

# 5

## DIRECT ASSESSMENT TECHNIQUES

**F**aculty can use a variety of methods to assess student learning. The next three chapters will review many strategies and should provide sufficient information for informed judgements concerning their use. As you read about each potential strategy, think about how you might use it to assess your learning objectives and related program characteristics. Strategies vary in complexity and in demands on student and faculty time. An assessment study may be as simple as including a question on an exam or as complicated as conducting multiple focus groups. As you plan assessment activities, consider practical constraints and select strategies that are realistic, focused, and manageable.

Direct assessments often involve quantitative measurements, and experts have developed a number of techniques for developing measurement procedures and evaluating their quality. Although good assessment can be conducted without expertise in such matters, faculty should be aware that a significant body of work on measurement theory has been developed, and they should seek advice when needed. Most campuses have faculty or staff with expertise in testing and measurement.

### PUBLISHED TESTS

An easy response to an assessment mandate is to use already existing, professionally developed tests. Most faculty have taken tests like the

Scholastic Aptitude Test (SAT; <http://www.collegeboard.com/>) and the American College Test (ACT; <http://www.act.org/aap/index.html>). These are called *standardized tests* because all students take them under identical conditions.

Many published tests exist, and most campuses have testing officers who have publishers' catalogs and related information. Most published tests were not created as program assessment instruments, so faculty should carefully consider them before using them for that purpose. For example, Graduate Record Examinations (GRE; <http://www.gre.org/>) were designed for making graduate admissions decisions, and items were created to identify high achievers. Such tests are not particularly sensitive to differences among average or below average students. The GRE General Test measures verbal, quantitative, and analytical skills. The GRE Subject Tests are available for specific disciplines, such as biochemistry, English literature, and psychology.

A variety of other tests for specific programs are available. For example, Major Field Tests were created to assess undergraduate student achievement in a number of disciplines, such as biology, business, chemistry, and history (<http://www.ets.org/hea/mft>). The Praxis Series tests (<http://www.ets.org/praxis/index.html>) are designed to assess aspects of teacher competence, and the publisher recommends their use by teacher education programs and state agencies responsible for certifying teachers.

Some standardized tests have been used to assess general education, such as the Academic Profile and Collegiate Assessment of Academic Proficiency (CAAP) tests. The Academic Profile (<http://www.ets.org/hea/acpro>) was designed to measure "college-level reading, critical thinking, writing, and mathematics in the context of material from the humanities, social sciences, and natural sciences." The CAAP (<http://www.act.org/caap/index.html>) test is "designed to help institutions measure the academic achievement levels of their students in selected core academic skills" (writing, reading, math, science reasoning, and critical thinking).

Computers have been used to score objective tests for a long time, and publishers are developing and using software that scores essay responses. Such programs are likely to become more common as test publishers reduce their reliance on multiple-choice items and collect more writing samples. Although expensive to develop, these programs

allow publishers to provide immediate scores to online test takers and reduce reliance on trained readers. ACCUPLACER (<http://www.collegeboard.com/highered/apr/accu/accu.html>) is designed to measure incoming students' competence in reading, writing, and mathematics. This test assesses writing samples with sophisticated software that "knows" when it can generate a score and when it should refer a test to a human for scoring. Research shows that ACCUPLACER's computer-generated scores correlate highly with scores generated by human readers, but sophisticated test takers can generate meaningless answers that can fool the software (Holst & Elliott, 2002). COMPASS e-Write (<http://www.act.org/e-write/index.html>) is another online writing test that provides electronic scoring of writing samples.

Faculty who are considering the use of a standardized test should review a *specimen set* (copy of the test, test manual, and other materials) and should consider professional reviews of the test, such as those available in the *Mental Measurements Yearbook* (Impara & Plake, 2001). A primary question concerns the alignment of the test with the program objectives and curriculum. Faculty might create an alignment matrix that relates test items or test subscales to their learning objectives, and they should determine if items are current and at the desired depth of processing. Sometimes tests only provide a single score, and this score may be too broad for effective assessment. Practicalities also cannot be ignored. Faculty should consider how much the test costs, if proctors require special training, how much time is needed for testing and scoring, and how they will motivate students to take the test seriously. For example, if they plan to give tests within courses that are scheduled for 50 minutes, they should verify that the test can be administered, completed, and collected in that time; and they should have some confidence that students will be motivated to demonstrate the extent of their learning. Another important consideration is faculty acceptance of the test. If faculty will not act on results, the test will not be useful.

Students are asked to take many tests during their college years, and they have the right to be treated fairly. The American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education created a Joint Committee on Testing Practices, and in 1994 this committee developed a Code of Fair Testing Practices in Education. This code specifies the

obligations of test developers and test users of educational tests, such as tests used for admissions, assessment, diagnosis, and placement. Assessment efforts often make use of data used for other purposes, such as admissions or course grading. This code was not meant to apply to routine classroom exams, but if faculty embed assessment within these exams, they should be aware of its principles. Many of these guidelines are reasonable for any student testing, but much of the concern in the Code is over *high-stakes testing*, such as testing that determines if individual students can be admitted into college or if remedial courses will be required.

The Code specifies that test developers should “strive for fairness” for all test takers. They should avoid content or language that might be offensive to individual test takers, eliminate test items that unfairly discriminate against groups of test takers, ensure that tests are valid, and provide reasonable accommodations to students with special needs. Test publishers are expected to provide information needed to make informed judgments to those who select and use tests. This includes explicit statements of what is being measured, who can be measured fairly, test development procedures, evidence of test effectiveness for recommended uses, limitations affecting test use, specialized skills needed for test administration and scoring, norm group characteristics, information needed to interpret scores accurately, and warnings about possible misuses of test scores. Test publishers are expected to provide specimen sets to professionals who select tests to enable them to make independent judgments concerning the appropriateness of the test for their use.

Published tests, like all assessment strategies, have strengths and limitations. They generally are carefully developed, highly reliable, professionally scored, and nationally normed. They frequently provide a number of norm groups, such as norms for community colleges, liberal arts colleges, and comprehensive universities. Online versions of tests are increasingly available, and some provide immediate scoring. In addition, some publishers allow faculty to supplement tests with their own items, so tests can be adapted to better serve local needs.

The Conference on College Composition and Communication Committee on Assessment (1995) developed a policy statement concerning the assessment of writing skills. They expressed concern about standard-

ized exams, especially those that do not examine writing samples: “. . . choosing a correct response from a set of possible answers is not composing” (Assumptions section, ¶ 8). In addition, they warn that reliance on such tests might mislead students into thinking that good writing is accomplished quickly and conforms to stylistic and grammatical rules without concern for meaning. They suggest that funds allocated for writing assessment might be better spent compensating local readers than by purchasing published tests because faculty development and curriculum reform are inevitable when faculty work together to do assessment.

Published tests are not useful as direct measures for program assessment if they do not align with program learning objectives. Most standardized tests rely heavily on multiple-choice items which often focus on specific facts, but program learning objectives more often emphasize higher-level skills. In addition, local curricula may not include coverage of the relevant content. If the test does not reflect the learning objectives that faculty value and the curricula that students experience, results are likely to be discounted and inconsequential.

Student motivation can be a problem. Staff at some campuses tell horror stories about expensive testing programs that yielded problematic data. Students were enticed to take exams by promises of free pizza or t-shirts, and some responded by randomly filling in answer sheets or completing three-hour exams in minutes. Possible solutions include providing scores to individuals so they can understand their relative strengths, offering prizes to students or groups of students with the highest scores, embedding tests in capstone courses, and making a passing score a graduation requirement.

Standardized tests also cost money, and department budgets may be stretched by ongoing reliance on them. If faculty determine that it is important to compare students to national norms, standardized tests are valuable tools, but the marginal gain from annual testing is questionable. Their occasional use to evaluate the impact of curricular change may be more reasonable, especially if faculty have the goal of helping graduates become more competitive on standardized exams. Although some object to “teaching to the test,” this criticism is less reasonable if the test measures the program’s learning objectives.

## LOCALLY DEVELOPED TESTS

Probably the best way to ensure that tests tap program learning objectives is to create them locally. Faculty could create single test items or groups of items that they embed within course exams, or they could create entire tests that are administered to groups of students, such as seniors in capstone courses. Figure 5.1 describes an ongoing project that uses locally developed exams to assess quantitative skills at University of Wisconsin-Madison. This example illustrates how assessment projects can serve multiple functions and can benefit current and future students.

Traditional test items have essay, multiple-choice, true-false, matching, or completion formats. Figure 5.2 summarizes some characteristics of these formats and suggestions for creating good items (Davis, 1993; McKeachie, 1999; Miller & Miller, 1997; Nilson, 1998). A common distinction is between items that require *recognition* and items that require *recall*. Multiple-choice and matching items require recognition of right answers. Completion items and essay questions require recall because test takers must generate the answer on their own. Faculty who create the test should consider if their interest is in assessing recognition, recall, or deeper levels of processing. Another common distinction is between *speed* and *power* tests. Speed tests measure how quickly students can do simple things, and power tests measure deep processing. The goal of testing is to assess how well students have mastered learning objectives. If higher-order thinking skills are being examined, power tests should be used because students will need time to reflect on and compose their answers. Test questions should be phrased in simple, direct language to ensure that students understand what is being asked of them, and tests should align with courses and curricula to guarantee that students have been exposed to appropriate learning opportunities. In addition, test makers should develop tests that have sampling validity; that is, the test items should cover the entire range of interest, rather than isolated segments.

Traditional testing methods have been criticized, especially their reliance on multiple-choice formats. For example, the Association of American Colleges and Universities (2002) warns that "Multiple choice tests, in particular, provide little evidence of the analytical power, creativity, resourcefulness, empathy, and abilities to apply knowledge and

### FIGURE 5.1 EMBEDDED ASSESSMENT OF QUANTITATIVE SKILLS

Faculty members in the Quantitative Assessment Committee at the University of Wisconsin-Madison found a useful way to help faculty and students while doing assessment (Robbin & Alvarez-Adem, 2001). Each semester they invite faculty who teach courses that build on a mathematical foundation to identify mathematical or statistical skills essential for successful course completion. Then they create a competency exam based on these identified skills, and faculty administer it early in the semester. Exams are scored by mathematics graduate students, and they provide formative feedback to each student and aggregated results to each faculty member. Many of the involved courses are in the general education program, such as economics, mathematics, and physics courses, but they also provide this service to faculty who teach junior-level major courses that have a quantitative basis, including courses in engineering, nursing, and the social sciences. This is not a small operation. Tests were administered to over 2,000 students in 16 courses during the 2000-2001 academic year.

Results have been enlightening for faculty. For example, they found that nearly half the students in teacher-preparation mathematics courses could not multiply and divide numbers with decimals, and this led to immediate curricular changes. Faculty also have discovered other uses for their tests. For example, food science faculty began with a test created for a required major course, then used it as a pretest for all new students. They provide a follow-up workshop for students with identified problems.

This effort has been sustained since 1990 and continues to grow each year because faculty find it useful. Faculty who teach the basic quantitative methods courses use project results to assess their program, and they have revised courses and refocused their teaching to reduce identified problems. Faculty on the Quantitative Assessment Committee are pleased that faculty outside of the mathematics department are becoming more aware of the mathematical basis for their courses and are examining the validity of their assumptions about students' incoming skills. There has also been increased, fruitful communication between mathematics faculty and faculty in other disciplines on how to develop student mathematical competence.

transfer skills from one environment to another that students will need for college success" (Barriers to Readiness section, ¶ 6), and they remind us that "Learning is more than the acquisition of discrete facts. . . . students need to know facts, but even more importantly how to interpret and what to do with those facts" (Barriers to Readiness section, ¶ 4).

Wiggins (1998) argues that "Conventional test questions, be they from national tests or the teacher down the hall, do not replicate the kinds of challenges that adults face in the workplace, in civic affairs, or

FIGURE 5.2  
COMMON TEST ITEM FORMATS

Item Type	Characteristics and Suggestions
Completion	<p>These items require students to fill-in-the-blank with appropriate terms or phrases. They appear to be best for testing vocabulary and basic knowledge, and they avoid giving students credit for guessing by requiring recall, rather than recognition. Scoring can be difficult if more than one answer could be correct. When writing these items, avoid giving cues to correct responses (such as using "a" versus "an" or "was" versus "were" and providing varying amounts of space for answers); create questions with answers that are in the middle or end of the item, rather than at the beginning; and avoid textbook quotes that lack meaning out of context.</p>
Essay	<p>Essay questions are very popular and can be used to assess higher-order thinking skills. They generally ask for explanations and justifications, rather than memorized lists. Key words in essay questions are <i>summarize, evaluate, contrast, explain, describe, define, compare, discuss, criticize, justify, trace, interpret, prove, and illustrate</i> (Moss &amp; Holder, 1988). Avoid questions that are too broad, for example, "Write everything you know about the Civil War." A format that you might find useful is to assign students a role, a task, and an audience. For example, a question might ask students to be an expert in forensic psychology (the role) who is to appear before a parole board of intelligent laymen (the audience) to explain psychological factors associated with recidivism (the task). Such questions can provide data for the authentic assessment of important learning objectives.</p>
Matching	<p>Usually these questions are presented as two columns, and students are required to associate elements in column B with elements in column A. Such items are easy to score, but they are relatively difficult to construct and they seem best suited for testing knowledge of factual information, rather than deeper levels of understanding. If each answer can be used only once, students can use elimination to select answers without knowing the material. When writing these items, keep the elements in the columns short, list the elements in a logical order (e.g., alphabetical or numerical order), tell students if responses can be used more than once, keep all segments of the question on the same page of the test, and consider including some extra elements in column B to reduce guessing by elimination.</p>

Item Type	Characteristics and Suggestions
Multiple-choice	<p>Multiple-choice questions are popular because they can measure many concepts in a short period of time, and they generally are better than other objective questions at assessing higher-order thinking. They are easy to score, and item banks associated with popular textbooks often are available. Writing good items takes time, and there is strong temptation to emphasize facts, rather than understanding. Multiple-choice items have three parts: the stem, the correct answer, and the incorrect alternatives (distractors). When writing items, create stems that are positively phrased (so double negatives do not confuse test takers), that express a complete thought, that avoid unnecessary detail, and that avoid giving clues (such as using "a" versus "an"). Avoid redundant phrasing in response options by integrating common phrasing into the stem. If response options have a natural order (e.g., from small to large), give them in order so students can easily see the range. Distractors should be designed to attract students who have common misconceptions. They should not be written to "trick" students into selecting the wrong answer.</p>
True-false	<p>True-false items are relatively easy to construct and grade, but they appear to be best at assessing factual knowledge, rather than deep understanding. When writing true-false questions, avoid cues. For example, "usually" and "often" generally suggest true statements, and "always" and "every" generally suggest false statements. Avoid textbook quotes which lack meaning out of context, and avoid response patterns. For example, a true answer should just as likely follow a true item as a false item.</p>

in their personal lives" (p. 22). He argues that faculty should embed more authentic testing in courses to provide ongoing, formative feedback on student progress toward understanding what they are learning. Figure 5.3 contrasts authentic and traditional tests. Wiggins does not argue for the elimination of traditional tests, but he strongly advocates less reliance on them. In addition, he acknowledges that traditional tests may have strong validity if they correlate highly with authentic measures, but they generally lack the formative and motivational benefits of authentic tests.

Authentic-style assessments need not be complicated. For example, faculty at Mary Washington College (2002) assess aspects of information

FIGURE 5.3  
AUTHENTIC VERSUS TRADITIONAL TESTS

Authentic Test	Traditional Test
Requires students to create solutions to complex, real-world problems by integrating and applying what they have learned.	More often requires students to recall or recognize correct answers.
Usually involves a single, complex task.	Test usually is composed of items that are unrelated to each other.
Can provide direct evidence of student mastery of complex learning objectives.	Usually provides indirect evidence of student mastery of complex learning objectives.
Scoring requires subjective judgment.	Scoring more often is mechanical.
Variety of answers may be acceptable, although some may be better than others.	Usually there is one correct answer.
Expectations and criteria may be known in advance; e.g., a rubric may be provided to students.	Scoring criteria generally are not known in advance.
Generally, formative feedback is provided to students.	More often summative feedback is provided to students, telling them what they don't know, rather than how to improve.
Opportunities may be available to redo or revise the product.	Usually there is one-chance testing, with no opportunity for revision.
May occur in a complex environment, such as a field placement or laboratory.	Usually done as a timed, paper-and-pencil test in a classroom.
May involve students working cooperatively or with other colleagues to construct a solution.	Usually students take the test by themselves.
May encourage deeper learning.	May encourage memorization and "cramming."

literacy by asking students to respond to email messages and to make prescribed changes in word processor and spreadsheet documents. Class assignments and course activities also provide opportunities to integrate authentic assessment into courses, and these will be described in the next section.

Locally developed tests have a number of strengths. They allow faculty to explicitly tie assessments to program objectives, and appropriate mixes of items allow this to be done efficiently. These tests are likely to be less reliable than published tests, but, if well constructed, they are likely to have good validity. Because local faculty write the test, they should be interested in the results and willing to use them, and the discussion of results should easily lead to reflection on student learning and program support for it.

Norm groups usually are not available for locally developed tests, but faculty from campuses with similar missions could cooperate to develop their own norms, and they could assess student work together or provide independent assessment of each other's student work. This might add to the credibility of findings because unbiased "outsiders" have contributed scoring decisions or benchmarking information. Creating and scoring exams does take time, but if exams are embedded within courses, this is time already included in routine faculty workloads. As with standardized tests, student motivation is important so that students display the extent of their learning.

### EMBEDDED ASSIGNMENTS AND COURSE ACTIVITIES

Standardized and locally developed tests generally are given to groups of students simultaneously, frequently with time restrictions, but assessments also can be embedded as in-class activities or homework assignments. For example, embedded assignments and course activities could be integrated into:

- Classroom assessment activities.
- Community service learning and other fieldwork activities.
- Culminating projects, such as senior theses and papers in capstone courses.
- Group projects and presentations.

- Homework assignments.
- In-class presentations.
- In-class writing assignments.
- Poster presentations and student research conferences.
- Senior recitals and exhibitions.

These assignments might provide authentic assessment of important learning objectives, especially if case studies (Honan & Rule, 2002), problem-based learning (Duch, Groh, & Allen, 2001), or other real-world activities are involved. Students taking community service learning or fieldwork classes (Heffernan, 2001) often are required to analyze relationships between their academic learning and what they learn in their placements, directly demonstrating their mastery of relevant learning objectives. In addition, community supervisors could be asked to assess students' abilities to deal effectively with clients, to respect ethical standards, and to communicate professionally; and their involvement should help them make better contributions to the cohesive curriculum because they are aware of its learning objectives.

Assessment projects can be routinely integrated into specific courses, and, like most embedded assessments, they can serve multiple functions. For example, the management program at the Central Missouri State University's Harmon College of Business Administration gives a battery of tests to students in their orientation course and again in their capstone course. Students learn about program objectives early in their academic careers, and they are given individual feedback on their entering skills so they can focus on identified deficiencies (Palomba & Palomba, 2001). This strategy allows faculty to examine both value-added and absolute attainment components of student learning.

Major assignments, such as senior projects or theses, can provide valuable assessment data. Assignments that require oral presentations or group work allow faculty to assess content mastery, and they also provide opportunities to assess communication and interpersonal skills. Some programs routinely have senior recitals or exhibitions, and these provide excellent assessment opportunities. For example, the studio arts program in the University of Colorado at Boulder's Department of Fine Arts (2000) invited two external reviewers to evaluate displays at their

annual exhibition of graduating seniors' work, turning this celebratory event into an opportunity to collect assessment data.

It is hard to imagine a college or university program that does not have some learning objectives related to students' written communication skills, and these generally are assessed using locally developed writing assignments. The Conference on College Composition and Communication Committee on Assessment (1995) position statement stresses the need to collect writing samples based on reasonable writing assignments for the students being assessed, to evaluate student papers fairly within their social context, and to recognize that writing is a complex process, so the quality of an individual's writing is expected to vary across writing assignments. Their position statement provides specific recommendations, such as collecting more than one writing sample for each student and providing students sufficient time to draft and edit documents. They argue that faculty should assume responsibility for defining and field testing writing tasks, developing scoring guides and reader training procedures, assessing the documents, and using results to improve curriculum and pedagogy.

Angelo and Cross (1993) describe a variety of classroom assessment techniques, and many could be adapted as embedded assessments. For example, the "one-sentence summary" asks students to briefly summarize an important concept, and this technique could be used to examine students' understanding of major concepts, events, or issues in the discipline. Students might be asked to write a one-sentence summary of the American constitution, the concept of "gravity," or Jungian theory. Another technique, "empty outlines," asks students to complete a partially filled out or blank outline that ties key concepts together, and this could be used to analyze students' overall understanding of the discipline. A "background knowledge probe" or "misperception/preconception check" could be used at the beginning of courses to analyze students' mastery of assumed content and skills, and they could be used again to verify progress. Although created to provide formative assessment for faculty who teach specific courses, such embedded assessments could provide quick program assessment information.

What differentiates embedded assessments from other class activities is that they are designed to collect information on specific program learning objectives. In addition, results are pooled across courses and instructors to indicate program accomplishments, not just the learning of students in specific course sections.

These assignments generally are graded, as usual, by course instructors. Individual faculty probably would vary in the criteria they apply when assigning grades, giving more credit for course-specific learning. Copies of student products generally are collected for later assessment. Sometimes small programs accumulate copies for a year or two before analyzing them to ensure that sufficient numbers of materials are examined. Someone usually removes identifying information for students and faculty, and reviewers analyze the work to assess specific objectives. Faculty generally develop specific scoring criteria targeting the learning objectives so readers focus on assessment, rather than grading. As with locally developed exams, faculty on campuses with similar missions might work together to assess materials and to develop norms.

Walvoord and Anderson (1998) suggest an alternative data collection strategy. Faculty develop a common scheme for assessing elements of embedded assignments, do the assessment as they grade, and pool these assessment data across courses. For example, program faculty could adopt a common scoring rubric for assessing critical thinking or written communication skills. Data could be accumulated and analyzed periodically to assess program objectives and to identify trends or responses to curricular changes.

Embedding assessments within course activities and assignments is good practice. Learning can be assessed using direct measures that are aligned with program objectives, and students generally are motivated to show the extent of their learning. As with locally developed tests, faculty who are involved in creating the assessment process are likely to be interested in the results and willing to use them. These assignments should fit easily into courses that are aligned with relevant program learning objectives, and grades should provide important feedback to students. Faculty discussion of scoring criteria should increase their common understanding of program objectives, and if they jointly review student work, their discussion of results and their implications for change can occur with the evidence immediately in mind and available to them.

## COMPETENCE INTERVIEWS

Competence interviews are exams which are orally administered. Interviewers can work alone or in groups, and they can interview single students or groups of students. Unlike written exams, interviews allow fac-

ulty to ask follow-up questions to clarify the breadth and extent of students' understanding. Competence interviews can be used to directly assess a number of learning objectives, such as knowledge of key terms, theories, and findings in the discipline; the ability to integrate information to discuss complex problems or issues; and communication, critical thinking, and interpersonal skills.

Competence interviews are common in foreign language programs, especially if the ability to converse in the language is among the program's learning objectives. Student conversational ability is assessed in an authentic process, and faculty are able to evaluate fluency as well as competency of language use, such as appropriate use of vocabulary, grammar, and syntax. Competence interviews also are common in professional programs, such as social work and nursing. Such interviews might simulate interactions with clients, or they may be more traditional in nature, based on a series of questions and answers.

Many of us have experienced competence interviews from the student's perspective—our dissertation defense—and we should remember these experiences when designing competence interviews for our own students. Procedures which are too threatening are likely to frustrate students and keep them from demonstrating the extent of their learning.

Competence interviews can be *structured* or *unstructured*. Structured interviews involve asking the same questions each time, and interviewers follow a well-rehearsed script. When conducting unstructured interviews, interviewers are allowed to vary their questions, and the process is more open. Skilled interviewers can solicit information on deep processing by encouraging students to elaborate on and explain ideas. More information on interviewing skills is provided in the next chapter.

Interviewers generally ask *open-ended questions*, rather than *closed-ended questions*. Open-ended questions invite respondents to generate longer, more thoughtful replies. Closed-ended questions invite respondents to provide brief answers that are likely to be either correct or wrong. "Name the current president of the United States" is a closed-ended question, while "Describe how checks and balances are built into our national government," "How well do these checks and balances work?" and "What evidence leads you to that conclusion?" are open-ended questions.

Faculty should develop clear understanding of the purpose of the interviews, how they will be conducted, and how they will be scored. It is



sometimes difficult to conduct interviews while scoring them, and interviewers sometimes work in teams, separating these roles. Faculty also could invite community professionals to participate. Their input could give fresh perspectives on student attainment, program objectives, and the curriculum.

Those who conduct competence interviews generally require training so that collected information can be aggregated meaningfully. Figure 5.4 describes Kansas State University's competence interviews for assessing general education objectives.

Practical issues require attention. Interviews take time to conduct and may be difficult to schedule. Interview protocols (scripts) must be developed and tested. Subjective judgments are used to assess learning, and their reliability and validity can be improved by developing explicit scoring criteria and by carefully training interviewers and raters. As with other procedures, student motivation is important.

Competence interviews are not the most efficient way to get some types of information, such as student knowledge of specific facts. Interviews could be combined with other assessment activities. For example, students could take written exams that assess their knowledge of facts, followed by interviews designed to assess their deeper levels of understanding.

## PORTFOLIOS

Portfolios are becoming increasingly popular for course grading and program assessment. Students are required to create compilations of their work, and they usually are required to reflect on their achievement of learning objectives and how the presented evidence supports their conclusions. Portfolio requirements engage students in the assessment process and encourage them to take responsibility for and pride in their learning. Students may develop better understanding of their own academic growth, and they may find their portfolios useful when applying for jobs or graduate programs. School principals in some states routinely expect teaching applicants to bring portfolios to interviews, and portfolios have a long history in other disciplines, such as architecture, art, design, and photography. Here are a few examples of campus portfolio experiences:

### FIGURE 5.4 GENERAL EDUCATION COMPETENCE INTERVIEWS

Kansas State University faculty use General Education Senior Interviews to assess general education objectives (IDEA Center, 1998). Faculty participate in a workshop to learn how to establish rapport with students, conduct the interview as a conversation rather than an inquisition, and complete evaluation forms. Teams of three faculty work together during two weeks of interviews, and each interview is scheduled for 45 to 50 minutes, followed by ten to 15 minutes for completing the assessment.

Senior students who began at the university as freshmen are randomly selected from each college, and they receive \$25 for participating. Instructions are distributed several weeks before the interview, and they encourage students to spend several hours considering their response to a broad question: "Please discuss a topic that you find personally interesting or important and that is not related to your major. Identify two that you would like to discuss. Although usually there is only time to talk about one topic, you can discuss the second if time allows" (IDEA Center, 1998, p. 8).

Students are allowed to discuss the question with friends and to bring notes with them to the interview. As the interview begins, they are assured that the interviewers are assessing the program, rather than individual students, and that their responses and names will be kept confidential. Faculty follow a structured interview protocol and ask clarifying questions to allow them to assess responses, such as, "I think I understand what you are saying, but could you tell us what your personal viewpoint is about this topic?" (IDEA Center, 1998, p. 21).

Faculty independently rate each student and attempt to reach consensus if there is disagreement. Thirteen rating scales are used for each student, and they assess a variety of dimensions, such as how well students display broad interest in topics outside of their discipline, demonstrate depth of understanding, provide evidence supporting their perspective, discuss alternative perspectives, and use oral communication and critical thinking skills.

- Faculty at Alverno College require students to develop a diagnostic digital portfolio, and it is used developmentally by students and advisors to track student growth and by faculty to improve courses and curricula. Student self-assessment is an important component of this process (Loacker, 2002).
- Faculty at Ferris State University (2002) collect writing portfolios from all campus writing courses, and teams of reviewers rate the attainment of their general education writing objectives. Reviewers also examine the types of writing that students do to ensure a cohesive curriculum.

- Indiana University-Purdue University Indianapolis (IUPUI) pilot tested the use of an electronic portfolio in 2000/2001, and they attempted to develop a model with broad campus support that satisfied concerns about security, privacy, and usefulness for assessment. Each student portfolio is organized into three sections: About Me (students describe their individual backgrounds), My Academic Goals and Plans (students develop plans in the freshman year then refine them as they progress), and Principles of Undergraduate Learning (students present evidence of their attainment of relevant learning objectives, including general education objectives). Preliminary work suggested the need to reassure faculty that portfolios are not to be used to evaluate faculty themselves, and that it is acceptable to share materials that are works-in-progress; and students required help to learn how to reflect on their own learning (Banta & Hamilton, 2002).
- New Century College requires all students to develop portfolios, and a faculty reviewer must approve the portfolio before a student can graduate. Each portfolio must include a self-assessment based on evidence presented for nine campus-wide competencies, a reflective essay, and a career development plan. Faculty evaluate portfolios using a rubric that examines the completeness and quality of the evidence and the self-analysis. Details are provided on their web site (New Century College, 2002).
- Olivet College requires students to build portfolios as they progress through the curriculum. Students must demonstrate satisfaction of lower-division learning goals before they can move into upper-division coursework, and they must demonstrate satisfaction of program goals before they graduate. Freshmen begin structuring their portfolios around general education objectives in a required portfolio course where they learn how to provide evidence and reflect on its meaning, and most majors provide portfolio seminars for their students (Petrulis, 2002). Faculty mentors evaluate all portfolios each semester, providing developmental support for each student (Olivet College, 2002).

Two basic types of portfolios are common: *showcase portfolios* and *developmental portfolios*. Showcase portfolios document the extent of

learning by featuring the student's best work. Developmental portfolios are designed to show student progress, and they include evidence of growth by comparing products from early and late stages of the student's academic career. Portfolios also come in different formats. Although traditional paper portfolios in binders or folders are common, some programs require *webfolios* that are submitted on web sites or compact discs (e.g., Sterken, 1999). As students become more proficient in the development of these electronic products, and as faculty become more comfortable accessing them, webfolios will become more popular because they help departments avoid the hassle of storing and distributing printed records.

Faculty must make a number of decisions before portfolios are assigned, and they should answer questions like these:

- What is the purpose of the requirement—to document student learning, to demonstrate student development, to learn about students' reflections on their learning, to create a document useful to students, to help students grow through personal reflection on their personal goals?
- When and how will students be told about the requirement, including what materials they need to collect or to produce for it?
- Will the portfolios be used developmentally or will they be submitted only as students near graduation?
- Will portfolios be showcase or developmental?
- Are there minimum and maximum lengths or sizes for portfolios?
- Who will decide which materials will be included in portfolios—faculty or students?
- What elements will be required in the portfolio—evidence only from courses in the discipline, other types of evidence, evidence directly tied to learning objectives, previously graded products or clean copies?
- Will students be graded on the portfolios? If so, how and by whom?
- How will the portfolios be assessed to evaluate and improve the program?

- What can be done for students who have inadequate evidence through no fault of their own?
- What will motivate students to take the portfolio assignment seriously?
- How will the portfolio be submitted—hard copy or electronic copy?
- Who “owns” the portfolios—students or the program?
- Who has access to the portfolios and for what purposes?
- How will student privacy and confidentiality be protected?

Faculty should have a clear idea of what information will be needed and how that information will be used. If their major intent is to directly assess student mastery of program learning objectives, they probably should ask students to organize their portfolios around these objectives and to reflect on their attainment. If their program objectives call for value-added information or if portfolios will be used for advising, developmental portfolios may be more useful, but if the program objectives emphasize absolute attainment, showcase portfolios probably are better.

Portfolio assignments should clarify faculty expectations for portfolio content, length, organization, and comprehensiveness. Students generally include products from their courses, such as term papers and exams, but they also may be allowed to use other types of evidence, such as documentation from work or volunteer experiences or products created in courses outside the program.

Portfolio assignments can be integrated into the curriculum, and they can be used during advising. For example, instructors could design assignments with portfolios in mind and remind students that these assignments should be saved for their portfolios. Faculty may require some specific assignments in the portfolio, such as capstone course papers or reports completed in research methods classes. Advisors could periodically review draft portfolios with their advisees and discuss their progress, integrating the portfolio into a developmental assessment process. Some departments ask students to analyze long-term goals, such as career goals, in their portfolios. Students could plan their use of elective courses to meet their personal goals, and they may be required to discuss how their education has helped them move toward their attain-

ment. In this way, the assessment process serves multiple functions—student development, as well as assessment.

Faculty also must decide how to encourage students to submit useful portfolios and how to structure the assignments to promote student learning. Most programs require students to submit portfolios as a course or graduation requirement. To encourage students to submit quality portfolios, some programs require students to meet a minimum standard of performance to graduate, while other programs grade portfolios and use that grade as all or part of a class grade. If students are not motivated to prepare quality portfolios, the process probably will not be useful, and students who put a lot of work into their portfolios may be resentful. Pass/fail grading probably is the easiest, but may not motivate students to submit quality portfolios. Some programs grade students on the total portfolio, including the evidence from prior classes, but this approach suffers from the critique of “double grading” prior work. If only the reflective essay is graded, there may be insufficient student motivation to submit a well-documented file. A popular procedure is to grade students on the reflective part of the portfolio and on the completeness of the evidence, but not on the quality of previously graded work.

Faculty should decide how they will handle students who have inadequate evidence for their portfolios. Students may have good reasons for not having appropriate materials. For example, transfer students may have taken classes at institutions that did not encourage them to save their work, and some students may have depended on storage media that failed. It seems reasonable to provide exceptions for such students and to allow them to demonstrate their learning in other ways.

Faculty in small programs may enjoy reading the handful of portfolios submitted each year, but this joy could diminish in larger programs that annually collect rooms full of thick portfolios. Scoring rubrics or other techniques that help faculty review materials efficiently and effectively are crucial to sustaining faculty involvement in portfolio use. Faculty also might consider sampling strategies, such as carefully analyzing a random sample of portfolios or only a few objectives each year or collecting portfolios from some, rather than all, students. For example, some programs annually invite a few representative students to develop portfolios, and these students are awarded stipends or course credit for their effort. Faculty considering the use of portfolios for program assessment

should consider student and faculty workload demands, alternative assessment strategies, how portfolios will be analyzed, how portfolio reviewers will be trained, and how projects will be sustained *before* they ask students to generate these products (Lopez, 1998).

Portfolios have a number of strengths. Because evidence generally is from experiences within the program, gaps in the curriculum are easily identified and the discussion of portfolio results focuses faculty on student learning and program support for it. The workload for faculty and students can be high, and preserving student privacy and confidentiality can be a challenge. Students should be aware that faculty will review the portfolios, and this may cause them to be cautious in criticizing the program or their own learning.

### COLLECTIVE PORTFOLIOS

Just reading the last section may have discouraged you from considering portfolios, especially if you think it might involve reviewing huge collections of materials for hundreds of students each term. An alternative exists. Rather than ask students to prepare individual portfolios, faculty can create collective portfolios, collections of student work that are created by faculty for assessment purposes.

Faculty decide which objectives are to be examined, identify relevant student materials (e.g., course exams and assignments), decide on a sampling scheme, then collect the materials and assess them. The sampling scheme might involve collecting materials from whole classes, random samples of students within classes, *systematic samples*, or *purposeful samples*. Systematic samples are collected using a systematic process, such as collecting products from every tenth student on a class list. Purposeful samples are created using predetermined criteria. For example, each instructor may select the work of the lowest, middle, and highest 10% of their class. In this way, faculty know that they are not just examining the strongest or weakest students. When interpreting results, faculty should consider how data were collected because the proportion of underachieving students might vary with sampling technique. Here are a few examples of collective portfolio projects:

- Faculty in the University of Colorado at Boulder's Department of Physics (2001) routinely assess lab projects from a junior-level

laboratory course and term papers and projects from their senior capstone course. They have found weaknesses in writing style and clarity and have made curricular changes to place more emphasis on the development of these skills.

- Faculty at Mary Washington College (2002) require all baccalaureate students to complete writing intensive courses in their majors. They periodically collect samples of papers in senior-level courses and use scoring rubrics to review learning objectives associated with writing. In their 2002 analysis, they found that nearly 98% of their students are competent writers.
- Johnson County Community College uses an institutional portfolio to assess its general education program. Faculty from multiple disciplines review student work to assess learning objectives associated with mathematics, writing, speaking, culture and ethics, modes of inquiry, and problem solving. Results, based on the application of holistic rubrics, are compiled by the Office of Institutional Research, and faculty review them and act on what they learn (Seybert, 2002).
- Faculty in California State University, Sacramento's Department of Sociology (2000) collected papers from samples of A, B, and C students in core upper-division courses in the major. They examined two learning goals, written communication skills, and the mastery of basic sociological concepts and theories; and they reviewed materials to identify outstanding work, satisfactory work, inadequate work, and ideas to address identified deficiencies. Their analysis led to a number of suggestions for improving writing assignments and a recognition of the need to implement these improvements while respecting individual faculty control over their courses.

Collective portfolios can be analyzed early in the assessment cycle, even before faculty examine curriculum alignment or plan carefully targeted embedded assignments. All that is needed is a group of faculty who are willing to share some of their student products and discuss what these products tell them about student learning. The process has more

potential if the course materials are designed to align with specific objectives. For example, if students are not told in an assignment to compare the relative usefulness of alternative theoretical approaches, faculty may be unable to effectively assess this type of objective. The process is likely to evolve as faculty discover limitations of the evidence available for review.

Faculty are so accustomed to grading that special attention is required to keep them focused on assessing objectives. If they are examining student documents to assess writing skills, they probably can make reasonably reliable and valid judgments quickly, without getting bogged down in other aspects of the papers. Scoring rubrics are particularly effective for this purpose.

Like traditional portfolios, collective portfolios can provide useful assessment information, and they impose no additional work on students. Assessment results are likely to have greater reliability and validity if the relevant exam questions and assignments are designed to assess specific learning objectives. Preplanning embedded assessment gives faculty more control of the process, aligns coursework with program learning objectives, and should allow faculty to get more benefit from the use of collective portfolios.

### SUMMARY OF DIRECT ASSESSMENT TECHNIQUES

Each of the direct assessment techniques described in this chapter has potential strengths and limitations, as summarized in Figure 5.5. As faculty select and refine assessment strategies, they should design projects that exploit the strengths and minimize the risks associated with the techniques they employ.

FIGURE 5.5  
STRENGTHS AND LIMITATIONS OF  
DIRECT ASSESSMENT TECHNIQUES

Technique	Potential Strengths	Potential Limitations
Published tests	<ul style="list-style-type: none"> <li>• Can provide direct evidence of student mastery of learning objectives.</li> <li>• They generally are carefully developed, highly reliable, professionally scored, and nationally normed.</li> <li>• They frequently provide a number of norm groups, such as norms for community colleges, liberal arts colleges, and comprehensive universities.</li> <li>• Online versions of tests are increasingly available, and some provide immediate scoring.</li> <li>• Some publishers allow faculty to supplement tests with their own items, so tests can be adapted to better serve local needs.</li> </ul>	<ul style="list-style-type: none"> <li>• If the test does not reflect the learning objectives that faculty value and the curricula that students experience, results are likely to be discounted and inconsequential.</li> <li>• Most published tests rely heavily on multiple-choice items that often focus on specific facts, but program learning objectives more often emphasize higher-level skills.</li> <li>• Test scores may reflect criteria that are too broad for meaningful assessment.</li> <li>• Students may not take the test seriously if test results have no impact on their lives.</li> <li>• Tests can be expensive.</li> <li>• The marginal gain from annual testing may be low.</li> <li>• Faculty may object to standardized exam scores on general principles, leading them to ignore results.</li> </ul>
Locally developed tests	<ul style="list-style-type: none"> <li>• Can provide direct evidence of student mastery of learning objectives.</li> <li>• Appropriate mixes of items allow faculty to address various types of learning objectives.</li> <li>• Can provide for authentic assessment of higher-level learning.</li> </ul>	<ul style="list-style-type: none"> <li>• These exams are likely to be less reliable than published exams.</li> <li>• Reliability and validity generally are unknown.</li> <li>• Creating effective exams requires time and skill.</li> <li>• Scoring exams takes time.</li> </ul>

(continued on page 100)

Technique	Potential Strengths	Potential Limitations
	<ul style="list-style-type: none"> <li>• Students generally are motivated to display the extent of their learning.</li> <li>• If well constructed, they are likely to have good validity.</li> <li>• Because local faculty write the exam, they are likely to be interested in results and willing to use them.</li> <li>• Can be integrated into routine faculty workloads.</li> <li>• Campuses with similar missions could decide to develop their own norms, and they could assess student work together or provide independent assessment of each other's student work.</li> <li>• Discussion of results focuses faculty on student learning and program support for it.</li> </ul>	<ul style="list-style-type: none"> <li>• Traditional testing methods may not provide authentic measurement.</li> <li>• Norms generally are not available.</li> </ul>
Embedded assignments and course activities	<ul style="list-style-type: none"> <li>• Can provide direct evidence of student mastery of learning objectives.</li> <li>• Out-of-class assignments are not restricted to time constraints typical for exams.</li> <li>• Students are generally motivated to demonstrate the extent of their learning.</li> <li>• Can provide authentic assessment of learning objectives.</li> <li>• Can involve ratings by field-work supervisors.</li> <li>• Can provide a context for assessing communication and teamwork skills, as well as other types of learning objectives.</li> </ul>	<ul style="list-style-type: none"> <li>• Requires time to develop and coordinate.</li> <li>• Requires faculty trust that the program will be assessed, not individual teachers.</li> <li>• Reliability and validity generally are unknown.</li> <li>• Norms generally are not available.</li> </ul>

Technique	Potential Strengths	Potential Limitations
	<ul style="list-style-type: none"> <li>• Can be used for grading as well as assessment.</li> <li>• Faculty who develop the procedures are likely to be interested in results and willing to use them.</li> <li>• Discussion of results focuses faculty on student learning and program support for it.</li> <li>• Data collection is unobtrusive to students.</li> </ul>	
Competence interviews	<ul style="list-style-type: none"> <li>• Can provide direct evidence of student mastery of learning objectives.</li> <li>• The interview format allows faculty to probe for the breadth and extent of student learning.</li> <li>• Can be combined with other techniques that more effectively assess knowledge of facts and terms.</li> <li>• Can involve authentic assessment, such as simulated interactions with clients.</li> <li>• Can provide for direct assessment of some student skills, such as oral communication, critical thinking, and problem-solving skills.</li> </ul>	<ul style="list-style-type: none"> <li>• Requires time to develop, coordinate, schedule, and implement.</li> <li>• Interview protocols must be carefully developed.</li> <li>• Subjective judgments must be guided by agreed-upon criteria.</li> <li>• Interviewer training takes time.</li> <li>• Interviewing using unstructured interviews requires expertise.</li> <li>• Not an efficient way to assess knowledge of specific facts and terms.</li> <li>• Some students may be intimidated by the process, reducing their ability to demonstrate their learning.</li> </ul>
Portfolios	<ul style="list-style-type: none"> <li>• Can provide direct evidence of student mastery of learning objectives.</li> <li>• Students are encouraged to take responsibility for and pride in their learning.</li> </ul>	<ul style="list-style-type: none"> <li>• Requires faculty time to prepare the portfolio assignment and to assist students in preparing portfolios.</li> <li>• Requires faculty analysis and, if graded, faculty time to assign grades.</li> </ul>

(continued on page 102)

Technique	Potential Strengths	Potential Limitations
	<ul style="list-style-type: none"> <li>• Students may become more aware of their own academic growth.</li> <li>• Can be used for developmental assessment and can be integrated into the advising process to individualize student planning.</li> <li>• Can help faculty identify curriculum gaps.</li> <li>• Students can use portfolios and the portfolio process to prepare for graduate school or career applications.</li> <li>• Discussion of results focuses faculty on student learning and program support for it.</li> <li>• Webfolios or CD-ROMs can be easily viewed, duplicated, and stored.</li> </ul>	<ul style="list-style-type: none"> <li>• May be difficult to motivate students to take the task seriously.</li> <li>• May be more difficult for transfer students to assemble the portfolio if they haven't saved relevant materials.</li> <li>• Students may refrain from criticizing the program if their portfolio is graded or if their names will be associated with portfolios during the review.</li> <li>• It may be difficult to protect student confidentiality and privacy.</li> </ul>
Collective portfolios	<ul style="list-style-type: none"> <li>• Can provide direct evidence of student mastery of learning objectives.</li> <li>• Students generally are motivated to display the extent of their learning.</li> <li>• Workload demands generally are more manageable than traditional portfolios.</li> <li>• Students are not required to do extra work.</li> <li>• Discussion of results focuses faculty on student learning and program support for it.</li> <li>• Data collection is unobtrusive to students.</li> </ul>	<ul style="list-style-type: none"> <li>• If assignments are not aligned with the objectives being examined, evidence may be problematic.</li> <li>• If sampling is not done well, results may not generalize to the entire program.</li> <li>• Reviewing the materials takes time and planning.</li> </ul>