Guerilla/Alamy Stock Photo

CHAPTER THIRTEEN

# Assessing Student Learning

## CHAPTER OUTLINE



### LEARNING OUTCOMES

At the end of this chapter, you should be able to:

**13.1** Identify well-constructed instructional objectives and explain how they are used effectively

**13.2** Differentiate among types of evaluations based on their purposes

**13.3** Describe how to write fair, effective tests and a variety of types of test items

**13.4** Explain how to evaluate student work using authentic, portfolio, and performance assessment

**13.5** Describe how assessment of student learning influences intentional teaching

## CHAPTER OUTLINE (CONTINUED)



Mr. Sullivan is having a great time teaching about the Civil War, and his eleventh-grade U.S. history class is having fun, too. Mr. Sullivan is relating all kinds of anecdotes about the war. He describes a battle fought in the nude (a group of Confederates were caught fording a river), the time Stonewall Jackson lost a battle because he took a nap in the middle of it, and several stories about women who disguised their gender to fight as soldiers. He tells the story of a Confederate raid (from Canada) on a Vermont bank. He passes around real minié balls and grapeshot, and wondered if they had killed anyone. In fact, Mr. Sullivan has gone on for weeks about the battles, the songs, and the personalities and foibles of the generals. Finally, after an interesting math activity in which students have to figure out how much Confederate money they would need to buy a loaf of bread, Mr. Sullivan has students put away all their materials to take a test.

The students are shocked. The only question is: What were the main causes, events, and consequences of the Civil War?

Mr. Sullivan's lessons are fun. They are engaging. They use varied presentation modes. They integrate skills from other disciplines. They are clearly accomplishing one important objective of social studies: building enjoyment of the topic. However, as engaging as Mr. Sullivan's lessons are, there is little correspondence between what he is teaching and what he is testing. He and his students are on a happy trip, but where are they going?

### USING YOUR EXPERIENCE

**COOPERATIVE LEARNING** In a group of four or five students, draw a value line from 1 to 100, with 1 representing poor teaching and 100 representing great teaching. Take turns marking where you would place Mr. Sullivan on this scale. Let each person explain his or her rating. Now review the ratings and change them as appropriate. Discuss better methods that Mr. Sullivan might use to teach and then assess his students.

In teaching lessons, units, and courses, how do you know where you are going and whether or not you and your students are getting there? This chapter discusses objectives and assessments, as well as the goals of teaching and ways of determining whether goals are being achieved. Objectives are the learning plan for what students should know and be able to do at the end of a course of study; lessons must be designed to accomplish these objectives. Evaluations of students must indicate the extent to which each student has actually mastered those objectives by the end of the course (Banks, 2012; McMillan, 2011; Spinelli, 2011). Every teacher should have a clear idea of where the class is going, how it will get there, and how to know whether it has arrived.

## WHAT ARE INSTRUCTIONAL OBJECTIVES AND HOW ARE THEY USED?

InTASC 7
Planning for Instruction

What do you want your students to know or be able to do at the end of today's lesson? What should they know at the end of a series of lessons on a particular subject? What should they know at the end of the course? Knowing the answers to these questions is one of the

most important prerequisites for intentional, high-quality instruction (Burke, 2009; Moss & Brookhart, 2012). A teacher is like a wilderness guide with a troop of tenderfeet. If you do not have a map or a plan for getting the group where it needs to go, the whole group will surely be lost. Mr. Sullivan's students are having a lot of fun, but because their teacher has no plan for how his lessons will give them essential concepts related to the Civil War, they are unlikely to learn those concepts.

Setting out objectives at the beginning of a course is an essential step in providing a framework into which individual lessons will fit (Moss, Brookhart, & Long, 2011; Reeves, 2011). Without such a framework it is easy to wander off the track, to spend too much time on topics that are not central to the course. One high school biology teacher spent most of the year teaching biochemistry; her students knew all about the chemical makeup of DNA, red blood cells, chlorophyll, and starch, but little about zoology, botany, anatomy, or other topics that are usually central to high school biology. Then in late May the teacher panicked because she realized that the class had to do a series of laboratory exercises before the end of the year. On successive days they dissected a frog, a sheep eye, a sheep brain, and a pig fetus! Needless to say, the students learned little from those hurried labs and little about biology in general. This teacher did not have a master plan and was deciding week by week (or perhaps day by day) what to teach, thereby losing sight of the big picture—the scope of knowledge that is generally agreed to be important for a high school student to learn in biology class. Few teachers follow a plan rigidly once they make it, but the process of making it is still very helpful.

An **instructional objective**, sometimes called a *behavioral objective,* is a statement of skills or concepts that students are expected to know at the end of some period of instruction. Typically, an instructional objective is stated in such a way as to make it clear how the objective will be measured (see Mager, 1997). Some examples of instructional objectives are as follows:

- Given 100 division facts (such as 27 divided by 3), students will give correct answers to all 100 in 3 minutes.
- When asked, students will name at least five functions that characterize all living organisms (respiration, reproduction, etc.).
- In an essay, students will be able to compare and contrast the artistic styles of van Gogh and Gauguin.
- Given the statement "Resolved: The United States should not have entered World War I," students will be able to argue persuasively either for or against the proposition.

**Connections 13.1**
For more on lesson planning and lesson objectives as components of effective instruction, see Chapter 7.

Note that despite varying enormously in the type of learning involved and the performance levels they address, these objectives have several things in common. Mager (1997), whose work launched the behavioral objectives movement, described objectives as having three parts: performance, conditions, and criteria. Explanations and examples are given in Table 13.1.

**TABLE 13.1 • Parts of a Behavioral Objectives Statement**

| | PERFORMANCE | CONDITIONS | CRITERION |
|---|---|---|---|
| *Definition* | An objective always says what a learner is expected to do. | An objective always describes the conditions under which the performance is to occur. | Whenever possible, an objective describes the criterion for acceptable performance. |
| *Question Answered* | What should the learner be able to do? | Under what conditions do you want the learner to be able to do it? | How well must it be done? |
| *Example* | Correctly use adjectives and adverbs. | Given 10 sentences with missing modifiers, . . . | . . . the student will correctly choose an adjective or adverb in at least 9 of the 10 sentences. |

## Planning Lesson Objectives

In practice, the skeleton of a behavioral objective is condition–performance–criterion. First, state the conditions under which learning will be assessed, as in the following:

- Given a 10-item test, students will be able to . . .
- In an essay, the student will be able to . . .
- Using a compass and protractor, the student will be able to . . .

The second part of an objective is usually an action verb that indicates what students will be able to do. For example (from Gronlund & Brookhart, 2009):

- Write
- Distinguish between
- Identify
- Match
- Compare and contrast

Finally, a behavioral objective generally states a criterion for success, such as the following:

- . . . all 100 multiplication facts in 3 minutes.
- . . . at least five of the nations that sent explorers to the New World.
- . . . at least three similarities and three differences between the U.S. government under the Constitution and under the Articles of Confederation.

Sometimes a criterion for success cannot be specified as the number correct. Even so, success should be specified as clearly as possible:

- The student will write a two-page essay describing the social situation of women as portrayed in *A Doll's House.*
- The student will think of at least six possible uses for an eggbeater other than beating eggs.

**WRITING SPECIFIC OBJECTIVES** Instructional objectives must be adapted to the subject matter being taught. When students must learn well-defined skills or information with a single right answer, specific instructional objectives can be written as follows:

- Given 10 problems involving addition of two fractions with like denominators, students will solve at least 9 correctly.
- Given 10 sentences lacking verbs, students will correctly choose verbs that agree in number in at least 8 sentences. Examples: My cat and I [has, have] birthdays in May. Each of us [want, wants] to go to college.
- Given a 4-meter rope attached to the ceiling, students will be able to climb to the top in less than 20 seconds.

Some material, of course, does not lend itself to such specific instructional objectives, and it would be a mistake in such cases to adhere to objectives that have numerical criteria. For example, the following objective could be written:

- The student will list at least five similarities and five differences between the situation of immigrants to the United States in the early 1900s and that of immigrants today.

Note that this objective asks for lists, which might not demonstrate any real understanding of the topic. A less specific but more meaningful objective might be the following:

- In an essay, the student will compare and contrast the situation of immigrants to the United States in the early 1900s and that of immigrants today.

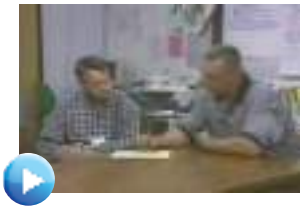
This general instructional objective would allow students more flexibility in expressing their understanding of the topic and promote comprehension rather than memorization of lists of similarities and differences.

**WRITING CLEAR OBJECTIVES** Instructional objectives should be specific enough to be meaningful. For example, consider the following objective concerning immigrants:

- Students will develop a full appreciation for the diversity of peoples who have contributed to the development of U.S. society.

This sounds nice, but what does "full appreciation" mean? Such an objective neither helps you prepare lessons nor helps students understand what is to be taught and how they will be assessed.

**PERFORMING A TASK ANALYSIS** In planning lessons, it is important to consider the skills required in the tasks to be taught or assigned. For example, you might ask students to use the Internet to write a brief report on a topic of interest. The task seems straightforward enough, but consider the separate skills involved:



MyEdLab

**Video Example 13.1**

Two teachers discuss plans for a unit on civilizations and their impact on history. Can you identify the prerequisites that they expect the students to know and the component skills they expect students to use and develop during the unit?

- Knowing how to find information on the Internet
- Knowing how to judge sources on the Internet according to their objectivity and accuracy
- Getting the main idea from expository material
- Planning or outlining a brief report
- Writing expository paragraphs
- Understanding language mechanics skills (such as capitalization, punctuation, and usage)

These skills could themselves be broken down into subskills. You must be aware of the subskills involved in any learning task to be certain that students know what they need to know to succeed. Before assigning the Internet report task, you would need to be sure that students knew how to use Internet resources and that they could comprehend and write expository material. You might teach or review these skills before sending students to their computers.

In teaching any new skill, it is important to consider all the subskills that go into it. Think of all the separate steps involved in adding fractions, in writing chemical formulas, or in identifying topic sentences and supporting details. For that matter, consider the skills that go into writing a business letter on MS Word, as illustrated in Figure 13.1.

This process of breaking tasks or objectives down into their simpler components is called **task analysis**. In planning a lesson, a three-step process for task analysis may be used.

1. ***Identify prerequisite skills.*** What should students already know before you teach the lesson? For example, for a lesson on adding fractions with unlike denominators, students need to know how to find least common multiples, how to multiply to find equivalent fractions, how to add fractions with like denominators, and how to simplify fractions.
2. ***Identify component skills.*** In the actual lesson, what subskills must students be taught before they can learn to achieve the larger objective? To return to the adding-fractions example, each of the steps must be planned for, taught, and assessed during the lesson.
3. ***Plan how component skills will be assembled into the final skill.*** The final step in task analysis is to reassemble the subskills into the complete process being taught. For example, students might be able to do each of the skills needed to add fractions with unlike denominators, but this does not necessarily mean they can put them all together to do the whole task. The subskills must be integrated into a complete process that students can understand and practice.

**BACKWARD PLANNING** Just as lesson objectives are more than the sum of specific task objectives, the objectives of a course of study are more than the sum of specific lesson objectives. For this reason it makes sense to start by writing broad objectives for the course as a whole, then objectives for large units, and only then specific behavioral objectives (see Gronlund & Brookhart, 2009). This is known as **backward planning**. For example, Mr. Sullivan would have done well to identify the objective of his Civil War unit as follows:

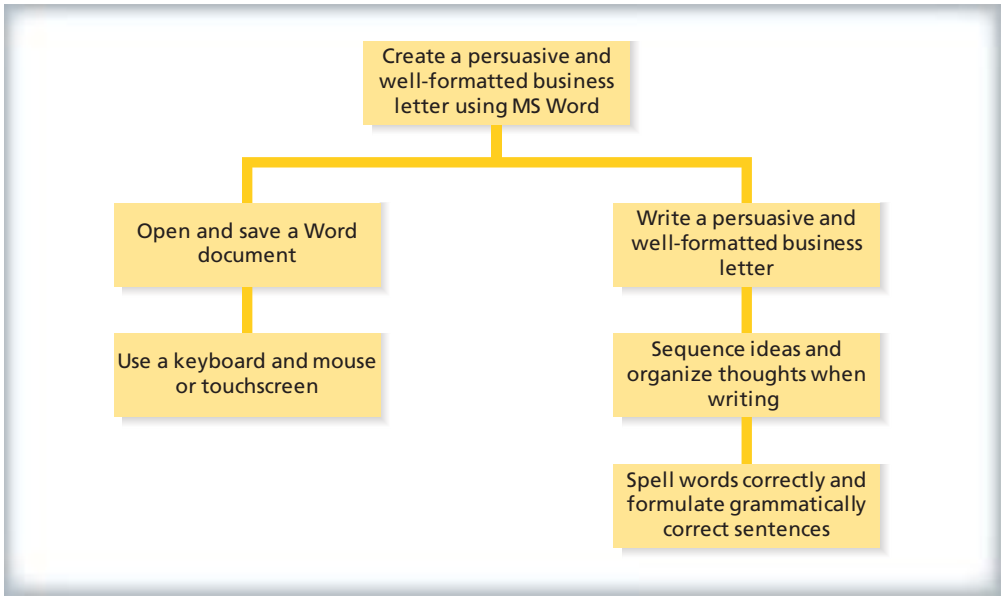



**FIGURE 13.1 • Example of a Skill Hierarchy**

Before students can practice the main skill (writing a business letter on MS Word), they must be able to use a computer and compose a letter. These skills must all be learned before the main skill can be mastered. They are independent of one another and can be learned in any order. Before composing a letter, students must be able to spell words and organize written ideas. Finally, to use a computer, the learner first has to learn how to use a mouse/touchscreen and how to open and save documents.

"Students will understand the major causes, events, and consequences of the Civil War." Next he might have written more detailed objectives related to causes, events, and consequences and then planned units and individual lessons around these objectives. A detailed example of the backward planning process is illustrated in Table 13.2 and described in the next Theory into Practice.

## TABLE 13.2 • Example of Objectives for a Course in Language Arts

Teachers can allocate instructional time for a course by (a) deciding what topics to cover during the year or semester, (b) deciding how many weeks to spend on each topic, (c) choosing units within each topic, (d) deciding how many days to spend on each, and (e) deciding what each day's lesson should be.

| COURSE OBJECTIVES (WEEKS ALLOCATED) | UNIT OBJECTIVES (DAYS ALLOCATED) | LESSONS |
|---|---|---|
| Writing fictional stories: 3<br>Writing nonfiction: 3<br>Writing persuasive essays: 2<br>. . . etc. | Parts of a story: 2<br>Details and elaboration: 3<br>Writing, revision, and the creative process: 3<br>. . . etc. | Lesson 1: Parts of a Story—Overview<br>Introduction<br>Setting<br>Protagonist<br>Antagonist<br>Plot<br>Conclusion<br>Lesson 2: Parts of a Story-Identification from Examples<br>Introduction<br>Setting<br>Protagonist<br>Antagonist<br>Plot<br>Conclusion |

## THEORY INTO PRACTICE

### Planning Courses, Units, and Lessons

In planning a course, it is important to set long-term, middle-term, and short-term objectives before starting to teach (Diamond, 2008; Dougherty, 2012; Fisher & Frey, 2014c; Reeves, 2015). Before the students arrive for the first day of class, you need to have a general plan of what will be covered all year, a more specific plan for what will be in the first unit (a connected set of lessons), and a very specific plan for the content of the first lessons (as shown in Table 13.2). All states and many districts have established standards for each subject, and these standards should help guide your planning.

Table 13.2 implies a backward planning process. First, the course objectives are established. Then unit objectives are designated. Finally, specific lessons are planned. The course objectives list all the topics to be covered during the year. You might divide the number of weeks in the school year by the number of major topics to determine what each will require. More or less time could be reserved for any particular topic, as long as adequate time is allowed for the others. A whole semester could be spent on any one of the topics in Table 13.2, but this would be inappropriate in a survey course on life science. You must make hard choices before the first day of class about how much time to spend on each topic to avoid spending too much time on early topics and not having enough time left to do a good job with later ones. Some history teachers always seem to find themselves still on World War I in mid-May and have to compress most of the 20th century into a couple of weeks!

Table 13.2 shows approximate allocations of weeks to each of the topics to be covered. These are only rough estimates to be modified as time goes on.

**Unit Objectives and Unit Tests** After course objectives have been laid out, the next task is to establish objectives for the first unit and to estimate the number of class periods to spend on each objective (Diamond, 2008). It is a good idea to write a unit test as part of the planning process. Writing a test in advance helps you focus on the important issues to be covered. For example, in a 4-week unit on the Civil War, Mr. Sullivan might have decided that the most important concepts students should learn are the causes of the war, a few major points about the military campaigns, the importance of the Emancipation Proclamation, Lincoln's assassination, and the history of the Reconstruction period. These topics would be central to the unit test on the Civil War. Writing this test could have helped him put into proper perspective the importance of the various issues that should be covered. It's not that he shouldn't have shared anecdotes and shown students Civil War weapons, but preparing the unit test would have helped him keep the big picture in mind.

The test that you prepare as part of your course planning might not be exactly the test that you give at the end of the unit. You may decide to change, add, or delete items to reflect the content you actually covered. But this does not diminish the importance of having decided in advance exactly what objectives you wanted to achieve and how you were going to assess them.

Many texts provide unit tests and objectives, making your task easier. Sample objectives and test items are available from state and local departments of education, and they can be found on the Internet. However, even if you have ready-made objectives and tests, it is still important to review their content and change them as necessary to match what you expect to teach.

**Certification Pointer**

For your teacher certification test, you may be asked to take a goal from a state curriculum standard and write a behavioral objective to meet that standard.

If you prepare unit tests from scratch, use the guide to test construction presented later in this chapter. Be sure that the test items cover the various objectives in proportion to their importance to the course as a whole (that is, the more important objectives are covered by more items), and include items that assess higher-level thinking as well as factual knowledge.

**Lesson Plans and Lesson Assessments** The final step in backward planning is to plan daily lessons. Table 13.2 shows how a given unit objective might be broken down into daily lessons. The next step is to plan the content of each lesson. A lesson plan consists of an objective, a plan for presenting information, a plan for giving students practice (if appropriate), a plan for assessing student understanding, and, if necessary, a plan for reteaching students (or whole classes) if their understanding is inadequate.

## Aligning Objectives and Assessment

InTASC 6

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 − 184 a2. 412 − 298 a3. 596 − 448 |
| 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:<br>c1. Water _____<br>c2. Carbon dioxide _____<br>c3. Coal _____<br>c4. Table salt _____ |

## Using Taxonomies of Instructional Objectives

**Connections 13.2**
For information on thinking skills and critical thinking, see Chapter 8.

**INTASC 5**
Application of Content

In writing objectives and assessments, it is important to consider different skills and different levels of understanding. For example, in a science lesson on insects for second-graders, you might want to impart both information (the names of various insects) and a set of attitudes (such as an appreciation of the importance of insects to the ecosystem and the idea that science is fun). In other subjects you might try to convey facts and concepts that differ by type. For example, in teaching a lesson on topic sentences in reading, you might have students first recall the definition of topic sentences, then identify topic sentences in paragraphs, and finally write their own topic sentences for original paragraphs. Each of these activities demonstrates a different kind of understanding of the concept "topic sentence," and this concept has not been adequately taught if students can do only one of these activities. These various lesson goals can be classified by type and degree of complexity. A taxonomy, or system of classification, helps you to categorize instructional activities.

MyEdLab
**Video Example 13.2**
In a lesson intended to encourage students to see relationships between geography and economy, Ms. Holmquest encourages her class to analyze information. It's clear from this lesson, as well as from the chart on the board, that previous lessons focused on knowledge and comprehension.

**BLOOM'S TAXONOMY** In 1956, Benjamin Bloom and some fellow researchers (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956) published a **taxonomy of educational objectives** that has been influential in the research and practice of education ever since. Bloom and his colleagues categorized objectives from simple to complex or from factual to conceptual. The following key elements (from simple to complex) constitute what is commonly called Bloom's taxonomy for the cognitive domain (Badgett & Christmann, 2009; Marzano & Kendall, 2007).

1. ***Knowledge (recalling information).*** The lowest level of objectives in Bloom's hierarchy, knowledge comprises objectives such as memorizing math facts or formulas, scientific principles, or verb conjugations.
2. ***Comprehension (translating, interpreting, or extrapolating information).*** Comprehension objectives require that students show an understanding of information as well as the ability to use it. Examples include interpreting the meaning of a diagram, graph, or parable; inferring the principle underlying a science experiment; or predicting what might happen next in a story.
3. ***Application (using principles or abstractions to solve novel or real-life problems).*** Application objectives require students to use knowledge or principles to solve practical problems. Examples include using geometric principles to figure out how many gallons of water to put into a swimming pool of given dimensions, or using knowledge of the relationship between temperature and pressure to explain why a balloon is larger on a hot day than on a cold day.
4. ***Analysis (breaking down complex information or ideas into simpler parts to understand how the parts relate or are organized).*** Analysis objectives require students to see the underlying structure of complex information or ideas. Examples of analysis objectives include contrasting schooling in the United States with education in Japan, or identifying the main idea of a short story.
5. ***Synthesis (creation of something that did not exist before).*** Synthesis objectives involve using skills to create completely new products. Examples include writing a composition, deriving a mathematical rule, designing a science experiment to solve a problem, or making up a new sentence in a foreign language.
6. ***Evaluation (judging something against a given standard).*** Evaluation objectives require making value judgments against some criterion or standard. For example, students might be asked to compare the strengths and weaknesses of two tablet computers in terms of flexibility, power, and available software.

Because Bloom's taxonomy is organized from simple to complex, some people interpret it as a ranking of objectives from trivial (knowledge) to important (synthesis, evaluation). However, this is not the intent of the taxonomy. Different levels of objectives are appropriate for different purposes and for students at different stages of development (Marzano & Kendall, 2007). For example, you want your physician to have a deep understanding of how the human body works, but you also hope she knows the names of all the body parts, medicines, and devices in her area of specialization, all knowledge-level objectives!

**TABLE 13.3 • Examples of Objectives in a Behavior Content Matrix**

A behavior content matrix can remind teachers to develop instructional objectives that address skills at various cognitive levels.

| TYPE OF OBJECTIVE | EXAMPLE 1: THE AREA OF A CIRCLE | EXAMPLE 2: MAIN IDEA OF A STORY | EXAMPLE 3: THE COLONIZATION OF AFRICA |
|---|---|---|---|
| Knowledge | Give the formula for area of a circle. | Define *main idea*. | Make a time line showing how Europeans divided Africa into colonies. |
| Comprehension | | Give examples of ways to find the main idea of a story. | Interpret a map of Africa showing its colonization by European nations. |
| Application | Apply the formula for area of a circle to real-life problems. | | |
| Analysis | | Identify the main idea of a story. | Contrast the goals and methods used in colonizing Africa by the different European nations. |
| Synthesis | Use knowledge about the areas of circles and volumes of cubes to derive a formula for the volume of a cylinder. | Write a new story based on the main idea of the story read. | Write an essay on the European colonization of Africa from the perspective of a Bantu chief. |
| Evaluation | | Evaluate the story. | |

The primary importance of Bloom's taxonomy is in its reminder that we want students to have many levels of skills. All too often, teachers focus on measurable knowledge and comprehension objectives and forget that students cannot be considered proficient in many skills until they can apply or synthesize those skills (see Iran-Nejad & Stewart, 2007). On the other side of the coin, some teachers fail to make certain that students are well rooted in the basics before heading off into higher-order objectives.

**USING A BEHAVIOR CONTENT MATRIX** One way to be sure that your objectives cover many levels is to write a **behavior content matrix**. This is simply a chart that shows how a particular concept or skill will be taught and assessed at different cognitive levels. Examples of objectives in a behavior content matrix appear in Table 13.3. Note that for each topic, objectives are listed for some but not all levels of Bloom's taxonomy. Some topics do not lend themselves to some levels of the taxonomy, and there is no reason why every level should be covered for every topic. However, using a behavior content matrix in setting objectives forces you to consider objectives above the knowledge and comprehension levels.

**OBJECTIVES BEYOND THE BASICS** Learning facts and skills is not the only important goal of instruction. Sometimes the feelings that students have about a subject or about their own skills are at least as important as how much information they learn. Instructional goals related to attitudes and values are called **affective objectives**. Many people would argue that a principal purpose of a U.S. history or civics course is to promote values of patriotism and civic responsibility, and that one purpose of any mathematics course is to give students confidence in their ability to use mathematics. In planning instruction, it is important to consider affective as well as cognitive objectives. Love of learning, confidence in learning, and development of prosocial, cooperative attitudes are among the most important objectives you should have for your students. Sternberg (2008) suggests that schools supplement objectives related to the 3R's (reading, 'riting, and 'rithmetic) with three more R's: reasoning, resilience, and responsibility (also see Rothstein & Jacobsen, 2009; Stiggins & Chappuis, 2012). In addition, creativity is an objective worth pursuing, even if measuring it is not straightforward (Brookhart, 2013a).

## Research on Instructional Objectives

Three principal reasons are given for writing instructional objectives. One is that this exercise helps to organize your planning. As Mager (1997) puts it, if you're not sure where you're going, you're liable to end up someplace else and not even know it. Another is that establishing instructional objectives helps to guide evaluation. Finally, it is hypothesized that development of instructional objectives improves student achievement.

Although it would be a mistake to overplan or adhere rigidly to an inflexible plan, most experienced teachers create, use, and value objectives and assessments that are planned in advance. Perhaps the most convincing support for the establishment of clear instructional objectives is indirect. Cooley and Leinhardt (1980) found that the strongest single factor predicting student reading and math scores was the degree to which students were actually taught the skills that were tested. This implies that instruction is effective to the degree to which objectives, teaching, and assessment are coordinated with one another. Specification of clear instructional objectives is the first step in ensuring that classroom instruction is directed toward giving students critical skills, those that are important enough to test.

It is important to make sure that instructional objectives that are communicated to students are broad enough to encompass everything the lesson or course is supposed to teach. There is some danger that giving students too narrow a set of objectives might focus them on some information to the exclusion of other facts and concepts.

## WHY IS EVALUATION IMPORTANT?

**InTASC 6**

**Assessment**

**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.

**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 13.4**
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.

**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 13.4 • 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 achievement (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).

**Connections 13.9**
For more on rewards that are too easy to attain, see Chapter 10.

## 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 science 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. Unlike 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).

**Connections 13.10**
For more on grades and standardized tests, see Chapter 14.

There are two keys to reliable summative assessment. First, you should use multiple assessment opportunities (Brookhart & Nitko, 2015; Popham, 2014a). No student should receive a grade based

### 21ST CENTURY LEARNING

The advent of the Common Core State Standards, discussed in detail in Chapter 14, has profound implications for classroom assessment. Assessments based on the Common Core will be used for accountability in most states. To be fair to students (and teachers), students should have regular opportunities to practice activities and assessments based on Common Core State Standards. The English/Language Arts standards emphasize analysis and integration of content, writing, use of technology, and collaboration, all of which can and should affect classroom assessments as well. The Mathematics standards emphasize problem solving, reasoning, constructing arguments, and collaboration. These would be valuable parts of regular assessment even if the Common Core didn't exist, but the widespread adoption of the Common Core gives teachers one more good reason to focus on deeper learning, integration of diverse content, technology use, and collaboration (Marzano et al., 2013; Zhao, 2015).

**Creativity and Authentic Problem Solving**

Since the progressive era began a hundred years ago, educators have advocated creativity and authentic problem-solving skills as key outcomes of education. Yet because these outcomes are difficult and time-consuming to measure reliably, they are often downplayed in comparison to relatively easy-to-assess facts and skills. New assessment software, especially adaptive testing, is beginning to make regular assessment of creativity and authentic problem solving more practical, as when computers are employed to

**InTASC 5**
Application of Content

pose complex, open-ended problems at the student's precise level of functioning. Such solutions, used as frequent benchmark assessments as well as summative assessments, may soon help teachers both to focus more on these essential 21st century skills and to monitor students' development as creative problem-solvers.

**QUESTION**

- What are the potential difficulties in judging creativity?
- Do you think it is ethical to penalize a student based on what is judged to be a lack of creativity?
- Do you think creativity should count toward grades?
- Is creativity a necessary life skill?
- How might a lack of creativity be a problem in the workplace?

on only one test, because too much can go wrong with only one assessment. Second, you should test learning when it is completed, not as it is developing. It is better to collect summative evaluation information as students complete instructional units, as well as to use major unit and final tests.

MyEdLab **Self-Check 13.2**

MyEdLab **Video Analysis Tool 13.1** Go to MyEdLab and click on the Video Analysis Tool to access the exercise "Formative assessment: teacher's perspective."

MyEdLab **Video Analysis Tool 13.2** Go to MyEdLab and click on the Video Analysis Tool to access the exercise "Formative assessment: revision and practice."

# HOW ARE TESTS CONSTRUCTED?

Once you know the concept domains to be assessed in a test of student learning, it is time to write test items. Writing good achievement tests is a critical skill for effective teaching. This section presents some basic principles of achievement testing and some practical tools for test construction (see Chappuis, 2015; Miller et al., 2013; Popham, 2014; Salend, 2016; Witte, 2012). Achievement testing is taken up again in Chapter 14 in relation to standardized tests.

**Connections 13.11**
For more on achievement testing in relation to standardized tests, see Chapter 14.

## Principles of Achievement Testing

Gronlund and Brookhart (2009) listed six principles to keep in mind in preparing achievement tests, paraphrased as follows:

**Connections 13.12**
For more on the characteristics and uses of standardized achievement tests, see Chapter 14.

1. ***Achievement tests should measure clearly defined learning objectives that are in harmony with instructional objectives.*** Perhaps the most important principle of achievement testing is that the tests should correspond with the course objectives and with the instruction that is actually provided (Lloyd et al., 2013; Squires, 2009; Thorndike & Thorndike-Christ, 2010; Waugh & Gronlund, 2013). An achievement test should never be a surprise for students; rather, it should assess the students' grasp of the most important concepts or skills the lesson or course is supposed to teach.
2. ***Achievement tests should measure a representative sample of the learning tasks included in the instruction.*** With rare exceptions (such as multiplication facts), achievement tests do not assess every skill or fact students are supposed to have learned. Rather, they sample from among all the learning objectives. If students do not know in advance what questions will be on a test, then they must study the entire course content to do well. However, the test items must be representative of all the objectives (contents and skills) that were covered. For example, if an English literature course spent 8 weeks on Shakespeare and 2 weeks on other Elizabethan authors, the test should have about four times as many items related to Shakespeare as to the others. Items that are chosen to represent a particular objective must be central to that objective. There is no place in achievement testing for tricky or obscure questions. For example, a unit test on the American Revolution should ask questions related

to the causes, principal events, and outcomes of that struggle, not about who rowed George Washington across the Delaware. (*Answer:* John Glover and his Marblehead Marines.)

3. ***Achievement tests should include the types of test items that are most appropriate for measuring the desired learning outcomes.*** Items on achievement tests should correspond as closely as possible to the ultimate instructional objectives (Banks, 2012; Schimmer, 2014; Witte, 2012). For example, in mathematics problem solving, one of your goals might be to enable students to solve problems like the ones they will encounter outside of school. Matching items or multiple-choice questions might be inappropriate for this kind of exam, because in real life we do not select from a menu of possible solutions to a problem.
4. ***Achievement tests should fit the particular uses that will be made of the results.*** Each type of achievement test has its own requirements. For example, a test that is used for diagnosis would focus on particular skills with which students might need help. A diagnostic test of elementary arithmetic might contain items on subtraction involving zeros in the minuend (e.g., 307 minus 127), a skill with which many students have trouble. In contrast, a test that is used to predict future performance might assess a student's general abilities and breadth of knowledge. Formative tests should be very closely tied to material that has recently been presented, whereas summative tests should survey broader areas of knowledge or skills.
5. ***Achievement tests should be as reliable as possible but should nevertheless be interpreted with caution.*** A test is reliable to the degree that students who are tested a second time fall in the same rank order. In general, writers of achievement tests increase test reliability by using relatively large numbers of items and by including few items that nearly all students get right or that nearly all students miss (O'Connor, 2009). The use of clearly written items that focus directly on the objectives that are actually taught also enhances test reliability. Still, no matter how rigorously reliability is built into a test, there will always be some error of measurement. Students have good and bad days or can be lucky or unlucky guessers. Some students are test-wise and usually test well; others are test-anxious and test far below their actual knowledge or potential. Therefore, no single test score should be viewed with excessive confidence. Any test score is only an approximation of a student's true knowledge or skills and should be interpreted as such.
6. ***Achievement tests should improve learning.*** Achievement tests of all kinds, particularly formative tests, provide important information on students' learning progress (Dueck, 2014; Sousa, 2016). Stiggins and Chappuis (2012), for example, urge that assessments *for* learning are more important than assessments *of* learning. Achievement testing should be seen as part of the instructional process and used to improve instruction and guide student learning (Chappuis, 2015; Russell & Airasian, 2012). This means that achievement test results should be clearly communicated to students soon after the test is taken; in the case of formative testing, students should be given the results immediately. Teachers should use the results of formative and summative tests to guide instruction, to locate strong and weak points in students' understandings, and to set an appropriate pace of instruction.

**Connections 13.13**
For more on the reliability of achievement tests, see Chapter 14.

**InTASC 6**

Assessment

## THEORY INTO PRACTICE

### Making Assessments Fair

Although fairness in assessment is something everyone believes in, defining fairness in assessment is not straightforward. Indeed, the latest edition of the *Standards for Educational and Psychological Testing* gives four definitions and acknowledges that many more appear in the literature (AERA/APA/NCME, 1999). Fairness means being honest, impartial, and free from discrimination. Besides being ethical, fairness makes good instructional sense. Fair testing encourages students to expend more effort on learning because they come to see that success depends only on what they know and can do (Oosterhof, 2009).

Fairness in assessment arises from good practice in four phases of testing: writing, administering, scoring, and interpreting assessments. Practices that lead to fairness in these areas are considered separately below.

***Writing Assessments*** Base assessments on course objectives. Students expect a test to cover what they have been learning. They also have a right to a test that neither "tricks" them into wrong answers nor rewards them with a high score for guessing or bluffing.

Avoid contexts and expressions that are more familiar or intriguing to some students than to others. One challenge in writing tests is to make sure none of your students are advantaged or disadvantaged because of their different backgrounds. For example, music, sports, or celebrity-related examples might be appealing to some students but not others. Language choices and specific topics should not be used if they are more well known or interesting to some students than to others. If avoiding such choices proves impossible, then at least make sure the items that favor some students are balanced by others that favor the rest.

***Giving Assessments*** Make sure students have had equal opportunities to learn the material on the assessment. Regardless of whether students have learned as much as they can, at least they should have had equal chances to do so. If some students are given extra time or materials that are withheld from others, those others probably will not feel they have been treated fairly.

Make sure students are familiar with the formats they will be using to respond. If some students are not comfortable with the types of questions on an assessment, they will not have an equal chance to show what they can do. If that might be the case, provide some practice with the format beforehand to help them succeed.

Give plenty of time. Most tests in education do not cover content that will eventually be used under time pressure. Thus, most assessments should reward quality instead of speed. Only by allowing enough time so that virtually all students have an opportunity to answer every question can you prevent haste from being a barrier to performance.

***Scoring Assessments*** Make sure the rubric used to score responses awards full credit to an answer that is responsive to the question asked, as opposed to requiring more information than requested for full credit. If the question does not prompt the knowledgeable student to write an answer that receives full credit, then it should be changed. It is unfair to reward some students for doing more than has been requested in the item; not all students will understand the real (and hidden) directions because they have not been told.

***Interpreting Assessments*** Base grades on summative end-of-unit assessments rather than formative assessments that are used to make decisions about learning as it is progressing. The latter are intended to be diagnostic and are used to help accomplish learning. Because grades certify attainment, they should be determined on the basis of assessments made after learning has taken place.

Base grades on several assessment formats (McTighe & Wiggins, 2013). Because students differ in their preferred formats, some are advantaged by selected-response tests, others by essay tests, others by performance assessments, and still others by papers and projects. Also, base grades on multiple assessments taken at different times. Make appropriate accommodations for English learners and for students with disabilities, such as giving more time on tests if they need it (Herrera, Cabral, & Murray, 2013; Voltz, Sims, & Nelson, 2010). Finally, make sure factors that might result in atypical performance for a student are recognized to minimize the importance of the student's score on that assessment. If it is known that a student has not done her or his best, then basing a grade or other important decision on that assessment is not only unfair but also inaccurate.

**ON THE WEB**

DiscoverySchool.com has rubrics for every kind of assessment imaginable, as well as lesson plans and other useful information for educators. For student project assessment rubrics, see eduscapes.com. For a website on creating a rubric, see http://rubistar.4teachers.org/ and http://elearningindustry.com/the-5-best-free-rubric-making-tools-for-teachers.

## Using a Table of Specifications

Achievement tests should measure well-specified objectives. The first step in the test development process is to decide which concept domains the test will measure and how many test items will be allocated to each concept. Waugh & Gronlund (2013) suggest that teachers make up a **table of specifications** for each instructional unit, listing the various objectives taught and the different levels of understanding to be assessed (also see Guskey, 2005). The levels of understanding might correspond to Bloom's taxonomy of educational objectives (Bloom et al., 1956; Marzano & Kendall, 2009). Bloom, Hastings, and Madaus (1971) recommend classifying test items for each objective according to six categories, as shown in Table 13.5, a table of specifications for a social studies unit.

The table of specifications varies for each type of course and is nearly identical to behavior content matrixes, discussed earlierin this chapter. This is as it should be; a behavior content matrix is used to lay out objectives for a course, and the table of specifications tests those objectives.

Once you have written items corresponding to your table of specifications, look over the test in its entirety and evaluate it against the following standards:

1. Do the items emphasize the same concepts you have emphasized in day-to-day instruction? (Recall how Mr. Sullivan, in the chapter-opening vignette, ignored this commonsense rule.)
2. Has an important area of content or any objective been overlooked or underemphasized?
3. Does the test cover all levels of instructional objectives included in the lessons?
4. Does the language of the items correspond to the language and reading level you used in the lessons?
5. Are the instructions clear, even for students who have difficulty with instructions?
6. Is there a reasonable balance between what the items measure and the amount of time that will be required for students to develop a response?
7. Did you write model answers or essential component outlines for the short essay items? Does the weighting of each item reflect its relative value among all the other items?

Evaluation that is restricted to information acquired from paper-and-pencil tests provides only certain kinds of information about students' progress in school. Other sources and strategies for appraisal of student work must be used, including checklists, interviews, classroom simulations, role-playing activities, and anecdotal records. To do this systematically, you may keep a journal or log to record concise and cogent evaluative information on each student throughout the school year.

## 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**.

## TABLE 13.5 • Table of Specifications for a Social Studies Unit on the Suffragists

This table of specification classifies test items and objectives according to six categories ranging from knowledge of terms to ability to apply knowledge.

| A. KNOWLEDGE OF TERMS | B. KNOWLEDGE OF FACTS | C. KNOWLEDGE OF RULES AND PRINCIPLES | D. SKILL IN USING PROCESSES AND PROCEDURES | E. ABILITY TO MAKE TRANSLATIONS | F. ABILITY TO MAKE APPLICATIONS |
|---|---|---|---|---|---|
| Suffrage | Make a timeline of important events in the suffrage movement | What laws did the suffragists break? | | Make a Venn diagram to compare and contrast the suffragists with other groups in America who were denied their right to vote. | Write a diary entry from the point of view of a teenage girl whose mother was put in jail during a suffragists rally. |
| Equality | Give three examples of ways women were not given rights that were equal to those given to men | Compare the rights of women in the United States in 1920 to the rights of women in ancient Athens. | How were the principles stated in the Declaration of Independence not in line with the laws for women, and especially for African-American women? | The class will form two teams who will debate about whether equality exists among all students within our school. | Research and write a report on gender inequality today. Contrast it to the gender inequalities that existed during the American suffrage movement. What do you think about the term "equality"—can there be degrees of equality? |
| Ballot | Write a short paragraph on what a ballot is | Pretend you are in charge of a ballot box in 1916. Make a list of rules you would post on it. | Name a situation where you'd use a ballot even now | After voters cast ballots, how are they used to determine who wins an election? | In many countries, people cannot vote. If we could not vote in our country, how might this change our government and people's lives? |
| Civic | Define civic | What does it mean to be civic-minded? | What actions can you take to be civic-minded? | Identify five people in history who are/ were civic minded and explain what they did | Develop a hypothetical civic-improvement project. |
| Picket | Explain what it means to picket | Is it legal to picket? | Make a picket sign about something serious you'd like to protest | Find a newspaper article about a situation where picketing was involved. Write a 1-2 paragraph opinion about whether you think that was the most effective way to protest in that situation. | Why is the right to picket an important part of our democracy? |

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.

## THEORY INTO PRACTICE

### Writing Multiple-Choice Tests (Format Suggestions)

The following guidelines are useful resources for constructing effective multiple-choice items (see Badgett & Christmann, 2009; McMillan, 2011).

**1.** Make the stem sufficiently specific to stand on its own without qualification. In other words, the stem should contain enough information to set the context for the concepts in it. The following example shows a stem for which insufficient context has been established:

Applied behavior analysis can be
a. classical conditioning.
b. punishment.
c. reinforcement contingencies.
d. self-actualization.

An improved version of this stem is the following:

What is the main emphasis of modern classroom use of applied behavior analysis?
a. Classical conditioning
b. Punishment
c. Reinforcement contingencies [*correct choice*]
d. Self-actualization

**2.** Avoid long and complicated stems unless the purpose of the item is to measure a student's ability to deal with new information or to interpret a paragraph. The stem should not be too wordy; a test is not the place to incorporate instruction that should have been given in the lessons.

**3.** The stem and every choice in the list of potential answers ought to fit grammatically. In addition, avoid repeating phrases or words to begin each of the alternatives; instead, these should be part of the stem. It is also a sound idea to have the same grammatical form (say, a verb) at the beginning of each choice, as in the following example.

The task of statistics is to
a. *make* the investigation of human beings more precise and rigorous.
b. *promote* the social sciences to a status of being as respectable as the physical sciences.
c. *predict* human behavior.
d. *reduce* large masses of data to an interpretable form. [correct choice]

**4.** Take special care in using no-exception words such as *never, all, none,* and *always.* These words are most commonly found in incorrect statements because the requirement of no exceptions usually makes statements wrong. In multiple-choice items these words often

give clues to the test-wise but concept-ignorant student. Words allowing qualification, such as *often, sometimes, seldom, usually, typically, generally,* and *ordinarily,* are most often found in correct statements (or responses that are true) and, along with no-exception words, such specific determiners should be avoided whenever possible, or at least distributed among both correct answers and distractors.

**5.** Avoid making the correct choice the only one that is qualified (e.g., by an "if" clause). Also, it should be neither the longest nor the shortest of the alternatives (the longest can otherwise be guessed to be right because absolutely correct answers often require qualification and precision). These features make a choice stand out.

**6.** Do not use an item that can be answered on the basis of information contained in another item on the same test.

**7.** Avoid overly inclusive options that contain other options. For example, the choices "dogs" and "setters" should not be in the same item because a setter is a type of dog. Similarly, be cautious in using "all of the above" as an alternative, because it often reduces the possible correct choices to one or two alternatives. The following example illustrates how a student might know very little and get the correct answer. By knowing that only one of the choices is incorrect, a student will reduce the number of plausible choices from four to two.

What type of research is best for investigating the effects of a new instructional program on mathematics achievement?

a. Correlational
b. Experimental [*correct choice*]
c. Historical
d. All of the above

The student who knows that "Historical" is not a good choice also knows that item d must be incorrect, so the answer must be a or b.

**8.** After a test, discuss the items with students, and note their interpretations of the wording of each. Students often understand certain phrases quite differently from the way you may have intended. Such feedback will help you revise items for the next test, as well as inform you about students' understandings.

**9.** Do not include a choice that is transparently absurd. All choices should seem plausible to a student who has not studied.

Besides these guidelines for writing multiple-choice items, here are some suggestions about format:

- List the choices vertically rather than side by side.
- Use letters rather than numerals to label the choices, especially on scientific and mathematical tests.
- Use word structures that make the stem agree with the choices according to acceptable grammatical practice. For example, a completion-type stem would require that each of the choices begin with a lowercase letter (unless it begins with a proper noun).
- Avoid repeating the same word or phrase in the stem and in only one alternative.
- Avoid overusing one letter position as the correct choice, as well as any patterns in the correct answers. Instead, correct choices should appear in random letter positions.

As an illustration of how test "wiseness" rather than knowledge can help students pass an exam, take the brief test in Figure 13.2.

**FIGURE 13.2 • A Test of "Test Wiseness"**

The following test is about a made-up country, Quizzerland. Use your test wiseness to guess the answers to these very bad items.

1. What is the main currency used in Quizzerland?
   - **a.** Dollar
   - **b.** Peso
   - **c.** Quark
   - **d.** Pound
2. Describe the pattern of annual rainfall in Quizzerland.
   - **a.** Mostly rainy in the highlands, dry in the lowlands
   - **b.** Rainy
   - **c.** Dry
   - **d.** Snowy
3. How many children are there in Quizzerlandian families?
   - **a.** Never more than 2
   - **b.** Usually 2–3
   - **c.** Always at least 3
   - **d.** None
4. What would be the correct response to any question asked here?
   - **a.**
   - **b.**
   - **c.**
   - **d.**

Answers:

1. c (process of elimination)
2. a (longer item with qualifications is usually correct)
3. b ("always" and "never" items are usually wrong)
4. d (this response hasn't been used yet)

**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

$$\begin{array}{r} 4037 \\ -\ 159 \end{array}$$

- **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. One rubric for high school math problem solving appears in Figure 13.3.

*Level 3*
The response indicates application of a reasonable strategy that leads to a correct solution in the context of the problem. The representations are essentially correct. The explanation and/or justification is logically sound, clearly presented, fully developed, and supports the solution, and does not contain significant mathematical errors. The response demonstrates a complete understanding and analysis of the problem.

*Level 2*
The response indicates application of a reasonable strategy that may be incomplete or undeveloped. It may or may not lead to a correct solution. The representations are fundamentally correct. The explanation and/or justification supports the solution and is plausible, although it may not be well developed or complete. The response demonstrates a conceptual understanding and analysis of the problem.

*Level 1*
The response indicates little or no attempt to apply a reasonable strategy or applies an inappropriate strategy. It may or may not have the correct answer. The representations are incomplete or missing. The explanation and/or justification reveals serious flaws in reasoning. The explanation and/or justification may be incomplete or missing. The response demonstrates a minimal understanding and analysis of the problem.

*Level 0*
The response is completely incorrect or irrelevant. There may be no response, or the response may state, "I don't know."

**FIGURE 13.3 • Generic Rubric for Brief Constructed-Response Items in High School Mathematics in Maryland**

*Source:* W. D. Schafer, G. Swanson, N. Bené, & G. Newberry, "Effects of Teacher Knowledge of Rubrics on Student Achievement in Four Content Areas," *Applied Measurement in Education, 14,* 2001, pp. 151–170.

## THEORY INTO PRACTICE

### Detecting Bluffing in Students' Essays

Students who are not well prepared for essay tests are likely to try to bluff their way through. Credit should not be given unless the question is specifically answered. Here are some common types of bluffing:

1. ***Student repeats the question in statement form (slightly paraphrased) and tells how important the topic is.*** ("The role of assessment in teaching is extremely important. It is hard to imagine effective instruction without it.")

2. ***Student writes on a topic he or she knows about and fits it to the question.*** (A student who knows testing well but knows little about performance assessment, and is asked to compare testing and performance assessment, might describe testing in considerable detail and frequently state that performance assessment is much superior for evaluating the type of learning measured by the test.)

3. ***Student liberally sprinkles the answer with basic concepts whether they are understood or not.*** (Asked to write about any assessment technique, a student responds by frequently mentioning the importance of "validity" and "reliability.")

4. ***Student includes the teacher's basic beliefs wherever possible.*** ("The intended learning outcomes must be stated in performance terms before this type of test is constructed or selected.")

## 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.

## THEORY INTO PRACTICE

### Peer Evaluations

An evaluation technique often used in cooperative learning, especially in creative writing and (less often) mathematics problem solving, is to have students rate each other's work on a specific set of criteria before the teacher rates them on the same criteria (Brookhart, 2013a; Erkens, 2015; Reynolds, 2009; Smith, 2009). The peer evaluation does not contribute to a student's score or grade but gives the student feedback to use in revising the composition or product. Figure 13.4 shows a peer response guide that might be used for a comparison–contrast writing assignment. The partner, and then the teacher, enters a check mark for each category in which the student has done an adequate job. The partner and the teacher also mark the student's paper to make suggestions for improvement. Peer evaluation provides a formative evaluation for the writer, but it also gives the evaluator an invaluable opportunity to take the teacher's perspective and gain insight into what constitutes good writing.

**EVALUATING PROBLEM-SOLVING ITEMS** Problem solving involves several important components that fit most disciplines, including understanding the problem to be solved, attacking the problem systematically, and arriving at a reasonable answer. Following is a detailed checklist of elements common to most problem solving that can guide the weighting of elements in your evaluation of a student's problem-solving abilities.

**Problem-Solving Evaluation Elements**

- ☐ **1.** Problem organization
  - ☐ a. Representation made by table, graph, chart, etc.
  - ☐ b. Representation shown fits the problem.
  - ☐ c. Global understanding of the problem demonstrated.
- ☐ **2.** Procedures (mathematical: trial and error, working backward, experimental process, empirical induction)
  - ☐ a. A viable procedure was attempted.
  - ☐ b. The procedure was carried to a final solution.
  - ☐ c. Computation (if any) was correct.
- ☐ **3.** Solution (mathematical: a table, number, figure, graph, etc.)
  - ☐ a. Answer was reasonable.
  - ☐ b. Answer was checked.
  - ☐ c. Answer was correct.
- ☐ **4.** Logic specific to the detail or application of the given information was sound.

If you wish to give partial credit for an answer that contains correct elements, or if you want to inform students about the value of their responses, you must devise ways to do this consistently. The following points offer some guidance.

*(continued)*

**1.** Write model responses before giving partial credit for such work as essay writing, mathematical problem solving, laboratory assignments, and any work that you evaluate according to the quality of its various stages.

**2.** Explain to students in sufficient detail the meaning of the grades you give to communicate the value of the work.

The following examples illustrate outlines of exemplary student work from mathematics and social studies or literature.

**FROM MATHEMATICS** Students are given the following problem:

In a single-elimination tennis tournament, 40 players are to play for the singles championship. Determine how many matches must be played.

**Evaluation**

- ☐ a. Evidence that the student understood the problem, demonstrated by depiction of the problem with a graph, table, chart, equation, etc. (*3 points*)
- ☐ b. Use of a method for solving the problem that had potential for yielding a correct solution—for example, systematic trial and error, empirical induction, elimination, working backward. (*5 points*)
- ☐ c. Arrival at a correct solution. (*3 points*)

The three components in the evaluation were assigned points according to the weight the teacher judged each to be worth in the context of the course of study and the purpose of the test. You can give full credit for a correct answer even if all the work is not shown in the response, provided that you know that students can do the work in their heads. But it is important to guard against the **halo effect**, which occurs when you know which student wrote which response and you alter the grading of the paper in accordance with your opinion of the student. The same response should receive the same score no matter who wrote it. Use of a detailed rubric, or scoring guide, in evaluation is a way to make scoring more objective and thus to avoid any halo effects.

**FROM SOCIAL STUDIES OR LITERATURE** Students are asked to respond with a 100-word essay to the following item:

Compare and contrast the development of Inuit and Navajo tools on the basis of the climates in which these two peoples live.

**Evaluation**

- ☐ a. The response gives evidence of specific and accurate recall of the climates in which the Inuit and Navajos live (*1 point*) and of Inuit and Navajo tools. (*1 point*)
- ☐ b. The essay develops with continuity of thought and logic. (*3 points*)
- ☐ c. An accurate rationale is provided for the use of the various tools in the respective climates. (*3 points*)
- ☐ d. An analysis comparing and contrasting the similarities and differences between the two groups and their tool development is given. (*8 points*)
- ☐ e. The response concludes with a summary and closure. (*1 point*)

These two examples should suggest ways to evaluate items in other subject areas as well. Giving partial credit for much of the work students do certainly results in a more thorough evaluation of student progress than does scoring the work as merely right or wrong. The examples show how to organize objective assessments for evaluating

work that does not lend itself to the simple forms of multiple-choice, true–false, completion, and matching items. Points do not have to be used to evaluate components of the responses. In many situations, some kind of evaluative descriptors might be more meaningful. **Evaluative descriptors** are statements describing strong and weak features of a response to an item, a question, or a project. In the mathematics example, a teacher's evaluative descriptor for item a might read, "You have drawn an excellent chart showing that you understand the meaning of the problem, and that is very good, but it seems you were careless when you entered several important numbers in your chart."

Note that each of these examples is much like a rubric and can be generalized to broad ranges of topics. If teachers and students discuss these during instruction, students will have a device that helps them understand what they are working toward, and both teachers and students will have a common language that they can use during instruction and in their formative assessments.

| *Criterion* | *Partner* | *Teacher* |
|---|---|---|
| **Content**<br>**1.** Shows how concepts are similar | | |
| **2.** Shows how concepts are different | | |
| **3.** Well organized | | |
| **4.** Good opening sentence | | |
| **5.** Good concluding sentence | | |
| **Mechanics**<br>**1.** Spelling correct | | |
| **2.** Grammar correct | | |
| **3.** Punctuation correct | | |
| **4.** At least 2 pages | | |

**FIGURE 13.4 • Example of a Partner Response Form for a Comparison–Contrast Assignment**

MyEdLab **Self-Check 13.3**

## WHAT ARE AUTHENTIC, PORTFOLIO, AND PERFORMANCE ASSESSMENTS?

**InTASC 6**

**Assessment**

After much criticism of traditional testing (e.g., Beers, 2011; McTighe & Curtis, 2015; Shepard, 2000; Zhao, 2015), critics have developed and implemented alternative assessment systems that are designed to avoid the problems raised by typical multiple-choice tests. The key idea behind the testing alternatives is that students should be asked to document their learning or demonstrate that they can actually do something real with the information and skills they have learned in school (Brookhart, 2015; Greenstein, 2012; Lewin & Schoemaker, 2011; McTighe & Wiggins, 2013). For example, students might be asked to keep a portfolio, design a method of measuring wind speed, draw a scale model of a racing car, or write something for a real audience. Such tests are referred to as *authentic assessments* or *performance assessments* (McTighe & Wiggins, 2013). One goal of these "alternative assessments" is to demonstrate achievement in realistic contexts. In reading, for example, the authentic assessment movement has led to the development of tests in which students are asked to read and interpret longer sections of text and show a deep understanding.

MyEdLab

**Video Example 13.4**
Portfolios are especially effective when teachers meet with students to review the materials and reflect on successes and goals for the future.

In science, authentic assessments might involve having students set up and carry out an experiment. In writing, students might be asked to write real letters or newspaper articles. In math, students might solve complex physical problems that require insight and creativity. Authentic tests sometimes require students to integrate knowledge from different domains—for example, to use algebra in the context of reading about and performing a science experiment and to write up the results.

## Portfolio Assessment

One popular form of alternative assessment is **portfolio assessment**: the collection and evaluation of samples of student work over an extended period (Brookhart, 2015; Greenstein, 2012; McMillan, 2011). You may collect student compositions, projects, and other evidence of higher-order functioning and use this evidence to evaluate student progress over time. For example, many teachers have students maintain portfolios of their writings that show the development of a composition from first draft to final product; portfolios can also be used for journal entries, book reports, artwork, computer printouts, or papers showing development in problem solving (Brookhart, 2015). Portfolios are increasingly being maintained on computers to supplement paper files (Diehm, 2004; Niguidula, 2005). Refer to Figure 13.5 for sample criteria in evaluating student writing portfolios.

**FIGURE 13.5 • Sample of Criteria for Evaluating Students' Writing Ability through Portfolio Assessment**

| Portfolio Evaluation Form | | | | |
|---|---|---|---|---|
| **Name:** | **Improvement Needed 1** | **2** | **3** | **Excellent 4** |
| 1. All assigned work is included | | | | |
| 2. Log sheet is completed | | | | |
| 3. Final reflection is completed | | | | |
| 4. Work demonstrates improvement on previous areas of weakness | | | | |
| 5. Writing incorporates teacher's feedback from earlier work | | | | |
| 6. Portfolio demonstrates improved writing overall | | | | |
| **Additional Criteria** | | | | |
| 7. | | | | |
| 8. | | | | |
| 9. | | | | |
| 10. | | | | |
| **Teacher Comments:** | | | | |

Portfolio assessment has important uses when you want to evaluate students for reports to parents or other within-school purposes. When combined with on-demand assessments and used with consistent and public rubrics, portfolios showing improvement over time can provide powerful evidence of change to parents and to students themselves (Burke, 2009).

**Certification Pointer**
A teacher certification question may ask you to respond to a case study by suggesting a way to implement portfolio assessment that would be appropriate for the case.

## ON THE WEB

For reports, newsletters, and other publications about assessment, particularly performance and portfolio assessment, visit cresst.org, the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), located at UCLA. To view articles and multimedia related to assessment and other topics in education, go to the website of the George Lucas Educational Foundation at edutopia.org. Also visit the Education Commission of the States website, which contains a list of assessment-related sites, at ecs.org.

## THEORY INTO PRACTICE

### Using Portfolios in the Classroom

#### PLANNING AND ORGANIZATION

- Develop an overall flexible plan for student portfolios. What purposes will the portfolios serve? What items will be required? When and how will they be obtained? What criteria will be applied for reflection and evaluation?
- Plan sufficient time for students to prepare and discuss portfolio items. Assessing portfolios takes more time and thought than correcting paper-and-pencil tests.
- Begin with one aspect of student learning and achievement, and gradually include others as you and the students learn about portfolio procedures. The writing process, for instance, is particularly well suited to documentation through portfolios.
- Choose items to be included in portfolios that will show developing proficiency on important goals and objectives. Items that address multiple objectives help to make portfolio assessments more efficient.
- Collect at least two types of items: required indicators (Arter & McTighe, 2001; Murphy & Underwood, 2000) or core items and optional work samples. Required or core indicators are items collected for every child that will show how each is progressing. Optional work samples show individual students' unique approaches, interests, and strengths.
- Place a list of goals and objectives in the front of each portfolio along with a list of required indicators, and also include a place for recording optional items, so that you and the students can keep track of contents.

#### IMPLEMENTATION

- In order to save time, to ensure that portfolio items are representative of students' work, and to increase authenticity, embed the development of portfolio items into ongoing classroom activities.

*(continued)*

- Give students responsibility for preparing, selecting, evaluating, and filing portfolio items and keeping portfolios up to date. Young children will need guidance with this.
- For selected portfolio items, model reflection and self-assessment for students to help them become aware of the processes they used, what they learned, what they have yet to learn, and what they might need to do differently next time.
- Be selective. A portfolio is not a haphazard collection of work samples, audio or video recordings, pictures, websites, and other products. It is a thoughtful selection of items that exemplify children's learning. Random inclusion of items quickly becomes overwhelming.
- Use information in portfolios to place learners on a sequence of developing skills.
- Analyze portfolio items for insight into students' knowledge and skills. As you do this, you will understand more of the students' strengths and needs, thinking processes, preconceptions, misconceptions, error patterns, and developmental benchmarks.
- Use portfolio information to document and celebrate students' learning, to share students' accomplishments with parents and other school personnel, and to improve and target classroom instruction. If portfolios are not linked to improving instruction, they are not working. (For guides to portfolio evaluation, see Brookhart, 2013; McTighe & Wiggins, 2013; Stiggins & Chappuis, 2012.)

## Performance Assessment

Tests that involve actual demonstrations of knowledge or skills in real life are called **performance assessments** (Brookhart, 2015; McMillan, 2011; Popham, 2014a; Shavelson, 2013). For example, ninth-graders might be asked to conduct an oral history project, reading about a significant recent event and then interviewing the individuals involved. The quality of the oral histories, done over a period of weeks, indicates the degree of the students' mastery of the social studies concepts involved. Wiggins (1993) describes assessments used in the last 2 weeks of school in which students must apply everything they have learned all year to analyze a sludge that mixes a variety of solids and liquids. Some schools are requiring elaborate "exhibitions," such as projects developed over many months, as demonstrations of competence. More time-limited performance assessments might ask students to set up experiments, respond to extended text, write in various genres, or solve realistic math problems. Technology enables students to set up complex experiments that require deep understanding of and insight into science or math, for example (Clarke-Midura, 2014).

## Effectiveness of Performance Assessments

One of the most important criticisms of traditional standardized tests is that they can focus teachers on teaching only a narrow range of skills that happen to be on the test (see Popham, 2004). How might performance assessments be better? At least in theory, it should be possible to create tests that would require such a broad understanding of subject matter that the test would be worth teaching to.

For example, consider the performance test in math shown in Figure 13.6. Imagine that your students will have to demonstrate their time-telling skills. The only way to teach to such a test will be to expose students to the various ways to tell time.

Beyond all the practical problems and expense of administering and scoring performance tests, it is not yet clear whether performance tests will solve all the problems of standardized testing. For example, Shavelson, Baxter, and Pine (1992) studied performance assessments in science.

**Capability to Be Assessed**

- *Goal to be assessed:* Establishes the time using a clock
- *Type of capability involved:* Rule

**Performance to Be Observed**

- *Domain of tasks associated with goal being assessed*
  - Tells time using a digital or analog clock
  - With an analog clock, tells time with numbers or other markings on the face
  - With digital or analog clock, tells time with or without seconds indicated
  - With digital or analog clock, tells time with different shapes, sizes, and colors used for the clock face
- *Description of task to be performed:* Student views the face of an analog clock and states the displayed time
  - *Focus on process or product?* Product
  - *Prerequisite skills to be verified: The student can read numbers.*
- *Required materials*
  - Clock face with movable hour and minute hands. The face should contain numbers to designate hours. The clock should not have a second hand.
- *Guidelines for administration*
  - Use eight different time settings, with the minute hand twice within each of the quarter hours.
  - Vary the hour hand through its full range.
  - The minute and hour hands should be distinctly visible in all settings.
- *Instructions to students*
  - Say to the student, "What time does this clock show?"

**Scoring Plan**

Time stated by student is correct within 1 minute.

**FIGURE 13.6 • Example of a Performance Assessment Activity: Telling Time**

*Source:* Oosterhof, Albert. *Developing and using classroom assessments* (4th ed.) (c) 2009, p. 186. Reprinted and electronically reproduced by permission of Pearson Education Inc., Upper Saddle River, NJ.

They found that student performance on such assessments could be reliably rated, but different performance assessments produced very different patterns of scores, and student scores were still related more closely to student aptitude than to what students were actually taught. Similar findings were reported in studies by the Educational Testing Service (1995), Linn (1994), and Supovitz and Brennan (1997).

## Scoring Rubrics for Performance Assessments

Performance assessments are typically scored according to rubrics that specify in advance the type of performance that is expected for each activity (Brookhart, 2013; Burke, 2009; Popham, 2014; Vagle, 2014). Figure 13.7 shows one rubric (from Taylor, 1994) that was developed for an essay on character development in stories that students had read.

Performance assessment tasks are similar to essay items in that students might approach them in multiple ways. It is, therefore, also important for performance assessments that students understand the criteria for scoring. One way to ensure this is to write a few generic rubrics that are flexible enough to apply to the full range of student performance. Figure 13.3 gave an example of a generic rubric that has been applied to outcomes in high school mathematics. It has been suggested that using rubrics such as this in classroom instruction can enhance student achievement (Schafer, Swanson, Bené, & Newberry, 2001).

Planning for performance assessments takes time, and avoiding the pitfalls of subjectivity in rating performances takes practice. However, a few well-thought-out, well-written items for a performance assessment could serve, for example, as a summative evaluation for all or most of your educational objectives for an entire unit (see Figure 13.8).

**Certification Pointer**

You may be asked on your teacher certification test to give an example of a performance goal and then to write a behavioral objective, an activity, and an assessment of student learning that would accomplish the goal.

**FIGURE 13.7 • Sample Scoring Rubric: Targeted Performance, Performance Criteria, and a Description of Performances at Different Score Points**

*Source:* Catherine Taylor, "Assessment for Measurement or Standards," *American Educational Research Journal*, *31*(2), pp. 231–262, 1994.

**Performance**

Essay on Character Development in Literature

**Performance Criteria**

- Character is identified.
- At least three aspects of the character's development during the course of the story are described.
- Appropriate support for each character aspect is given, using excerpts from the story.
- Character's contribution to the story's plot is described.
- At least three excerpts from the story are given as support for the writer's ideas about the character's contribution to the story.
- Text references used for support are appropriate.

**Scoring Rubric**

| | |
|---|---|
| 4 points | Essay is complete, thorough, and insightful in describing the character's development and contribution to the story. Adequate support is given to encourage us to consider the writer's point of view. All excerpts from the text enhance our understanding of the writer's view of the character. |
| 3 points | Essay is complete in describing the character's development and contribution to the story. Adequate support is given to encourage us to consider the writer's point of view. Most excerpts from the text enhance our understanding of the writer's view of the character. |
| 2 points | Essay is complete in its description of either the character's development or the character's contribution to the story. Some support is given to help us consider the writer's point of view. Most excerpts from the text enhance our understanding of the writer's view of the character for the element described. |
| 1 point | Essay is mostly complete in its description of either the character's development or the character's contribution to the story. Support is given for the writer's point of view, but it is not always convincing. Few excerpts from the text enhance our understanding of the writer's view of the character for the element described. |
| 0 points | The written essay was not completed, is significantly lacking in performance of all criteria, or is off task. |

## Assessment through Digital Games and Simulations

Computers have long been used to assess students' learning, but until recently they have only provided easy, rapid scoring and recordkeeping. Today, however, computers are beginning to be used to assess students' performance as they participate in games, simulations, and other activities (Schaaf, 2015). For example, students working together on a simulated lab exercise in science might be assessed digitally on the basis of their personal contributions to the lab (Erkens, 2015). Students playing games against the computer might have fun and learn, but at the same time their responses can be recorded and evaluated against standards. Some day, perhaps students will no longer take tests that are separate from the learning activities they do every day (McTighe & Curtis, 2015).

*Psychology Fair Rubric*

✓ **Our Psychology Project . . .**

☐ (25 pts) Provides background information—it cites other studies, explains our interest in the topic, and presents a rationale for the topic. [The better the background information, the more detailed it is, and the more it "fits," the more points you earn.]

Brief example: "We are interested in how dress affects behavior. We always felt better when we were dressed up and also thought that less violence occurs between people who 'dress up.' Cohen and Cohen (1987) found that students who wore uniforms performed 10 percent better on exams and were cited for fewer office referrals. Thus we wanted to look into this topic further."

☐ (25 pts) Gives a description of the study (an abstract—a *general* statement in 100 words or fewer about your project).

Brief example: "This study investigates the relationship between academic performance and the use of uniforms in schools. Three schools in Derry County, Pennsylvania, were surveyed on issues of academic performance and office referrals. The use of uniforms showed an increase in achievement and decreased behavioral problems."

☐ (40 pts) Has a measurable hypothesis with specific variables defined and identified.

☐ (25 pts) Includes at least one graph, chart, or other visual aid that *summarizes the data*. Someone should be able to look at your graph/chart and clearly see what the variables and results were.

☐ (10 pts) Includes a clean copy of any survey or other scale that was used to gather data.

☐ (40 pts) Includes a written procedure that tells the observer *exactly* what we did.

Brief example: "We took 3 days to survey 100 students and 30 teachers."

☐ (30 pts) Includes a section that explains the data and tells whether the hypothesis was accurate.

Brief example: "Our data reflect that our hypothesis was correct: The 30 percent increase in scores reflects the improved achievement of students while . . ." [Again, develop your explanation. If you say only, "We were right" or "Our hypothesis is correct/wrong," you will not receive more than half the points.]

☐ (30 pts) Includes a section explaining the significance of the study—why it is important.

Brief example: "This is an important study because it reflects a bias that many people may not be aware of, as well as a way in which students can improve scores and reduce their own behavioral problems. It further . . ."

☐ (50 pts) Is interactive. That is, observers can take the test, view the screen, do the quiz, and so on. [This can be done in a variety of ways. For instance, if the test is long, have observers do part of it or show a video of your procedure.]

Total Points Possible: 275

**FIGURE 13.8 • Semester-Long Assessment**

*Source:* Mr. Charles Greiner, Lusher Charter School. New Orleans/Tulane University.

## HOW ARE GRADES DETERMINED?

One of the most perplexing and often controversial tasks you face is grading student work (Brookhart & Nitko, 2015; Reeves, 2015; Quinn, 2012; Schimmer, 2016; Scriffiny, 2008). Is grading necessary? It is clear that some form of summative student evaluation is necessary, and grading of one kind or another is the form predominantly used in most schools.

### Establishing Grading Criteria

Many sets of grading criteria exist, but regardless of the level of school that teachers teach in, they generally agree on the need to explain the meaning of grades they give (Brookhart & Nitko, 2015; Stiggins & Chappuis, 2012; Vatterott, 2015). Grades should communicate at least the relative value of a student's work in a class. They should also help students to understand better what is expected

of them and how they might improve. They can also serve as a basis for a productive conversation with students and parents (Webber & Wilson, 2012).

Teachers and schools that use letter grades attach the following general meanings to the letters:

A = superior; exceptional; outstanding attainment

B = very good, but not superior; above average

C = competent, but not remarkable work or performance; average

D = minimum passing, but serious weaknesses are indicated; below average

E = failure to pass; serious weaknesses demonstrated

## Assigning Letter Grades

All school districts have a policy or common practice for assigning report card grades. Most use A-B-C-D-F letter grades, but many (particularly at the elementary school level) use various versions of outstanding–satisfactory–unsatisfactory (Brookhart & Nitko, 2015; Reeves, 2015; Schimmer, 2016). Some simply report percentage grades. The criteria on which grades are based vary enormously from district to district. Secondary schools usually give one grade for each subject taken, but most elementary schools and some secondary schools include ratings on effort or behavior as well as on performance.

The criteria for giving letter grades might be specified by a school administration, but grading criteria are most often set by individual teachers using very broad guidelines. In practice, few teachers could get away with giving half their students A's or with failing too many students; but between these two extremes, teachers may have considerable leeway (see Guskey, 2014; Tomlinson & Moon, 2013).

**ABSOLUTE GRADING STANDARDS** Grades may be given according to absolute or relative standards. Absolute grading standards might consist of preestablished percentage scores required for a given grade, as in the following example:

| Grade | Percentage Correct |
|---|---|
| A | 90–100 percent |
| B | 80–89 percent |
| C | 70–79 percent |
| D | 60–69 percent |
| F | Less than 60 percent |

In another form of absolute standards, called *criterion-referenced grading,* you decide in advance which performances constitute outstanding (A), above-average (B), average (C), below-average (D), and inadequate (F) mastery of the instructional objective.

Absolute percentage standards have one important disadvantage: Student scores might depend on the difficulty of the tests they are given. For example, a student can pass a true–false test (if a passing grade is 60 percent) by knowing only 20 percent of the answers and guessing on the rest (getting half of the remaining 80 percent of the items correct by chance). On a difficult test where guessing is impossible, however, 60 percent could be a very high score. For this reason, use of absolute percentage criteria should be tempered with criterion-referenced standards. That is, you might use a 60–70–80–90 percent standard in most circumstances but establish (and announce to students) higher standards for tests that students are likely to find easy and lower standards for more difficult tests.

Another disadvantage is that the ranges of the grades are typically different, especially for F's. A student who receives an F may be very close to a D or may be hopelessly far from "passing." This is true for the other grades, too, but the large range of F (0 to 60 percent) emphasizes the uncertainty. Moreover, the consequences of an F are often quite severe.

**RELATIVE GRADING STANDARDS** A **relative grading standard** is a system whereby a teacher gives grades according to the students' rank in their class or grade. The classic form of relative grading is specifying what percentage of students will be given A's, B's, and so on. A form of this practice is called *grading on the curve* because students are given grades on the basis of their position on a predetermined distribution of scores.

Relative grading standards have the advantage of placing students' scores in relation to one another without regard to the difficulty of a particular test. However, relative grading standards also have serious drawbacks (see Guskey, 2014; O'Connor, 2009). One is that because they hold the number of A's and B's constant, students in a class of high achievers must get much higher scores to earn A's or B's than students in low-achieving classes—a situation that is likely to be seen as unfair. Teachers often deal with this problem by giving relatively more A's and B's in high-achieving classes than in others. Another disadvantage of relative grading is that it creates competition among students; when one student earns an A, this diminishes the chances that others may do so. Competition can inhibit students from helping one another and hurt social relations among classmates (Guskey, 2014).

Strict grading on the curve and guidelines for numbers of A's and B's have been disappearing in recent years. For one thing, there has been a general grade inflation; more A's and B's are given now than in the past, and C is no longer the expected average grade but often indicates below-average performance (Goodwin, 2011; Pattison, Grodsky, & Muller, 2013). Anderson (1994) summarized a national survey of eighth-graders who were asked to report their English grades since sixth grade. The results were as follows:

Mostly A's: 31 percent

Mostly B's: 38 percent

Mostly C's: 23 percent

Mostly D's: 6 percent

Mostly less than D's: 2 percent

Results were similar in mathematics, and grades were only slightly lower in high-poverty schools than in middle-class schools. It is likely that these self-reported grades are somewhat higher than those students actually received, but it is nevertheless likely that the average grade today is B, not C.

The most common approach to grading involves looking at student scores on a test, taking into account test difficulty and the overall performance of the class, and assigning grades in such a way that about the "right number" of students earn A's and B's and the "right number" fail. Teachers vary considerably in their estimates of what these right numbers should be, but schools often have unspoken norms about how many students should be given A's and how many should fail.

## Performance Grading

One of the most important limitations of traditional grades is that although they may give some indication of how students are doing in comparison to others, they provide no information about what students know and can do. A student who gets a B in English might be disappointed or breathe a sigh of relief, depending on what she expected. However, this grade does not tell her or her parents or teachers what she can do, what she needs to do to progress, or where her strengths or weaknesses lie (Marzano & Heflebower, 2011; Quinn, 2012). Furthermore, giving a single grade in each subject can reinforce the idea that students are more able or less able, or perhaps more motivated or less motivated, rather than the idea that all students are growing.

Some schools have responded to these limitations with an alternative approach to grading called *performance grading* (Guskey, 2014), in which teachers determine what children know and can do and then report this in a way that is easy for parents and students to understand (Guskey, 2014). Figure 13.9 (from Wiggins, 1994) shows one page of a language arts assessment keyed to fifth-grade exit standards, or expectations of what a fifth-grader should know. A parent who receives a form like this can see how the student is progressing toward the kind of performance the school district has defined as essential. Note that although the form does provide information on how the student is doing in comparison to other students, the emphasis is on growth over time.

**Cherry Creek School District**
**Polton Community Elementary**
**School Fairplay Progress Report**
**(Language Arts Section)**

**Student Name** _____ **Grade 3** _____ **4** _____
**Teacher** _____ **School Year** _____

Performance-based graduation requirements focus on student mastery of the proficiencies. The curriculum and written progress report are geared toward preparing students for this task. A date (for example, 3/12) indicates where a student is performing on a continuum of progress based on the fifth-grade exit standards.

| | Basic | Proficient | Advanced |
|---|---|---|---|
| **Language Arts Proficiency 1** | | | |
| Listens, interpreting verbal and nonverbal cues to construct meaning. | Actively listens, demonstrates understanding, and clarifies with questions and paraphrasing. | Actively listens for purpose, demonstrates understanding, and clarifies with questions and paraphrasing. | Actively listens for purpose, demonstrates understanding, clarifies with questions and paraphrasing, classifies, analyzes, and applies information. |
| **Language Arts Proficiency 2** ◄——————————————————► | | | |
| Conveys meaning clearly and coherently through speech in both formal and informal situations. | Appropriately speaks to inform, explain, demonstrate, or persuade. Organizes a speech and uses vocabulary to convey a message. | Appropriately speaks to inform, explain, demonstrate, or persuade. Organizes a formal speech and uses vocabulary to convey a message. | Appropriately speaks to inform, explain, demonstrate, or persuade. Organizes a formal speech with details and transitions adapting subject and vocabulary. Uses eye contact, gestures, and suitable expression for an audience and topic. |
| **Language Arts Proficiency 3** ◄——————————————————► | | | |
| Reads to construct meaning by interacting with the text, by recognizing the different requirements of a variety of printed materials, and by using appropriate strategies to increase comprehension. | Reads varied material, comprehends at a literal level. Recalls and builds knowledge through related information. Begins to use strategies to develop fluency, adjusting rate when reading different material. | Reads varied material, comprehends literally and interpretively. Synthesizes and explores information, drawing inferences. Critiques author's intent, analyzes material for meaning and value. Applies strategies to increase fluency, adjusting rate when reading different material. | Reads varied material, comprehends and draws inferences, recalls and builds knowledge through related information. Applies strategies to increase fluency, adjusting rate when reading different material. |
| **Language Arts Proficiency 4** ◄——————————————————► | | | |
| Produces writing that conveys purpose and meaning, uses effective writing strategies, and incorporates the conventions of written language to communicate clearly. | Appropriately writes on assigned or self-selected topics. Clear main ideas, few details. Weak elements in the beginning, middle, end. Sentence structure lacks variety and contains errors. | Appropriately writes on assigned or self-selected topics. Clear main ideas, interesting details, clear organization, sequencing, varied sentence structure, edits to reduce errors. Appropriate voice and word choice. | Appropriately writes on assigned or self-selected topics. Connects opinions, details, and examples. Effective organization and sequencing, meaningful sentence structure, edits to eliminate most errors. Appropriate voice and word choice. |

**As compared to the class in the area of Language Arts, your child**
Note: The teacher places a check in one box per marking period to indicate child's status in language arts.

| 1 | 2 | 3 | Marking Periods |
|---|---|---|---|
| | | | Displays strong performance |
| | | | Demonstrates appropriate development |
| | | | Needs practice and support |

**FIGURE 13.9 • Sample Performance Grading Criteria**

*Source:* From "Toward Better Report Cards," *Educational Leadership*, by Grant Wiggins. Copyright © 1994 reprinted with permission of Grant Wiggins.

**SCORING RUBRICS FOR PERFORMANCE GRADING** A key requirement for the use of performance grading is collection of work samples from students that indicate their level of performance on a developmental sequence. Collecting and evaluating work that students are already doing in class (such as compositions, lab reports, or projects) is called *portfolio assessment* (Brookhart, 2013, 2014, 2015; McTighe & Wiggins, 2013), discussed earlier in this chapter. An alternative is to give students tests in which they can show their abilities to apply and integrate knowledge, skills, and judgment. Most performance grading schemes use some combination of portfolios and on-demand performance tests. In either case, student performance is usually evaluated against rubrics, which describe, for example, partially proficient, proficient, and advanced performance, or which indicate a student's position on a developmental sequence.

## Other Alternative Grading Systems

Several other approaches to grading are used in conjunction with innovative instructional approaches. In the system called *contract grading,* students negotiate a particular amount of work or level of performance that they will achieve to receive a certain grade. For example, a student might agree to complete five book reports of a given length in a marking period to receive an A. **Mastery grading** involves establishing a standard of mastery, such as 80 or 90 percent correct on a test. All students who achieve that standard receive an A; students who do not achieve it the first time receive corrective instruction and then retake the test to try to achieve the mastery criterion (Fisher, Frey, & Pumpian, 2011; Guskey, 2014). Finally, many teachers give grades based on improvement or effort, usually in combination with traditional grades. In this way a student who is performing at a low level relative to others can nevertheless receive feedback indicating that he or she is on a path leading to higher performance (see Tomlinson & Moon, 2013).

**LETTING STUDENTS RETAKE TESTS** Many teachers allow students to retake tests, especially if they failed the first time (Dueck, 2011; Wormeli, 2011). This can be a good idea if it gives students an opportunity to do additional studying and master the material the class is studying. For example, a student might be given 2 days to study the content that was tested and then take an alternative form of the test. (Giving the same test to the student is not recommended because that would allow the student to study only the questions that were asked.) The student might then be given a grade that is one letter grade lower than he or she scored on the second test, because the student had an advantage in having an extra opportunity to study. There is some danger that if students know they can retake tests, they might not study until after attempting the first test, but in general, giving students a second chance is a good way to allow those who are willing to put in extra effort to improve a poor grade. Some schools give grades of A, B, C, or incomplete, permitting additional time and support until all students are able to earn at least a C (Kenkel, Hoelscher, & West, 2006).

## Assigning Report Card Grades

Most schools give report cards four or six times per year—that is, every 6 or 9 weeks. Report card grades are most often derived from some combination of the following factors (Brookhart & Nitko, 2015; Reeves, 2015):

- Scores on quizzes and tests
- Scores on papers and projects
- Scores on homework
- Scores on seatwork
- Class participation (academic behaviors in class, answers to class questions, and so on)
- Deportment (classroom behavior, tardiness, attitude)
- Effort

These are listed in order from the most formal and reliable measures of achievement to those considered least valid as a learning indicator. The first two factors listed are summative assessments, and virtually everyone would consider them appropriate for grading. The next two are typically formative and thus indicate how learning is progressing when it is still incomplete. They are less appropriate because they do not convey information about status at the end of instructional units.

MyEdLab

**Video Example 13.5**

Textbook author Bob Slavin presents a story about his first experience grading students' papers. Why do you think a teacher might feel he or she "failed" when a student fails a class? What can you do to make sure you can always justify the grades you give students?

The final three might contribute to achievement, but they are not achievement. Basing grades on them could communicate misinformation to others about students (Guskey, 2014). Teachers often give different weights to various factors, stating (for example) that grades will be based 30 percent on quizzes, 30 percent on a final test, 20 percent on homework, and 20 percent on class participation. This helps communicate to students what is most important to the teacher.

One important issue arises when scores are to be combined for grading—how to treat missing work, such as homework assignments (O'Connor, 2009; Reeves, 2006, 2015). Some teachers assign a "zero" to missing work. Similarly, they may assign a zero to tests for other reasons. But a zero can be devastating (it is so far from even a passing grade that it is virtually impossible for the student to recover). This practice can only be viewed as punitive. A better strategy might be to use a system whereby grades are converted to a reasonable set of numerical grades (e.g., A = 4, B = 3, etc.), with 0 given for any missing work. To illustrate the difference in these two strategies, consider a student who misses one assignment out of five. If she were given a zero for the missing work and her scores for the assignments were 92, 86, 0, 73, and 91, her average score would be 68.4, or a D in a 60–70–80–90 grading scheme. Converting the scores using the letter grades, on the other hand, would give her a mean of 2.6, which would be a solid C. Another solution used in some schools is to award a minimum grade of 50 on a 100-point scale, to avoid having a single zero make success impossible. A study of this in high schools found that a minimum grade did not cause grade inflation and had benefits in terms of course passing (Carey & Carifio, 2012).

Sometimes a student's performance on a test or a quiz seems unusually poor for him or her. Such atypical assessments might occur for nonacademic reasons, such as a disruption at home or in school. A private conversation with the student about the test or quiz might uncover a problem that should be looked into, and the student might be given an opportunity to retake the test. Some teachers drop the lowest score a student receives on quizzes to avoid penalizing the student for one unusual slippage.

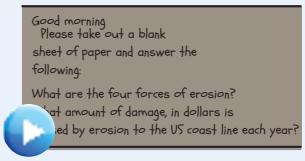
## THE INTENTIONAL TEACHER

### Using What You Know about Assessing Student Learning to Improve Teaching and Learning

Intentional teachers assess student learning in ways that align with both their goals and their instruction. They use assessment results to adjust their instruction and to provide important feedback to students, families, and communities. Intentional teachers know that no one measure is ideal for every circumstance, and they implement a range of assessments that fits their purposes and circumstances.

- They plan courses, units, and lessons around essential objectives.
- They carefully align their assessments of students' learning with these objectives.
- They use taxonomies of instructional objectives to be sure that they teach all types of objectives, not just knowledge and comprehension.
- They use formative assessments continually to find out what students have learned so far, and then use that information both to inform students and to adjust the level and pace of their teaching.
- They create tests and quizzes that touch on all types of learning and focus on key unit objectives.
- They create assessments that reliably determine whether students have or have not mastered essential concepts and skills.
- They use a variety of response formats in tests, including constructed-response as well as multiple-choice and fill-in-the-blanks questions.
- They assess higher-order skills, such as problem solving and creativity.

- For appropriate content, they collect and evaluate portfolios of student work so they can determine how students are progressing on authentic tasks. Examples include compositions, problem solving, art projects, and music performances.
- They give grades fairly and reliably based on students' achievement of standards, and explain to students and their parents what the grades were based on and what needs to be done to improve them.
- They proactively engage students and parents in discussions about grading, with an emphasis on what has been accomplished and what remains to be accomplished.
- They give students opportunities to improve their grades by retaking similar tests after doing additional studying, or use a mastery grading system in which students have multiple opportunities to meet standards.




MyEdLab
**Application Exercise 13.1**
In the Pearson etext, watch a classroom video. Then use the guidelines in "The Intentional Teacher" to answer a set of questions that will help you reflect on and understand the teaching and learning presented in the video.

One important principle in report card grading is that grades should never be a surprise. Students should always know how their grades will be computed, whether classwork and homework are included, and whether class participation and effort are taken into account. Being clear about standards for grading helps you avert many complaints about unexpectedly low grades and, more important, lets students know exactly what they must do to improve their grades (Guskey, 2014; O'Connor & Wormeli, 2011; Reeves, 2015).

Many schools give an "interim" grade in the middle of a marking period, which gives students an early idea of how they are doing and a warning if they seem headed for trouble. A variation on this practice is to provide an interim grade only if students are heading for a D or F. Adding comments to the grade to explain what the student needs to do to earn a higher grade can be very helpful in maintaining motivation and improving performance (Dueck, 2014).

Another important principle is that grades should be private. There is no need for students to know one another's grades; making grades public only invites invidious comparisons among students. Finally, it is important to restate that grades are only one method of student evaluation. Written evaluations can provide useful information to parents and students (Marzano, Yanoski, Hoegh, & Simms, 2013). Computerized gradebooks are now widely available and widely used. Guskey (2014), however, warns that you should be careful when using this time-saving software and avoid letting the program make decisions that you should make yourself (Mertler, 2014).

MyEdLab **Self-Check 13.4**

# 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.

# KEY TERMS

Review the following key terms from the chapter.

affective objectives 347
assessment 345
backward planning 343
behavior content matrix 347
completion items 361
criterion-referenced interpretations 350
distractors 357
evaluation 348
evaluative descriptors 367
foils 357
formative evaluations 350
halo effect 366
instructional objective 340
learning objectives 345

long essay item 361
mastery grading 377
matching items 360
multiple-choice items 356
norm-referenced interpretations 350
performance assessments 370
portfolio assessment 368
problem-solving assessment 364
relative grading standard 375
selected-response items 356
short essay item 361
stem 356
summative evaluations 350
table of specifications 356
task analysis 342
taxonomy of educational objectives 346
teaching objectives 345
true–false items 360

## SELF-ASSESSMENT: PRACTICING FOR LICENSURE

Directions: The chapter-opening vignette addresses indicators that are often assessed in state licensure exams. Reread the chapter-opening vignette and then respond to the following questions.

1. Mr. Sullivan is having a difficult time connecting what he is teaching and what he is testing. Which of the following evaluation tools is most likely to help Mr. Sullivan make the connection?
   a. Multiple-choice test
   b. Instructional objectives
   c. Traditional teaching strategies
   d. Open-book testing
2. Mr. Sullivan might use a chart showing how a concept or skill will be taught at different cognitive levels in relation to an instructional objective. What is this chart called?
   a. Task analysis
   b. Backward planning
   c. Behavior content matrix
   d. Table of specifications
3. Mr. Sullivan might improve the connection between what he teaches and what he tests by following which of the following pieces of advice?
   a. Include all instructional content in the test.
   b. Make a test that includes all item types: true–false, multiple-choice, matching, short answer, essay, and problem-solving items.
   c. Be free from the confines of instructional objectives.
   d. Design a test that fits the particular uses that will be made of the results.
4. Which of the following types of evaluation is Mr. Sullivan using?
   a. Summative
   b. Aptitude
   c. Affective
   d. Task analysis
5. Why would Mr. Sullivan choose to construct a table of specifications?
   a. To indicate the type of learning to be assessed for different instructional objectives
   b. To measure a student's performance against a specified standard
   c. To make comparisons among students
   d. To identify conditions of mastery
6. In a brief essay, explain why evaluation is important.
7. Write instructional objectives, create a table of specifications using Bloom's taxonomy, develop a lesson plan, and write a short test for a topic of study.

MyEdLab **Licensure Exam 13.1** Answer questions and receive instant feedback in your Pearson eText in MyEdLab.