Assessment Guidance

Oregon Department of Education

255 Capitol Street NE Salem, OR 97310-0203

Phone: 503.947.5600

Fax: 503.378.5156

10/20/2014

The Oregon Department of Education (ODE) recognizes that educators at all levels of leadership are faced with many complex decisions around assessment planning. This guidance includes the *Oregon Criteria for High Quality Assessments*, which districts can use to evaluate the quality of assessments whether they are obtained from an outside source or developed internally.

Assessment

Assessment Guidance

Let Us Know What You Think

ODE invites you to contact Cristen McLean, Assessment Policy Analyst, at cristen.mclean@state.or.us to provide feedback or to discuss how you are using this resource.

Oregon Criteria for High Quality Assessments

User Guide for Criteria

WHICH ASSESSMENTS DO THE CRITERIA APPLY TO?

The criteria are written to be broad enough to apply to assessments that were developed locally, obtained from an open educational resource or commercial source, or assessments that include both locally developed content and content obtained from outside the school or district. ODE recommends using these criteria in the following ways:

- to evaluate assessments and assessment practices already in use in the district or school;
- to evaluate assessments that you are considering obtaining from either a commercial source or an open-education resource;
- as a reference to consult—alongside assessment development training materials—during local assessment development[1].

HOW TO READ THE CRITERIA TABLES

WHAT ARE THE CRITERIA?

1. Clear Purpose: *Why am I assessing?*
2. Clear Learning Target(s): *What am I assessing?*
3. Quality Assessment: *How can I assess it well?*
4. Proper Test Administration: *How will I ensure test conditions do not interfere with a student's ability to perform well on a test? And how do I maximize a student's ability to perform well on a test?*
5. Effective Communication of Results: *How will I share results for maximum impact?*

---

[1] The criteria are not intended to be used a process for assessment development.

Criteria	1. CLEAR PURPOSE
Summary	Why am I assessing? Everyone using the assessment should have a clear understanding of the purpose of the assessment, including how results will be used.
Significance	Without a clear purpose for the assessment, it is not possible to evaluate the quality of the assessment. In order to determine whether the assessment has fulfilled its purpose, the purpose must be known from the beginning.

Key Concepts	A clear and focused purpose for the assessment is critical for the assessment to be high quality. Likewise, the more purposes an assessment is designed to fulfill, the less likely it is that the assessment will serve any of them well.
	
	The majority of assessments fulfill one of these two primary purposes, to:
	1. inform and improve instruction; or
	2. measure outcomes
	    o assessments to certify student mastery, in which case the assessment would be designed for reliable determination of a specific achievement level (meeting the threshold for demonstrating "mastery") and could include fewer questions with questions clustered around that achievement level.
	    o assessments to measure student growth, in which case the assessments would include many questions across the distribution of student performance in order to determine student's movement along a learning progression.
	
	Following is partial list of how assessment results may be used:
	• To monitor student progress in attaining knowledge and skills measured on the statewide assessment
	• To certify student mastery of skills in order to meet certain requirements, such as the Essential Skills graduation requirements
	• To support instructional planning and ensure that the instruction is matched to the student's knowledge and skills
	• To measure student growth in order to evaluate educator effectiveness

Expectations	The following information should be clearly stated in the test documentation and should be known by everyone using assessment: • Who is included in the intended audience for the assessment results? • Are the assessment intended to inform instruction or measure outcomes? • What is the intended use of the assessment results?

Criteria	1. CLEAR PURPOSE
Guiding Questions	- What is the purpose for the assessment? - What is the target population for this assessment? - Who will use the results? - What decisions will they use the results to inform? - How will the assessment results be used? - Is the assessment to inform and improve instruction? Or is it to measure outcomes?

Criteria	2. CLEAR LEARNING TARGETS
Summary	What am I assessing? Clearly stated learning targets clarify the scope of the assessment. A well-designed assessment mirrors the learning targets embedded in local curriculum.
Significance	When the assessment mirrors the local curriculum, it ensures that all students have had the opportunity to learn the content prior to being assessed on it and that the assessment format or content does not provide an advantage to certain groups of students based on their background knowledge.
Key Concepts	Learning targets should be derived from the local curriculum which has been aligned to adopted standards. These learning targets can be further informed by the learning progressions in each content area. Learning progressions can identify the knowledge from each course that is *essential* for students to know in order to be successful in subsequent courses in that content area. Learning targets for content knowledge include, but are not limited to, the state standards. Learning targets for skills and strategies may also come from other sources. A test blueprint can help ensure that a test aligns to the local curriculum. Also known as test specifications, these blueprints typically include a matrix or chart identifying the targets to be tested and the number of items and/or percentage of the test designed to measure each learning target.
Expectations	• The learning targets measured by the assessment should be clearly stated. • The number of questions covering a particular learning target should be proportional to the significance of the standard in the learning progression and the amount of time spent on that standard in the local curriculum. • The assessment should include enough questions to allow for a generalization about student learning relative to the target. • When assessments are used to measure growth, it is important that those assessments contain an adequate range of items and scoring options to capture the growth of all students, including the highest performing students.
Guiding Questions	• What evidence is collected to assess whether a student has met or exceeded the intended learning target(s)? • Do the learning targets arise from local curriculum standards? • Are the learning targets clearly stated and universally understood and accepted?

Criteria	3. QUALITY ASSESSMENT
Summary	How can I assess the learning target(s) well? The assessment questions should allow students to show what they know and can do with regards to the learning target(s).
Significance	A rigorous scoring protocol for constructed response questions and performance tasks will make it easier to establish credibility of scores.
Key Concepts	Questions are also known as items and there are various item types, including selected response (multiple-choice, true/false, fill-in-the-blank, matching); constructed response questions (e.g., essays, graphing, short answer written responses, etc.); performance tasks; and observations/interviews. In order to identify the appropriate item type for the learning target (so that the item effectively represents the content, complexity, and depth of knowledge) the developer needs to have a very strong understanding of the learning target.

The rigor of the scoring procedure should increase in proportion to the stakes attached to the assessment and the complexity of the item. There are four main options for scoring constructed response questions, in order of rigor from lowest to highest.
- Teachers directly score their own students' work.
- Teachers score a randomized and anonymous selection of student work.
- Teachers provide the first score on student work, and a second teacher or separate, objective third party provides a second score (anywhere from 5% to 100% of the student work); discrepancies may be managed in a variety of ways, including discussion between raters or a third review by a more senior scorer.
- Teachers do not score any of their own students' work; student work is outsourced to a separate, objective third party for single or double-scoring with provisions in place for the resolution of discrepancies if a system of double-scoring is utilized.

Regardless of whether the scoring is completed by the teacher or a third party, the scorer should have a strong understanding of the criteria and be qualified to score the student work. When there are multiple scorers on a single student response, a process should be established for initial and ongoing calibration of raters so that criteria for determining scores and achievement levels is applied consistently and equitably.

Bias and sensitivity review is important to ensure that the question does not unintentionally favor or disadvantage students from specific groups and that the question does not include information or context that might offend a large audience. Some types of bias that may be present in the vocabulary and reading material include racial bias, gender bias, socio-economic bias. |

Criteria	3. QUALITY ASSESSMENT
Expectations	• The question format and options for how the student answers the question should measure the content and the skills described in the learning target. • The question format, options for how the student answers the question, and the way answers are scored should all measure the full range of complexity and depth of knowledge required in the learning target. • Together, the question format, the options for how the student answers the question, and the way answers are scored provide a reasonable estimation of student mastery of the learning target. • The connections between questions and their scoring criteria should be apparent to anyone familiar with the learning target. • The higher the stakes attached to the assessment results, the more rigorous the scoring procedure should be. • The more complex the questions, the more rigorous the scoring procedure should be. Vocabulary and reading material used in the questions should have been reviewed through a bias/sensitivity/accessibility lens.
Guiding Questions	• Has a proper scoring procedure(s) been selected given the target? • Have high quality questions been developed? • Are assessment items clearly aligned to the standards? • Have sources of bias been anticipated and minimized? • Have sound scoring procedures been developed?

Criteria	4. PROPER TEST ADMINISTRATION
Summary	How will I ensure test conditions do not interfere with the assessment? And how do I maximize a student's ability to perform well on a test? Students should be able to demonstrate their mastery or growth relative to the learning target without any extraneous factors interfering; they should also be able to access all allowable supports that do not alter the degree to which the test measures the learning targets.
Significance	Poor test administration conditions have the potential to interfere with students demonstrating their full potential and puts at risk the accuracy and usefulness of the data.
Key Concepts	Support is a broad term that includes the concepts of accommodations and other accessibility features. Some examples of supports include changes to the assessment to make it more accessible. Supports that compromise the construct being tested would not be appropriate, though, because they could affect the integrity/accuracy of the results.
Expectations	- Supports that make the assessment easier for students to access and that do not reduce the connection between the question and the learning target should be offered where feasible and appropriate, - The testing interface, whether paper- or technology-based, should not interfere with student performance. - Supports should be available to permit English learners and students with disabilities to demonstrate their knowledge and abilities - Assessment should not contain features that unnecessarily prevent English learners and students with disabilities from accessing the content of the questions. - When assessment results will be used together--for instance pre-test results are used along with post-test results to show growth--it is important that the testing environment and test administration conditions (amount of time available, accommodations available, etc.) are the same. Plans should be established for secure management of assessments and assessment data, so that no one individual gains access to unauthorized information. - Strategies should be in place to detect testing irregularities before, during, and after testing. Procedures should be developed and adopted for how to address testing irregularities that may occur. - Data should be managed securely to protect student privacy, reflecting compliance with all applicable federal and state laws and requirements.
Guiding Questions	- Is the test administered in a way that maintains the integrity of the measure? - What procedures are in place to ensure equitable testing conditions? - Has a plan been established to maintain the test and the data securely?

Criteria	5. EFFECTIVE COMMUNICATION OF RESULTS
Summary	How will I share results for maximum impact? Educators, students, and parents should be able to look at the results from the assessment and understand individual student achievement relative to the learning targets.
Significance	Assessments should be designed to provide information about student achievement of the learning target. If the information about student achievement of the learning target (the results) are not usable, then the assessment was a waste of both resources and instructional time. If the results from the assessment are not timely, they could become irrelevant before they are received.
Key Concepts	No matter how impressive the formatting or how quick the delivery, foundational to the usefulness of the results is that the assessment meets criteria 1-3 above.
Expectations	• Reports should provide information about student achievement relative to the learning target. • Reports should be instructionally valuable, easy to understand, and available in time to be useful. • For growth measures, reports should illustrate growth accomplished.
Guiding Questions	• Are timely and useful results assured? • When serving a formative purpose, has communication been tailored to that application? • When the purpose is summative, has communication been designed for that purpose? • Are score reports easily understood by parents? Students?

Assessment Vocabulary

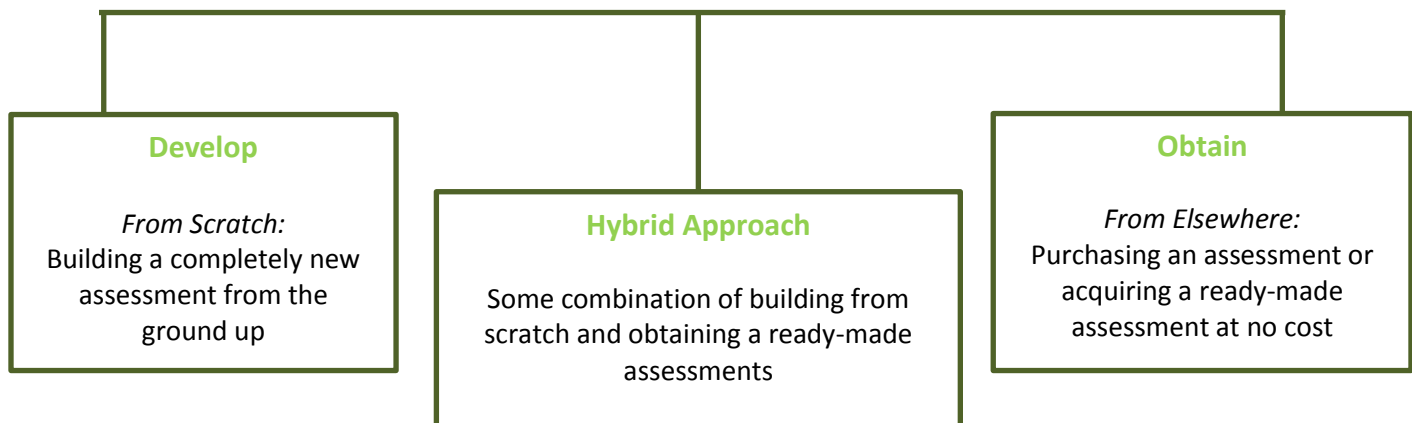| TYPE	TIMING	PURPOSE	TYPES OF DECISIONS TO BE INFORMED
Formative	during instruction	Process used to diagnose where students are in their learning Used to identify how teachers can adjust instruction to increase learning	• Where can we improve instruction right away? • Which students need specific help?
Interim	(between summative and formative)	Used to provide interim evidence of performance against a defined learning target Results can be meaningfully aggregated and reported alongside summative	• Predictive: What performance can be anticipated on the summative assessment? • Evaluative: Is the curricular program or strategy effective? • Instructional: How should the instruction and curriculum be adapted to better meet student needs?
Summative	end of year or unit	Used to provide culminating evidence of performance against a defined learning target	• What standards has each student mastered? • What grade does each student receive?

Consideration: *Determining Whether to Develop or Obtain Assessments*

For each assessment, educators need to determine the source of the assessment. The *Assessment Source Options* diagram illustrates the three options. Districts will need to determine whether they will:

- develop the assessment locally;
- obtain an assessment through an open educational resource (or other non-commercial source), or through a commercial source; or
- implement a hybrid approach (e.g., adapting an assessment obtained from an open educational resource or purchasing an item bank and supplementing it).

ASSESSMENT SOURCE OPTIONS

Develop

From Scratch:
Building a completely new assessment from the ground up

Hybrid Approach

Some combination of building from scratch and obtaining a ready-made assessments

Obtain

From Elsewhere:
Purchasing an assessment or acquiring a ready-made assessment at no cost

Regardless of whether the assessment is locally developed or developed by someone outside of the district, the assessment needs to mirror the local curriculum and the educators need to have a strong understanding of the assessment. In order for this to occur, the following factors are critical: assessment literacy and transparency of assessments that were obtained from an external source.

Assessment Literacy

Broadly, a person with assessment literacy will have enough understanding of measurement principles to identify and describe the features of a high quality assessment and use the assessment results appropriately (make accurate, test-based decisions). When planning their professional learning time, districts should budget time to build assessment literacy. Additionally, in order for assessment literacy to become

embedded in the school culture, classroom teachers should have a lead role in assessment conversations that relate to the importance of assessment literacy.

When there is a high degree of assessment literacy across a district, that district will be well positioned to make strong decisions about assessment planning, implementation, and use of results. For instance, such a district would be able to,:

- critically evaluate assessment vendor presentations in order to make the best decisions about a procurement for their district;
- try out assessments and effectively determine whether the assessment is providing value to the district;
- make accurate inferences based on the assessment data;
- reflect upon local assessment practices to identify what is working well and what needs to be revised; and
- develop an assessment plan that coordinates the timing and use of different assessment types across a year.

Transparency in Assessments

Assessments that are not developed locally need to be transparent, meaning that:

- information is provided about how questions were created,
- example questions are available for review,
- information is provided about how assessments are assembled and scored,
- test blueprints are available for review.

A high degree of transparency in terms of test design enables districts, schools, and educators at all levels of leadership to evaluate the assessments to make sure that they align to the local curriculum and they meet the Oregon High Quality Assessment Criteria. An example of an assessment with a high-degree of transparency would be an assessment for which the questions and test blueprints are available for educators to review and engage with, such that educators may edit the existing assessments and/or build their own assessment.

References

Jones, M., & Vickers, D. (2011). *Considerations for performance scoring when designing and developing next generation assessments: White paper.* Retrieved from http://www.pearsonassessments.com/research/researchpub/researchlist.type. White_Papers.html

Pinkus, L.M. (Ed.). (2009). *Meaningful measurement: The role of assessments in improving high school education in the twenty-first century*. Washington, DC: Alliance for Excellent Education.

Popham, W.J. (2008). *Teach better, test better.* Alexandria, VA: ASCD.

Rhode Island Department of Elementary and Secondary Education. (n.d.). *Comprehensive assessment system: Rhode Island criteria and guidance.* Retrieved from https://www.ride.ri.gov/InstructionAssessment/Assessment/Comprehensive AssessmentSystemCAS.aspx

Smarter Balanced Assessment Consortium. (n.d.) *Bias and sensitivity guidelines*. Retrieved from http://www.smarterbalanced.org/wordpress/wp content/uploads/2012/05/TaskItemSpecifications/Guidelines/BiasandSensitivity/Bias andSensitivityGuidelines.pdf

Chappuis, J., Stiggins, R. J., Chappuis, S., & Arter, J., (2012). *Classroom assessment for student learning: Doing it right—using it well*, 2/e. Saddle River, NJ: Pearson Education Inc.

Stiggins, R. J. (2014). *Defensible teacher evaluation: Student growth through classroom assessment.* Thousand OAKS, CA: Corwin.

The Council of Chief State School Officers. (2010). *Building an interim assessment system: A workbook for school districts.* Washington, DC: Author.