Oregon Technical Skill Assessment Part 2
Approving Technical Skill Assessments

This document is **Part 2** of a two part series about technical skill assessment.
--Part 1 addresses the general framework for technical skill assessment in Oregon.
--Part 2 focuses on identifying or developing technical skill assessments that meet standards of validity and reliability.

Introduction

The Perkins reauthorization of 2006 shifts the measure of technical skill assessment from grades or GPA in Career and Technical Education (CTE) programs and courses to measurement by assessments that are valid and reliable. In response to those requirements, the Oregon Department of Education (ODE) will provide a list of valid and reliable assessments that accommodate the diversity in CTE Programs of Study and meet the federal requirements. The list will include commercially available assessments as well as approved locally developed assessments. This document outlines a process for identifying appropriate technical skill assessments whether they are commercially or locally produced.

Phasing in technical skill assessments

No later than the beginning of the 2012-2013 school year all approved CTE Programs of Study **must** include technical skill assessments that are valid and reliable. Assessments need to be in place for both secondary and post-secondary components of the CTE Program of Study. Prior to 2012, assessments will be phased in as CTE Programs of Study are approved and appropriate assessments become available. Specific secondary and post-secondary assessments are identified on the application for the secondary component of a CTE Program of Study when it is submitted for approval. The secondary and post-secondary assessments do not need to be identical. If an appropriate assessment is not available at the time of program approval, a plan for identifying or developing an assessment **must** accompany the CTE Program of Study application.

Some students may need an alternative form of assessment in order to accurately demonstrate their technical skill attainment. These students must not be excluded from taking the assessment. Alternative forms of an assessment may be available from a vendor or an instructor can provide an appropriate form of assistance to the student. In either case, the results of the assessment should accurately reflect how well a student has met the same standards required of all students taking the assessment.

Approval of Commercially Available Technical Skill Assessments

All ODE approved technical skill assessments need to meet the criteria listed in Appendix A of this document. The following types of commercially produced assessments are usually able to meet these criteria.

1. Assessments used as a national or state licensure exam.
2. Assessments used as an industry-recognized certification exam.
3. A nationally validated assessment produced by a recognized organization (i.e. NOCTI, SkillsUSA)

A request can be made to the Oregon Department of Education to add commercially available assessments to the approved list. The request should be sent by e-mail to the ODE contact listed in this document for review by ODE staff. In the e-mail identify:

- The organization that distributes the assessment.
- A website address or e-mail contact.
- The name of the specific assessment.
- The CTE Program of Study where the assessment will be used.

Requests will be evaluated and notification of the outcome will be sent to the individual making the request. A link to approved assessments will be posted on the Oregon Skill Sets web pages.

In some cases a commercially produced assessment may not meet the basic criteria outlined in Appendix A. This can occur if:

1. The assessment needs to have significant modification for local use.
2. The assessment is used as part of a text or curriculum and does not have evidence of validity or reliability.
3. The assessment has been developed commercially to meet Oregon needs, but has not been pilot tested.

If the commercially produced assessment does not meet the requirements for validity and reliability, it will be treated as a locally produced assessment and ODE will require further data before including it on the approved list.

At times, commercially available assessments may not report performance levels but they may provide normed data. Normed results indicate how well a student performs on the assessment compared to other similar students. The results of the assessment are usually given as a percentile. For the purpose of Oregon technical skill assessment, scores that are above the 50th percentile would be considered proficient until an appropriate cut score can be determined.

Approval of Locally Developed Technical Skill Assessments

ODE will recognize locally produced technical skill assessments that meet the same criteria required for commercially available technical skill assessments. The developer of the assessment is responsible for submitting adequate documentation to demonstrate how the requirements are being met. Locally produced assessments may be advantageous because:

1. Administration costs are often lower than commercially produced assessments.
2. Assessments can better reflect the actual curriculum.
3. Assessments may reflect alternative approaches such as evaluation of performance or portfolios.

These advantages need to be balanced against the costs of producing a local technical skill assessment. To assure validity and reliability, locally produced assessments will require a significant amount of time and effort to develop and pilot test.

General Principles for Producing an Assessment

A technical skill assessment can provide data to help determine whether students are meeting standards. This is usually called a criterion-referenced assessment. Typically criterion-referenced assessments are tests with multiple choice questions. Alternative approaches that measure student performance on a specific task or set of tasks can also be criterion-referenced. A criterion-referenced assessment needs to be developed carefully or the validity and reliability of the assessment can suffer.

Tests

When writing a test it is important to write questions and responses carefully. Test questions that are too easy, too difficult, or ambiguous can affect the reliability of the test. Questions that require extended responses, such as an essay question, should have clear scoring criteria.

An alternative to writing test questions is to use a test item database. These databases usually contain items that are properly written, however, they still need to be evaluated carefully for validity and reliability. Another alternative is to use tests that are provided with specific curriculum. Although these tests may meet the basic validity and reliability criteria, copyright restrictions can limit the use of the test. ODE will not promote violation of copyright law. If the proposed test is copyrighted, ODE may request proof that the educational institutions have the rights to reproduce and use the test.

Performance assessments must also be designed carefully. A performance assessment uses a task or multiple tasks to determine student proficiency. The assessment should also provide sufficient evidence that a student can perform consistently. The tasks need to be clear to the student and all materials related to the tasks must be available. Performance assessments are usually scored with a rubric or scoring guide. A scoring guide should have enough detail so that student performance can be consistently scored.

Determining Validity

There are two acceptable routes for determining validity of a locally produced assessment. One route is to have a systematic and formal review of the test or performance by a group of professionals that represent the appropriate industry. The industry group validates that the assessment reflects appropriate industry standards.   Another approach is to create an assessment framework or table of specifications. The framework or table needs to identify the match between specific test items and specific industry-recognized standards. There should be an appropriate number of items or tasks representing each of the standards being tested. In either case, it would be appropriate to have industry representatives involved in the assessment development process to insure a high level of validity.

Determining Reliability

Tests

Reliability relates to the ability of an assessment to consistently measure the same thing. Reliability is frequently determined through a statistical analysis of data gathered during a pilot test. For a typical test the following statistical measures are commonly used to measure reliability.

- Chronbach's alpha
- Kuder-Richardson Formula 20
- Pearson Product-Moment Correlation Coefficient (test-retest reliability)
- Spearman Brown prediction formula (split-half reliability)

Each statistic produces a reliability coefficient that has a highest possible value of 1.0. Generally values of 0.6 to 0.7 are considered adequate and 0.8 or higher is good. Reliability of a test can be improved by clarifying questions or lengthening the test. Analysis of individual test questions may reveal why a reliability coefficient is low. Test questions with multiple correct answers can lower reliability.

Assessments using student performance or portfolios require a different method for determining reliability. The most common method is called inter-rater reliability. This method demonstrates reliability if multiple "judges" use the same scoring guide with the same piece student work and independently assign the same score. There are several statistical approaches to determine inter-rater reliability. Most approaches may not be practical for locally developed performances or portfolios. A rater training plan can substitute for the statistical approach. The purpose of the training plan is to calibrate the scores given by any rater against those of a set of expert raters. Calibration requires examples of performance or portfolios that have been previously scored along with appropriate scoring guides. The goal of calibration is to have each rater provide a similar score that is close to the expertly determined score. Training should take place immediately before evaluating the performance.

The conditions under which the test is being administered can affect reliability. This is particularly true for performance or portfolios assessments. The conditions under which the assessment is being conducted should be clearly documented. This includes a description of any tasks students are completing, any materials they should have, the type of assistance that can be provided, and the time that is allowed.

Determining Proficiency Levels

Tests

A technical knowledge and skills test can help determine whether a student is proficient as measured against a set of standards. Proficiency levels are usually determined by a panel of content experts. A typical way to do this is to rank questions on a test from easiest to most difficult and have the experts place a "bookmark" at the border between basic and proficient. Another "bookmark" is placed between proficient and advanced. This will provide a range of scores that are considered proficient.

Performance Assessments

In a performance assessment, the proficiency level can be built into the scoring guide used for the task or portfolio. Scoring guides may have two or more levels. At a minimum, a scoring guide needs to distinguish between someone who is proficient when performing the task and someone who is not. At times it is useful to have levels in a scoring guide that are beyond proficient. This would be especially true in a high school assessment where students would not be expected to meet industry-level proficiency. Scoring guides are usually classified as analytic or holistic. An analytic scoring guide breaks a task or portfolio into very specific measurable criteria that can be rated. For example, an analytic scoring guide for a

construction task might include some specific ranges of measurement that are allowable on each aspect of the task. An analytic scoring guide is often expressed as a check list based on very specific tasks. A holistic scoring guide, on the other hand, groups criteria into a larger category. In the construction example, there may be a category of overall accuracy of measurement that includes error, methods of taking measurements, and ability to correctly read values from a set of plans. The score for the category would come from a holistic look at the criteria

Contact Information

Information on technical skill assessment can be found on the Oregon Department of Education website at:

http://www.ode.state.or.us/search/page/?id=1656

Questions about technical skill assessment can be sent to:

Tom Thompson, Education Specialist
Tom.thompson@state.or.us
(503) 947-5790

Appendix A

Criteria for Approval of Oregon Technical Skill Assessments

Technical skill assessments used for reporting data on technical skill attainment in CTE Programs of Study must be approved by the Oregon Department of Education. All approved assessments must meet the following criteria.

1) Student performance data is available to instructors for reporting purposes as individual student data that can be used to calculate the percent of students who are proficient.

2) Student data will be available to instructors in time to be reported to ODE.

3) Produced using recognized guidelines for construction of quality assessments.

4) Demonstrated **validity** through any of the following

   a) Recognition by a state, national, or regional industry group as a valid assessment.

   b) An assessment framework, table of specifications, or other set of assessment specifications tied to industry-based standards.

5) Demonstrated **reliability** through a clearly defined set of assessment administration guidelines and any of the following

   a) Evidence of pilot testing conducted to determine reliability.

   b) Statistical test of reliability appropriate to the type of assessment.

   c) A rater training plan that would assure consistent scoring of performance assessments.

6) Linked to a particular cluster or focus area for which there are existing CTE Programs of Study in Oregon.

7) Alignment tools such as test specifications or assessment frameworks are available to assessment administrators.

8) Provides an age-appropriate proficiency score that has been established through any of the following means

   a) Determined by a panel of industry experts.

   b) Determined using normed data based on populations related to those being tested.

   c) Required for post-secondary program entry.

   d) Required for an industry recognized certificate or degree

Appendix B

The most recent version of the Locally Developed Assessment Application can be found on the following page:
http://www.ode.state.or.us/search/page/?id=1656