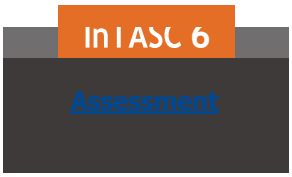
CHAPTER THIRTEEN

Assessing Student Learning

Aligning Objectives and Assessment



Because instructional objectives are stated in terms of how they will be measured, it is clear that objectives are closely aligned with **assessment**, which consists of measuring the degree to which students have learned the objectives set out for them. Most assessments in schools are tests or quizzes or informal verbal assessments such as questions in class. However, students can also show their learning by writing an essay, creating a multimedia presentation, painting a picture, doing a car tune-up, or baking a pineapple upside-down cake.

One critical principle of assessment is that assessments and objectives must be clearly linked (Martone & Sireci, 2009; McAfee, Leong, & Bodrova, 2016). Students learn some proportion of what they are taught; the greater the overlap between what was taught and what is tested, the better students will score on the test and the more accurately any need for additional instruction can be determined (Lloyd et al., 2013; Popham, 2014a; Russell & Airasian, 2012; Squires, 2009). Teaching should be closely linked to instructional objectives, and both should clearly relate to assessment (Buhle & Blachowicz, 2008/2009). If any objective is worth teaching, it is worth testing, and vice versa.

As noted earlier, one way to specify objectives for a course is to actually prepare test questions before the course begins (see Waugh & Gronlund, 2013). This allows you to write general **teaching objectives** (clear statements of what students are expected to learn through instruction) and then to clarify them with very specific **learning objectives** (specific behaviors students are expected to exhibit at the end of a series of lessons), as in the following examples.

Teaching Objective	Specific Learning Objective (Test Questions)
a. Ability to subtract three-digit numbers regrouping once or twice	a1. 237 a2. 412 a3. 596 2184 2298 2448
b. Understanding use of language to set mood in Edgar Allan Poe's "The Raven"	b1. How does Poe reinforce the mood of "The Raven" after setting it in the first stanza?
c. Ability to identify the chemical formulas for common substances	Write the chemical formulas for the following: c1. Water _____ c2. Carbon dioxide _____ c3. Coal _____ c4. Table salt _____

WHY IS EVALUATION IMPORTANT?

Evaluation, or assessment, consists of all the means used in schools to formally measure student performance (Lloyd et al., 2013; McMillan, 2011; Popham, 2014; Waugh & Gronlund, 2013). These include quizzes and tests, written evaluations, and grades. Student evaluation usually focuses on academic achievement, but many schools also assess behaviors and attitudes. Many elementary schools provide descriptions of student behaviors (such as "follows directions," "listens attentively," "works with others," "uses time wisely"). In upper elementary, middle, and high school the prevalence of behavior reports diminishes successively, but even many high schools rate students on such criteria as "works up to ability," "is prepared," and "is responsible."

Why do teachers use tests and grades? You use them because, one way or another, you must periodically check and communicate about students' learning. Tests and grades tell teachers, students, and parents how students are doing in school. You can use tests to determine whether your instruction was effective and to find out which students need additional help. Students can use tests to find out whether their studying strategies are paying off. Parents need grades to learn how their children are doing in school; grades usually serve as the one consistent form of communication between school and home. Schools sometimes need grades and tests to make student placements. States and school districts need tests to evaluate schools and, in some cases, teachers. Ultimately, colleges use grades and standardized test scores to decide whom to admit, and employers use grade-based evidence of attainment, such as diplomas and other credentials, in hiring decisions. Teachers must therefore evaluate student learning; few would argue otherwise. Research on the use of tests finds that students learn more in courses that test students than in those that do not (Dempster, 1991; Haynie & Haynie, 2008).

Student evaluations serve six primary purposes (see Waugh & Gronlund, 2013):

1. Feedback to students
2. Feedback to teachers
3. Information to parents
4. Information for selection and certification
5. Information for accountability
6. Incentives to increase student effort

Evaluation as Feedback

Connections 13.3
For more on feedback as a component of effective teaching, see Chapter 7.

Imagine that a store owner tried several strategies to increase business—first advertising in the newspaper, then sending fliers to homes near the store, and finally holding a sale. However, suppose that after trying each strategy, the store owner failed to record and compare the store's revenue. Without taking stock this way, the owner would learn little about the effectiveness of any of the strategies and might well be wasting time and money. The same is true of teachers and students. They need to know as soon as possible whether their investments of time and energy in a given activity are paying off in increased learning.

FEEDBACK FOR STUDENTS Like the store owner, students need to know the results of their efforts (Fisher & Frey, 2014c; Marzano, Yanoski, Hoegh, & Simms, 2013). Regular evaluation gives them feedback on their strengths and weaknesses. For example, suppose you had students write compositions and then gave back written evaluations. Some students might find out that they needed to work more on content, others on the use of modifiers, and still others on language mechanics. This information would help students to improve their writing much more than would a grade with no explanation (Brookhart & Nitko, 2015; Chappuis, Stiggins, Chappuis, & Arter, 2012).

To be useful as feedback, evaluations should be as specific as possible (Quinn, 2012). For example, Cross and Cross (1980/1981) found that students who received written feedback in addition to letter grades were more likely than other students to believe that their efforts, rather than luck or other external factors, determined their success in school.

FEEDBACK TO TEACHERS One of the most important (and often overlooked) functions of evaluating student learning is to provide feedback to teachers on the effectiveness of their instruction. You cannot expect to be optimally effective if you do not know whether students have grasped the main points of your lessons. Asking questions in class and observing students as they work gives you some idea of how well students have learned; but in many subjects brief but frequent quizzes, writing assignments, and other student products are necessary to provide more detailed indications of students' progress. Well-crafted questions can help you understand students' thinking and uncover misconceptions (Brookhart, 2014; McTighe & Wiggins, 2013; Salend, 2016; Wiliam & Leahy, 2015). Evaluations also give information to the principal and the school as a whole, which can be used to guide overall reform efforts by identifying where schools or subgroups within schools are in need of improvement (McTighe & Curtis, 2015; Mertler, 2014; Schimmer, 2016). Electronic whiteboards with digital response devices can provide teachers with immediate information on how many students have understood each objective the teachers have taught and assessed (Magaño & Marzano, 2014).

Evaluation as Information

A report card is called a report card because it reports information on student progress. This reporting function of evaluation is important for several reasons.

INFORMATION TO PARENTS First, routine school evaluations of many kinds (test scores, stars, and certificates as well as report card grades) keep parents informed about their children's schoolwork. For example, if a student's grades are dropping, the parents might know why and may be able to help the student get back on track. Second, grades and other evaluations set up informal home-based reinforcement systems. Recall from Chapter 11 that many studies have found that reporting regularly to parents when students do good work and asking parents to reinforce good reports improve student behavior and achievement. Without much prompting, most parents naturally reinforce their children for bringing home good grades, thereby making grades important and effective as incentives.

Connections 13.4
For more on information for parents, see Chapter 11.

MyEdLab
Video Example 13.3
This teacher provides evaluative feedback individually to the student and then later to his mother. In both conferences, she asks how they can work together to help support the student's learning.

INFORMATION FOR SELECTION Some sociologists see the sorting of students into societal roles as a primary purpose of schools: If schools do not actually determine who will be a butcher, a baker, or a candlestick maker, they do substantially influence who will be a laborer, a skilled worker, a white-collar worker, or a professional. This sorting function takes place gradually over years of schooling. In the early grades, students are sorted into reading groups. Later some eighth-graders take algebra, whereas others take prealgebra or general mathematics. In high school, students are often steered toward advanced, basic, or remedial levels of particular courses, and a major sorting takes place when students are accepted into various colleges and training programs. Moreover, throughout the school years, some students are selected into special-education programs, into programs for the gifted and talented, or into other special programs with limited enrollments.

Closely related to selection is certification, the use of tests to qualify students for promotion or for access to various occupations. For example, many states and local districts have tests that students must pass to advance from grade to grade or to graduate from high school. Bar exams for lawyers, board examinations for medical students, and tests for teachers such as the National Teachers' Examination are examples of certification tests that control access to professions.

INFORMATION FOR ACCOUNTABILITY Often, evaluations of students serve as data for the evaluation of teachers, schools, districts, or even states. Every state has some form of statewide testing program that allows the states to rank every school in terms of student performance (Banks, 2012;

Connections 13.5
For more information on ability grouping, see Chapter 9.

Miller, Linn, & Gronlund, 2013). These test scores are also often used in evaluations of principals, teachers, and superintendents. Consequently, these tests are taken very seriously.

Evaluation as Incentive

One important use of evaluations is to motivate students to give their best efforts (Dueck, 2014; Vagle, 2014). In essence, high grades, stars, and prizes are given as rewards for good work. Students value grades and prizes primarily because their parents value them. Some high school students also value grades because they are important for getting into selective colleges.

MyEdLab **Self-Check 13.1**

HOW IS STUDENT LEARNING EVALUATED?

InTASC 6
Assessment

Evaluation strategies must be appropriate for the uses that are made of them (McMillan, 2011; Penuel & Shepard, 2016; Salend, 2016). To understand how assessments can be used most effectively in classroom instruction, it is important to know the differences between formative and summative evaluations and between norm-referenced and criterion-referenced interpretations.

Formative and Summative Evaluations

Certification Pointer
For your teacher certification test, you may be given a case illustrating an evaluation of student performance, and you will need to categorize that evaluation as formative or summative.

Assessments can be divided into two categories: formative and summative. Essentially, a formative evaluation asks, "How well are you doing and how can you be doing better?" A summative evaluation asks, "How well did you do?" A **formative evaluation** is designed to tell teachers whether additional instruction is needed and to tell students whether additional learning is needed (Gewertz, 2015; Heritage, 2011; Higgins, 2014; Marzano et al., 2013; Tomlinson & Moon, 2013). Formative, or diagnostic, tests are given to discover strengths and weaknesses in learning and to make midcourse corrections in pace or content of instruction (Fisher & Frey, 2014a). Formative evaluations might even be made "on the fly" during instruction, through oral or brief written learning probes, or by listening to students during groupwork. Increasingly, computerized exercises and games are being used to give teachers and students immediate feedback on students' learning (Phillips & Popovíc, 2012). Formative evaluation is useful to the degree that it is informative, closely tied to the curriculum being taught, timely, and frequent (Dunn & Mulvenon, 2009; Fogarty & Kerns, 2009; McMillan, 2011; Popham, 2014a; Spinelli, 2011). For example, frequent quizzes that are given and scored immediately after specific lessons might serve as formative evaluations, providing feedback to help both teachers and students improve students' learning. Effective uses of formative assessments in lessons were discussed in Chapter 7.

In contrast, **summative evaluation** refers to tests of student knowledge at the end of instructional units (such as final exams). Summative evaluations may or may not be frequent, but they must be reliable and (in general) should allow for comparisons among students. Summative evaluations should also be closely tied to formative evaluations and to course objectives (Gronlund & Brookhart, 2009; Schimmer, 2016).

Connections 13.6
For more on standardized testing, see Chapter 14.

Norm-Referenced and Criterion-Referenced Evaluations

Certification Pointer
Your teacher certification test may require you to evaluate when it would be more appropriate to use a criterion-referenced test and when to use a norm-referenced test.

Interpretation in order to attach a degree of value to a student's performance is an important step in an evaluation. The distinction between norm referencing and criterion referencing is based on the way students' scores are interpreted.

Norm-referenced interpretations focus on comparisons of a student's scores with those of other students. Within a classroom, for example, grades commonly are used to give teachers an idea of how well a student has performed in comparison with classmates. A student might also have a grade-level or school rank (Guskey, 2014); and in standardized testing, student scores might be compared with those of a nationally representative norm group.

Criterion-referenced interpretations focus on assessing students' mastery of specific skills, regardless of how other students did on the same skills. Criterion-referenced evaluations are best if they are closely tied to specific objectives or well-specified domains of the curriculum being taught. Table 13.4 compares the principal features and purposes of criterion-referenced and norm-referenced testing (see also Waugh & Gronlund, 2013; Popham, 2014b; Thorndike & Thorndike-Christ, 2010).

TABLE 13.G • Comparison of Two Approaches to Achievement Testing

Norm-referenced tests and criterion-referenced tests serve different purposes and have different features.

FEATURE	NORM-REFERENCED TESTING	CRITERION-REFERENCED TESTING
Principal use	Survey testing	Mastery testing
Major emphasis	Measures individual differences in achievement	Describes tasks students can perform
Interpretation of results	Compares performance to that of other individuals	Compares performance to a clearly specified achievement domain
Content coverage	Typically covers a broad area of achievement	Typically focuses on a limited set of learning tasks
Nature of test plan	Table of specifications is commonly used	Detailed domain specifications are favored
Item selection procedures	Items selected to provide maximum discrimination between individuals (to obtain high score variability); easy items typically eliminated from the test	Includes all items needed to adequately describe performance; no attempt is made to alter item difficulty or to eliminate easy items to increase score variability
Performance standards	Level of performance determined by *relative* position in some known group (e.g., student ranks fifth in a group of 20)	Level of performance commonly determined by *absolute* standards (e.g., student demonstrates mastery by defining 90 percent of the technical terms)

Source: Gronlund, Norman E., *How to make achievement tests and assessment*, 5th Edition, © 1993. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.

Formative evaluation is almost always criterion referenced. In formative testing, teachers want to know, for example, who is having trouble with Newton's laws of thermodynamics, not which student is first, fifteenth, or thirtieth in the class in physics knowledge. Summative testing, in contrast, can be either criterion referenced or norm referenced. Even if it is criterion referenced, however, teachers usually want to know, on a summative test, how each student did in comparison with other students.

Matching Evaluation Strategies with Goals

Considering all the factors discussed up to this point, what is the best strategy for evaluating students? The first answer is that there is no one best strategy (Penuel & Shepard, 2016; Popham, 2014a, b). The best means of accomplishing any one objective of evaluation might be inappropriate for other objectives. Therefore, you should choose different types of evaluation for different purposes. At a minimum, two types of evaluation should be used: one directed at providing incentive and feedback, and the other directed at ranking individual students relative to the larger group.

EVALUATION FOR INCENTIVE AND FEEDBACK Traditional grades are often inadequate as incentives to encourage students to give their best efforts and as feedback to teachers and students (Tomlinson & Moon, 2014; Wiliam, 2014). The principal problems are that grades are given too infrequently, are too far removed in time from student performance, and are poorly tied to specific student behaviors. Research has found that achievement is higher in classrooms where students receive immediate feedback on their quizzes than in classrooms where feedback is delayed (Duckor, 2014; Tomlinson, 2014a; Wiggins, 2012).

Another reason why grades are less than ideal as incentives is that they are usually based on comparative standards. In effect, it is relatively easy for high-ability students to achieve A's and B's but very difficult for low achievers to do so. As a result, some high achievers do less work than they are capable of doing, and some low achievers give up. As was noted in Chapter 10, a reward that is too easy or too difficult to attain, or one that is felt to be a result of ability rather than of effort, is a poor motivator (Chapman & King, 2005; Wigfield & Eccles, 2000).

For these reasons, traditional grades should be supplemented by evaluations that are better designed for incentive and feedback. For example, teachers might give daily quizzes of 5 or 10 items that are scored in class immediately after completion, or they might have students write daily "mini-essays" on a topic the class is studying. These give both students and teachers the information they need to adjust their teaching and learning strategies and to rectify any deficiencies revealed by the evaluations (Shepard, 2005). If teachers make quiz results important by having them count toward course grades or by giving students with perfect papers

Connections 13.7
Rewards and motivation are discussed in Chapter 5.

Connections 13.8
For more on which rewards make poor motivators, see Chapter 10.

special recognition or certificates, then quiz scores also serve as effective incentives, rewarding effective studying behavior soon after it occurs. It is important to have a clear and objective set of criteria that student work is compared with so students can see exactly why they scored as they did. If the criteria are illustrated using a rubric that has descriptions of different levels of achievement (scores) as well as examples of student work that is at the highest levels of achieve- ment (or better yet, that is typical of each possible score students might receive according to the rubric), then students can see exactly how their achievement compares with the criteria (Stiggins & Chappuis, 2012).

Evaluation for Comparison with Others

There are times when you need to know and to communicate how well students are doing in comparison to others. This information is important to give parents (and students themselves) a realistic picture of student performance. For example, students who have outstanding skills in sci- ence ought to know that they are exceptional, not only in the context of their class or school, but also in a broader state or national context. In general, students need to form accurate perceptions of their strengths and weaknesses to guide their decisions about their futures.

Comparative evaluations are traditionally provided by grades and by standardized tests. Un- like incentive/feedback evaluations, comparative evaluations need not be conducted frequently. Rather, the emphasis in comparative evaluations must be on fair, unbiased, reliable assessment of student performance.

To be fair, comparative evaluations and other summative assessments of student performance must be firmly based on the objectives established at the beginning of the course and consistent with the formative incentive/feedback evaluations in format as well. No teacher wants a situation in which students do well on week-to-week assessments but then fail the summative evaluations because there is a lack of correspondence between the two forms of evaluation. For example, if the summative test uses essay questions, then the formative tests leading up to it should also include essay questions (Tileston & Darling, 2008).

Writing Selected-Response Test Items

Test items that can be scored correct or incorrect without the need for interpretation are referred to as **selected-response items**. Multiple-choice, true–false, and matching items are the most common forms. Note that the correct answer appears on the test and the student's task is to select it. There is no ambiguity about whether the student has or has not selected the correct answer. Each type, however, has its own advantages and disadvantages.

MULTIPLE-CHOICE ITEMS Considered by some educators to be the most useful and flexible of all test forms (Badgett & Christmann, 2009; Waugh & Gronlund, 2013), **multiple-choice items** can be used in tests for most school subjects. The basic form of the multiple-choice item is a **stem** followed by choices, or alternatives. The stem may be a question or a partial statement that is completed by one of several choices. No perfect number of choices exists, but using four or five is most common—one correct response and wrong but plausible answers that are referred to as **distractors** or **foils**.

The following examples demonstrate two types of multiple-choice items, one with a question stem and the other with a completion stem:

1. What color results from the mixture of equal parts of yellow and blue paint?
   a. Black
   b. Gray
   c. Green [*correct choice*]
   d. Red
2. U.S. presidents are actually elected to office by
   a. all registered voters.
   b. our congressional representatives.

c. the Electoral College. [*correct choice*]
d. the Supreme Court.

When writing a multiple-choice item, keep two goals in mind. First, a capable student should be able to choose the correct answer and not be distracted by the wrong alternatives. Second, you should minimize the chance that a student who is ignorant of the subject matter can guess the correct answer. To achieve this, the distractors must look possible to the uninformed; their wording and form must not identify them readily as bad answers. Hence, one of the tasks in writing a good multiple-choice item is to identify two, three, or four plausible, but not tricky, distractors.

TRUE–FALSE ITEMS Another type of multiple-choice question is the **true–false item**. The main drawback of the true–false format is that students have a 50 percent chance of guessing correctly. For this reason, it should rarely be used.

MATCHING ITEMS As commonly presented, **matching items** usually take the form of two lists, say *A* and *B.* For each item in list *A,* the student has to select one item in list *B.* The basis for choosing must be clearly explained in the directions. Matching items can be used to cover a large amount of content; that is, a large (but not unmanageably large) number of concepts should appear in the two lists. Each list should cover related content (use more than one set of matching items for different types of material). The primary cognitive skill that matching exercises test is recall.

Matching items can often be answered by elimination because many teachers maintain a one-to-one correspondence between the two lists. To engage students in the content, not the format, teachers should either include more items in list *B* than in list *A* or allow reuse of the items in list *B.*

Writing Constructed-Response Items

Constructed-response items require the student to supply rather than select the answer. The simplest form is fill-in-the-blank items, which can often be written to reduce or eliminate ambiguity in scoring. Still, unanticipated responses might lead to ambiguous answers, raising questions in the mind of the instructor on how to score them. Constructed-response items also come in short essay and long essay forms.

FILL-IN-THE-BLANK ITEMS When there is clearly only one possible correct answer, an attractive format is completion, or "fill in the blank," as in the following examples.

1. The largest city in Germany is _____.
2. What is 15 percent of $198.00? _____
3. The measure of electric resistance is the _____.

The advantage of these **completion items** is that they can reduce the element of test-wiseness to near zero. For example, compare the following items:

1. The capital of Maine is _____.
2. The capital of Maine is
   - **a.** Sacramento
   - **b.** Augusta
   - **c.** Juneau
   - **d.** Boston

A student who has no idea what the capital of Maine is could pick Augusta from the list in item 2 because it is easy to rule out the other three cities. In item 1, however, the student has to know the answer. Completion items are especially useful in math, because multiple-choice items may help to give the answer away or reward guessing, as in the following example

4037
2 159

- **a.** 4196
- **b.** 4122
- **c.** 3878 [*correct answer*]
- **d.** 3978

If students subtract and get an answer other than any of those listed, they know that they have to keep trying. In some cases they can narrow the alternatives by estimating rather than knowing how to compute the answer.

It is critical to avoid ambiguity in completion items. In some subject areas this can be difficult because two or more answers can reasonably fit a fragment that does not specify the context, as in the next two examples.

1. The Battle of Hastings was in _____. [Date or place?]
2. "H_2O" represents_____. [Water or two parts hydrogen and one part oxygen?]

If there is any ambiguity possible, it is probably best to move to a selection type of item such as multiple choice.

Writing and Evaluating Essay Tests

Short essay questions allow students to respond in their own words. The most common form for a **short essay item** provides a question for the student to answer. The answer may range from a sentence or two to a page of, say, 100 to 150 words. A **long essay item** requires more length and more time, allowing greater opportunity for students to demonstrate organization and development of ideas. Although they differ in length, the methods available to write and score them are similar.

The essay form can elicit a wide variety of responses, from giving definitions of terms to comparing and contrasting important concepts or events. These items are especially suited for assessing students' ability to analyze, synthesize, and evaluate. Hence, you might use them to appraise students' progress in organizing data and applying concepts at the highest levels of instructional objectives. Of course, these items depend heavily on writing skills and the ability to

phrase ideas, so exclusive use of essays might cause the teacher to underestimate the knowledge and effort of a student who has learned the material but is a poor writer.

One of the crucial mistakes teachers make in writing essay items is failing to specify clearly the approximate detail required in the response and its expected length. Stating how much weight an item has relative to the entire test is generally not sufficient to tell students how much detail must be incorporated in a response. The following example illustrates this point.

Poor Essay Item

Discuss the role of the prime minister in Canadian politics.

Improvement

In five paragraphs or less, identify three ways in which the Canadian prime minister and the U.S. president differ in their obligations to their respective constituencies. For each of the three, explain how the obligations are different.

Note that the improved version expresses a length (five paragraphs or less), the aspect to be treated (differences between the prime minister and the president), the number of points to be covered (three; if you write "at least three," that would introduce ambiguity into the task), how the points should be selected (differ in their obligations to their respective constituencies), and the direction and degree of elaboration needed (explain how the obligations are different). This item points the student toward the desired response and gives you a greater opportunity to explain the criteria by which student responses will be judged.

An essay item should contain specific information that students are to address. Some teachers are reluctant to name the particulars they want students to discuss, because they believe that supplying a word or phrase in the instructions is giving away too much information. But if an item is ambiguous, different students will interpret it differently. Consequently, they will be responding to different questions, and the test will almost surely not be fair to all of them.

Essay items have a number of advantages in addition to letting students state ideas in their own words. For example, essay items are not susceptible to correct guesses. They can promote efficient assessment by requiring students to combine several concepts in one response. They can also be used to measure creative abilities, such as writing talent or imagination in constructing hypothetical events, as well as assessing organization and fluency.

On the negative side is the problem of reliability in scoring essay responses. Some studies demonstrate that independent marking of the same essay response by different teachers can result in appraisals ranging from excellent to a failing grade (Popham, 2014a). A second drawback is that essay responses take considerable time to evaluate. The time you save by writing one essay item instead of several other kinds of items must be paid back when grading the essays. Third, essay items in general take considerable response time from students. Consequently, they typically cannot be used to cover broad ranges of content. Nevertheless, essay items enable teachers to see how well students can use the material they have been taught. Breadth is sacrificed for depth.

The following suggestions provide additional guidelines for writing effective essay items.

1. As with any item format, match the items with the instructional objectives.
2. Do not use such general directives in an item as "discuss," "give your opinion about," or "tell all you know about." Rather, carefully choose specific response verbs such as "compare," "contrast," "identify," "list and define," and "explain the difference."
3. Write a response to the item before you give the test to estimate the time students will need to respond. About four times your response time is a fair estimate.
4. Rewrite the item to point students clearly toward the desired response.
5. Require all students to answer all items. Even though it seems attractive to allow student choice in which items to answer, that is fundamentally an unfair practice. First, students differ in their ability to make the best selections. Second, the items will not be of equivalent difficulty. And third, students who know they will have a choice can increase their score by studying very carefully only part of the material.

After writing an essay item—and clearly specifying the content that is to be included in the response—you must have a clear idea of how you will score various elements of a student's response. The first step is to write a model response or a detailed outline of the essential elements students are being directed to include in their responses to which you can compare students' responses. If you intend to use evaluative comments but not letter grades, your outline or model will serve as a guide for pointing out to students any omissions and errors in their responses, as well as the good points of their answers. If you are using letter grades to score the essays, you should compare elements of students' responses with the contents of your model and give suitable credit to responses that match the relative weights of elements in the model.

If possible, you should ask a colleague to assess the validity of the elements and their weights in your model response. Going a bit further and having the colleague apply the model criteria to one or more student responses will increase the reliability of your scoring (see Langer & Colton, 2005). Be sure to offer to do the same for them!

One issue related to essay tests is whether and how much to count grammar, spelling, and other technical features. If you do count these factors, give students separate grades in content and in mechanics so that they will know the basis on which their work is being evaluated.

A powerful use of assessment in instruction is to generate one or more scoring rubrics that can be shared with students well in advance of the test. The rubrics, like the example, should be generic, in that they can be applied to a broad range of essays. Students can see which aspects of their achievement will contribute to a positive evaluation and can practice to make sure their work illustrates those critical elements. You might show students (anonymous) essays from previous years to illustrate the rubric.

Writing and Evaluating Problem-Solving Items

Connections 13.14
For more on problem solving, see Chapter 8.

INTASC 5
Application of Content

INTASC 6
Assessment

In many subjects, such as mathematics and the physical and social sciences, instructional objectives include the development of skills in problem solving, so it is important to assess students' performance in solving problems (Badgett & Christmann, 2009; McMillan, 2011). A **problem-solving assessment** requires students to organize, select, and apply complex procedures that have at least several important steps or components. It is important to appraise the students' work in each of these steps or components.

The following example shows a seventh-grade-level mathematical problem and a seventh-grader's response to it. The discussion of evaluating problem solving that follows can be applied to any discipline.

PROBLEM

Suppose two gamblers are playing a game in which the loser must pay an amount equal to what the other gambler has at the time. If Player A won the first and third games, and Player B won the second game, and they finished the three games with $12 each, with how much money did each begin the first game? How did you get your answer?

A student's response:

After game	A had	B had
3	$12.00	$12.00
2	6.00	18.00
1	15.00	9.00
In the beginning	$ 7.50	$16.50

When I started with Game 1, I guessed and guessed, but I couldn't make it come out to 12 and 12.

Then I decided to start at Game 3 and work backward. It worked!

How will you objectively evaluate such a response? As in evaluating short essay items, you should begin your preparation for appraising problem-solving responses by writing either a model response or, perhaps more practically, an outline of the essential components or procedures that are involved in problem solving. As with essays, problem solving may take several different yet valid approaches. The outline must be flexible enough to accommodate all valid possibilities.

SUMMARY

What Are Instructional Objectives and How Are They Used?

Research supports the use of instructional, or behavioral, objectives, which are clear statements about what students should know and be able to do at the end of a lesson, unit, or course. These statements also specify the conditions of performance and the criteria for assessment. In lesson planning, task analysis contributes to the formulation of objectives, and backward planning facilitates the development of specific objectives from general objectives in a course of study. Objectives are closely linked to assessment. Bloom's taxonomy of educational objectives classifies educational objectives from simple to complex, including knowledge, comprehension, application, analysis, synthesis, and evaluation. A behavior content matrix helps to ensure that objectives cover many levels.

Why Is Evaluation Important?

Formal measures of student performance or learning are important as feedback for students and teachers, as information for parents, as guidance for selection and certification, as data for assessing school accountability, and as incentives for increasing student effort.

How Is Student Learning Evaluated?

Strategies for evaluation include formative evaluation; summative evaluation; norm-referenced evaluation, in which a student's scores are compared with other students' scores; and criterion-referenced evaluation, in which students' scores are compared to a standard of mastery. Students are evaluated through tests or performances. The appropriate method of evaluation depends on the goal of evaluation. For example, if the goal of testing is to find out whether students have mastered a key concept in a lesson, a criterion-referenced formative quiz or a performance would be the most appropriate.

How Are Tests Constructed?

Tests are constructed to elicit evidence of student learning in relation to the instructional objectives. Achievement tests should be constructed in keeping with six principles: They should (1) measure clearly defined learning objectives, (2) examine a representative sample of the learning tasks included in instruction, (3) include the types of test items most appropriate for measuring the desired learning outcomes, (4) fit the uses that will be made of the results, (5) be as reliable as possible and interpreted with caution, and (6) improve learning. A table of specifications helps in the planning of tests that correspond to instructional objectives. Types of test items include multiple-choice, true–false, completion, matching, short essay, and problem-solving items. Each type of test item has optimal uses, along with advantages and disadvantages. For example, if you want to learn how students think about, analyze, synthesize, or evaluate some aspect of course content, a short essay test might be most appropriate, provided that you have time to administer it and evaluate students' responses.

What Are Authentic, Portfolio, and Performance Assessments?

Portfolio assessment and performance assessment avoid the negative aspects of pencil-and-paper multiple-choice tests by requiring students to demonstrate their learning through work samples or direct real-world applications. Performance assessments are usually scored according to rubrics that specify in advance the type of performance expected.

How Are Grades Determined?

Grading systems differ in elementary and secondary education. For example, informal assessments might be more appropriate at the elementary level, whereas letter grades become increasingly important at the secondary level. Grading standards might be absolute or relative (grading on the curve). Performance grading is a way for teachers to determine what children know and can do. A key requirement for performance grading is judicious collection of work samples from students that indicate level of performance. Another approach is to give students tests in which they can show their abilities. Other systems include contract grading and mastery grading. Report card grades typically average scores on tests, homework, seatwork, class participation, deportment, and effort.