

Oregon Technical Skill Assessment Part 1

A Conceptual Framework

This document is **Part 1** 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.

Purposes for Technical Skill Assessments

- ✚ Document attainment of industry-based technical knowledge and skills;
- ✚ Document evidence of career development and preparation;
- ✚ Document readiness for secondary-to-postsecondary CTE Program of Study transition; and
- ✚ Evaluate CTE programs for continuous improvement

Using Valid and Reliable Assessments to Align with Industry

Perkins IV calls for CTE Programs of Study to use technical skill assessments that are valid, reliable and aligned with industry standards as a measure of a CTE concentrator's technical skill attainment. Technical skill assessments are one source of data to evaluate whether CTE Programs of Study effectively prepare students to pursue their career interests. An effective technical skill assessment approach should contribute to the improvement of instruction.

Not all assessments are created equal. An assessment that fits the need of one program may not fit another. In all cases, assessments must be valid and reliable.

Valid - the extent to which an assessment measures what it is intended to measure. For example, if a student performs well on a reading test, how confident are we that the student is a good reader? A valid technical skill assessment is aligned with the industry-based standards intended to be measured, provides an accurate estimate of a CTE concentrator's performance relative to the standard, and is fair.

Reliable – consistency in the measurement and scoring of assessments; specifically, the extent to which separate administrations of the same technical skill assessment approach score students in the same way. Reliability is often determined using a statistical test that either compares how a student performs on different parts of a test or how well two observers can agree to student performance on a specific task.

Guidelines for Choosing an Appropriate Assessment

1. **Select assessments that are appropriate for the industry-based standards within a program of study.** Appropriateness here is defined as how easily the assessment can be aligned with, and integrated into a program. It may be appropriate to implement multiple approaches to adequately measure technical skill attainment.
2. **Ensure assessments have an explicit relationship to the program's content and instruction.** The potential exists that an assessment choice may adversely affect curriculum alignment with industry standards. This can occur if the scope of knowledge and skills an assessment covers does not reflect what industry requires. Similarly, assessment choices can negatively affect a program's instructional methods. If an assessment covers only terminology, definitions and concepts or "soft" employability skills, this may result in a program only using instructional methods that teach this level of knowledge, but fall short of promoting technical skill attainment and learning transfer.
3. **Establish a defined industry role for the validation and recognition of assessments.** Using assessments aligned with industry standards will only succeed if industry is involved and publicly endorse the assessment in some tangible way. Employers can help "set the bar" by reviewing the assessment and determining qualification scores for meeting technical skill attainment. Most national commercially produced technical skill assessments incorporate industry input. Local industry involvement in selection of specific assessments can help validate the importance of technical skill attainment and determine the connection to local employment needs.
4. **Ensure that all students have an equal opportunity to demonstrate technical skill attainment.** 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.

Specific Criteria to be Considered when Selecting and Implementing Technical Skill Assessments

A. Assessment Construction & Administration—*what are the implementation issues?*

- ☐ Adopt, adapt or align assessments of specific technical knowledge and skills based on validated industry-based standards such as the Oregon Skill Sets.
- ☐ If locally developed, industry-aligned assessments are being implemented, the construction of the assessments should reach beyond just the individual classroom teacher and include stakeholders such as other CTE Program of Study instructors and industry partners.
- ☐ Identify the use and role of cognitive-level assessment.
- ☐ Identify the use and role of performance-based assessment.
- ☐ What is the cost/benefit analysis of developing assessments vs. purchase and use of existing technical skill assessments aligned to CTE Program of Study content standards? Is the selected assessment strategy financially sustainable?
- ☐ Assure that technical skill assessments can be administered at the end of the secondary and postsecondary CTE Program of Study component.
- ☐ Assure that data from technical skill assessments is available for required reporting.

B. Assessment Requirements—*what are the requirements for using the assessment?*

- ☐ Student Prerequisites—students have met the designation of a CTE concentrator and are ready to be assessed.
- ☐ Program Requirements—alignment of content and instruction with adopted, state-recognized CTE content standards.
- ☐ Instructional Requirements—instructor and program coordinator have access to reports on assessment development and validation. Assessment blueprints and objectives are easily available to inform content alignment and document validity.

C. Assessment Use—*how are assessments going to be used?*

- ☐ Accountability & Continuous Program Improvement—use of assessments allows for determination of progress towards specific targets that inform performance-driven decisions and requires access to student-level performance data.
- ☐ Support—assessments are used to diagnose and monitor student progress towards meeting industry-based standards.
- ☐ Bridge—assessments are used to bridge the secondary and postsecondary components of a CTE Program of Study and to bridge education with industry.

D. Assessment Impact—*what is the impact of using assessments on curriculum?*

- Classroom Impact—use of assessments has a positive cost/benefit impact on curriculum and instruction changes.
- Cross Level Integration—use of assessment require secondary and postsecondary levels of a program of study standards to be highly aligned and articulated.
- Professional Development—experience with, and the results from, assessments are used to inform an on-going professional development cycle focused on the improvement of CTE standards-based curriculum and instruction.

Reporting Student Data

There are a number of reasons students should be assessed on technical skill attainment. The data can help individual teachers evaluate the program in order to strategically use available resources for improvement. The data can also help students understand when they are prepared for their next steps toward a career. While individual schools or consortia may wish to assess a large group of students to evaluate programs, only data from a certain population of students needs to be reported for Perkins accountability. Use the following guidelines when deciding which secondary and post-secondary students must be assessed for technical skill attainment.

1. Secondary students should be concentrators, having completed at least one credit in a CTE Program of Study, and are ready to be assessed. A secondary student who is ready to be assessed should have completed their required/foundation courses in a CTE Program of Study. Some students may be ready before they have completed all the courses. An instructor can identify additional students who are ready to be assessed.
2. Post-secondary students should have completed their CTE program.

Since assessments will vary in type and scoring, it is only necessary to report whether the student met or did not meet the level of proficiency defined by the assessment. Actual assessment scores are not reported.

Using Assessments to Determine Proficiency

Assessments are usually designed to be criterion referenced or norm referenced. Norm referenced assessments compare the performance of one student with that of another student. They do not indicate whether or not the performance is acceptable according to a set of standards. Criterion referenced assessments compare the performance of each student against a set of standards. The student is usually determined to be proficient if they meet or exceed a certain cut score. Technical skill assessments used for Perkins reporting should be criterion referenced.

A proficiency level, or cut score, is often determined by a panel of experts in the topic being assessed. When the panel is identifying the proficiency level, they consider the characteristics of the people being assessed and what constitutes an acceptable performance. A proficiency level determined for a person entering a profession is not necessarily an appropriate proficiency level for someone who has just begun their education and training. When selecting an assessment make sure that proficiency level is appropriate for the experience of the students.

There may be little relationship between the cut scores of two assessments even if they cover similar content. Different assessments can create significant changes in the number of students identified as proficient. One way to avoid problems is to pilot a new assessment and compare the results with other sources of data such as test scores, work samples, or course grades. An appropriate assessment should produce scores that can be predicted based on this other data.

Proficiency and Course Grades

Prior to Perkins IV, course grades were used as a measure of technical skill attainment. Course grades often reflect more than just proficiency in technical skill attainment. ODE and the Perkins IV legislation do not require the use of technical skill assessments to determine grades for courses. Instructors may use technical skill assessments and the data generated through those assessments for other purposes. An assessment could be a viable component of an articulation agreement. It also could serve as a comprehensive course exam.

Comparison of Assessment Types

There are multiple ways to assess student attainment of technical skills. The Oregon Department of Education encourages schools and regions to select a type of assessment that will provide data that accurately reflect student technical knowledge and skills. Choosing an assessment requires balancing validity and reliability with other criteria such as cost and ease of implementation. The six types of assessments described below are appropriate for measuring technical skill attainment. Starting in January 2009, ODE will provide information on assessments that meet the requirements of Perkins IV and fit into one or more of these categories.

1) Standards-based criterion-referenced tests

Description: Specific measure of elements of state CTE content standards through a standardized instrument

Reliability: Generally as high as standardized achievement tests

Alignment with Industry-based Standards: High, in theory, although specific tests vary based on how well they sample the content domain represented by the standards. No single test can measure performance for the full set of standards.

Validity as a Measure of What Is Taught in Individual Classrooms: Varies depending on the degree to which teachers have aligned their curriculum to standards. In practice, varies from high to not at all.

Cost: High initial cost to develop or purchase a test that meets psychometric standards. Ongoing costs depend on how often new forms are implemented. If forms are released annually, a new test must be developed or purchased each year, which increases the cost.

Professional Development Requirements: Implies significant professional development for teachers who need to understand the standards and how to teach to them.

Conditions of Administration: Administered under highly standardized conditions, generally during class time (reduces instructional time)

Diagnostic Properties: Varies based on characteristics of the test. If the test has enough items to produce subscales, then reasonably specific diagnostic information relative to standards can be provided. Otherwise, results are more general to achievement within a subject area.

Sensitivity to Changes in Instruction: Theoretically, highly sensitive. However, if the test does not map well onto the content domain, sensitivity is decreased because instruction may not necessarily cover areas assessed on the test.

Effect on Teacher Instructional Practice: Can serve as a framework for instruction, but can also drive instruction toward repetitive basic skills instruction, depending on what is tested and how it is tested.

Adaptability to Reflect Local Practices and Priorities: Very limited, except to the degree that local practice has been adapted to reflect the standards or the standards have been developed to reflect existing local practice.

Accountability Uses: School-to-school comparisons within a state, growth models, as a component of a CTE performance measurement and reporting.

Primary Limitations: As noted, cannot assess the full knowledge domain of the CTE standards, which leads to a narrowing of how the standards are taught. Format and time limitations restrict ability to assess higher order thinking skills and more complex skills.

2) End-of-CTE-course/CTE program exams

Description: Designed on a course-by-course or program basis to assess the key elements of that particular course or program.

Reliability: Can be developed to achieve high levels of reliability

Alignment with Industry-based Standards: Can be well aligned with state standards, assuming the state standards align with the content of a particular course, or, conversely, that a particular course is designed to align with state standards.

Validity as a Measure of What Is Taught in Individual Classrooms: Very high. The alignment should be nearly complete. Teachers should be teaching precisely what is on the test and delving more deeply into closely related content and topics to strengthen student understanding.

Cost: High to develop initial test forms. Can be moderate thereafter, depending on how often the forms are updated. If content topics remain constant across forms and items are selected or purchased from a pool, costs can be constrained. Cost is greater if the test includes any performance tasks, such as a problem or paper that is scored externally.

Professional Development Requirements: Materials need to be developed that allow teachers to be trained in the content and pedagogy of the course. Professional development, however, is much more focused than with standardized tests. Teacher networks can be utilized, and teachers can share materials as a form of professional development.

Conditions of Administration: Normally in class, near the end of the school year, during class time. If the results are incorporated into the student's grade for the course, the exam becomes something like a final rather than a separate test.

Diagnostic Properties: Highly diagnostic, but not useful for students who take it because it is taken at the end of the year. Provides teachers with very detailed information on how their students did on a wide range of content that is specifically taught in the course.

Sensitivity to Changes in Instruction: Highly sensitive. If teacher deviates from specified content to any significant degree, student performance is likely to be affected. Conversely, teachers who maintain fidelity to course framework should generally witness better student results.

Effect on Teacher Instructional Practice: High degree of influence of test on instruction. As noted, teachers statewide who teach the course can share materials, lesson plans, etc., in ways that help them teach particular elements of the course better.

Adaptability to Reflect Local Practices and Priorities: Not highly adaptive. Because test is consistent regionally, statewide, or nationally it does not support local decisions on curriculum and textbooks that are inconsistent with the state course framework.

Accountability Uses: Can be used as a component of individual student grades, can provide information for course placement in high school and college, and allows comparisons among schools within a state.

Primary Limitations: Limited by number of students who take the course within the school and statewide. Schools may have an incentive to keep certain students out of courses for which a statewide end-of-course exam is mandated. Can lead to poor instruction if the test is not designed carefully to promote better instruction.

3) On-demand performance tasks

Description: Complex task designed to elicit a more complete and authentic student response that demonstrates knowledge and skills not easily captured and measured on standardized assessments. This can be a state task, such as writing prompt, math problem solving prompt, or science experiment.

Reliability: Varies based on the type of task and scoring method utilized. Sufficiently reliable scoring can be achieved on performance tasks. However, a related issue is task difficulty, which can be more problematic to make consistent across tasks and administrations.

Alignment with Industry-based Standards: High alignment with state standards can be achieved, particularly with standards that address more complex aspects of learning.

Validity as a Measure of What Is Taught in Individual Classrooms: Moderate to high, assuming teachers have aligned their instruction with state standards. However, some tasks assume teachers have skills to teach material in ways that require sophistication and attention to complex cognitive skill development, which may not always be the case.

Cost: Task development costs are moderate, but ongoing costs to score the tasks can be significant. However, costs can be controlled when scoring is done through virtual means where scorers do not have to be assembled physically. Some states and contractors are instituting such methods for writing samples in particular.

Professional Development Requirements: Moderate, particularly if the task expects teachers to develop student knowledge and skill in unfamiliar areas or in ways that teachers are not accustomed to teaching.

Conditions of Administration: Standardized in terms of when and how tasks are administered, generally within class during class time, although performance tasks on tests such as CTE student organization competitive events are administered under secure testing conditions.

Diagnostic Properties: Moderate due to time lag involved in scoring tasks and limited number of score points on the scoring guide. Can provide useful general information about overall student skill level on each scoring point.

Sensitivity to Changes in Instruction: Highly sensitive. If a teacher does not address the component content and skill elements of the task, the student has little chance of doing well based simply on test-taking savvy. A good performance task should be designed in a way that it is highly representative of instruction. However, a poorly designed task will not be reflective of instruction on the knowledge and skills being assessed.

Effect on Teacher Instructional Practice: Tends to have a high amount of influence because the task is administered in class, and it is difficult to ignore students' performance on the task. This does not mean that all teachers integrate the task into their teaching, only that they generally pay attention to it.

Adaptability to Reflect Local Practices and Priorities: Very low. The task must be administered consistently statewide for it to be scored consistently. This limits

dramatically the possibility of adapting the task to local conditions, curriculum, or teaching practices.

Accountability Uses: Can be a good independent measure of student learning or a correlate with another measure because the task in theory should be measuring something different than a standardized norm-referenced or criterion-referenced test. However, due to somewhat lower reliability, it is not as good of a measure of student-level performance.

Primary Limitations: Expensive, difficult to develop comparable tasks from year to year, scoring provides limited number of data points or dimensions, viewed as an outside imposition into classroom teaching, may not fit with teacher's existing curriculum and lesson plans, some teachers may not want to or be able to teach the desired skills.

4) Extended performance tasks

Description: A task that takes several weeks to complete and may involve multiple attempts but that must follow criteria, guidelines, and rules that are consistent across all tasks of the same type. Is scored using a common scoring tool. The Oregon diploma Extended Application may be an example of an extended performance task.

Reliability: Moderate. Scoring is done using a common scoring guide, but task prompts may vary. Students may produce different products, for example, while the scoring may be on the technical skills they utilized.

Alignment with Industry-based Standards: Potentially high. Individual tasks can be designed to respond to specific standards and content knowledge. The problem is that no individual task can cover more than a few standards.

Validity as a Measure of What Is Taught in Individual Classrooms: Variable, particularly if students select the topic. Can be higher if the task is standardized to a greater degree. Tends to be most valid as a measure of key cognitive skills the task emphasizes rather than of mastery of particular content knowledge because task content varies.

Cost: Low initial cost, mainly to develop procedures, scoring tools, sample topics. Low ongoing costs if tasks are scored locally, higher if all or some are scored off-site. It is not necessary to rescore all tasks off-site unless they are used for high stakes decisions.

Professional Development Requirements: Moderate to high. Teachers must know how to help students select topics or projects, must be able to organize instruction so that students are developing the necessary skills to complete the task (e.g., research and technical skills), and must be capable of scoring the task reliably.

Conditions of Administration: In the classroom as an assignment that is part of the course or that is conducted in a particular course. Semi-standardized in terms of conditions of administration.

Diagnostic Properties: Not designed to be diagnostic for individual students because the task tends to be culminating in nature, and the amount of time it takes to score the task means that students have often completed the course or school year by the time they receive their scores, which are relatively general in nature.

Sensitivity to Changes in Instruction: Highly sensitive. If key cognitive skills and strategies are not developed, students will be unable to undertake the task with any degree of success. Instruction must emphasize the key cognitive strategies and capabilities over a sustained period of time for students to be able to demonstrate them on the extended performance task.

Effect on Teacher Instructional Practice: Significant. Teachers must be committed to developing student abilities necessary to complete the extended performance task and to embed the task within their instruction. Teachers must provide scaffolding and support for students to complete the task, including periodic checking of work and help with access to necessary resources.

Adaptability to Reflect Local Practices and Priorities: Moderate. The topics or projects selected by students can be adapted significantly to the local context. However, the skills developed and the means to score student performance cannot.

Accountability Uses: Most useful in assessing cognitive skills and other more complex capabilities. Can provide useful information for postsecondary placement purposes as well as for accountability measures of school performance.

Primary Limitations: Complicated, requires teachers to be able to manage and support a process that takes weeks and includes significant classroom and independent student work. Requires a common scoring guide and state monitoring of quality control through the rescoring of student work (moderation). Considerable variance in the content knowledge basis for the task is likely to exist across all tasks if students are allowed to select their own topic or project.

5) Portfolios or Collections of Evidence

Description: Variation on moderated school-based assessment in which results from teacher scoring of classroom-based assessments are not necessarily moderated. Relies on the collection of evidence by classroom teachers who follow prescribed criteria to gather information on individual student performance, but can contain numerous other forms of evidence that are standards-based, such as criterion-referenced tests, performance demonstrations, and other measures of proficiency.

Reliability: Reliability of teacher scoring is a function of training and of the quality of scoring guides utilized. Reliability is low if teachers score on their own criteria or school-based criteria. Reliability increases if scoring standards are the same across a wider range of sites, districts, or state and teachers have access to training that includes example collections and annotated exemplars.

Alignment with Industry-based Standards: Potentially quite high if the collections must specifically contain evidence related to the standards. The collection has the potential to cover a much wider range of state standards than any on-demand assessment and to contain evidence of larger, more complex skills, such as higher-order cognitive strategies.

Validity as a Measure of What Is Taught in Individual Classrooms: Very high. The collection is almost a direct representation of what is taught in the classroom. Although teachers must be sure to address standards that must be included in the portfolio (if such standards are specified), the collection clearly expresses the assignments teachers give and, by inference, the instruction students receive in those areas.

Cost: Larger cost implications for districts, particularly in regard to training costs. Other issues include storage of student work, time for teachers to score items in the portfolio, and transcribing the elements of the portfolio so that information moves with the student across classes and grades.

Professional Development Requirements: Moderate to high. Teachers have to learn how to select student work that meets the requirements of the collection, how to score the work, and how to align instruction with the required components of the collection. They also need to develop necessary instructional strategies to enable students to produce quality work for the collection.

Conditions of Administration: Locally, in the classroom. The collection can be assembled in one classroom or in more than one classroom and in more than one subject over multiple years; for example, if the collection is going to be used in the graduation determination process. Collections work best when they are accompanied by guidelines on what is to be included, how much teachers can help students prepare the work, how much of the work must be done in class under direct teacher supervision, and other conditions of administration that guarantee uniformity and quality control.

Diagnostic Properties: Highly diagnostic. Students receive feedback on a range of skills and knowledge areas as they assemble the collection and have a continuous goal for their work in mind throughout the process.

Sensitivity to Changes in Instruction: Highly sensitive. Collections reflect changes in teacher instruction very quickly as students produce new pieces for the collection or improve existing pieces.

Effect on Teacher Instructional Practice: Significant. The collection becomes a focal point for instruction and classroom assignments.

Adaptability to Reflect Local Practices and Priorities: Can incorporate a wide range of instructional practices and priorities because teacher and student can select the specific pieces of work for the collection and there is no requirement that work necessarily be uniform in terms of the topics covered as long as standards are addressed.

Accountability Uses: Potential uses for classroom-based grading, promotion, college admission, employment and placement. Could also provide general data on a state level because all students could conceivably be required to assemble portfolios. The standards covered in each student portfolio would need to be uniform even if the evidence is not.

Primary Limitations: Very complex and hard for teachers to integrate into instruction without support and encouragement (or requirement). Most teachers do not have sufficient knowledge of assessment to design or score student work in a manner necessary to ensure the collection maintained its integrity. Teachers tend to view scoring process as extra work. Maintenance of data, transcribing, and other logistical issues often overwhelm local school administrators and teachers.

6) Industry Skill Certification Exams

Description: Examinations developed and administered by industry groups independently of school systems. Generally tests of knowledge and skill specific to the job classification.

Reliability: Highly reliable. Standardized tests. Can involve performance demonstrations in simulated work settings as well.

Alignment with Industry-based Standards: High. Overlap with industry-based standards in any given area, but may require mastery of a wider range of standards than those explicitly measured.

Validity as a Measure of What Is Taught in Individual Classrooms: Very high in Career/Technical Education courses that are well aligned with the exam.

Cost: Significant if the institution or state pays for the student to take the exam.

Professional Development Requirements: Significant for CTE teachers whose courses are supposed to align with the exam (sometimes industry groups provide this training to teachers).

Conditions of Administration: Standardized, often online, if not, then paper-and-pencil test or performance demonstration (e.g., NOCTI).

Diagnostic Properties: High. Provide good information on what the student knows related to performing the job or task the exam is designed to measure.

sensitivity to Changes in Instruction: No sensitivity to high sensitivity, depending on whether the teacher is in a course that leads to the exam. In such courses, curriculum must be carefully designed to prepare students for the exam. Otherwise, instruction is divorced from the exam.

Effect on Teacher Instructional Practice: Negligible to high, depending on alignment of course with exam.

Adaptability to Reflect Local Practices and Priorities: Low. Exams are standardized. Therefore, local classes should not be adapted.

Accountability Uses: Achieving certification in distinct occupational areas. Examples: network administrator, automotive technician, paramedic. Hundreds of such exams exist in a wide variety of areas.

Primary Limitations: Students taking these exams are well prepared for a particular job, many of which are well paying and lead to career tracks. However, the exams themselves do not reflect mastery of state standards per se.

The six types of assessment are excerpts of—

POLICY ANALYSIS: Oregon Student Assessment System

Presented to the State Board of Education for Initial Discussion on August 23, 2007

Full paper available at: <[Conley: Policy Analysis of Oregon Assessment System](#)>

Prepared by David T. Conley, Ph.D.

Professor, University of Oregon

Director, Center for Educational Policy Research

Chief Executive Officer, Educational Policy Improvement Center

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