# TYPES OF ASSESSMENT

Interest in alternative types of assessment has grown rapidly during the 1990s, both as a response to dissatisfaction with multiple-choice and other selected-response tests and as an element in a systemic strategy to improve student outcomes. Alternative assessments range from written essays to hands-on performance tasks to cumulative portfolios of diverse work products. This chapter describes four types of alternative assessment that might meet the needs of vocational educators and summarizes assessments in use in the cases selected for our study. The chapter concludes with a brief discussion of the advantages and disadvantages of different types of assessment.

## ALTERNATIVES TO SELECTED-RESPONSE ASSESSMENT

The most familiar form of assessment is one in which the test-taker is asked to select each response from a set of specified alternatives. Because the test-taker chooses an option rather than creating an answer from scratch, such an assessment is called a *selected-response* assessment. Such assessments include multiple-choice, matching, and true-false tests.

Alternatively, an assessment can require a student to develop his or her own answer in response to a stimulus, or prompt. An assessment of this form, such as one that requires an essay or a solution to a mathematical problem, is called a *constructed-response* assessment. Neither the prompts nor the responses need be written, however. Responses commonly include any form whose quality can be judged accurately, from live performances to accumulated work products.

For this reason, constructed-response assessments are also called *performance assessments*.  In our study, we also used the less technical term *alternative assessment* as a synonym for both of these terms.

A major distinguishing feature of all constructed-response assessments is that humans must score the responses.[1]  Someone must review each answer (be it an essay, performance, project, or portfolio), compare it to a standard, and decide whether it is acceptable. Human scoring is slower and more expensive than machine scoring. Furthermore, as the answers grow more complex, the scoring judgments are more difficult and subject to greater error.

There are a variety of ways to classify assessments (Hill and Larson, 1992; Herman, Aschbacher, and Winters, 1992).  In fact, since the range of constructed-response types and situations is limitless and more formats are being developed all the time, it is unlikely that there will be a single best system of classification.  For our purposes, we used categories developed by the National Center for Research in Vocational Education (NCRVE) that are clearly relevant to vocational educators (Rahn et al., 1995).  There are four major categories of assessment strategies:  written assessments, performance tasks, senior projects, and portfolios.  As Table 4 shows, the written assessment category includes both selected- and constructed-response assessments, whereas the other three categories involve only constructed-response assessments.

The classification system is based primarily on format—how the questions are presented and how responses are produced.  However, selected-response and constructed-response assessments differ in many other ways, including the complexity of their development, administration, and scoring; the time demands they place on students and teachers; their cost; and the cognitive demands they make on students.  These differences are explored in the remainder of this chapter and Chapter Four.

_____

[1]There have been recent advances in computerized scoring of constructed-response assessments, but these systems are still in the research phase and will not be widely available for years.

**Table 4**

**Broad Categories of Assessment**

| Category | Response Type | |
| --- | --- | --- |
| | Selected | Constructed |
| Written assessments | | |
|   Multiple choice, true-false, matching | ✓ | |
|   Open ended | | ✓ |
|   Essay, problem based, scenario | | ✓ |
| Performance tasks | | ✓ |
| Senior projects (research paper, project, oral presentation) | | ✓ |
| Portfolios | | ✓ |

## Written Assessments

Written assessments are activities in which the student selects or composes a response to a prompt. In most cases, the prompt consists of printed materials (a brief question, a collection of historical documents, graphic or tabular material, or a combination of these). However, it may also be an object, an event, or an experience. Student responses are usually produced "on demand," i.e., the respondent does the writing at a specified time and within a fixed amount of time. These constraints contribute to standardization of testing conditions, which increases the comparability of results across students or groups (a theme that is explored later in Chapters Four and Five).

Rahn et al. (1995) distinguish three types of written assessment, one of which involves selected responses and two of which involve constructed responses. The first type is multiple-choice tests,[2] which are commonly used for gathering information about knowledge of facts or the ability to perform specific operations (as in arithmetic). For example, in the Laborers-AGC programs, factual knowledge of environmental hazards and handling procedures is measured using multiple-choice tests. The Oklahoma testing program uses multiple-choice tests of occupational skills and knowledge derived from statewide job analyses.

---

[2]Matching and true-false tests are also selected-response written assessments.

Multiple-choice tests are quite efficient. Students answer numerous questions in a small amount of time. With the advent of optical mark sensors, responses can be scored and reported extremely quickly and inexpensively. Such tests provide an efficient means of gathering information about a wide range of knowledge and skills. Multiple-choice tests are not restricted to factual knowledge; they can also be used to measure many kinds of higher-order thinking and problem-solving skills. However, considerable skill is required to develop test items that measure analysis, evaluation, and other higher cognitive skills.

The other two types of written assessment both involve constructed responses. The first consists of open-ended questions requiring short written answers. The required answer might be a word or phrase (such as the name of a particular piece of equipment), a sentence or two (such as a description of the steps in a specific procedure), or a longer written response (such as an explanation of how to apply particular knowledge or skills to a situation). In the simplest case, short-answer questions make very limited cognitive demands, asking students to produce specific knowledge or facts. In other cases, open-ended assessments can be used to test more complex reasoning, such as logical thinking, interpretation, or analysis.

The second type of constructed-response written assessment includes essays, problem-based examinations, and scenarios. These items are like open-ended questions, except that they typically extend the demands made on students to include more complex situations, more difficult reasoning, and higher levels of understanding. Essays are familiar to most educators; they are lengthy written responses that can be scored in terms of content and/or conventions. Problem-based examinations include mathematical word problems and more open-ended challenges based on real-life situations that require students to apply their knowledge and skills to new settings. For example, in KIRIS, groups of three or four twelfth-grade students were given a problem about a Pep Club fund-raising sale in which they were asked to analyze the data, present their findings in graphical form, and make a recommendation about whether the event should be continued in the future. Scenarios are similar to problem-based examinations, but the setting is described in greater detail and the problem may be less well formed, calling for greater creativity. An example is the scenario portion of C-TAP, which requires students

to write an essay evaluating a real-life situation and proposing a solution (such as determining why a calf is sick and proposing a cure).

## Performance Tasks

Performance tasks are hands-on activities that require students to demonstrate their ability to perform certain actions. This category of assessment covers an extremely wide range of behaviors, including designing products or experiments, gathering information, tabulating and analyzing data, interpreting results, and preparing reports or presentations. In the vocational context, performance tasks might include diagnosing a patient's condition based on a case study, planning and preparing a nutritionally balanced meal for a vegetarian, or identifying computer problems in an office and fixing them. Performance tasks are particularly attractive to vocational educators because they can be used to simulate real occupational settings and demands. Our cases included many examples of performance tasks. For instance, each Oklahoma vocational student had to complete two tasks designed and scored by his or her teachers. The VICA competitions primarily involved lifelike simulations, such as an emergency team responding to an accident victim.

The skills that must be demonstrated in performance tasks can vary considerably. Some tasks may demand that a student demonstrate his or her abilities in a straightforward way, much as was practiced in class (e.g., adjusting the spark plug gap). One health trainee assessment involved changing hospital bed sheets while the bed was occupied, a skill that participants had practiced frequently. Other tasks may present situations demanding that a student determine how to apply his or her learning in an unfamiliar context (e.g., figuring out what is causing an engine to run roughly). Teachers participating in the NBPTS certification process must respond to unanticipated instructional challenges presented during a day-long series of assessment exercises.

As assessments become more open ended and student responses become more complex, scoring grows more difficult. A variety of methods have been developed to score complex student performances, including both holistic and analytic approaches. In some cases, students are assessed directly on their performance; in other cases, assessment is based on a final product or oral presentation. For

example, in the VICA culinary arts contest, students prepare platters of cold food and a multicourse meal of cooked food using the ingredients and equipment provided. Judges assess both the procedures used (by rating organizational skills, sanitation, and safety) and the final product (by rating presentation and taste). Similarly, in the KIRIS interdisciplinary performance events, students work together in groups on open-ended activities and then produce individual products. The group work is not judged, just the individual responses.

Traditionally, vocational educators have relied on performance-based assessment strategies to judge student mastery of job-specific skills. For example, an automotives teacher judges whether a student can change the oil of a car by asking him or her to perform the task. However, other strategies may be required if that teacher wants to assess a student's ability to understand the technical principles underlying an automotive engine.

Recently, researchers have developed performance tasks that can be administered and scored by computer. Such computer-based performance assessment systems are in the experimental stage, but the results of recent research are promising. Vocational educators may be able to add computer-based tools to their list of assessment alternatives in the not too distant future.

Two types of computerized assessment tools deserve attention. First, computers are being used to simulate interactive, real-world problems. For example, O'Neil, Allred, and Dennis (1992) developed a simulation of negotiation skills in which students interact with a computer as if they were negotiating with another individual. The researchers found strong evidence that the simulation provided a valid measure of interpersonal negotiation skills within the workplace context. It is easy to imagine other occupational skills that might be assessed using computer simulations. Second, expert computer systems are being developed that can review and score constructed responses. For example, Bennett and Sebrechts (1996) developed a computer system that scored student responses to algebra word problems. This system was as accurate as human judges in determining the correctness or incorrectness of student responses, although it was less effective in classifying student errors. Similar prototype systems have been used to score answers to

programming problems, to analyze architectural design problems, and to identify student misconceptions in subtraction (Bennett and Sebrechts, 1996).

Although these results are encouraging, it will take considerable time before computer-based assessment tools are widely available.  None of the cases we studied used computer-based assessments, and, with the exception of this brief look at the topic, we did not include them in our analyses.  If this study were reconducted five years from now, we would expect much more attention to be given to these alternatives.

## Senior Projects

Senior projects are distinct from written assessments and performance tasks because they are cumulative, i.e., they reflect work done over an extended period rather than in response to a particular prompt.  The term *senior project* is used here to identify a particular type of culminating event in which students draw upon the skills they have developed over time.  It has three components:  a research paper, a product or activity, and an oral presentation, all associated with a single career-related theme or topic.  The format is designed to be motivating, to permit involvement of people from business or community, and to encourage integration of academic and vocational ideas.  For this reason, the process of implementing senior projects in a school often involves collaboration between teachers in many subjects who agree to guide the student's selection and accept the work for credit in more than one course.

All three components of a senior project are organized around a single subject or theme, such as a traditional method of making furniture, the creation of an appealing store window display, or a fashion show.  To complete the research paper, the student must conduct research about aspects of the subject he or she has not previously studied.  The student draws upon library and other resources and produces a formal written paper.  The student then creates a product or conducts an activity relevant to the subject.  This might include making something or doing community volunteer work for an extended period and documenting it.  The purpose is to demonstrate knowledge or skills relevant to the subject.  Finally, the student presents his or her work orally to a committee or at a public forum.

The length and complexity of the senior project make evaluation difficult. Schools that have implemented this type of assessment have spent a fair amount of time deciding how to judge the quality of the various elements. Their scoring guides reflect concerns about content, technical knowledge, organization and time management, the extension of knowledge outside traditional school domains, communication skills, and even appearance (Rahn et al., 1995, p. U3-12). These all involve subjective judgments, so great care must be taken to ensure that scores are accurate and meaningful.

## Portfolios

Like a senior project, a portfolio is a cumulative assessment that represents a student's work and documents his or her performance. However, whereas a senior project focuses on a single theme, a portfolio may contain any of the forms of assessments described above plus additional materials such as work samples, official records, and student-written information. For example, in the C-TAP portfolio, students not only provide an artifact (or evidence of one if it is not portable) but give a class presentation that is evaluated as part of their project. Records may include transcripts, certificates, grades, recommendations, resumes, and journals. Portfolios also often contain a letter of introduction to the reader from the student explaining why each piece has been included. They may contain career development materials, letters from supervisors or employers, completed job applications, test results, and samples of work products. The contents may reflect academic accomplishment, industrial or career-related accomplishments, and personal skills.

Some portfolios are designed to represent the student's best work, others are designed to show how the student's work has evolved over time, and still others are comprehensive repositories for all the student's work. Both the KIRIS portfolios (for writing and mathematics) and the C-TAP portfolios (for a vocational area) are built around a selection of the student's best work. The C-TAP portfolio adds other types of assessment such as records (a resume) and a work artifact (a writing sample).

Portfolios present major scoring problems because each student includes different pieces. This variation makes it difficult to develop

scoring criteria that can be applied consistently from one piece to the next and from one portfolio to the next. States that have begun to use portfolios on a large scale have had difficulty achieving acceptable quality in their scoring (Stecher and Herman, 1997), but they are making progress in this direction. One approach is to set guidelines for the contents of the portfolios so that they all contain similar components. Specific learner outcomes can be identified for each component and then techniques can be developed for assessing student performance in terms of these outcomes.

Table 5 shows the range of assessment types being used in the sites selected for this study.

## COMPARING SELECTED-RESPONSE AND ALTERNATIVE ASSESSMENTS

For decades, selected-response tests (multiple-choice, matching, and true-false) have been the preferred technique for measuring student achievement, particularly in large-scale testing programs. In one form or another, selected-response measures have been used on a large scale for seventy-five years. Psychometricians have developed an extensive theory of multiple-choice testing, and test developers have accumulated a wealth of practical expertise with this form of assessment.

Nevertheless, there are limitations to using multiple-choice and other selected-response measures. First, these traditional forms of assessment may not measure certain kinds of knowledge and skills effectively. For example, it is difficult to measure writing ability with a multiple-choice test. Similarly, a teacher using cooperative learning arrangements in a classroom may find that selected-response measures cannot address many of the learning outcomes that are part of the unit, including teamwork, strategic planning, and oral communication skills. In these cases, multiple-choice tests can only provide indirect measures of the desired skills or abilities (e.g., knowledge of subject-verb agreement, capitalization, and punctuation, and the ability to recognize errors in text may serve as surrogates for a direct writing task). Users of the test results must make an inference from the score to the desired domain of performance.

**Table 5**

**Types of Assessments in the Sample**

| Assessment System | Selected Response Multiple Choice | Constructed Response Written Open Ended | Written Essay, etc. | Performance | Senior Project | Portfolio |
|---|---|---|---|---|---|---|
| Career-Technical Assessment Program (C-TAP) | | | ✓ | ✓ | | ✓ |
| Kentucky Instructional Results Information System (KIRIS) | | | ✓ | ✓ | | ✓ |
| Laborers-AGC environmental training and certification programs | ✓ | | | ✓ | | |
| National Board for Professional Teaching Standards (NBPTS) certification program | | | ✓ | ✓ | | ✓ |
| Oklahoma competency-based testing program | ✓ | | | ✓ | | |
| Vocational /Industrial Clubs of America (VICA) national competition | ✓ | | | ✓ | | |

Second, when used in high-stakes assessment programs, multiple-choice tests can have adverse effects on curriculum and instruction. Many standardized multiple-choice tests are designed to provide information about specific academic skills and knowledge. When teachers focus on raising test scores, they may emphasize drill, practice, and memorization without regard to the students' ability to transfer or integrate this knowledge. Instruction may focus on narrow content and skills instead of broader areas, such as critical thinking and problem solving (Miller and Legg, 1993). In addition, many think multiple-choice tests emphasize the wrong behaviors

given that few people are faced with multiple-choice situations in their home or work lives (Wiggins, 1989).

During the past few years, constructed-response assessment approaches have gained popularity as tools for classroom assessment and large-scale use. Proponents of alternative forms of assessment believe they will alleviate some of the problems presented by multiple-choice tests. It is easier to measure a broader range of skills and ability using constructed-response approaches than selected-response approaches. To measure writing ability, one asks students to write; to test oral communication, one has students give oral reports. In addition, alternative assessments permit the use of complex, realistic problems instead of the narrow or decontextualized problems that appear on many multiple-choice tests. Because of this, teaching to alternative assessments is desirable, because good test preparation will be good instruction.

However, alternative assessments are not without problems. In fact, they may have many of the same flaws cited for multiple-choice tests. Critics argue that poorly designed alternative assessments can also be very narrow, so that teaching to them may also be undesirable. For example, mathematics portfolios may overemphasize "writing about mathematics" at the expense of learning mathematical procedures. In addition, alternative assessments have practical problems, including high cost, administrative complexity, low technical quality, and questionable legal defensibility (Mehrens, 1992). These flaws are of greatest concern when assessments are used to certify individuals for work or to reward or sanction people or systems. (These issues are discussed in greater detail in Chapters Four and Five.) Table 6 compares selected-response and constructed-response measures in terms of a number of important features.

**Table 6**

**Features of Selected- and Constructed-Response Measures**

| Feature | Selected Response | | | Constructed Response | | |
|---|---|---|---|---|---|---|
|  | Rarely | Some-times | Usually | Rarely | Some-times | Usually |
| Easy to develop |  |  | ✓ | ✓ |  |  |
| Easy to administer |  |  | ✓ |  | ✓ |  |
| Easy to score |  |  | ✓ | ✓ |  |  |
| Similar to real world in per-formance demands ("authentic") | ✓ |  |  |  |  | ✓ |
| Efficient (requires limited time) |  |  | ✓ | ✓ |  |  |
| Credible to stakeholders |  |  | ✓ |  | ✓ |  |
| Embodies desired learning activities | ✓ |  |  |  |  | ✓ |
| Sound basis for determining quality of scores |  |  | ✓ |  | ✓ |  |
| Effective for factual knowledge |  |  | ✓ |  |  | ✓ |
| Effective for complex cognitive skills (e.g., problem solving) |  | ✓ |  |  |  | ✓ |