODE to support districts in implementing balanced approaches to assessment, including formative assessment practices, interim assessment systems, and summative assessments (see [A New Path for Oregon: System of Assessment to Empower Meaningful Student Learning](#), [Statewide Assessment Secretary of State Audit](#), and the Every Student Succeeds Act Standards and Accountability Workgroup). The New Path document concludes with a quotation from Chapuis (2009):

Assessment for learning is a gift we give our students. It is a mirror we hold up to show them how far they have come. It is a promise that we will use assessment, not to punish or reward, but to guide them on their learning journey.

ODE is committed to supporting partners in understanding how to use the **Right Assessment for the Right Purpose** in support of equitable decision-making for all of Oregon's students. This balance, conveyed in the graphic below, forms an aspirational target for our statewide approach to assessment:



THE RATIONALE FOR A BALANCED APPROACH

A balanced assessment system is fundamental to our public education system. To date, Oregon's statewide assessment system has been composed almost exclusively of annual summative assessments, which are administered annually and intended to provide an estimate of learning primarily across schools, districts, and the state. While this approach allows comparison of performance across grade levels and/or schools, etc., it does not fully meet the needs of educators in the classroom. Classroom educators need evidence of learning that they can act on within an instructional cycle (best provided by interim assessments), as well as evidence on learning that is occurring in the moment throughout classroom instruction (best from formative assessment practices). In short, Oregon needs a balanced approach to assessment.

Only a balanced approach to assessment can provide evidence of learning that addresses questions at different points of time within the education process and across the educational system, from student, to classroom, to school, to district, to state level determinations.

COMPONENTS OF A BALANCED ASSESSMENT SYSTEM

FORMATIVE ASSESSMENT

There is widespread misunderstanding of formative assessment. Many researchers and vendors speak as if the terms “formative assessment” are synonymous with tests. This is a misunderstanding of the research that supports the practice of formative assessment.

Formative assessment is a rigorous and exacting approach to high impact instructional practices, not a test.

The Formative Assessment for Students and Teachers (FAST) group of experts from the Council of Chief State School Officers (CCSSO) defines [formative assessment](#) as a process of educator-to-student and, more importantly, student-to-student interactions and behaviors that increase student ownership of the learning process. Most importantly, the process helps students to develop skills in self-assessment:

Formative assessment is a planned, ongoing process used by all students and educators during learning and teaching to elicit and use evidence of student learning to improve student understanding of intended disciplinary learning outcomes and support students to become self-directed learners. Effective use of the formative assessment process requires students and educators to integrate and embed the following practices in a collaborative and respectful classroom environment:

- *Clarifying learning goals and success criteria within a broader progression of learning;*
- *Eliciting and analyzing evidence of student thinking;*
- *Engaging in self-assessment and peer feedback;*
- *Providing actionable feedback; and*
- *Using evidence and feedback to move learning forward by adjusting learning strategies, goals, or next instructional steps.*

INTERIM ASSESSMENT SYSTEMS

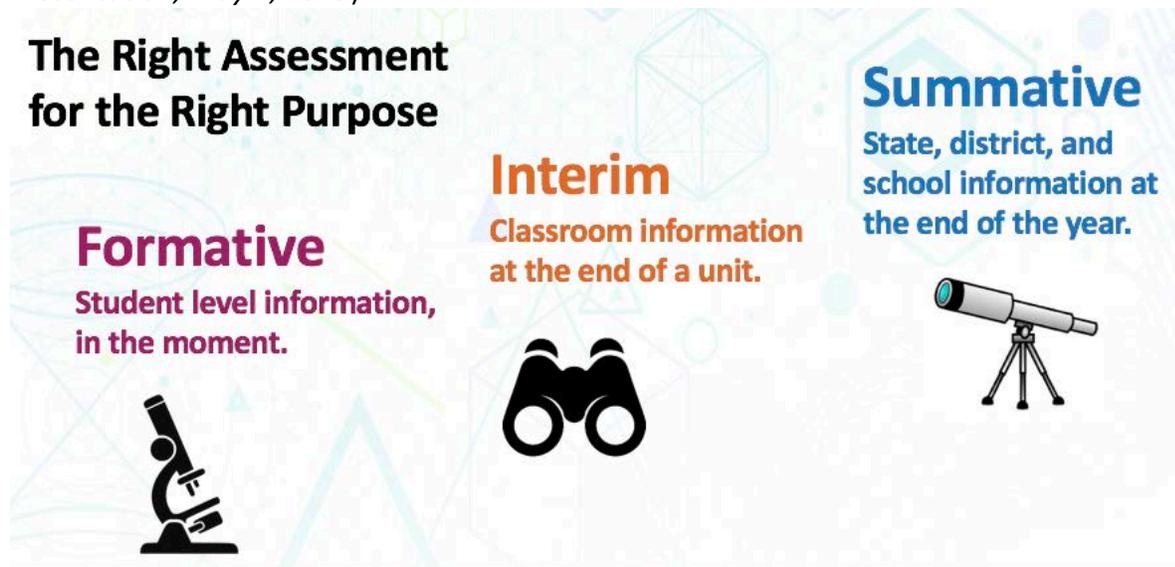
Interim assessments are periodic standards-based assessments that target specific units of content. Because interim assessment results can be used for multiple purposes, they are flexible but also subject to confusion. When administered in a standardized manner at the school or district level, districts will often elect to have students across an entire grade take the same interim assessment at the same time. This allows for a common, within-school or district view of how students are progressing toward understanding Oregon’s content standards. However, some of the most important uses of interim assessments are defined by classroom educators to inform ongoing instructional questions, such as, “Did my students understand that that concept?” and, “What aspects do I probably need to revisit and reframe to support their learning?” Teachers also can leverage the systems in non-standardized ways to see what scaffolds or supports students might need to aid comprehension. Robust interim assessment

systems provide teachers and students with access to the questions to support substantial evaluation of student performance. Excellent interim systems also connect student performance to instruction, outlining not just where student learning is needed, but providing classroom educators with instructional examples of what to work on next.

SUMMATIVE ASSESSMENTS

Summative assessments are standards-based assessments of proficiency administered after instruction has occurred, generally at the end of an academic school year. State-level summative assessments are typically used for school accountability, program evaluation, and to estimate groups of students' achievement levels. Summative assessments are reflective of the full depth, breadth, and complexity of grade level content standards to which they are aligned. They must also reflect the full range of the student population that participates in each grade level assessment, called the full performance continuum.

A balanced approach to assessment includes summative assessment as part of a larger assessment plan as shown in the graphic below (*Source: Ways & Means Subcommittee Presentation, May 1, 2019*).



Formative assessment practices are feedback discussions between teachers and students, and students and peers in the moment. As the graphic conveys, they offer a very close, discrete look at student learning that highlights important details at the student level, much like a microscope. **Formative assessment offers the greatest power for improving student learning.**

Interim assessments may be selected by classroom educators to meet several instructional purposes. Interim assessments are typically used to determine whether students are on track, if they have mastered shorter units or lessons within the school year that have been the focus of a period of instruction. They are likened here to binoculars, as the results are a little farther away from the student, but landscape details and patterns are still noticeable.

Summative assessments, like a telescope, are useful for looking at large systems from afar and identifying patterns that might not be visible at a finer grain of observation. Summative

assessments are administered at the end of the year and designed to provide systems-level information for annual state, district, and school-level decision-making.

It is easy to see why a telescope cannot be used to see cells, nor can a microscope be used to determine how many moons Saturn has. We would not use binoculars for either of those purposes. Similarly, each assessment serves a specific purpose and should be used only for that purpose.

ODE'S PLAN TO SUPPORT DISTRICTS

IMPLEMENTATION PLAN

ODE will continue to develop our [student-centered assessment system](#) by supporting the development of formative practices in classrooms across Oregon and providing an interim assessment system statewide. We will shrink the footprint of our summative assessment by decreasing costs, decreasing time spent testing (such as increasing efficiencies in test scheduling and administration), and clearly communicating about the purposes of each type of assessment and appropriate uses of the results from each assessment approach. Our previous efforts in this area include:

2014-15 [Berkeley Evaluation and Assessment Research Center](#) (BEAR): Developing, Scoring, and Evaluating Interim assessments – Though the legislature funded interim assessments for 2014-15, ODE was precluded from procuring resources at the statewide level. This resulted in great variance in district decision-making and limited our options in supporting this work to providing guidance and evaluation frameworks. ODE provided the resources in the BEAR project to districts to help them make better decisions in selecting interim assessment systems.

2016-17 [Oregon Formative Assessment for Students and Teachers](#) (OFAST) project – ODE worked with Dr. Margaret Heritage and nine Oregon districts to develop a comprehensive course containing six training modules on formative assessment practices that is available to all via a Creative Commons licensing agreement (CC-BY-NC-SA). The course was designed to be implemented through Communities of Practice (CoPs) in 90-minute, monthly sessions with substantial support from researchers. We are currently working on re-scaling the resources from this work such that they can be successfully implemented within one year in a Professional Learning Community (PLC) structure, with weekly meetings and competing priorities for teacher time as part of the implementation lens.

2016-17 [Performance Assessment Development Sites](#) (PADS) – ODE worked with seven districts on performance assessment development. The PADS project was about building and assessing deeper learning. Educators drew upon performance assessments from the Performance Assessment Resource Bank (<https://www.performanceassessmentresourcebank.org/>) or other performance assessment banks that were vetted by Center for Collaborative Education and publicly

available. Performance assessments were available in math, English, science, and social science educators in elementary through high school grades.

2016-17 & 2018-19 K-3 Formative Assessment Resources was a project led by North Carolina Department of Instruction that ODE worked on with three districts. The project developed learning progressions that could be used to evaluate student progress toward outcomes in several areas, including, but not limited to, literacy, numeracy, perseverance, gross motor skills, and social-emotional learning. Resources from this project will be posted on the Assessment Team website by fall 2019.

2016-17 [Building Educator Assessment Literacy](#) (BEAL) – Guides participants through the process used to score performance tasks from our summative English language arts and mathematics assessments. Resources provided include workshop instructions and a sample agenda, Handout Books, presentation slides, facilitator’s guide, and activities and evaluation of said activities.

2016-17 English Language Proficiency Standards (See [Module 1](#))– the ELPA21 Consortium, one of Oregon’s non-profit assessment partners, provides seven training modules for understanding our English language proficiency (ELP) standards, how they are assessed, and how they can be woven into PLCs and instructional processes using formative assessment practices.

Ongoing Performance Assessment Resources – in support of Oregon’s local performance assessment requirement, we have developed and make available several performance assessments and resources online. A guidance document called, Quality Performance Assessment: A Guide for Schools and Districts, elaborates the tenets of the process, including alignment to standards, design, and interpretation of results to support equity and excellent outcomes for all students. If you would like a copy of this sent to you, email ode.asmt-team@ode.state.or.us.

2018-19 Smarter Balanced Interim/Benchmark and Digital Library Pilot – Worked with three schools in North Santiam, Eugene 4J, and Grants Pass on a pilot of the SBAC interim system in order to learn how Oregon educators can most effectively use these assessment tools and resources to inform instruction and gauge student learning periodically throughout the school year. While the Legislature did not fund this initiative for the 2019-21 biennium, the pilot provides us with information about what resources and procedures are needed for a possible statewide implementation in future biennia.

2018-19 Smarter Balanced Interim/Benchmark District Purchase – Several districts/charter schools independently purchased the SBAC interim system. We supported their training needs in collaboration with SBAC and AIR in these applications.

ODE continues to support district implementation of formative assessment practices through ESDs and as part of the District and School Improvement process. We’re committed to finding ways of supporting districts in this arena. The following efforts outline current and future projects ODE is conducting to help foster increases in assessment literacy.

2018-19 [Starting Smarter Website](#) – The Assessment Team worked with SBAC to develop a website with content that parents and other community partners can use to interpret individual student score reports. The website includes a section that connects the student’s scale score to an achievement level descriptor so parents can interpret what the scores they receive mean. Sample score reports and where to find information that is presented and how to interpret that information is also provided, as well as information about our content standards, student accessibility, and parent roadmaps that explain our ELA and mathematics standards and how parents can support their child at home at all grade levels K-12 in both English and Spanish.

2019-20 SBAC Interim/Benchmark and Digital Library – Working with our vendors to provide resources for districts who wish to make school- or district-wide purchases.

2019-20 Parent Assessment Literacy Training Module – Working with researchers at the University of Oregon to develop a set of training modules to increase parental understanding of our statewide assessment system and local district assessments. We’re partnering with FACT Oregon and the Oregon Parent Teacher Association in the development process. These resources include an animated Parent Training Module that answers the following five questions that parents have consistently regarded as critical according to a literature review conducted by colleagues at Northwest Education Association (NWEA):

1. How are Oregon’s assessments fair?
2. What types of assessments does my child take and why?
3. How did my child do on assessments?
4. How do I support learning at home?
5. How can I get help if I have questions?

All materials will be made available on ODE’s website in the fall of 2019 as [Creative Commons](#) resources that allow others to modify and build upon the resources as long as attribution is made and the materials are not used for commercial purposes. Such materials must also be licensed in the same manner everywhere they are shared (CC-BY-NC-SA). The parent training modules are Phase 1 of a three phase project. Phase 2 will include the development of training modules for educators, while Phase 3 will target school and district administrators.

PROBLEMS OF PRACTICE

SUMMATIVE ASSESSMENT PARENT COMMUNICATION RESOURCES

ODE operates within the provisions of both [ORS 329.479](#) and federal accountability rules around participation, and our guidance to districts reflects this. ODE promotes and honors the role of adult students and parents in making decisions regarding participation in Oregon’s statewide assessments. In the fall of 2018 we added language to the [Test Administration Manual](#) to better support districts in ensuring that parents and adult students are properly

notified of and understand their decision-making role regarding participation in state assessments. ODE's guidance to districts is that there be no persuasion, positive or negative, aimed at convincing families or adult students to choose to either opt out or participate. Instead, the focus of ODE's guidance directs districts to provide the information required under ORS 329.479 so that families and adult students can then decide for themselves. We are also developing the assessment literacy resources listed above to help ensure that parents have access to the information they need to make a fully-informed decision about their child's participation in statewide assessments.

ASSESSMENT AND ACCOUNTABILITY

ODE is commonly asked if we have considered using formative evidence or interim assessment results in lieu of our summative assessments within our accountability system (sometimes referred to as "innovative" assessment designs). The short answer is, we use each assessment type for the purpose it is best designed to fulfill. Using formative evidence or interim assessment results for accountability raises several concerns:

1. Instruments used for statewide accountability must be administered in a standardized manner in order to be comparable. Comparability is a strength of summative assessments.
2. Using interim assessments for accountability accidentally undermines local responsiveness to student needs:
 - a. Basing accountability on fall, winter, and spring interim assessments would result in a *de facto* scope and sequence for curricula and reduce local flexibility to meet student needs.
 - b. Educator and student access to the interim items would be prohibited because of test security concerns. This would in turn limit the use of interims to increase understanding of the standards, guide instruction, and pinpoint students' strengths and weaknesses.
 - c. Educators would no longer be able to independently select assessments that match the content that students have recently digested in the classroom. Nor would they be able to conduct non-standardized administrations in search of specific data (such as leading the entire class in a Performance Task activity to test comprehension of a targeted skill).
3. Formative assessment is a system of high-leverage instructional practices that empower students to take charge of their own learning. It is not a test. The evidence of learning collected is often student-specific, making comparison impossible. Formative assessment practices cannot be used within an accountability system and simultaneously maintain their intended instructional purpose.

APPROPRIATE USE OF TIME SPENT TESTING

Strategies

Student instructional time should be maximized throughout the school year. When used appropriately, assessment enhances instruction. Conversely, inappropriate uses of assessment may actually interfere with instructional opportunities, as in the following examples:

- Test administration schedules which lead to inefficient use of instructional time.

- Delaying or interrupting the instruction of students who are not being assessed while their peers are taking an assessment.
- Omitting instructional opportunities for students who have completed their assessments while waiting for their peers to finish.

Although obstacles such as test administration constraints, test security management, coordinating staff for test administration, and computer lab scheduling may be difficult to negotiate, all students—both those who are testing and those who finish early—benefit from intentional pre-planning. Some suggestions include:

- Creating an assessment plan based on an individual student’s needs and opportunities.
- Maximizing efficiency in test administration. For example, schedules should factor in time to move classrooms, log into the system, and complete testing.
- Considering the student attention span, which generally increases with age.
- Enabling shared test administration and instructional responsibilities through team teaching or combining grade level classrooms.
- Providing computer-based instruction to students who have completed the assessment while other students are testing, taking necessary precautions to avoid compromising the validity of tests still in progress. Examples include sufficient spacing, placement, and visual barriers.
- Arranging volunteers to assist classrooms (consistent with district policies and best practices of high quality instruction).

Inflection Point

ODE staff researched the impact of testing time upon student outcomes in order to determine whether we could provide empirically-based observations about test performance patterns as they relate to the time spent testing. We evaluated student performance in relation to the time they are taking to complete our assessments and noticed the following patterns, which held across all tested grade levels. We are calling the time point identified an *inflection point*. Up to this inflection point, additional time spent testing results in substantive increases in performance. The *Inflection Point* is the point in time after which most students experience only minimal gains in performance, in the form of score points, for additional time spent testing.

Information on the inflection point is intended to be used as a guideline. Students should never be told to stop testing, nor summarily dismissed from testing if they are meaningfully engaged in the assessment process. Some students, such as students with disabilities or students who are English learners, have different inflection points than the average listed below. Their scores continue to grow with time spent testing, so long as they remain engaged with this test. Inflection points are provided to guide discussions about time spent testing. Districts concerned about time spent testing should have these discussions and develop attendant policies well in advance of test administration. Educators should also work with students and families to ensure that the time that they are investing in test participation is time wisely expended.

Content Area	Average Time for Test Administration	Inflection Point
English language arts (Combined CAT & PT)	~ 190 Minutes (3 hours, 10 minutes)	~ 300 Minutes (5 hours)
Mathematics (Combined CAT & PT)	~ 125 Minutes (2 hours, 5 minutes)	~ 200 Minutes (3 hours, 20 minutes)
Science	*Pending	*Pending

*Our summative science assessment results will not be finalized until fall 2019. We plan to publish inflection point guidance for science before the 2020-21 school year.

ASSESSMENT LITERACY

APPROPRIATE USES OF ASSESSMENT RESULTS

Assessment literacy means differentiating between appropriate and inappropriate uses of different types of assessments and assessment results. Assessment literate educators know how, when, and why to assess student learning. They are able to identify, select, and/or create appropriate, efficient, and precise assessments that engage students in demonstrating their knowledge and abilities relative to targeted learning goals. Educators who have developed assessment literacy evaluate the quality of the evidence that has been collected. They accurately analyze, interpret, and use data, from both quantitative and qualitative sources, to help drive instruction and increase student learning. These principles are encapsulated in the following three tenets:

1. Assessments are typically developed for a singular purpose and should only be used for that purpose.
2. Assessments should generally not be used for any purposes for which the test developer or user does not have sufficient validity evidence.
3. No single assessment result should be used in isolation to make high-stakes decisions at the individual (i.e., student) level.

MISUSES OF ASSESSMENT RESULTS

The following examples illustrate practices that are inconsistent with our three foundational measurement tenets. Each example is followed by a suggestion of an appropriate alternative practice.

- Using statewide assessment results at the classroom, school, or district levels as a central component in a teacher evaluation system (see [American Statistical Association](#)). **INSTEAD** – Educators can work from the understanding that there are many factors that influence student learning and academic achievement, and identifying the unique contributions of an individual educator is difficult at best. Cohort effects, out-of-school opportunities, systems-level impacts (as opposed to educator-level impacts), and even previous educator effects all differentially impact students. Use of statewide assessment results in teacher evaluation systems is an untenable practice if they play more than a supplemental role. This is particularly true given that assessments are administered on an annual basis, and changes in test scores are thus evaluated over a 12-month period, but students generally do not attend schools during the summer months. This makes it

impossible to disentangle differences between teachers from differences in out-of-school opportunities. Educators, principals, and other administrators can review statewide assessment results as one part of a larger body of evidence to help inform discussions about the quality of educational experiences that students receive, but it is critical that they be mindful of the inherent limitations of the data, even when collected across multiple years.

- Using the results from a single assessment, such as a statewide summative assessment, as a “gatekeeper” to exclude a student from any district or school programs, such as talented and gifted, International Baccalaureate, or Advanced Placement programs.
INSTEAD - Such decisions should be informed by multiple measures to reduce concerns related to measurement error, which is present in all forms of psychological and educational testing. Generally, the tests should be administered via differing *modes*. When students are assessed in only one way, or one mode, their performance may reflect the mode of assessment rather than their ability.
- Overuse and misuse of foundational skills assessments.
INSTEAD – Educators can work from an understanding that implementation of these types of assessments can be important for students who are struggling to meet grade level expectations, particularly in the early grades, and are also useful for early identification of support needs. Administration of such assessments is not always an appropriate use of instructional time, however, particularly in situations where students who are at or above expectations are being assessed more than three times per year, or when test administration is not scheduled efficiently. Educators should also be aware that there are important limitations to the level at which foundational skills assessments can inform instruction. They do not accurately summarize performance relative to the full depth, breadth, and complexity of our state content standards and benchmarks. For this reason, they can lead to false impressions that students are doing quite well because they are connected to content that has limited scope and complexity.
- Confusion about the lack of consistency in results between our statewide summative assessments and foundational skills assessments.
INSTEAD – Educators can work from the understanding that our statewide summative assessments are reliable measures of student learning of the full depth, breadth, and complexity of our standards. They include performance demands across all levels of cognitive complexity [Depth of Knowledge (DOK) 1 through 4; note: these categories do not apply directly to our new NGSS-aligned science assessment].
 - DOK 1 – Recall and reproduction
 - DOK 2 – Basic skills and concepts
 - DOK 3 – Strategic thinking and reasoning
 - DOK 4 – Extended thinking

Foundational skills assessments are, by design, composed of narrower content and lower-level cognitive demand, typically composed of items at DOK 1 and 2. Thus, when educators identify a discrepancy between foundational skills assessments and Oregon’s summative assessments, they can attribute this to the lack of full alignment between

foundational skills assessments and our complex and challenging content standards (to which our summative assessments are aligned).

UNDERSTANDING OREGON'S ASSESSMENT SYSTEM

The first step in building assessment literacy is to develop an understanding of the purpose of Oregon's Statewide Assessment System (OSAS). Common questions about the OSAS are included in **Appendix A**. In order to further develop assessment literacy, educators need to agree to a common set of terms employed in discussions around student assessments in the K-12 public school context in Oregon. ODE has worked with state, district, and higher education partners to develop an Assessment Definitions document in **Appendix B**. The document cites source definitions where available. Where no source definition was available, definitions were developed and vetted. One of the most critical learning opportunities that our teams garnered from this work was accurately describing the differences between foundational skills assessments, which are reflective of a limited set of standards at lower levels of cognitive complexity, and standards-based assessments, which are designed to be reflective of the full depth, breadth, and complexity of our adopted content standards.

APPENDIX A

Rationale and Questions about the Oregon Statewide Assessment System (OSAS)

RATIONALE

Our statewide assessment system is a set of tools that ODE uses to identify opportunity gaps and patterns in achievement and underachievement that can be used to design more efficient approaches to district and school improvement efforts at the systems level. We aim to ensure equitable opportunity and excellent outcomes for all of Oregon's students, while also adhering to our professional regulatory obligations. This section frames why the OSAS exists, why it is important for all of Oregon's students, and why ODE is dedicated to the system as part of our commitment to equity. We address three critical questions:

1. Why is it important to assess learning for all of Oregon's students?
2. Why is it important to use standardized administration and instruments?
3. Why is it important to publicly report the results of these assessments at the school, district, and state levels?

WHY IS IT IMPORTANT TO ASSESS LEARNING FOR ALL OF OREGON'S STUDENTS?

At heart, Oregon's statewide assessment system and student participation expectations are grounded in equity. ODE is committed to equity and sees our statewide assessments as important tools in the work of ensuring that all of Oregon's students experience excellent educational outcomes. Our foundational assumption is that every Oregon student deserves a meaningful opportunity to learn, including students of color, students from culturally and linguistically diverse backgrounds, students who identify as lesbian, gay, bi-sexual, trans-sexual, or queer (LBGTQ), students experiencing poverty, and students with disabilities. We regard learning to be a natural human state; when learning is not occurring, we search for environmental factors that are impeding learning and support districts in removing those hurdles.

Pursuant to our dedication to equity and our Constitutional commitment to provide an excellent public education for all of Oregon's students (See [Article VIII](#)), our statewide assessments serve as a quality assurance process that gauges how effectively Oregon's education system is preparing students to graduate from high school ready for college, career, and civic life. Statewide assessment helps to ensure that all students are provided with a consistent and viable education, first by maintaining high standards through a rigorous assessment system and second by ensuring that student participation and performance are publicly reported. The system has been evaluated within the most rigorous validity frameworks that is available, the United States Peer Review of Title 1 Assessment Systems process, and has been deemed to meet or exceed all but a few minor requirements, receiving a *Substantially Meets* designation (see [January 28, 2019 USEd Peer Review letter](#)).

When large groups of students do not participate, it limits educators' and policy makers' ability to get a clear and accurate picture of the health of Oregon's education system. Both geographical gaps (e.g., at the school or district level) and population gaps (e.g., demographic groups) cloud the view of how all districts are doing at serving all students, evaluate effectiveness of supports, and make decisions around resource allocation.

In addition, Oregon law requires that each district administer the state tests to all students enrolled in a public school located within the district's boundaries, including public charter schools and alternative programs. Federal requirements demand that 95% of all students

participate, both within each district and across the state, including 95% of students who are English Learners, students with individualized education programs (IEPs), and students with the most significant cognitive disabilities who take an alternate assessment based on alternate academic achievement standards.

Specific to Oregon's English Learners, Oregon's English Language Proficiency Assessment (ELPA) assesses English Learners' English proficiency in the language domains of listening, speaking, reading, and writing. Under federal law, 100% of students eligible for English language development services are expected to participate annually in the state's ELPA. Educators and policy makers at the local, state, and federal level use the results of the ELPA to evaluate the effectiveness of services provided to support English Learners in acquiring English language proficiency so that they can successfully access and engage with academic instruction, meet the state standards, and graduate college- and career-ready.

Educators must remain mindful that many students have been excluded from our school buildings and our educational processes historically. For example, there are numerous stories of students with disabilities being asked not to come to school during testing prior to the implementation of the No Child Left Behind Act. Practices like this may occur in areas where we are not ensuring that at least 95% of our students are participating in the statewide assessments.

WHY IS IT IMPORTANT TO USE STANDARDIZED ADMINISTRATION AND INSTRUMENTS?

The *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014) define standardization of test administration as, "maintaining consistent testing environment and conducting tests according to detailed rules and specifications, so that testing conditions are the same for all test takers on the same and multiple occasions (p.224)." Standardization with regard to scoring ensures that student test responses are scored in the same manner, a manner which is predetermined and systematically applied.

Standardization is what makes assessment results comparable. We cannot imagine a situation where it is not important to give students equitable opportunities to demonstrate what they know and can do. We cannot imagine a situation where it is not important that educators and the public are able to compare student performance to a defined level of mastery.

Let's consider a simple example. You are taking a math assessment. The results will count toward your course grade that you need to receive a diploma. The instructor determines that your test will cover the final chapter in the unit, which is the most challenging. In addition, the professor determines that you will need to get a 95% to get an A. Your classmates, however, are taking a different test that covers only the first chapter in the unit, which is easier. They only need to get a 90% to get an A. These are examples of an assessment administration and scoring procedure that are not standardized. It is this fact that makes them unfair.

ODE is aware that the words "standardized assessment" can be emotionally and politically charged for some community partners; however, the stories shared with us typically have more to do with misuses of the test results than they have to do with standardized assessments. We certainly have more work to do in building understanding about assessment literacy, including

test types, purposes, and appropriate uses of the data. For example, many community partners are not aware that all of the assessments in the OSAS, Oregon’s “standardized assessments,” are “Oregonized” assessments, that is, assessments developed in conjunction with Oregon educators (teachers, TOSAs, and administrators) to ensure that the assessments are truly aligned with classroom practices throughout Oregon as well as meeting rigorous psychometric standards set forth by the AERA, APA, and NCME *Standards*.

WHY IS IT IMPORTANT TO PUBLICLY REPORT THE RESULTS OF THESE ASSESSMENTS AT THE SCHOOL, DISTRICT, AND STATE LEVELS?

Publicly reporting student test results is important because it creates transparency and builds trust between the state and the public. The fact that academic outcomes are made available to the public holds state and local agencies accountable for equitably providing all students in their communities with a high quality educational experience, requiring us to prove that we’re really providing the services that we claim. It also galvanizes us to act when the data reveal that not all students have an equitable opportunity to learn. It is important to report not just overall averages, as these statistics can effectively hide how specific student groups, such as students from diverse backgrounds or students with exceptionalities, are doing. ODE’s commitment to equity requires that we must draw attention and scrutiny to how we are serving our students from historically underserved backgrounds. At the local level, the results of the state tests provide educators and administrators with information about what educational approaches are working and where additional resources are needed. At the state level, the state test results help policy makers evaluate the effectiveness of state-supported programs and identify schools and districts that need additional supports to ensure more students are meeting the state’s standards for what students need to know and be able to do.

Likewise, public reporting is an important ingredient in a working democracy, as it gives the citizenry important information they may use to vote for representatives and helps them understand what is being done for (or to) them and their children.

WHY IS IT APPROPRIATE TO USE SUMMATIVE ASSESSMENTS AS EVIDENCE FOR THE DETERMINATION OF SCHOOL AND DISTRICT ACCOUNTABILITY RATINGS IF THOSE ASSESSMENTS SHOULDN’T BE USED FOR EDUCATION DECISIONS REGARDING INDIVIDUAL STUDENTS?

The foundation of this response hinges on a basic understanding of reliability, which increases with the number of test items. An assessment’s reliability increases directly in proportion to the number of items and the amount of information each item captures. An assessment that was reliable for high-stakes decisions at the student level would be composed of more items than educators and parents would be willing to allow for, given the impact upon instructional time. However, aggregated summative results from shorter tests demonstrate greater reliability after aggregation. In general, state results are the most reliable, followed by district, then school, then classroom, then student. Consider this example: a student may answer only a couple questions that are aligned to a given standard; we’re not sure that the student does/doesn’t understand the standard based on that limited evidence. If 30-40 students answer those same two questions and a pattern is evident, then we have a sufficient reason to make instructional changes.

Results from summative assessments are designed primarily to inform program and systems level decisions at the school, district, and state levels. They are less useful at the classroom or student level, not only due to reliability constraints, but also because of the fact that they can

influence decisions only toward the end of a school year (and for this reason are best-situated to inform instructional practices for the following year). For individual students, summative assessment results should only be used to supplement classroom-based evidence or as a general guide for instructional planning on an annual basis.

WHY IS IT APPROPRIATE TO USE SUMMATIVE ASSESSMENTS AS EVIDENCE FOR THE ESSENTIAL SKILLS IF SUMMATIVE ASSESSMENTS SHOULDN'T BE USED FOR EDUCATION DECISIONS REGARDING INDIVIDUAL STUDENTS?

There are four ways that students can demonstrate proficiency relative to Oregon's Essential Skills: 1) statewide summative assessment; 2) other standardized assessment; 3) work samples; and, 4) local assessment option. Statewide summative assessments are currently used by the majority of Oregon's students to meet Essential Skills requirements in reading, writing, and mathematics. However, students have three other options. Providing multiple opportunities in multiple assessment formats reduces the risk that students may be denied a diploma despite having acquired sufficient proficiency in the required Essential Skills.

It must also be noted that graduation requirements are not limited to assessment of Essential Skills. In addition to the assessment requirements, students must also provide evidence of their skills as described by courses, grades, credits, and career-based projects. Consultation with practitioners demonstrates that Essential Skills are rarely the singular barrier to graduation for most students. Rather, there are additional complexities that typically interfere with credit attainment.

APPENDIX B

Assessment Definitions

Oregon Department of Education (ODE) K-12 Student Assessment Practices and Test Types 2019-20

Purpose

This document was developed to set a foundation for discussions around assessment literacy in Oregon, to help ensure that we are speaking the same language and using common, shared definitions of terms that have caused great confusion due to inconsistent usage across research, education, and vendor contexts. ODE is committed to developing understanding of what it means to use **The Right Assessment for the Right Purpose** in all educational contexts as part of our goal of helping districts implement balanced assessment systems. Developing assessment literacy is the first step in that process. This document will evolve over time, but it is hoped that it sets the stage for better understanding of our assessment system as experienced by district users.

Terms

Assessments versus tests—Most practitioners use the terms assessment and test synonymously. However, assessment casts a much wider net and can involve any process where evidence is systematically collected and reviewed in order to guide decision-making (e.g., observations, rating scales, performance, tests). In addition, as noted below, formative assessment is not a test in any way, but a planned, systematic, student-driven process that involves a great deal of self and peer assessment, but does not include tests. In this document, assessment is generally referred to as the process of evidence collection and review, whereas tests are formal instruments that are developed for a singular purpose or use.

Foundational Assumptions

Data from tests should only be used for intended purposes. Validity evidence provided by test developers, including test vendors, must document that the results can be used for all students (i.e., students of color, students with disabilities, students who are English learners) or identify the groups of students for whom the test is not an appropriate tool. Vendors should clearly demarcate intended uses for their tests and also discourage use of test results for purposes for which the test is not designed. In addition, it is not appropriate for a single test result to be used for a high stakes decision for a student; high stakes decisions require multiple sources of information. Use of the assessment practices or tests listed below for any purposes outside of those listed within the “How are Data Used Appropriately?” column is inappropriate and not supported by the respective test designs or relevant validity evidence.

Reliability, Validity, and Fairness

There are three foundational concepts that all educators must understand in any effort aimed at increasing assessment literacy: 1) reliability, 2) validity, and 3) fairness. The Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014), the definitive source for all measurement terms and practices in the United States, defines those terms in the following way:

Reliability “The degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable and consistent for an individual test taker; the degree to which scores are free of random errors of measurement for a given group” (p. 222);

Validity “The degree to which accumulated evidence and theory support a specific interpretation of test scores for a given use of a test. If multiple interpretations of a test score for different uses are intended, validity evidence for each interpretation is needed” (p. 225);

Fairness “The validity of test score interpretations of intended use(s) for individuals from all relevant subgroups. A test that is fair minimizes the construct-irrelevant variance associated with individual characteristics and testing contexts that otherwise would compromise the validity of scores for some individuals” (p. 219).

It is important to note that all of these concepts are questions of degree, requiring expertise and professional judgment in order to determine sufficiency. It is important to leverage local expertise of measurement experts when making difficult interpretation decisions.

Formative Assessment

The Formative Assessment for Students and Teachers (FAST) group of experts from the Council of Chief State School Officers (CCSSO) defines [formative assessment](#) as a process of educator to peer and, more importantly, peer-to-peer interactions and behaviors that lead to increased student ownership of the learning process:

Formative assessment is a planned, ongoing process used by all students and educators during learning and teaching to elicit and use evidence of student learning to improve student understanding of intended disciplinary learning outcomes and support students to become self-directed learners. Effective use of the formative assessment process requires students and educators to integrate and embed the following practices in a collaborative and respectful classroom environment:

- Clarifying learning goals and success criteria within a broader progression of learning;
- Eliciting and analyzing evidence of student thinking;
- Engaging in self-assessment and peer feedback;
- Providing actionable feedback; and
- Using evidence and feedback to move learning forward by adjusting learning strategies, goals, or next instructional steps.

Test scaling and reporting

Norm-referenced Scales --Provide information about how each student performs compared to other students

Imagine a print out with a learning target at the top and the names of student in a numbered list (ranked)—this test tells you how student performance compares to others. This kind of test does not tell you whether a student met a learning target but rather compared to other students how the student did. The comparison is always relative to the performance of the “norm group,” which is the group of students who participated when the test was developed (selected such that they match population demographics).

Criterion-referenced Scales –Provide information about how each student performs in comparison to predetermined criteria (in standards-based contexts, the criteria are defined by achievement levels that elaborate student knowledge and skills and different performance levels; in informal diagnostic tests contexts, the term refers to skills-based milestones). Imagine a print out with a learning target at the top and next to each student’s name is information about whether the student fully mastered this learning target, partially mastered it, or did not master it—this kind of test tells you whether students know the content and/or skills regardless of whether their peers have demonstrated such knowledge. The comparison is not relative to other students, but to the level of performance set as passing or meeting a critical milestone.

Note: many tests compare performance to other students with normed scales (e.g., percentiles) and to a criterion (e.g., an achievement level). Some confusion results when vendors use the phrase “criterion-referenced” to refer to a normative benchmark, such as meeting the 75th percentile as a “criterion.” This is not an appropriate label, as a criterion-referenced scale must refer to content.

Standardization

Standardized tests— The Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014) define standardization of test administration as, “maintaining consistent testing environment and conducting tests according to detailed rules and specifications, so that testing conditions are the same for all test takers on the same and multiple occasions (p.224).” Standardization with regard to scoring ensures that student test responses are scored in the same manner, a manner which is predetermined and systematically applied. This does not mean that every student’s testing experience is exactly the same (because there are lots of ways the testing experience can be individualized in ways that support the student showing what he or she knows without giving the student an unfair advantage—for example, a student who has a reading processing disability may be allowed to have directions or prompts on a mathematics examination read aloud to them, as the content that is being measured is math, not reading).

Non-standardized tests— In some cases, tests are administered or scored in non-standardized conditions, and this makes the results impossible to compare. The results can be used to understand the specific student’s performance at that singular point in time that the evidence is collected. Collecting information about student knowledge and skills in non-standardized situations is student-specific and thus the results from the test cannot be aggregated or combined together because the way each student was measured is different. From all the tests and practices addressed here, formative assessment practices are the only type that are appropriately administered in a non-standardized manner.

Standards-Based Tests

Standards-based tests provide information about how well an individual student is doing in terms of the full depth, breadth, and complexity of the standards. Depth is typically defined by how much knowledge a student has about a singular standard (topic/idea). Breadth is a concept that defines the range of the different types of knowledge and skills a student needs to demonstrate understanding of the standards. Complexity is defined as the level of cognitive demand that is required by the standards (for educators, think Bloom’s Taxonomy). These levels in testing environments are typically defined by four levels: Level 1) Recall & Reproduction; 2) Basic Skills & Concepts; 3) Strategic Thinking and Reasoning; and, 4) Extended Thinking (Webb’s Model). Standards-based tests are used for a variety of purposes at the state, school, educator team, classroom, and individual student level and are an integral aspect of the standards-based reform movement.

Because instruction and learning experiences are based on state standards—and formative assessment practices are integrated into instruction—formative assessment enables educators and students to co-regulate learning so that they are partners in both building knowledge and skills and demonstrating knowledge and skills. By measuring specific state standards, the interim, classroom summative, and performance tasks let you know whether, after a period of instruction, students are able to demonstrate the knowledge and skills they were taught. This information should be used for reflection by both educator in terms of instruction and by the student in terms of effort and engagement.

Foundational Skill Tests

Foundational skill tests provide information about how well students are doing in developing the basic skills that provide a foundation for learning, including literacy, numeracy, motor skills, and behavior. The tests that are used for screening and progress monitoring are designed to determine how well an individual student is performing relative to foundational skill standards or expectations for that grade and time of year and can also be used to compare performance to other students. By looking at a large population of students and how those students do in reference to the norm group, these tests help determine which students may need additional attention or support. In addition, these test results, when looking across groups of students at the class, grade, school and district levels, can be utilized to help determine a system’s basic health. Foundational skills tests are typically predictive of future achievement, as well, as they measure critical skills that are commonly taught and learned in Oregon classrooms.

Foundational skill tests are not designed to tell you how well an individual student performs on more complex grade-level standards, however. This is because the tests focus on foundational skills (e.g., decoding, computation); they focus on the critical foundational skills which, in turn, enable a student to approach more complex learning tasks. They can help indicate whether a student is likely to be able to approach more rigorous tasks with ease or with difficulty. Because the tests do not include the full range of cognitive complexity of the content standards, they cannot be considered as aligned to grade-level standards outside of skills-based Reading Foundations standards at the early grades, and thus, for instructional planning, results from standards-based tests must be drawn upon to provide a complete picture.

The following tables convey the variety of assessment practices or test types in Oregon, as well as descriptions and examples. There are supplemental columns that convey the population of students expected to participate, the timing relevant to the situation, the purpose of each assessment, who is involved in testing, whether the assessment is standardized or not, whether the relevant scale is norm- or criterion-referenced, who the typical data users are, and, most importantly, how the data should be used (NOTE: uses beyond those listed are not supported by sufficient validity evidence and should be avoided). The information is divided into two major categories; assessments that are aligned to content standards and assessments that align to foundational skills.

Non-Standardized Assessment Practices

Assessment Practice	Description and Examples (Not All-Inclusive)	Who is Tested?	When?	For What Purpose?	Who Is Involved in the Testing?	Standardized?	Norm-referenced or Criterion-referenced?	Who Typically Interacts with the Data?	How are Data Used Appropriately?
Formative Assessment Practices	Formative assessment practices include: Clear learning goals Clear criteria for all student work Using tasks/activities that elicit evidence of student learning, coupled with higher-order questioning and discourse Descriptive feedback Peer feedback Self-assessment Collaborative culture of learning Using evidence to inform instruction	All students	Every day and throughout each class	A continuous process of gathering information about what students know and can do in order to move learning forward for each individual student.	Teams of Educators Educator Student	No	Never norm-based. Can be criterion-referenced (Student Knowledge & Skills with Performance Levels)	Educators Students	Data derived from formative assessment practices have a short shelf life and must be used in the moment at the individual student or small-group level in order to be effective.

Standardized Assessments
Classroom, Standards-Based Tests

Assessment Test Type	Description and Examples (Not All-Inclusive)	Who is Tested?	When?	For What Purpose?	Who Is Involved in the Testing?	Standardized?	Norm-referenced or Criterion-referenced?	Who Typically Interacts with the Data?	How are Data Used Appropriately?
Interim Tests	Educator-selected or developed unit tests Common school/district-developed tests Common educator-team-developed tests Smarter Balanced Interim Assessments STAR Benchmark tests MAP Benchmark tests	All students	At specific points during a school year (benchmark) or at the end of an instructional unit (common)	Provide data that can be aggregated across classrooms, schools, or the district to understand what students have learned.	Teams of Educators Educator	Depends upon use	Criterion-referenced (Student Knowledge & Skills with Performance Levels) OR Norm-referenced	School & District Leadership Teams of Educators Educators Students Parents	Professional Learning Communities, Grade Level or Data Team (Data Cycle) Meetings with grade level educators using the data plan grade-level core content and instruction.
Classroom Summative	Classroom Final Examinations (when all course content is covered) Unit tests for a specific class (covering content during a specific set of lessons or a unit)	All students	After a period of instruction	A formal attempt, by an individual educator, to determine what students have learned.	Educator	Yes, usually	Criterion-referenced (Student Knowledge & Skills with Performance Levels)	Educators Students Parents (The data are not generally aggregated beyond the class or educator.)	Generally used by educators to monitor their own instruction. Also useful for informing student achievement and grading practices.

Assessment Test Type	Description and Examples (Not All-Inclusive)	Who is Tested?	When?	For What Purpose?	Who Is Involved in the Testing?	Standardized?	Norm-referenced or Criterion-referenced?	Who Typically Interacts with the Data?	How are Data Used Appropriately?
Curriculum-Embedded	Work Samples Performance assessment Locally developed authentic tasks or projects scored against a rubric In-program Mastery Tests, Check-Outs, etc.	All students	Periodically	Measure students' skills on recently taught content using tasks that require students to demonstrate what they know, understand, and can do.	Educator	Depends For example, Work Samples are standardized; student-specific projects are not.	Criterion-referenced (Student Knowledge & Skills with Performance Levels OR Skills-Based Milestones) Note: rubric scoring is criterion-referenced	Educators Students (The data are not generally aggregated beyond the class or educator.)	Useful for testing of a variety of applied content knowledge and skills

Statewide Standards-Based Tests

Assessment Practice or Test Type	Description and Examples (Not All-Inclusive)	Who is Tested?	When?	For What Purpose?	Who administers the Testing?	Standardized?	Norm-referenced or Criterion-referenced?	Who Typically Interacts with the Data?	How are Data Used Appropriately?
Statewide Summative Content Assessments	Oregon's Summative ELA Oregon's Summative Mathematics Oregon's Summative Science	All students at tested grades (Grades 3-8 & 11 for ELA and Math and Grades 5, 8, & 11 for Science) other than those with significant cognitive disabilities who take the ORExt	Once per year in the spring, after sufficient instruction has occurred	Provide evaluative data that can be aggregated and used to summarize what groups of students have learned after a substantial period of instruction.	Educators or other district-approved personnel	Yes	Criterion-referenced (Student Knowledge & Skills with Performance Levels)	School & District Leadership Educators Researchers Legislators Various Policy Community partners Parents Students	Results are useful for guiding state policy decisions and making systems-level decisions at the state, district, and school levels. Districts use state assessment results to inform improvement planning and to assist with systems questions, such as efficacy of a new district curriculum, new professional development plan, allocation of resources etc.

Assessment Practice or Test Type	Description and Examples (Not All-Inclusive)	Who is Tested?	When?	For What Purpose?	Who administers the Testing?	Standardized?	Norm-referenced or Criterion-referenced?	Who Typically Interacts with the Data?	How are Data Used Appropriately?
									*Data should only be used within a broader scope of evidence when making decisions about classrooms or individual students.
Alternate Statewide Summative Content Assessments	Oregon Extended (ORExt)	Students with significant cognitive disabilities	Once per year in the spring, after sufficient instruction has occurred	Provide evaluative data that can be aggregated and used to summarize what groups of students have learned after a substantial period of instruction.	Educators or other district - approved personnel	Yes	Criterion-referenced (Student Knowledge & Skills with Performance Levels)	School & District Leadership Researchers Legislators Various Policy Community partners Parents Students	Results are useful for guiding state policy decisions and making systems-level decisions at the state, district, and school levels. Districts use state assessment results to inform improvement planning and to assist with systems questions, such as efficacy of a new district curriculum, new professional development plan, allocation of resources etc. *Data should only be used within a broader scope of evidence when making decisions about classrooms or individual students.
Summative English Language Proficiency Assessment	Oregon’s Summative English Language Proficiency (ELPA Summative)	Students identified as English Learners	Once per year in the spring	Tests are used to monitor English proficiency and determine when students are ready to exit English Learner status.	English Language Development educators or other district - approved personnel	Yes	Criterion-referenced (Student Knowledge & Skills with Performance Levels)	School & District Leadership English language coordinators Educators Specialists Parents Students	Results are used to inform the degree to which student has demonstrated English language proficiency that is sufficient to access grade-level content and to inform structure of ELD services to meet individualized student’s needs. *Data should only be used within a broader scope of evidence when making decisions about classrooms or individual students.

State-Required Screeners

Assessment Practice or Test Type	Description and Examples (Not All-Inclusive)	Who is Tested?	When?	For What Purpose?	Who administers the Testing?	Standardized?	Norm-referenced or Criterion-referenced?	Who Typically Interacts with the Data?	How are Data Used Appropriately?
English Language Proficiency Screener	Oregon’s English Language Proficiency Screener (ELPA Screener)	Students whose home language use survey and supplemental documentation suggest that a language other than English has had a significant impact on English language development.	When a student’s profile demonstrates that a language other than English may be impacting English language development.	Tests are used to identify a student’s English language proficiency levels.	ELD educators or other district -approved personnel	Yes	Criterion-referenced (Student Knowledge & Skills with Performance Levels)	School & District Leadership English language coordinators Educators Specialists Parents Students	Results are used to determine whether student is eligible for English language development services.
Dyslexia Screener	ODE-approved dyslexia screeners that address these areas in kindergarten: phonological awareness letter/sound correspondence rapid naming	All incoming Kindergarteners (or 1 st Graders if the student enters at 1 st Grade)	As per test developer guidelines (typically at beginning, middle, and end of the school year)	To determine a student’s risk for reading difficulty and the need for intervention	Kindergarten educators or other district-approved personnel	Yes	Norm-referenced & Criterion-referenced (Skills-based Milestones)	School & District Leadership Special education coordinators Educators Parents Specialists	Data from screening should be used to make informed decisions about evidence-based interventions and the progress monitoring that should follow. Students who don’t respond to intervention may require further evaluation (diagnostics) for concerns related to dyslexia.

State-Required Entry Test

Assessment Practice or Test Type	Description and Examples (Not All-Inclusive)	Who is Tested?	When?	For What Purpose?	Who administers the Testing?	Standardized?	Norm-referenced or Criterion-referenced?	Who Typically Interacts with the Data?	How are Data Used Appropriately?
Kindergarten Baseline	Kindergarten Assessment	All incoming Kindergarten students	Within the first six weeks of the Kindergarten school year	Tests are used to give a snapshot of the skills of incoming Kindergarten students and to determine where to allocate early learning resources	Kindergarten Educators or other district -approved personnel	Yes	Not applicable (the KA generates only raw scores at present; no scaling is performed)	School & District Leadership Early learning coordinators Kindergarten Educators Parents Specialists	Results are used to help determine if early learning systems are generally effective, as a piece of evidence within a larger scope of data to determine which students may need academic intervention(s), and to guide resource allocation at the state and regional levels.

National and International Tests

Assessment Practice or Test Type	Description and Examples (Not All-Inclusive)	Who is Tested?	When?	For What Purpose?	Who administers the Testing?	Standardized?	Norm-referenced or Criterion-referenced?	Who Typically Interacts with the Data?	How are Data Used Appropriately?
National and International Large-Scale Summative	NAEP TIMSS PISA PIRLS	Samples of students from participating jurisdictions	Periodically (e.g. TIMSS is every five years, PISA is every three years)	Provide comparable data for participating jurisdictions over time to benchmark performance	Contractors hired and trained by the test vendor	Yes	Criterion-referenced (Student Knowledge & Skills with Performance Levels)	U.S. Congress, U.S. Dept. of Education, State Departments of Education, Researchers	Results are useful for guiding national and state policy decisions and making systems-level decisions at the national and state levels.

Foundational Skill Tests

Assessment Practice or Test Type	Description and Examples (Not All-Inclusive)	Who is Tested?	When?	For What Purpose?	Who administers the Testing?	Standardized?	Norm-referenced or Criterion-referenced?	Who Typically Interacts with the Data?	How are Data Used Appropriately?
Screening	Screening tests are designed to identify students who are either eligible for a program or are at risk for not meeting benchmarks for critical foundational skills appropriate for that grade and time of year. Screening tests include clear mastery targets. DIBELS easyCBM Aimsweb District screeners (e.g., for TAG eligibility, AP courses, etc.)	All students	Varies depending on the purpose; typically part of a universal screening process in the fall, winter, and spring.	<ul style="list-style-type: none"> • Determine risk status at the individual student level • Determine basic health of the system • Determine the most appropriate starting point for instruction • Begin to plan for instructional groups. 	Educators or other district -approved personnel	Yes	Norm-referenced & Criterion-referenced (Skills-based Milestones)	School & District Leadership Educators	<p>Data from screening should be used to make informed decisions about evidence-based interventions and the progress monitoring that should follow. Data can be used for systems level planning. Individual students may be identified for further evaluation (diagnostics).</p> <p>District screeners should be supported by validity documentation that demonstrates that the tool(s) used are of sufficient technical adequacy pursuant to the purpose.</p>

Assessment Practice or Test Type	Description and Examples (Not All-Inclusive)	Who is Tested?	When?	For What Purpose?	Who administers the Testing?	Standardized?	Norm-referenced or Criterion-referenced?	Who Typically Interacts with the Data?	How are Data Used Appropriately?
Progress Monitoring	<p>To determine whether students are making progress toward proficiency in a specific and identified skill. These instruments should be aligned to student learning goals and can be used to determine whether current instructional practices or student supports need to be adjusted. The instruments provide information regarding a student’s performance on critical foundational skills, but may not be able to tell you how well the student is doing in reference to grade-level content standards.</p> <p>easyCBM iLearn AIMSweb STAR</p>	Students receiving additional supports or interventions	Frequently (weekly or bi-weekly), dependent upon student needs relative to the timeframe and frequency of the intervention.	Measure how the student is responding to skills-based support. Determine if progress is adequate or if more (or different) intervention is required.	Classroom educator and/or educator providing the instructional support/ intervention	Yes.	Norm-referenced & Criterion-referenced (Skills-based Milestones)	School & District Leadership Educators School Psychologists Related Service Providers Specialists	<p>Group data is examined to determine if the intervention is working.</p> <p>Individual student data may be compared to other students receiving the same group intervention to examine the effectiveness of the intervention for the individual.</p>

Assessment Practice or Test Type	Description and Examples (Not All-Inclusive)	Who is Tested?	When?	For What Purpose?	Who administers the Testing?	Standardized?	Norm-referenced or Criterion-referenced?	Who Typically Interacts with the Data?	How are Data Used Appropriately?
Informal Diagnostic*	Completing in-depth skills development inventory in a narrow skill area, pursuant to areas of concern identified by screening. Examples for instructional planning: Phonological awareness tests Phonics/decoding surveys Reading records	Students who show risk on initial screening. Students who do not respond to initial intervention.	Following universal screening or for students who do not respond to initial intervention	The goal is to answer the question: is there one or more foundational skills that the student needs to receive additional support for? What are the specific skills that are not mastered? This information is used to further individualize intervention.	Classroom educator and/or educator providing the instructional support/ intervention	Yes	Norm-referenced & Criterion-referenced (Skills-based Milestones)	Educators Specialists School Leadership Parents	Identify a student's specific skill deficits to make targeted, data-driven changes to further individualize intervention.

Assessment Practice or Test Type	Description and Examples (Not All-Inclusive)	Who is Tested?	When?	For What Purpose?	Who administers the Testing?	Standardized?	Norm-referenced or Criterion-referenced?	Who Typically Interacts with the Data?	How are Data Used Appropriately?
Formal Diagnostic*	Typically as part of a formal evaluation for special education eligibility (examples only): Developmental tests (e.g., play-based) Expressive/receptive language tests Adaptive Behavior tests Diagnostic reading evaluations Academic achievement tests Intellectual ability tests	Students who do not make adequate progress when provided with instructional support/intervention.	As needed	The goal is to answer the question: Does the student require more pervasive and intensive instructional supports?	Educators Student Assistance Teams School Psychologists Related Service Providers	Yes	Norm-referenced & Criterion-referenced (Skills-based milestones)	Educators Specialists School & District Leadership Parents	Teams make decisions about which individual students need more pervasive and intensive supports Appropriate to interpret only when a defensible case can be made that students have been provided with a meaningful opportunity to learn (in special education contexts)

*The term diagnostic is used in many ways in practice. These uses generally fall into two themes: 1) In terms of curriculum, the term is used to refer to tests that evaluate students before learning has taken place to identify appropriate curricular placement. We do not use this definition, as these types of tests are more accurately referred to as placement tests; 2) To identify an individual’s skills-based learning strengths and weaknesses and likely source of academic problems to make targeted, individualized changes to the student’s intervention program (informal diagnostic) and to determine if a student profile fits the eligibility criteria pursuant to the development of an individualized education program or Section 504 Plan (formal diagnostic).

Please use this space to record any observations, questions, or relevant resources.

NOTES: