

ASSESSMENT OF CLASSROOM LEARNING

LEARNING OBJECTIVES

After studying this chapter, you will be able to . . .

- L01** Explain the various purposes of assessment and the kinds of evaluative judgments that derive from each purpose.
- L02** Provide examples of the different methods by which student learning can be assessed.
- L03** Explain the distinction between norm-referenced and criterion-referenced grading.
- L04** Provide examples of how technology can support assessment practices and evaluation practices, and how assessment information and evaluative judgments can be communicated effectively.

visit 4ltrpress.cengage.com

Criterion ref vs. Norm Ref tests

Earlier parts of this book discussed three major aspects of the teacher's role: understanding student differences and how to address them properly, understanding the learning process and how to use that knowledge to formulate effective approaches to instruction, and understanding motivation to learn and how to engage motivation to establish a positive learning environment. Now we turn to assessing performance, which is an equally significant aspect of the teacher's role. Virtually everyone connected with public schools, from students to teachers and administrators to state education officials to members of the U.S. Congress, is keenly interested in knowing how much and how well students have learned.

In this and the next chapter, we describe a twofold process for assessing student learning: using teacher-made measures to assess mastery of the teacher's specific objectives and using professionally prepared standardized tests to measure the extent of a student's general knowledge base and aptitudes. Although the items that make up teacher-made and standardized assessments can be very similar, if not identical, these two types of assessments differ significantly in their construction, the conditions under which they are administered, and the purposes for which they are used. In short, standardized tests are designed to highlight where students, classrooms, schools, and districts stand with respect to one another in terms of general levels of performance in various skills and subject areas. Period. Think of this as assessment *of* learning. Teacher-made assessments, by contrast, are designed to highlight students' strengths and weaknesses, to give students timely feedback about the effectiveness of their study habits, and to provide teachers with timely information that can help them make more effective instructional decisions. These assessments may or may not look like traditional tests. Think of them as assessment *for* learning (Stiggins, 2007).

LO1 The Role of Assessment in Teaching

The role of assessment is to enhance learning. Therefore, assessment is both integral and critical to effective teaching. To help a student learn, a teacher must know what the student knows and is able to do. To determine if a lesson worked, a teacher must know if learning

objectives were met. We have examined a number of theories throughout this book to see how they apply to teaching. In order to apply those ideas, however, you will need to assess student learning. Consider, for a moment, the discussion in Chapter 2 of the Vygotskian zone of proximal development. Suppose, as a teacher, you decide that scaffolding—which research has shown to be an effective teaching strategy—should be employed to help your students learn. In order to scaffold a student's learning, you will first need to know the student's zone of proximal development; you will need to assess the student's knowledge and understanding in order to determine what the student knows and can do alone in order to determine the target of the necessary scaffolding. Assessment yields information about learning. Said another way, assessment informs learning. Because teaching is an effort to enhance learning, assessment also informs teaching.

Assessing student learning is a task that many teachers dislike and few do extremely well. One reason is that many lack in-depth knowledge of assessment principles (Guskey, 2003; Stiggins, 2002; Trevisan, 2002). Another reason is that the role of assessor is seen as being inconsistent with the role of teacher (or helper). A third reason is that many teachers think of assessment as merely grading rather than as instruction (a misconception we will address shortly). As a consequence, high quality assessment practices are too often not part of the culture of classrooms (Moss & Brookhart, 2009). This is unfortunate because, in fact, well-designed classroom assessment schemes contribute to student learning (Popham, 2006, 2011; Stiggins, 2002; Stiggins, Arter, Chappuis, & Chappuis, 2007).

A basic goal of this chapter is to help you understand how to use knowledge about assessment to enhance, rather than work against, your effectiveness as teacher. Toward that end, we will begin by defining what we mean by the term *assessment* and by two key elements of this process: *measurement* and *evaluation*.

What Is Assessment?

Broadly conceived, classroom assessment involves two major types of activities: first, collecting information about how much knowledge and skill students have learned (measurement); and then, making judgments about the adequacy

measurement

The assignment of value to products, performances, or attributes of people according to a rule-governed system.

evaluation

In assessment, the use of a rule-governed system to make judgments about the value or worth of a set of measures.

or acceptability of each student's level of learning (evaluation).

Some teachers, focusing heavily on the judgments they must make, tend to overlook the measurements that are used to make those judgments and how such information can help them teach more effectively. Those are the teachers that tend to think "grading" when they hear the word assessment. But as we will see, both aspects of classroom assess-

ment are critical to understanding student learning, and understanding student learning is critical to enhancing student learning.

Both measurement and evaluation activities can be accomplished in a number of ways. The most common ways that teachers measure learning is to have students take quizzes or exams, respond to oral questions, do homework exercises, write papers, solve problems, create products, and make oral presentations. Teachers can then evaluate the scores (i.e., the measurements taken) from those activities by comparing them either with one another or with an absolute standard (such as an A equals 90% correct). In this chapter, we will explain and illustrate the various ways in which you can measure and evaluate student learning with assessments that you create and administer regularly in your classroom (Airasian & Russell, 2008; Nitko & Brookhart, 2011).

Measurement For educational purposes, **measurement** is defined as the assignment of either a number (such as the score from a test) or a rating (such as the designation "excellent" or "exceeds standards" from a performance assessment) to certain attributes of people according to a rule-governed system. For example, we can measure someone's level of keyboarding proficiency by counting the number of words the person accurately types per minute. For an oral presentation, we might measure the quality by using a guide called a scoring rubric (Arter & Chappuis, 2008). In a classroom or other group situation, the rules that are used to assign the numbers or provide the rating ordinarily create a ranking that reflects how much of the attribute different people possess (Airasian & Russell, 2008; Nitko & Brookhart, 2011).

Evaluation Although related to measurement, evaluation is a distinct process that makes use of measurements. **Evaluation** involves using a rule-governed system to make judgments about the value or worth of a set of measures (Airasian & Russell, 2008; Nitko & Brookhart, 2011). What does it mean, for example, to say that a student answered 80 out of 100 earth science questions correctly? Depending on the rules that are used, it could mean that the student has learned that body of knowledge exceedingly well and is ready to progress to the next unit of instruction or, conversely, that the student has significant knowledge gaps and requires additional instruction.

Why Should We Assess Students' Learning?

As implied earlier, the short answer to the question of why to assess student learning is: To enhance student learning. Although that simple answer frames assessment as crucial to student learning—and it is—it leads

to other questions about how to use assessment to enhance

learning. There are two general ways in which assessment data (measurements) can be used to make judgments (evaluations). Thus, there are two kinds of evaluative judgments: summative and formative.

The distinction between summative and formative evaluation originated in the work of Michael Scriven (1967). In

the years since Scriven described how assessment serves different evaluative purposes, research on classroom assessment has evolved from an emphasis on the technical aspects of assessment (called psychometric theory) to research on how assessment practice contributes to a wide range of developmental, learning, and especially motivational outcomes for students. As a consequence of this shift, the terms *summative* and *formative evaluation* have been replaced by *summative* and *formative assessment* (Brookhart, 2009a).

After briefly describing summative and formative assessment, we will compare and contrast the two uses of assessment information in order to focus on assessment that can guide classroom teaching and learning day to day, hour to hour, or even minute to minute.



© flyfloor/Stockphoto

Summative Assessment (Assessment of Learning)

The first, and probably most obvious, reason for assessment is to provide to all interested parties a clear, meaningful, and useful summary or accounting of how well a student has met the teacher's objectives. When testing is done for the purpose of assigning a letter or numerical grade, it is often called **summative assessment** because its primary purpose is to provide an assessment of learning, to sum up how well a student has performed over time and at a variety of tasks.

Formative Assessment (Assessment for Learning)

A second reason for assessing students is to monitor their progress. The main things that teachers want to know from time to time are whether students are keeping up with the pace of instruction and are understanding all of the material that has been covered so far. For students whose pace of learning is either slower or faster than average, or whose understanding of certain ideas is faulty, instructional accommodations may be needed (recall the techniques discussed in Chapter 6). Because the purpose of such assessment is to facilitate, or form, learning and not to assign a grade, it is usually called **formative assessment.**

Assessment of Learning Compared to Assessment for Learning

A number of scholars have referred to assessment that leads to summative judgments as "assessment of learning" and assessment that is used to make formative judgments as "assessment for learning" (Moss & Brookhart, 2009; Stiggins, 2002, 2007; Tomlinson, 2007/2008). Two of these scholars, Connie Moss and Susan Brookhart, are with the Center for Advancing the Study of Teaching and Learning in the School of Education at Duquesne University. They have been conducting research for several years on how assessment practices advance teaching and learning. It is important to note that their research has been conducted *with* teachers, administrators and students in the Armstrong School District in Pennsylvania. Because they have worked closely with practicing educators, their conclusions are informed by practicing educators. Table 14.1 is based on Moss and Brookhart's work (2009): how they compare and contrast the characteristics of assessment of learning and assessment *for* learning.

The distinctions between summative assessment and formative assessment in Table 14.1 help us see that the process of assessment—first measuring and then evaluating—is tied closely to the underlying reason or purpose for assessing. As we will see later in this chapter (and in Chapter 15), summative assessment is not only legitimate, it is necessary. As a teacher, you will be required to assign grades to assignments, performances, quizzes, and tests. You will also be required to examine

the grades that you have assigned to each assignment, each performance, each quiz, and each test and issue a summary judgment: a final grade for a course or a grading period—a summative assessment of learning.

Along the way to those summative assessments of learning, you will be engaged in helping students meet the learning goals and objectives that define successful learning in your classroom. That is where formative assessment comes into play.

Formative assessments are conducted more or less continuously during an instructional unit, using both formal and informal assessment techniques. Periodic quizzes, homework assignments, in-class worksheets, oral readings, responses to teacher questions, and behavioral observations are all examples of formative assessments—if the results are used to generate timely feedback about what students have learned, what the source of any problems might be, and what might be done to prevent small problems from becoming major ones later in the year. Unlike summative assessment, which is a one-time event conducted only after instruction is finished, formative assessment has a more dynamic, ongoing, interactive relationship with teaching. The results of formative assessments affect instruction, which affects subsequent

summative assessment Testing done for the purpose of assigning a letter or numerical grade to sum up a student's performance at a variety of tasks over time.

formative assessment A type of assessment that monitors a student's progress in order to facilitate learning rather than to assign a grade. Also called *formative evaluation.*

TABLE 14.1 Characteristics That Distinguish Summative and Formative Assessment

Assessment of Learning (Summative Assessment)	Assessment for Learning (Formative Assessment)
Purpose is to summarize and audit learning	Purpose is to improve student learning
Conducted periodically to capture what learning has occurred	Conducted continuously while learning is in progress
Focus is on products of learning	Focus is on learning in process
Often perceived as an activity that occurs after the teaching-learning process has finished	Perceived as integral to the teaching-learning process
Teacher directed	Collaboration of teacher and student
Performance measures (e.g., scores on an exam) show where a student has arrived	Performance measures (e.g., scores on an exam) show a student's journey
Teachers use evidence to make a summary decision about success or failure (e.g., grades)	Teachers and students use evidence to make adjustments
Teachers take on the role of auditor; students are the audited	Teachers join with students as "intentional learners"

SOURCE: Adapted from Moss & Brookhart (2009).



© Mary Kate Demmy/Photo Edit

Classroom assessments can provide both summative and formative information. The former—assessments of learning—can tell the teacher (and others) what knowledge and skills a student has acquired. The latter—assessments for learning—can tell the teacher and the student what adjustments need to be made in order to take the next steps for the student's learning.

performance, and so on. Think back to our discussion of response to intervention (RTI) in Chapter 6. The RTI approach is sometimes characterized as “teach–test–teach.” The teacher instructs (or intervenes to help the student learn), student learning is measured and evaluated (to see if the intervention worked), and the teacher then modifies instruction based on the assessment. RTI can be thought of as a type of formative assessment.

Assessment as Learning Moss & Brookhart (2009) defined formative assessment as “an active and intentional learning process that partners the teacher and the students to continuously and systematically gather evidence of learning with the express goal of improving student achievement” (p. 6). In other words, they view assessment for learning as learning. Their claim is that unless both students and teachers are learning from the process, formative assessment is not occurring in the classroom. Whatever is happening might be something like formative assessment, but true formative assessment requires that the teacher and students are learning with the intention to improve student achievement. Their argument is that assessment for learning is successful only when teachers and students are becoming better and better because they are learning from their assessment activities in the classroom.

As we have indicated, learning about assessment is necessary to becoming an effective teacher. Mistilina

Sato, Ruth Chung Wei, and Linda Darling-Hammond (2008), for example, studied how the assessment practices of math and science teachers changed as a function of pursuing National Board Certification. Teachers who are candidates for certification by the National Board for Professional Teaching Standards go through an extensive assessment procedure and, at the end of the process, a summative judgment about whether they have or have not met the standards of the National Board. One becomes a National Board Certified Teacher or one does not. When teachers who sought National Board Certification were compared with those who did not, the authors found significant differences. National Board candidates exhibited substantial changes “in the variety of assessments used and the way assessment information was used to support student learning.” National Board candidates attributed changes in practice to the National Board standards and assessment tasks” (Sato, Wei, & Darling-Hammond, p. 669). In being assessed themselves, the teachers learned how to improve their assessment practices with their own students. Thinking back to Moss and Brookhart’s view of formative assessment, it is critical that assessment informs learning of both teachers and students.

In an essay entitled “Learning to Love Assessment,” Carol Ann Tomlinson (2007/2008) reflected on her long teaching career and, in particular, what she learned about and from the practice of classroom assessment. She concluded her essay as follows:

Lorna Earl (2003) distinguishes between assessment of learning, assessment for learning, and assessment as learning. In many ways, my growth as a teacher slowly and imperfectly followed that progression. I began by seeing assessment as judging performance, then as informing teaching, and finally as informing learning. In reality, all those perspectives play a role in effective teaching. The key is where we place the emphasis.

Certainly a teacher and his or her students need to know who reaches (and exceeds) important learning targets, thus summative assessment, or assessment of learning, has a place in teaching. Robust learning generally requires robust teaching, and both diagnostic and formative assessments, or assessments for learning, are catalysts for better teaching. In the end, however, when assessment is seen as learning for students as well as for teachers, it becomes most informative and generative for students and teachers alike. (p. 13)

Having made the argument that assessment is integral to teaching and critical for learning, we turn first to ways of measuring student learning and then to ways of evaluating student learning.

L02 Ways to Measure Student Learning

Just as measurement can play several roles in the classroom, teachers have several ways to measure what students have learned. Which type of measure you choose will depend, of course, on the objectives you have stated. For the purposes of this discussion, objectives can be classified in terms of two broad categories: knowing *about* something (for example, that knots are used to secure objects, that dance is a form of social expression, that microscopes are used to study things too small to be seen by the naked eye) and knowing *how to do* something (for example, tie a square knot, dance the waltz, operate a microscope). Measures that attempt to assess the range and accuracy of someone's knowledge are usually called written tests. And measures that attempt to assess how well somebody can do something are often referred to as performance assessments. Keep in mind that both types have a legitimate place in a teacher's assessment repertoire. Which type is used, and to what extent, will depend on the purpose or purposes you have for assessing students. In the next two sections, we will briefly examine the nature of both types.

Pause & Reflect

Over the past 10 to 12 years, you have taken probably hundreds of classroom tests. What types of tests best reflected what you learned? Why?

Written Tests

As we indicated at the beginning of this chapter, teachers spend a substantial part of each day assessing student learning, and much of this assessment activity involves giving and scoring some type of written test. Most written tests are composed of one or more of the following categories and item types: selected response (multiple choice, true-false, and matching) and constructed response (short answer and essay). In all likelihood, you have taken hundreds of these types of tests in your school career thus far.

Written tests measure the degree of knowledge about a subject.

In the next couple of pages, we will briefly describe the main features, advantages, and disadvantages of each test. As you read, bear in mind that what we said about the usefulness of both written tests and performance assessments applies here as well. No one type

of written test will be equally useful for all purposes. You are more likely to draw correct inferences about students' capabilities by using a variety of selected and constructed-response items.

Selected-Response Tests Selected-response tests are so named because the student reads a relatively brief opening statement (called a stem) and selects one of the provided alternatives as the correct answer. Selected-response tests are typically made up of multiple-choice, true-false, or matching items. Quite often all three item types are used in a single test. Although guidelines exist for writing selected-response items (see, for example, the 31 guidelines for writing multiple-choice items discussed by Haladyna, Downing, & Rodriguez, 2002), many of these guidelines have not been validated by research. Hence, test-item writing is currently as much an art as a science.

Selected-response tests are objectively scored and efficient but usually measure lower levels of learning and do not reveal what students can do.

Characteristics Selected-response tests are sometimes called *objective* tests because they have a simple and set scoring system. If alternative B of a multiple-choice item is keyed as the correct response and the student chooses alternative D, the student's answer is marked wrong, and the teacher's desire for a correct response cannot change the result. Selected-response tests are typically used when the primary goal is to assess what might be called foundational knowledge. This knowledge comprises the basic factual information and cognitive skills that students need in order to do such high level tasks as solve problems and create products (Stiggins, 2007).

Advantages A major advantage of selected-response tests is efficiency: A teacher can ask many questions in a short period of time. Another advantage is ease and reliability of scoring. With the aid of a scoring template (such as a multiple-choice answer sheet that has holes punched out where the correct answer is located), many tests can be quickly and uniformly scored. Moreover, there is some evidence that selected-response tests,

when well written, can measure higher level cognitive skills as effectively as constructed-response tests (Nitko & Brookhart, 2011; Stiggins, 2007).

Disadvantages Because items that reflect the lowest level of Bloom's taxonomy (verbatim knowledge) are the easiest to write, most teacher-made tests (and many standardized tests as well) are composed almost entirely of knowledge-level items (a point we made initially in Chapter 13). As a result, students focus on verbatim memorization rather than on meaningful learning. Another disadvantage is that, although we get some indication of what students know, such tests reveal nothing about what students can do with that knowledge. A third disadvantage is that heavy or exclusive use of selected-response tests leads students to believe that learning is merely the accumulation of universally agreed upon facts (Nitko & Brookhart, 2011; Martinez, 1999).

Short-Answer Tests As their name implies, short-answer tests require a brief written response from the student.

Characteristics Instead of selecting from one or more alternatives, the student is asked to supply from memory a brief answer consisting of a name, word, phrase, or symbol. Like selected-response tests, short-answer tests can be scored quickly, accurately, and consistently, thereby giving them an aura of objectivity. They are primarily used for measuring foundational knowledge.

Advantages Short-answer items are relatively easy to write, so a test, or part of one, can be constructed fairly quickly. They allow either broad or in-depth assessment of foundational knowledge because students can respond to many items within a short space of time. Because students have to supply an answer, they have to recall, rather than recognize, information.

★ { **Short-answer tests are easy to write but measure lower levels of learning.** }

Disadvantages Short-answer tests have the same basic disadvantages as selected-response tests. Because short-answer items ask only for short verbatim answers, students are likely to limit their processing to that level; thus these items provide no information about how well students can use what they have learned. In addition, unexpected but plausible answers may be difficult to score.

Essay Tests Essay items require students to organize a set of ideas and write a somewhat lengthy response to a broad question.

Characteristics The student is given a somewhat general directive to discuss one or more related ideas according to certain criteria. An example of an essay question is, "Compare operant conditioning theory and information-processing theory in terms of basic assumptions, typical research findings, and classroom applications."

Advantages Essay tests reveal how well students can recall, organize, and clearly communicate previously learned information. When well written, essay tests call on such higher level abilities as analysis, synthesis, and evaluation. Because of these demands, students are more likely to try to meaningfully learn the material on which they are tested (Nitko & Brookhart, 2011; Stiggins, 2007).

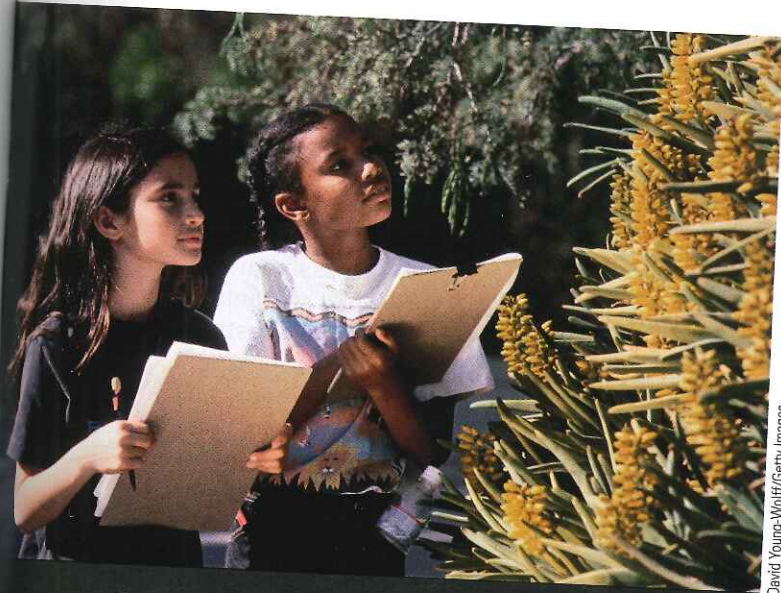
Disadvantages Consistency of grading is likely to be a problem. Two students may have essentially similar responses yet receive different letter or numerical grades because of differences in vocabulary, grammar, and style. These test items are also very time consuming to grade. And because it takes time for students to formulate and write responses, only a few questions at most can be given (Nitko & Brookhart, 2011; Liu, 2010). But recent developments in essay scoring by computer programs may drastically reduce or even eliminate these disadvantages in the near future (Myers, 2003).

Constructing a Useful Test Understanding the characteristics, advantages, and disadvantages of different types of written test and knowing how to write such test items are necessary but not sufficient conditions for creating an instructionally useful test. James Popham (2006, 2011), a noted measurement scholar, maintains that a useful classroom test has the following five attributes:

- **Significance.** The test measures worthwhile skills (such as the last four levels of Bloom's taxonomy—application, analysis, synthesis, and evaluation) and substantial bodies of important knowledge.
- **Teachability.** Effective instruction can help students acquire the skills and knowledge measured by the test.
- **Describability.** The skills and knowledge measured by the test can be described with sufficient clarity that they make instructional planning easier.
- **Reportability.** The test produces results that allow a teacher to identify areas of instruction that were probably inadequate.
- **Nonintrusiveness.** The test does not take an excessive amount of time away from instruction.

Performance Assessments

In recent years, many teachers, learning theorists, and measurement experts have argued that the typical



David Young-Wolff/Getty Images

Performance assessments provide students the opportunity to demonstrate what they can do with what they know.

written test should be used far less often than it is because it reveals little or nothing of the depth of students' knowledge and the ways students use their knowledge to work through questions, problems, and tasks. These individuals argue that because we are living in a more complex and rapidly changing world than was the case a generation ago, schools can no longer be content to hold students accountable for just how well they can learn, store, and retrieve information in more or less verbatim form. Instead, we need to teach and assess students for such capabilities as framing problems, formulating and carrying out plans, generating hypotheses, finding information that is relevant to the solution to a problem, and working cooperatively with others, because those are the types of skills that are necessary to cope successfully with the demands of life after school in the twenty-first century (Calfee, 2009; Cunningham, 2001; Fredrick, 2009).

In addition, the learning standards of such professional groups as the National Council of Teachers of Mathematics (<http://standards.nctm.org>), the National Council for the Social Studies (www.socialstudies.org/standards), the National Council of Teachers of English (www.ncte.org/standards), and the National Research Council (www.nap.edu/catalog/10256.html) call for students to develop a sufficiently deep understanding of subject matter that they can demonstrate their knowledge in socially relevant ways. One way to address these concerns is to use performance assessments.

What Are Performance Assessments? **Performance assessments** require students to use a wide range

of knowledge and skills over an extended period of time to complete a task or solve a problem under more or less realistic conditions. At the low end of the realism spectrum, students may be asked to construct a map, interpret a graph, or write an essay under highly standardized conditions. Everyone in the class completes the same task in the same amount of time and under the same conditions. At the high end of the realism spectrum, students may be asked to conduct a science experiment, produce a painting, or write an essay under conditions that are similar to those of real life. For example, students

may be told to produce a compare-and-contrast essay on a particular topic by a certain date, but the resources they choose to use, the number of revisions they make, and the schedule on which they work on the essay are left unspecified. When performance assessment is conducted under such realistic conditions, it is also called *authentic assessment* (Nitko & Brookhart, 2011; Gronlund & Waugh, 2009; Janesick, 2001).

Perhaps the clearest way to distinguish between traditional paper-and-pencil tests (such as multiple-choice tests) and performance assessments is to say that the former measure how much students know, whereas the latter measure what students can do with what they know. In the sections that follow, we will define four different types of performance assessments and then look at their most important characteristics.

Types of Performance Assessments There are four ways in which the performance capabilities of students are typically assessed: direct writing assessments, portfolios, exhibitions, and demonstrations.

Direct Writing Assessments Students are asked to write about a specific topic ("Describe the person whom you admire the most, and explain why you admire that person") under a standard set of conditions. Each essay is then scored by two or more people according to a set of defined criteria.

Portfolios A **portfolio** contains one or more pieces of a student's work, some of which demonstrate different stages of completion. For example, a student's writing portfolio may contain business letters; pieces of fiction; poetry; and an outline, rough draft, and final draft of a research paper. Through the inclusion of various stages of a research paper, both the process and the end product can be assessed. Portfolios can also be constructed for

performance assessment An assessment device that attempts to gauge how well students can use basic knowledge and skill to perform complex tasks or solve problems under more or less realistic conditions. Also called *performance-based assessment* and *authentic assessment*.

portfolio A collection of one or more pieces of a person's work, some of which typically demonstrate different stages of completion.

math and science, as well as for projects that combine two or more subject areas.

Either the student alone or the student in consultation with the teacher decides what is to be included in the portfolio. The portfolio is sometimes used as a showcase to illustrate exemplary pieces, but it also works well as a collection of pieces that represent a student's typical performances. In its best and truest sense, the portfolio functions not just as a housing for these performances but also as a means of self-expression, self-reflection, and self-analysis for an individual student (Chang, 2009; Lam & Lee, 2010).

Exhibitions Exhibitions involve just what the label suggests: a showing of such products as paintings, drawings, photographs, sculptures, videos, and models. As with direct writing assessments and portfolios, the products a student chooses to exhibit are evaluated according to a predetermined set of criteria.

Demonstrations In a demonstration, students are required to show how well they can use previously learned knowledge or skills to solve a somewhat unique problem (such as conducting a scientific inquiry to answer a question, interpreting a graph, or diagnosing the cause of a malfunctioning engine and describing the best procedure for fixing it) or to perform a task (such as reciting a poem, performing a dance, or playing a piece of music). Figure 14.1 shows a performance item for graph interpretation, the partially correct response of a student, and the corrective feedback offered by two classmates.

Characteristics of Performance Assessments

Performance assessments are different from traditional written tests in that they require the student to make an

active response, are more like everyday tasks, contain problems that involve many variables, are closely related to earlier instructional activities, use scoring guides that clearly specify the criteria against which responses will be evaluated, emphasize formative assessment, and are probably more responsive to cultural diversity.

Emphasis on Active Responding As we pointed out previously, the goal of performance assessment is to gain some insight into how competently students can carry out various tasks. Consequently, such tests focus on processes (that is, the underlying skills that go into a performance), products (observable outcomes, such as a speech or a painting), or both. For example, an instrumental music teacher may want to know whether students can apply their knowledge of music technique and theory to use the correct fingering and dynamics when playing a woodwind or piano (R. E. Clark, 2002).

Degree of Realism Although performance assessments strive to approximate everyday tasks, not every test needs to be—or can be—done under the most realistic circumstances. How realistic the conditions should be depends on such factors as time, cost, availability of equipment, and the nature of the skill being measured. Imagine, for example, that you are a third grade teacher and that one of your objectives is that students will be able to determine how much change they should receive after making a purchase in a store. If this is a relatively minor objective or if you do not have a lot of props available, you might simply demonstrate the situation with actual money and ask the students to judge whether the amount of change received was correct. If, however, you consider this to be a major objective and



Portfolio Assessment: Elementary Classroom

Go to the Education CourseMate website and watch the video, study the artifacts in the case, and reflect upon the following questions:

1. Portfolios can be time consuming for teachers to develop and grade. Based on this Video Case, do you think that using portfolios would be worth the extra time and effort? Explain your answer.
2. Based on this Video Case, why might some teachers find portfolios to be a more accurate representation of student abilities (as compared with other forms of assessment)?

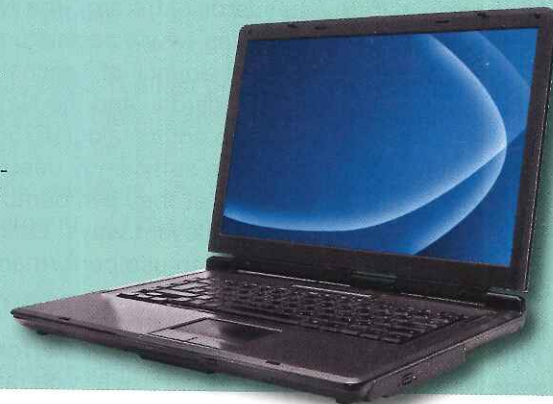


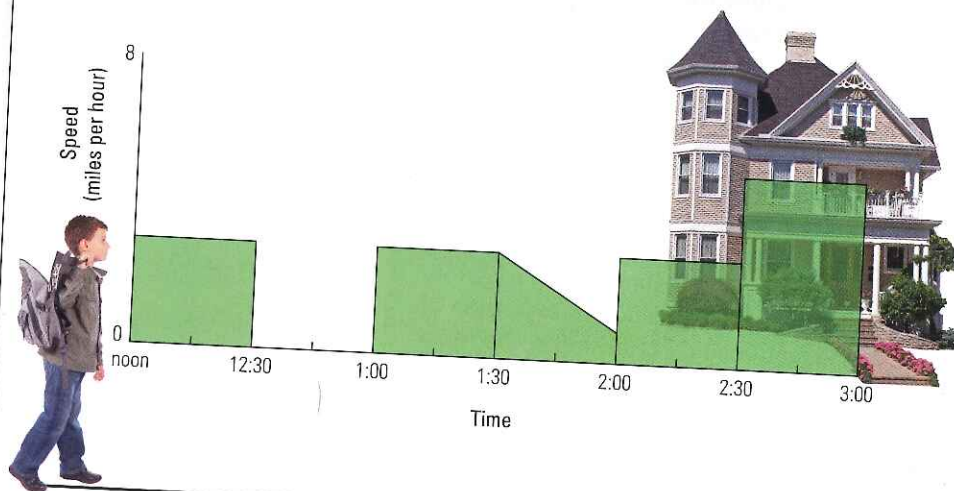
FIGURE 14.1 Example of a Performance Assessment: Interpreting a Graph

SOURCE: Parke & Lane (1997).

INTERPRETING A GRAPH

Use the following information and the graph to write a story about Tony's walk.

At noon, Tony started walking to his grandmother's house. He arrived at her house at 3 P.M. The graph below shows Tony's speed in miles per hour throughout his walk.



Write a story about Tony's walk. In your story, describe what Tony might have been doing at the different times.

Tony left his grandmother's house at noon. He made a stop at 12:30 to get a bite to eat Taco Bell. From 12:30 to 1:00 he stopped to rest on a bench at the park. From 1:00 to 1:30 he continued his walk. So Tony decided to stop at a friend house from 1:30 to 2:00. From 2:00 to 2:30 he continued his walk. Finally he arrived back at his grandmothers.

Students critique the response:

Because while he was at a friends house he was going 3 miles per hour. Also he didn't say when he sped up they didn't say.

Watch the chart when you write and put in minor details as well. Plus you can watch your speed.

you have the props available, you might set up a mock store and have each student make a purchase using real money (Gronlund & Waugh, 2009).

An example of a task that is realistic in content and intellectual demands but not in its setting (it takes place in the classroom) is "Read All About It!" Playing the roles of newspaper staff writers and editorial board members, students put together a special series for

their local newspaper that compares and contrasts the five major wars in which the United States was involved during the 1900s (World War I, World War II, Korean War, Vietnam War, and Persian Gulf War). Writing assignments include feature articles, opinion columns, and letters to the editor. Editorial responsibilities include story editor, photo editor, mock-up editor, layout editor, and copy editor (Moon, 2002).

Performance tests may vary in degree of realism.

Emphasis on Complex Problems To assess how well students can use foundational knowledge and skills in a productive way, the questions and problems they are given should be sufficiently open-ended and ill structured (Stiggins, 2007). The problems contained in the Quest Atlantis program that we described in earlier chapters are good examples of complex and somewhat ill-structured tasks. They have several interrelated parts,

rubric A scoring guide used in performance assessment that helps define and clarify levels of student performance from poor to exemplary.

provide few cues as to how they might be solved, and contain some uncertainty about what constitutes an appropriate solution.

Close Relationship Between Teaching and Testing

All too often students walk out of an exam in a state of high frustra-

tion (if not anger) because the content and format of the test seemed to have little in common with what was covered in class and how it was taught. It's the old story of teaching for one thing and testing for something else. Performance assessment strives for a closer match between teaching and testing. Often a performance assessment can be a variation or extension of a task used during instruction. For example, the mock store assessment mentioned earlier could follow an instructional activity in which students practiced making change.

This close relationship between assessment and instruction is not automatic, however; the teacher must deliberately establish it. For example, if in giving an oral book report a student is expected to speak loudly and clearly enough for everyone to hear, speak in complete sentences, stay on the topic, and use pictures or other materials to make the presentation interesting, the student needs to be informed of these criteria, and classroom instruction should be organized around them. One proponent of performance assessment cited the old farm adage "You don't fatten the cattle by weighing them" to make this point. He then went on to note, "If we expect students to improve their performance on these new, more authentic measures, we need to engage in 'performance-based instruction' on a regular basis" (McTighe, 1996/1997, p. 7).

By the same token, the assessment of students' performances should be limited to just the criteria emphasized during instruction (Nitko & Brookhart, 2011). One reason that proponents of performance

assessment push for this feature is that it has always been a standard part of successful programs in sports, the arts, and vocational education. Football coaches, for example, have long recognized that if they want their quarterback to know when during a game (the equivalent of a final exam) to attempt a pass and when not to, they must provide realistic opportunities for the quarterback to practice making this particular type of decision. Perhaps you recall our mentioning in Chapter 10 that realistic and varied practice are essential if students are to transfer what they learn in an instructional setting to an applied setting.

Pause & Reflect

Have you ever taken any kind of performance assessment as a student? Did you feel that it accurately reflected what you had learned? To what extent would you use performance measures for such academic subjects as writing, math, science, and social studies? Why?

Use of Scoring Rubrics A **rubric** is a scoring guide that specifies the capabilities students should exhibit (also known as content standards), describes the qualitative levels or categories into which the responses will be sorted (also known as performance standards), and specifies how the responses will be scored (as separate elements or holistically). For writing tasks, which are probably the most common performance assessments, some commonly used content criteria are clarity of purpose, organization, voice, word choice, grammatical usage, and spelling (Arter & Chappuis, 2008; Arter & McTighe, 2001). An example of a scoring rubric for an oral report is provided in Table 14.2.

Creating and using scoring rubrics and providing them to students at the beginning of a task are highly desirable for at least three reasons:

1. They increase the objectivity, consistency, and efficiency of scoring.

TABLE 14.2 Scoring Rubric for a Group Oral Presentation

Level	Content	Audiovisual Components	Group Members	Audience Members
Excellent	Accurate, specific, research based, retold in own words	Are unique, add to presentation quality of materials used, are neat, present a clear message	All equally involved in presentation and well informed about the topic	Maintain eye contact with presenters, ask many questions
Good	Less detailed, lacking depth, using a limited number of sources	Support topic but do not enhance presentation, exhibit some attempts at originality, present a clear message	Most are active; most are informed about the topic	Some not attending; ask limited or off-topic questions
Minimal	Limited information, general, straying from topic, not presented in own words	Inappropriate, no originality, detract from presentation, present a confusing message	One or two dominate; some do not seem well prepared or well informed	Not attending; ask no questions or off-topic questions

SOURCE: K. Montgomery (2000).

2. They help teachers match their instructional activities to the demands of the performance measure, the goal we discussed in the previous section.
3. They provide students with verbal descriptions and examples of the desired performance or product, allowing teachers to clearly communicate to students the types of behaviors that represent the range from unacceptable to exceptional performance and help students better monitor their progress and make productive changes in the quality of their work (Arter & Chappuis, 2008; Whittaker, Salend, & Duhaney, 2001). Students with learning disabilities are likely to experience the greatest benefit from being given a scoring rubric and being shown how to use it (Heacox, 2009; A. W. Jackson & Larkin, 2002).

Bear in mind, however, that scoring rubrics have their limitations. Although the rubrics used by two teachers to score writing samples may have some of the same content standards (such as clarity of purpose, organization, and grammar), they may differ as well (for instance, in the presence or absence of idea development, use of detail, and figurative use of language), because there are different ways to define good writing. Thus any one rubric is not likely to represent the domain of writing fully and may provide few or no opportunities for scorers to reward certain desirable writing skills (Osborn Popp, Ryan, & Thompson, 2009).

Use of Formative Assessment As we pointed out earlier, and as we will see in this chapter's Case in Print, tests can be used as a source of feedback to help students improve the quality of their learning efforts. Because many real-life performances and products are the result of several feedback and revision cycles,

performance assessment often includes this feature as well. Anyone who has ever done any substantial amount of writing can tell you (and we are no exception): A satisfactory essay, story, or even personal letter is not produced in one attempt. Usually, there are critical comments from oneself and others and subsequent attempts at another draft. If we believe that the ability to write well, even among people who do it for a living, is partly defined by the ability to use feedback profitably, why should this be any different for students (Stiggins, 2001)? Some specific forms of formative assessment are dress rehearsals, reviews of writing drafts, and peer response groups (Gronlund & Waugh, 2009).

Responsiveness to Cultural Diversity Traditional written tests have been criticized over the years for being culturally biased. That is, they are thought to underestimate the capabilities of many ethnic minority students, as well as students of low socioeconomic status, because they rely on a narrow range of item types (mainly selected response) and on content that mostly reflects the experiences of the majority culture (C. C. Johnson & Fargo, 2010; Stobart, 2005; J. Wang, Spalding, Odell, Klecka, & Lin, 2010). This criticism is based in large part on the constructivist view of learning: that meaningful learning occurs within a cultural context with which one is familiar and comfortable. If this is so, say the critics, then tests should be more consistent with the cultural context in which learning occurs. Performance assessments have been promoted as a way to assess more fairly and accurately the knowledge and skills of all students, and particularly minority students, because of their realism (including group problem solving) and closer relationship between instruction and assessment (H. Hart, 2009; Santamaria, 2009).



Performance Assessment: Student Presentation

Go to the Education CourseMate website and watch the video, study the artifacts in the case, and reflect upon the following questions:

1. During each presentation in Ms. Mosman's class, students assess their peers' performance. How do the peer assessments contribute to the learning of both the presenters and the assessors?
2. In this Video Case, Ms. Mosman claims that performance assessment allows students to demonstrate their learning in ways that tests may not. How do such demonstrations affect student learning and motivation?



Challenging Assumptions

Practice Assessment for Learning

The classroom assessments that teachers devise are among the most powerful influences on the quality of students' learning, largely due to their effect on self-efficacy, interest, and the types of learning strategies that students construct. Whether these assessments have positive or negative effects on students depends on how they are constructed and what purpose they are primarily intended for.

As we have noted in this and other chapters, classroom assessments can be used both to sum up what students have learned (summative assessment) and to provide information about the effectiveness of instruction and students' specific strengths and weaknesses (formative assessment). All too often, unfortunately, the formative type of assessment tends to be overshadowed by the summative type.

Many teachers are more concerned with giving students grades than with using information gained from assessment to improve their instruction. Although both types of assessment are legitimate, we encourage you to emphasize formative assessment because of its potential to positively shape students' learning.

To ensure that assessments serve as a positive force for learning, we believe teachers should take the following steps:

- Make sure that you are knowledgeable about, understand, and use the basic measurement concepts and practices described in this chapter. Don't fall into the trap that so many teachers have fallen into of treating classroom assessment as a necessary evil.
- Recognize that the most accurate and useful assessments of learning are composed of multiple and varied measures. Use the full range of assessments (written tests, performance assessments, checklists, rating scales) available to you.
- Align the content of your assessments with your objectives, and fully inform students about the content and demands of your assessments.
- Finally, use the results to learn how to work even more productively with your students.

What Do You Think?

Have you been in classes in which formative assessment was downplayed in favor of summative assessment? If so, what effect did that have on your motivation to learn? If not, how did the emphasis on formative assessment affect you?



Some Concerns About Performance Assessments

There is no question that alternative assessment methods have excited educators and will be used with increasing frequency. But some of the same features that make these new assessment methods attractive also create problems. One problem concerns the increased emphasis on standardized tests. Standardized tests are typically used as summative assessments (assessments of learning). They are given at the end of the academic year and are used solely to rank and compare students, schools, and school districts. Performance-based classroom tests, on the other hand, lend themselves to formative assessment (assessment for learning). They are given periodically to provide teachers, students, and parents with relevant information about the current level of student learning and to generate ideas about how performance might be improved. The challenge for the teachers is not to let the school district's preoccupation with high stakes tests (those on which poor performance has significant consequences for students, teachers, and administrators) crowd out their use of performance-based tests for formative evaluation purposes (Hargreaves, Earl, & Schmidt, 2002; Nichols & Berliner, 2007).

There are also questions about the reliability (how consistently the test performs) and validity (how accurately the test measures its target) of performance measures (Bachman, 2002). Susan Brookhart (2009b) described the "sort of" phenomenon that can occur with performance assessments: In such a case the performance "sort of taps a learning outcome but also requires extraneous skills or doesn't require all of the relevant skills" (p. 59). To illustrate, imagine a teacher who required her students to create and then perform a skit in order to demonstrate their understanding of a chapter in a novel the class was studying. Care would need to be taken to ensure that the assignment—and the assessment of the performance—focused on demonstrating understanding of the characters and the events in the chapter. In this situation, the teacher might be distracted by the dramatic performance and fail to assess the target of the performance.

LO3 Ways to Evaluate Student Learning

Once you have collected all the measures you intend to collect—for example, test scores, quiz scores,

homework assignments, special projects, ratings of products and performances, and laboratory experiments—you will have to give the data some sort of value (the essence of evaluation). As you probably know, this is most often done by using an A-to-F grading scale. There are two general ways to approach this task. One approach is making comparisons among students. Such forms of evaluation are called *norm referenced* because students are identified as average (or normal), above average, or below average. An alternative approach is called *criterion referenced* because performance is interpreted in terms of defined criteria. Although both approaches can be used, we favor criterion-referenced grading for reasons we will mention shortly.

Norm-Referenced Grading

A **norm-referenced grading** system assumes that classroom achievement will naturally vary among a group of heterogeneous students because of differences in such characteristics as prior knowledge, learning skills, motivation, and aptitude (to be discussed in Chapter 15). Under ideal circumstances (hundreds of scores from a diverse group of students), this variation produces a bell-shaped, or “normal,” distribution of scores that ranges from low to high, has few tied scores, and has only a very few low scores and only a very few high scores. For this reason, norm-referenced grading procedures are also referred to as “grading on a curve.”

Norm-referenced grading is based on the absence of external criteria.

The Nature of Norm-Referenced Grading Course grades, like standardized test scores, are determined through a comparison of each student’s level of performance with the normal, or average, level of other, similar students in order to reflect the assumed differences in amount of learned material. The comparison may be with all other members of the student’s class that year, or it may be with the average performance of several classes stretching back over several years. It is probably better for teachers to use a broad base of typical student performance, made up of several classes, as grounds for comparison than to rely on the current class of students. Doing so avoids two severe distorting effects: (1) When a single class

contains many weak students, those with more well-developed abilities will more easily obtain the highest grades; and (2) when the class has many capable students, the relatively weaker students are virtually predestined to receive low or failing grades (Brookhart, 2009b; Nitko & Brookhart, 2011; Gronlund & Waugh, 2009; Kubiszyn & Borich, 2010).

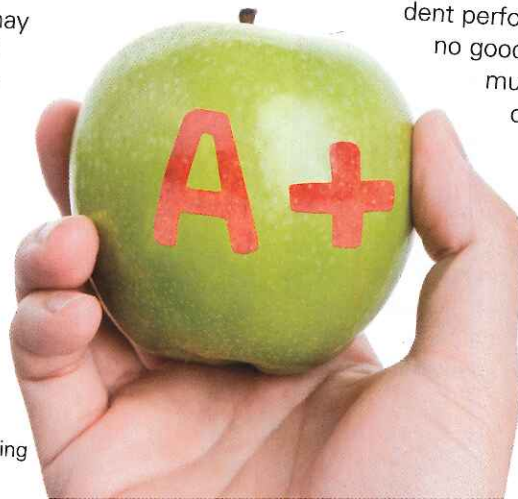
The basic procedure for assigning grades on a norm-referenced basis involves just a few steps:

1. Determine what percentage of students will receive which grades. If, for example, you intend to award the full range of grades, you may decide to give As to the top 15%, Bs to the next 25%, Cs to the middle 35%, Ds to the next 15%, and Fs to the bottom 10%.
2. Arrange the scores from highest to lowest.
3. Calculate which scores fall in which category, and assign the grades accordingly.

Many other arrangements are also possible. How large or small you decide to make the percentages for each category will depend on such factors as the nature of the students in your class, the difficulty of your exams and assignments, and your own sense of what constitutes appropriate standards. Furthermore, a norm-referenced approach does not necessarily mean that each class will have a normal distribution of grades or that anyone will automatically fail. For example, it is possible for equal numbers of students to receive As, Bs, and Cs if you decide to limit your grading system to just those three categories and award equal numbers of each grade. A norm-referenced approach simply means that the grading symbols being used indicate one student’s level of achievement relative to other students.

Proponents of norm-referenced grading typically point to the absence of acceptable external criteria for use as a standard in evaluating and grading student performance. In other words, there is no good way to determine externally how much learning is too little, just enough, or more than enough for some subject. And if there is no amount of knowledge or set of behaviors that all students must master, then grades may be awarded on the basis of relative performance among a group of students (Gronlund & Waugh, 2009).

norm-referenced grading A system of grading that assumes classroom achievement will vary among a group of heterogeneous students because of such differences as prior knowledge, learning skills, motivation, and aptitude, and so compares the score of each student to the scores of other students in order to determine grades.



© Feng Yu/Shutterstock Images

Case in Print

How Am I Doing?

Formative assessments are conducted more or less continuously during an instructional unit, using both formal and informal assessment techniques. . . . Unlike summative assessment, which is a one-time event conducted only after instruction is finished, formative assessment has a more dynamic, ongoing, interactive relationship with teaching. The results of formative assessments affect instruction, which affects subsequent performance, and so on. (p. 309)

Remote Tracking: Teachers Click in Quickly on Students' Performance

AMY HETZNER

Milwaukee Journal Sentinel, 4/22/06

Hartland—When social studies teacher Maria Fricker wants to see how much her students remember about the Bosnian war or whether they know how many electoral votes are needed to elect a U.S. president, she has them take out their remote controls.

Forget asking questions and calling on the few students who raise their hands.

With a product called the Classroom Performance System, which allows students to interact with a computer program through infrared response pads, teachers such as Fricker at North Shore Middle School in Hartland can see with the press of a finger whether their students are following the lesson.

The instant feedback lets teachers know if they need to spend more time on a topic or if they're dwelling on something their students already understand and can move on.

"It shows not only the strengths and weaknesses of our kids," Fricker said, "but the strengths and weaknesses of me."

With the Classroom Performance System and other remote control-like products that have entered the educational market in recent years, technology has come to the call of schools and teachers clamoring for new ways to gauge what their students are learning.

North Shore library media specialist Sue Klopp first witnessed the Classroom Performance System at a Wisconsin Educational Media Association conference. She said she liked the instant feedback it provided and how it transfixed everyone who walked by.

Given that North Shore was looking for more ways to assess students, Klopp thought the system would be a good fit for the school's goals.

North Shore Principal Dale Fisher did, too. He authorized purchasing two systems, for about \$2,000 each, for the current school year.

"The possibilities were quite endless for us to use it in the classroom," he said. "Traditional multiple-choice tests are given. The teacher can present the questions up on the board and then immediately get data on how their students understand the information. . . . You can differentiate your instruction better. It's better time management, as I see it."

The Classroom Performance System consists of a set of response pads and a software package.

Combined with a personal computer and a projector, it allows teachers to draft a set of questions that can be beamed onto a screen and then answered by students using the pads at their desks.

The pads, shaped and run like remote controls, have only eight buttons on them so the questions generally have to be crafted for multiple-choice or yes-no answers.

Because the computer program keeps track of how many students are keying in correct answers, as well as who is getting what wrong, teachers can get immediate feedback on how their entire class is doing.

Everyone Can See It

Not far from North Shore Middle School, Merton Intermediate School acquired a set of the Classroom Response System.

The hand-held responders, which look like the clunky cell phones of yore, contain a set of keys where students can answer multiple-choice questions, a calculating device and a small screen where teachers can transmit questions.

"What I really like is the potential for it to be real-time assessment—I'm talking to you and I ask questions as the lesson's going on. . . . I can see it, you can see it, we're getting the concept here," Merton Intermediate Principal Jon Wagner said.

A fourth-grade teacher is training herself on the system, which she plans to test with students next school year, Wagner said.

Help for Struggling Students

The Kettle Moraine School District recently purchased two sets of the responders for use with its special education students.

Strengths and Weaknesses of Norm-Referenced Grading

There are at least two circumstances under which it may be appropriate to use norm-referenced measurement and evaluation procedures:

1. *Evaluating advanced levels of learning.* You might, for example, wish to formulate a two-stage instructional plan in which the first stage involves helping all students master a basic level of knowledge and skill in a particular subject. Performance at this stage would be measured and evaluated against a predetermined standard (such as 80% correct on an exam). Once this

has been accomplished, you could supply advanced instruction and encourage students to learn as much of the additional material as possible. Because the amount of learning during the second stage is not tied to a predetermined standard and because it will likely vary due to differences in motivation and learning skills, a norm-referenced approach to grading can be used at this stage. This situation also fits certain guidelines for the use of competitive reward structures (discussed in Chapter 13) because everyone starts from the same level of basic knowledge.

Carol Smiley, who teaches cognitively disabled students at Kettle Moraine High School, said the devices are beneficial because they allow her to immediately direct her attention to struggling students.

But, while students like them, they haven't been trouble-free. Any assessment given on the responders has to be limited to a single class period, she said, and the technology doesn't allow for graphics that might be more helpful for certain students.

"It's nice because it does eliminate paper-pencil kind of things," Smiley said. "And also the anonymity—not everybody can look over your shoulder and see what you're doing."

The Merton and North Shore principals say the systems meet twin goals for their schools—incorporating technology into lessons and giving teachers data to help guide their instruction.

Recently in Fricker's class at North Shore, the teacher was able to see how well her students remembered their lessons about the workings of government by giving them a sample citizenship test used by the government.

While she controlled it through a computer set up in the middle of the classroom, the screen at the front of Fricker's room moved through a series of questions from who was Martin Luther King Jr. to how many branches of government there are at the federal level.

After students had keyed in their choices from the three she offered in response to each question, she clicked a button that placed a check next to the correct selection and showed how many students answered it right.

At the same time, a computer program recorded individual scores so Fricker could check back to see how students did individu-

ally, which she promised to do on more than one question.

"Oh boy, I'm going to find out who thinks Jim Doyle is the vice president of the United States," Fricker said at one point during the quiz. "I'm going to look it up later, guys."

Lets Students Respond

In addition to giving her quick feedback on how her students are grasping the subject matter, Fricker said the anonymity of the Classroom Performance System also could come in handy in dealing with controversial areas where students otherwise might be afraid to speak their minds.

She said she could poll them on how they feel about subjects such as the death penalty and abortion, then pursue why students feel certain ways without embarrassing ones that might not want to share their beliefs.

Fricker's students are enthusiastic about the system. "I think it's better than writing, and you get to see the answers right away and whether you got it right or wrong," said Jocelyn Budzien, 13, a seventh-grader at the school.

Teachers use the systems often for pre-tests and have been creative about coming up with fun ways to use them, such as "Jeopardy!"-style quizzes, said Luke Zarling, 12.

"It's just a good way to remember certain things on the test, and it's funner," he said. "In a way, it's almost like a game. But you're learning while you play the game."

Source: MILWAUKEE JOURNAL SENTINEL (ONLY STAFF-PRODUCED MATERIALS MAY BE USED). Copyright 2006 by JOURNAL/SENTINEL, INC. Reproduced with permission of JOURNAL/SENTINEL, INC. in the format Textbook via Copyright Clearance Center.

Questions and Activities

1. Formative evaluation means that teachers use assessments in order to make judgments about students' progress. It is also important that formative feedback be provided to students in a timely fashion. What advantages accrue to students who receive formative feedback immediately? What advantages accrue to teachers? To what extent do the advantages of the personal response system rest on a teacher's ability to formulate good questions?
2. The personal response system described in the article seems to provide useful information for students and helps teachers make instructional decisions, manage time, and maintain records of student performance. The system is also flexible in the sense that it allows teachers to tailor its use to specific lessons. But what are the disadvantages or limitations of the system described in the article? How might those disadvantages be minimized or overcome?
3. The students in the article report that using the response system is "better than writing" and "funner." Thinking back to our earlier study of motivation and social cognitive theory, how might the use of the system described in the article affect a student's sense of self-efficacy? Perhaps you or some of your friends have used a personal electronic response system in one or more of your classes. If so, reflect on your own reactions or ask your friends how they felt about their experiences with the system. To what extent did the system account for your reaction and to what extent did the questions used in the system account for your reaction?

2. *Selection for limited-enrollment programs.* Norm-referenced measurement and evaluation are also applicable in cases in which students with the best chances for success are selected for a limited-enrollment program from among a large pool of candidates. One example is the selection of students for honors programs who have the highest test scores and grade-point averages (Gronlund & Waugh, 2009).

The main weakness of the norm-referenced approach to grading is that there are few situations in

There are few appropriate uses for norm-referenced grading in classrooms.

criterion-referenced grading A system in which grades are determined on the basis of whether each student has attained a defined standard of achievement or performance.

which the typical public school teacher can appropriately use it. Either the goal is not appropriate (as in mastery of certain material and skills by all students or diagnosis of an individual student's specific strengths and weaknesses) or the basic conditions cannot be met (classes are too small or homogeneous or both).

When a norm-referenced approach is used in spite of these weaknesses, communication and motivation problems are often created.



Pause & Reflect

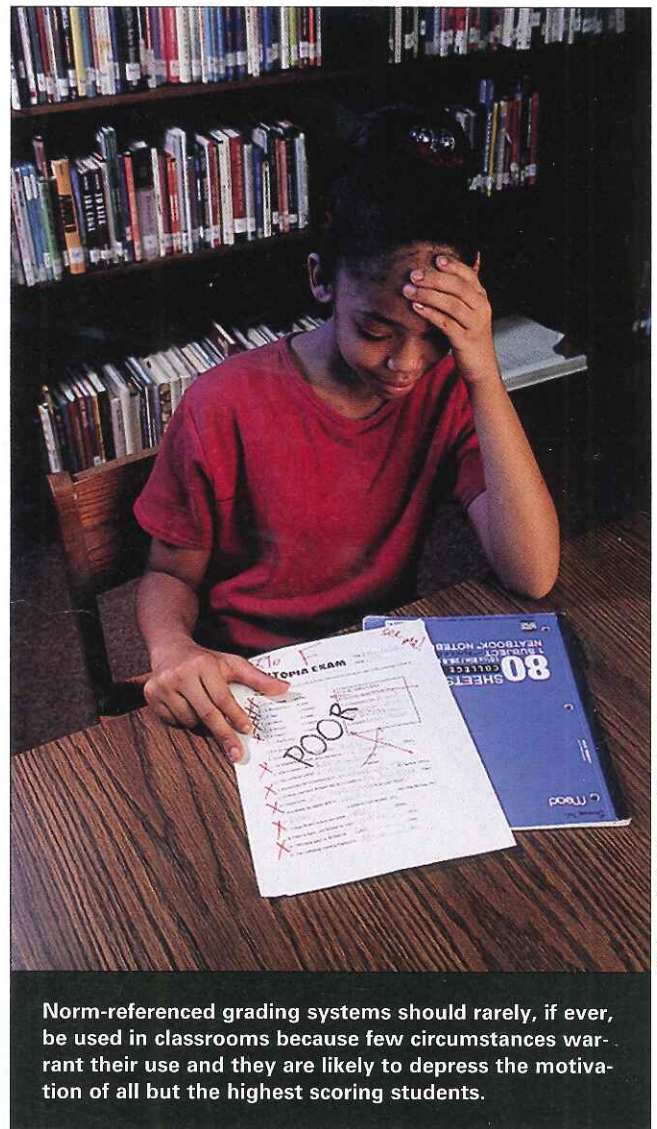
Have you ever taken a class that was graded on a curve? Did you feel that your grade accurately reflected how much you had learned? If not, why was the grade too low or too high?

Consider the example of a group of high school sophomores having a great deal of difficulty mastering German vocabulary and grammar. The students may have been underprepared, the teacher may have done a poor job of organizing and explaining the material, or both factors may be at work. The top student averaged 48% correct on all of the exams, quizzes, and oral recitations administered during the term. That student and a few others with averages in the high 40s will receive the As. Although these fortunate few may realize their knowledge and skills are incomplete, others are likely to conclude falsely that these students learned quite a bit about the German language, as a grade of A is generally taken to mean superior performance.

At the other extreme, we have the example of a social studies class in which most of the students are doing well. Because the students were well prepared by previous teachers, used effective study skills, were exposed to high quality instruction, and were strongly motivated by the enthusiasm of their teacher, the final test averages ranged from 94% to 98% correct. And yet the teacher who uses a norm-referenced scheme would assign at least As, Bs, and Cs to this group. Not only does this practice seriously damage the motivation of students who worked hard and performed well, but it also miscommunicates to others the performance of students who received Bs and Cs (Airasian & Russell, 2008).

Criterion-Referenced Grading

A **criterion-referenced grading** system permits students to benefit from mistakes and improve their level of understanding and performance. Furthermore, it establishes an individual (and sometimes cooperative)



Norm-referenced grading systems should rarely, if ever, be used in classrooms because few circumstances warrant their use and they are likely to depress the motivation of all but the highest scoring students.

Bob Daemrich/Bob Daemrich Photography

reward structure, which fosters a greater motivation to learn than other systems.

The Nature of Criterion-Referenced Grading Under a criterion-referenced system, grades are determined by the extent to which each student has attained a defined standard (or criterion) of achievement or performance. Whether the rest of the students in the class are successful or unsuccessful in meeting that criterion is irrelevant. Thus any distribution of grades is possible. Every student may get an A or an F, or no student may receive these grades. For reasons we will discuss shortly, very low or failing grades may occur less frequently under a criterion-referenced system.

A common version of criterion-referenced grading assigns letter grades on the basis of the percentage of test items answered correctly. For example, you may decide to award an A to anyone who correctly answers at least 85% of a set of test questions, a B to anyone

who correctly answers 75% to 84%, and so on down to the lowest grade. To use this type of grading system fairly, which means specifying realistic criterion levels, you would need to have some prior knowledge of the levels at which students typically perform. You would thus be using normative information to establish absolute, or fixed, standards of performance. However, although both norm-referenced and criterion-referenced grading systems spring from a normative database (that is, from comparisons among students), only the former system uses those comparisons to directly determine grades.

Criterion-referenced grades provide information about strengths and weaknesses.

Strengths and Weaknesses of Criterion-Referenced Grading Criterion-referenced grading systems (and criterion-referenced tests) have become increasingly popular in recent years, primarily because of the following advantages:

- Criterion-referenced tests and grading systems provide more specific and useful information about student strengths and weaknesses than do norm-referenced grading systems. Parents and teachers are more interested in knowing that a student received an A on an earth science test because she mastered 92% of the objectives for that unit than they are in knowing that she received an A on a test of the same material because she outscored 92% of her classmates.

- Criterion-referenced grading systems promote motivation to learn because they hold out the promise that all students who have sufficiently well-developed learning skills and receive good quality instruction can master most of a teacher's objectives (Gronlund & Waugh, 2009). The motivating effect of criterion-referenced grading systems is likely to be particularly noticeable among students who adopt mastery goals (which we discussed in Chapter 11) because they tend to use grades as feedback for further improvement (Moss & Brookhart, 2009; Stiggins, 2007).

One weakness of the criterion-referenced approach to grading is that the performance standards one specifies (such as a grade of A for 90% correct) are arbitrary and may be difficult to justify to parents and colleagues. (Why not 87% correct for an A? Or 92%?) A second weakness is that although a teacher's standards may appear to be stable from one test to another (90% correct for an A for all tests), they may in reality fluctuate as a result of unnoticed variation in the difficulty of each test and the quality of instruction (Gronlund & Waugh, 2009).

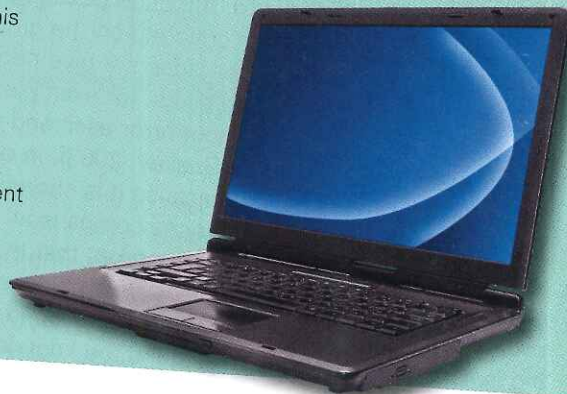
Finally, we would like to alert you to a characteristic of criterion-referenced evaluation that is not a weakness but is an unfortunate fact of educational life that you may have to address. In a variety of subtle and sometimes not so subtle ways, teachers are discouraged from using a criterion-referenced approach to grading because it tends to produce higher test scores and grades than a norm-referenced approach does. The reason for the higher scores is obvious and quite justified: When test items are based solely on the specific instructional objectives that teachers write and when



Assessment in the Middle Grades: Measurement of Student Learning

Go to the Education CourseMate website and watch the video, study the artifacts in the case, and reflect upon the following questions:

1. Evaluate the assessment practices used by Mr. Somers in this Video Case. Which ones did you find particularly effective? Support your answers using information from the textbook about effective assessment practices.
2. In this Video Case, Mr. Somers uses a written test to evaluate student understanding of course material. What are some other forms of assessment that could be used to gauge student understanding of a math unit?



those objectives are clear and provided to students, students know what they need to learn and what they need to do to meet the teacher's objectives. Also, because students' grades depend only on how well they perform, not how well their classmates perform, motivation for learning tends to be higher. The result is that students tend to learn more and score higher on classroom tests. So why should this happy outcome be a cause for concern? Because individuals who are not well versed in classroom measurement and evaluation may believe that the only reason large numbers of students achieve high grades is that the teacher has lower standards than other teachers. Consequently, you may find yourself in a position of having to defend the criteria you use to assign grades. Tom Kubiszyn and Gary Borich pointed out that although there is a great call for excellence in education, most people are hesitant to embrace marking systems that place excellence within every student's reach (2010).

Be aware of and avoid faulty measurement and grading practices.



A Mastery Approach A particular criterion-referenced approach to grading is often referred to as a mastery approach because it allows students multiple opportunities to learn and demonstrate their mastery of instructional objectives. This approach stems in large part from the work of John Carroll (1963) and Benjamin Bloom (1968, 1976) on the concept of *mastery learning*. The basic idea behind mastery learning is that most students can master most objectives if they are given good quality instruction and sufficient time to learn and are motivated to continue learning (Lalley & Gentile, 2008).

In a mastery approach, tests are used for formative as well as summative evaluation purposes. Thus students whose scores indicate deficiencies in learning are given additional instruction and a second chance to show what they have learned. Although pedagogically sound, this approach is often criticized on the grounds that life outside of school often does not give people a second chance. Surgeons and pilots, for example, are expected to do their jobs without error each and every time (Anders Ericsson, 2009; Guskey, 2003). In our view, this criticism is flawed because it is shortsighted and involves an apples and oranges comparison. First, even surgeons and pilots made mistakes that they were allowed to correct. Surgeons made their mistakes on cadavers and pilots on flight simulators. Second, schooling is about helping students acquire the knowledge

and skills they need to move from novices to experts and become self-directed learners.

Using a mastery approach in your classroom means focusing on formative assessment: the ideas surrounding assessment *for* learning discussed earlier in this chapter. In addition to thinking about the characteristics of formative assessment (seen in Table 14.1), the mastery approach to assessment is closely aligned with the practices of response to intervention (RTI) discussed earlier in this chapter and in Chapter 6. Like the mastery approach, RTI is focused on helping students master the material, not necessarily on getting it right the first time. In both the mastery approach and RTI, there is a continuous cycle of teaching followed by testing followed by more teaching, but now the teaching is informed by the results of the testing. Figure 14.2 is an example of a mastery-oriented, criterion-referenced approach to grading. Note that students are given more than one opportunity to demonstrate their mastery on each of the three exams.

Improving Your Grading Methods: Assessment Practices to Avoid

Earlier in this chapter, we noted that the typical teacher has little systematic knowledge of assessment principles and as a result may engage in a variety of inappropriate testing and grading practices. We hope that the information in this chapter will help you become more proficient at these tasks. (In addition, we strongly encourage you to take a course in classroom assessment if you have not already done so.) To reinforce what you have learned here, we will describe some of the more common inappropriate testing and grading practices that teachers commit. The following list is based largely on the observations of Susan Brookhart and Anthony Nitko (2008), Thomas Haladyna (1999), and Thomas Guskey (2002).

1. **Worshiping averages.** Some teachers mechanically average all scores and automatically assign the corresponding grade, even when they know an unusually low score was due to an extenuating circumstance. Allowances can be made for physical illness, emotional upset, and the like; a student's lowest grade can be dropped, or he can repeat the test on which he performed most poorly. Although objectivity in grading is a laudable goal, it should not be practiced to the extent that it prevents you from altering your normal procedures when your professional judgment indicates an exception is warranted.

Another shortcoming of this practice is that it ignores measurement error. No one can construct

FIGURE 14.2 Page From a Teacher's Grade Book and Instructions to Students: A Mastery Approach

	1st Exam		2nd Exam		3rd Exam		Exam Total Points	Projects			Extra Project	Grade
	1st Try	2nd Try	1st Try	2nd Try	1st Try	2nd Try		1	2	3		
Adams, Ann	16	18	17	18	18			P	P	P		
Baker, Charles	13	14	14		10	14		P				
Cohen, Matthew	14	16	18	16	17			P	P			
Davis, Rebecca	19		19		20			P	P	P		
Evans, Deborah	16	18	17	18	16	18		P	P	P		
Ford, Harold	18	16	17		15			P	P			
Grayson, Lee	10	13	12	14	12	15		P				
Hood, Barbara	16		17		15			P	P			
Inghalls, Robert	16	18	16		15			P	P			
Jones, Thomas	11	14	12	16	15			P				
Kim, David	18		19		19			P	P	P		
Lapine, Craig	14	16	18		16			P	P			
Moore, James	17		17		17			P	P			
Nguyen, Tuan	17	18	19		16	17		P	P	P		
Orton, John	10	10	11		9							
Peck, Nancy	14		15		14			P				
Quist, Ann	16	18	17	18	18			P	P	P		
Richards, Mary	16		17		15			P	P			
Santos, Maria	13		15		14			P				
Thomas, Eric	15	16	18	17	15			P	P			
Wong, Yuen	14		15		16			P				
Vernon, Joan	11	14	13	14	12	14		P				
Zacharias, Saul	16	18	17		16	19		P	P	P		



INSTRUCTIONS FOR DETERMINING YOUR GRADE IN SOCIAL STUDIES

Your grade in social studies this report period will be based on three exams (worth 20 points each) and satisfactory completion of up to three projects.

Here are the standards for different grades:

- A—Average of 18 or more on three exams, plus three projects at Pass level
- B—Average of 16 or 17 on three exams, plus two projects at Pass level
- C—Average of 14 or 15 on three exams, plus one project at Pass level
- D—Average of 10 to 13 on three exams
- F—Average of 9 or less on three exams

Another way to figure your grade is to add together points as you take exams. This may be the best procedure to follow as we get close to the end of the report period. Use this description of standards as a guide:

- A—At least 54 points, plus three projects at Pass level
- B—48 to 53 points, plus two projects at Pass level
- C—42 to 47 points, plus one project at Pass level
- D—30 to 41 points
- F—29 points or less

If you are not satisfied with the score you earn on any exam, you may take a different exam on the same material in an effort to improve your score. (Some of the questions on the alternate exam will be the same as those on the original exam; some will be different.) Projects will be graded P (Pass) or DO (Do Over). If you receive a DO on a project, you may work to improve it and hand it in again. You may also submit an *extra* project, which may earn up to 3 points of bonus credit (and can help if your exam scores fall just below a cutoff point). As you take each exam and receive a Pass for each project, record your progress on this chart.

First Exam		Second Exam		Third Exam		Project			Extra Project	Grade
1st Try	2nd Try	1st Try	2nd Try	1st Try	2nd Try	1	2	3		

the perfect test, and no person's score is a true indicator of knowledge and skill. Test scores represent estimates of these characteristics. Accordingly, giving a student with an average of 74.67 a grade of D when 75 is the minimum needed for a C pretends that the test is more accurate than it really is. This is why it is so important to conduct an item analysis of your tests. If you discover several items that are unusually difficult, you may want to make allowances for students who are a point or two from the next highest grade (and modify the items if you intend to use them again).

2. Using zeros indiscriminately. The sole purpose of grades is to communicate to others how much of the curriculum a student has mastered. When teachers also use grades to reflect their appraisal of a student's work habits or character, the validity of the grades is lessened. This occurs most dramatically when students receive zeros for assignments that are late (but are otherwise of good quality), incomplete, or not completed according to directions, or for exams on which they are suspected of cheating. This is a flawed practice for two reasons:

- First, and to repeat what we said in point 1, there may be good reasons why projects and homework assignments are late, incomplete, or different from what was expected. You should try to uncover such circumstances and take them into account.

- Second, zeros cause communication problems. If a student who earns grades in the low 90s for most of the grading period is given two zeros for one or more of the reasons just mentioned, that student could easily receive a D or an F. Such a grade is not an accurate reflection of what was learned.

If penalties are to be given for work that is late, incomplete, or not done according to directions, and for which there are no extenuating circumstances, they should be clearly spelled out far in advance of the due date and should not seriously distort the meaning of the grade. For students suspected of cheating, for example, a different form of the exam can be given.

Pause & Reflect

Because students in American schools feel considerable pressure to obtain high grades, a significant number of them feel driven to cheat. What might you do to reduce your students' tendency to cheat?

3. Providing insufficient instruction before testing. For a variety of reasons, teachers occasionally spend more time than they had planned on certain topics. In an effort to "cover the curriculum" prior to a scheduled exam, they may significantly increase the pace of instruction or simply tell students to read the remaining material on their own. The low grades that typically result from this practice will unfortunately be read by outsiders (and this includes parents) as a deficiency in students' learning ability, when in fact they more accurately indicate a deficiency in instructional quality.
4. Teaching for one thing but testing for another. This practice takes several forms. For instance, teachers may provide considerable supplementary material in class through lecture—thereby encouraging students to take notes and study them extensively—but base test questions almost entirely on text material. Or if teachers emphasize the text material during class discussion, they may take a significant number of questions from footnotes and less important parts of the text. A third form of this flawed practice is providing students with simple problems or practice questions in class that reflect the knowledge level of Bloom's taxonomy but then giving complex problems and higher level questions on a test. Remember what we said earlier in this book: If you want transfer, then teach for transfer.
5. Using pop quizzes to motivate students. If you recall our discussion of reinforcement schedules, you will recognize that surprise tests represent a variable interval schedule and that such schedules

produce a consistent pattern of behavior in humans under certain circumstances. Being a student in a classroom is not one of those circumstances. Surprise tests produce an undesirable level of anxiety in many students and cause others to simply give up. If you sense that students are not sufficiently motivated to read and study more consistently, consult Chapter 11 for better ideas on how to accomplish this goal.

6. Keeping the nature and content of the test a secret. Many teachers scrupulously avoid giving students any meaningful information about the type of questions that will be on a test or the material that test items will cover. The assumption that underlies this practice is that if students have been paying attention in class, diligently doing their homework, and studying at regular intervals, they will do just fine on a test. But they usually don't—and the main reason can be seen in our description of learning strategies (see Chapter 9). A good learning strategist first analyzes all of the available information that bears on attaining a goal. But if certain critical information about the goal is not available, the rest of the strategy (planning, implementing, monitoring, and modifying) will suffer.
7. Keeping the criteria for assignments a secret. This practice is closely related to the previous one. Students may be told, for example, to write an essay on what the world would be like if all diseases were eliminated, and to give their imagination free rein in order to come up with many original ideas. But when the papers are graded, equal weight is given to spelling, punctuation, and grammatical usage. If these aspects of writing are also important to you and you intend to hold students accountable for them, make sure you clearly communicate that fact.
8. Shifting criteria. Teachers are sometimes disappointed in the quality of students' tests and assignments and decide to change the grading criteria as a way to shock students into more appropriate learning behaviors. For example, a teacher may have told students that mechanics will count for one third of the grade on a writing assignment. But when the teacher discovers that most of the papers contain numerous spelling, punctuation, and grammatical errors, she may decide to let mechanics determine half of the grade. As we indicated before, grades should not be used as a motivational device or as a way to make up for instructional oversights. There are far better ways to accomplish these goals.
9. Combining apples and oranges. Students' grades are supposed to indicate how much they have learned in different subject matter areas. When

factors such as effort and ability are combined with test scores, the meaning of a grade becomes unclear. Consequently, measurement experts routinely recommend that teachers base students' grades solely on how well they have performed on written tests and performance assessments. Assessments of effort and ability should be reported separately (Gronlund & Waugh, 2009). Nevertheless, many teachers do not follow this recommendation. A survey (McMillan, Myran, & Workman, 2002) of just over 900 third through fifth grade teachers revealed that 36% factored a student's level of effort into a grade either quite a bit, extensively, or completely and that 47% used a student's ability level to help determine a grade either quite a bit, extensively, or completely. Many high school teachers engage in this practice as well.

A number of technological formats and products have been developed to make the task of classroom assessment easier, more informative, and less prone to error. The next section describes several of these formats and products.

LO4 Technology for Classroom Assessment

Assessment activities can account for about one third of a teacher's time. This large investment in time is partly due to the importance of assessment in both teaching and learning, but it is also related to the fact that many assessment activities involve time-consuming methods of creating, administering, and scoring tests and analyzing and recording scores. Fortunately, computer-based technology supports many of the assessment functions teachers must execute (Beatty & Gerace, 2009; Cardwell, 2000). As the technological tools for learning and assessment have grown, so have the expectations that aspiring teachers will demonstrate their capacity to use the technology (Yao, 2006).

Electronic Grade Books and Grading Programs

Electronic grade books can store records of student test performance, compute test averages and cumulative averages, weight scores, note students with particular scores or characteristics, and print grade reports with standard as well as specific student comments. Combining digital grade books with grading software allows teachers be consistent with the point-based grading systems that are used by most middle and high school teachers. These programs can scan and mark students' choices to selected-response test items (true-false, matching, multiple choice) and allow teachers to track, summarize, and present student performance in a variety of ways. But the efficiency and seeming objectivity of such programs mask a serious potential drawback: They can lead to unfair assignment of grades when used uncritically. The challenge of accurately assigning grades usually involves more than just mathematical precision.

To demonstrate the complex nature of grading and the need for professional judgment to supplement the use of computerized grading programs, consider the example offered by Thomas Guskey (2002) in Table 14.3.

The table represents a group of seven students, each of whom has been graded using three methods: calculating the simple average of all scores, calculating the median or middle score, and calculating the average with the lowest score deleted. Using the simple arithmetic average produces a grade of C for all students despite the differences in their grade patterns. Student 1, for example, started slowly but gradually improved. Student 2 exhibited the opposite pattern. Student 3's performance was consistently around the average. Student 4 failed the first two unit tests but scored near or at the top for the last three. Student 5 exhibited the opposite pattern from student 4. Student 6 had an unexcused absence for the first test and was given a

TABLE 14.3 Summary Grades Tallied by Three Different Methods

Student	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	Avg Score	Grade	Median Score	Grade	Avg With Lowest Deleted	Grade
1	59	69	79	89	99	79	C	79	C	84	B
2	99	89	79	69	59	79	C	79	C	84	B
3	77	80	80	78	80	79	C	80	B	79.5	B
4	49	49	98	99	100	79	C	98	A	86.5	B
5	100	99	98	49	49	79	C	98	A	86.5	B
6	0	98	98	99	100	79	C	98	A	98.8	A
7	100	99	98	98	0	79	C	98	A	98.8	A

Grading Scale: 90–100% = A; 80–89% = B; 70–79% = C; 60–69% = D; 59% or lower = F.
SOURCE: Guskey (2002).

score of 0 but scored near or at the top for the last four tests. Student 7 had virtually perfect scores for the first four tests but was caught cheating on the last one and received a score of 0. If giving all seven students the same grade strikes you as inappropriate, note that using the median score or the average with the lowest score deleted produces grades that range from A through C, and that students 4 and 5 could receive an A, B, or C, depending on which method is used.

Our purpose here is not to tell you which of these methods to use, as that will depend on other information that teachers typically have about student capabilities and their own beliefs about the appropriateness of different grading methods, but to remind you that computerized grading books should not be allowed to substitute for your professional judgment in awarding grades.

Technology-Based Performance Assessment

As you may recall from our earlier discussion, performance assessments give students the opportunity to demonstrate how well they can use the knowledge and skills that were the focus of an instructional unit to carry out realistic and meaningful tasks. Computer-based technology is an excellent vehicle for this purpose. For example, simulations are likely to be more effective than traditional paper-and-pencil tests for determining how well students understand and can carry out the process of scientific inquiry (planning an investigation, collecting data, organizing and analyzing the data, forming conclusions, and communicating findings). Recent research on the development of performance assessment procedures in complex task domains such as engineering design and medical diagnosis promises even greater effectiveness of technology-based performances in the future (Spector, 2006). A web-based simulation that lends itself to the assessment of scientific inquiry is the GLOBE environmental science education program

(www.globe.gov). Students who participate in GLOBE collect environmental data at a local site and submit it to a scientific database on the web. About 4,000 schools from countries around the world participate in this program. Teachers could use the GLOBE database to assess how well students analyze and interpret climate data by having them use a set of climate-related criteria (such as temperature at different altitudes, amount of sunshine, and amount of snow) to determine in which of several cities the next Winter Olympics should be held (Means & Haertel, 2002). Multimedia tools with text, audio, video, and graphics, such as the Quest Atlantis program mentioned in other chapters, also offer opportunities for students to demonstrate their ability to solve real-world problems in a number of content areas.

Digital Portfolios

What Is a Digital Portfolio? Digital portfolios (also called *electronic portfolios*) are similar in purpose to the more traditional portfolios, but they extend beyond paper versions because they can include sound effects, audio and video testimonials, voice-over explanations of a student's thinking process as a project is worked on, and photographs of such products as drawings, paintings, and musical compositions (Siegle, 2002).

The Components and Contents of Digital Portfolios

Because the purposes for having students construct a digital portfolio (such as to assign grades, assess students' strengths and weaknesses, evaluate a program or curriculum) are not always the same, the portfolio structure will vary somewhat across teachers and school districts. But some components, such as those in the list that follows, are frequently recommended (for instance, Beatty & Gerace, 2009; Goldsby & Fazal, 2001; Janesick, 2001) and should always be seriously considered for inclusion:

- The goals the student was attempting to achieve
- The guidelines that were used to select the included material
 - Work samples
 - Teacher feedback
 - Student self-reflection
- The criteria that were used to evaluate each entry
- Examples of good quality work



© basilemar23/Shutterstock Images

Just as the general components of a digital portfolio may vary, so may the particular media that are used. Following are some specific examples of the types of media a student may use and information that would be represented by each medium (Barrett, 2000; Gatlin & Jacob, 2002; Janesick, 2001; O'Lone, 1997; Siegle, 2002):

- *Digitized pictures and scanned images:* Photos of the student or objects the student has created, artwork, models, science experiments over time, fax exchanges with scientists, spelling tests, math work, self-assessment checklists.
- *Documents:* Electronic copies of student writing, reflection journals, publications, copies of web pages created, teacher notes and observations.
- *Audio recordings:* Persuasive speeches, poetry readings, rehearsals of foreign language vocabulary, readings of select passages, self-evaluations, interviews or voice notes regarding the rationale for work included.
- *Video clips:* Short videos showing the student or teams of students engaged in science experiments and explaining their steps, showing student performances in physical education or the performing arts, or students generating and/or presenting interdisciplinary projects.

Rubrics for Digital Portfolios and Presentations With all the information a digital portfolio might contain, how can a classroom teacher fairly and efficiently assess student learning? First, electronic writing, just like paper-based compositions, can be assessed holistically in a general impression rating. It can also be analyzed with specific criteria such as whether the work is insightful, well organized, clear, focused, relevant, sequentially flowing, persuasive, inspirational, and original (see Arter & Chappuis, 2008). There are also rubrics for analyzing the quality of a portfolio that has been posted to a website. One uses a 4-point scale (exceeds requirements, meets requirements, close to meeting standards, clearly

does not meet standards) to assess the design and aesthetics of the website, its usability, and the presence and clarity of the portfolio's contents (Goldsby & Fazal, 2001). The website 4teachers (www.4teachers.org) contains, among other things related to technology, a tool called RubiStar that provides templates for creating rubrics for several types of digital products.

Performance and Portfolio Assessment Problems

We would be remiss not to point out the problems often associated with technology-based performance and portfolio assessment. High quality performance assessments require multiple assessments (for both formative and summative purposes), extensive time, electronic equipment, careful planning, and continued modification (McGrath, 2003). Electronic portfolios can become large, complex, and time consuming to grade fairly and thus can overload teachers with work (A. Goodman, 2008; Pope, Hare, & Howard, 2002).

Staff development and teacher training are additional barriers to effective use of performance assessment and digital portfolios. But with proper training, teachers may begin to find ways in which technology-based school and classroom assessment plans are practical, cost effective, and qualitatively better than traditional assessments.

The Suggestions for Teaching that follow should help you properly implement the assessment concepts and research findings presented in this chapter.



© vgstudio/Shutterstock Images

Suggestions for Teaching

Communicating Assessment and Evaluative Judgments

Teaching is a profession; its practitioners are professionals. As professionals, teachers are called on to make decisions, take actions based on those decisions, and justify both their decisions and their actions. Another way of saying it is that teachers, because they are professionals, are accountable. High on the list of decisions and actions for which they must account are assessment and evaluation.

In the first chapter of this book, we mentioned the National Board for Professional Teaching Standards, and we will see the five key propositions underlying those standards in the last chapter. To become a National Board Certified Teacher, one must account—clearly and convincingly—for her or his teaching practice, including the proposition that “teachers are responsible for managing and monitoring student learning.”

Lee S. Shulman, who was instrumental in creating the National Board for Professional Teaching Standards, is president emeritus of the Carnegie Foundation for the Advancement of Teaching. In 2007, Dr. Shulman published an essay entitled “Counting and Recounting: Assessment and the Quest for Accountability.” He opened that essay by describing how his daughter had just returned from her first MBA class in managerial accounting and he had asked her how the first class went. He described his reaction as follows:

Imagine my surprise when [my daughter] responded that accounting was unexpectedly interesting because, she now realized, it should be understood as a form of narrative, a kind of drama. Within the ethical and technical rules of the field, the task of the accountant is to figure out . . . the stories [that] should be told. Accounting is basically about creating the plot, characters, and setting of the story. As the instructor explained to the class, “Your task is to render an account: to tell the facts of the case, the story . . . in an accurate and yet ultimately persuasive way.” (p. 20)

Telling a story, rendering an account, requires the author of the story to communicate clearly in order to be persuasive. As a teacher, engaged in the practice of assessment and evaluation, you will author many stories of learning. Here are some sugges-

tions that may help you render informative, useful, and convincing accounts of student learning. (The suggestions are adapted from a section in Shulman’s essay headed “Seven Pillars of Assessment and Accountability.”)

1 Be clear about the learning story you need to tell and why it needs to be told.

Chapter 13 began by examining learning goals, instructional objectives, and taxonomies of learning outcomes. The idea in planning instruction is to *begin with the end in mind*, to determine what your students need to learn and why.

a. Use taxonomies.

What are the learning goals for your students? What learning standards are they supposed to meet? What are your goals for them? Use the taxonomies discussed in Chapter 13 to think about the standards that your district and state will expect your students to meet. The taxonomies can help you move from general education standards to more specific learning goals—and help you justify why those learning goals are important for your students to meet. Once you have identified the kinds of cognitive, affective, and psychomotor outcomes your students should demonstrate, you can move to generating the specific instructional objectives that your students need to attain and that will guide the learning activities you employ in your classroom.

b. Write clear objectives.

Use either Mager’s or Gronlund’s approach (see Chapter 13) to specify the objectives your students need to meet. Either approach—correctly applied—yields clear, specific statements of what your students should know and be able to do.

Specifying objectives will not tell you how your story ends—that comes later—but with clear, specific instructional objectives in place, you will know how a successful story of learning *should* end.

2 Be clear about what you are assessing and what you are not assessing.

With your objectives for a unit of study specified, you are now ready to start building the assessments that will help you tell the learning story of each student. To make sure that your story is true and honest, you need to be clear about what you are and are not assessing.



© Bob Daemmerich/Photo Edit

For students to plan effectively how they will master your objectives, they need to know as early as possible how many test they will have to take, when the tests will occur, what types of items each test will contain, and what content they will be tested on.

One tool that can be very helpful in this regard is a *table of specifications*. Such a table specifies what objectives are being assessed and how those objectives contribute to the various taxonomic outcomes you are seeking for your students.

Take a look at the example in Figure 14.3. It identifies the major topics of this chapter and the taxonomic outcomes in the cognitive domain. The objectives can be organized by the major topics of the unit of study and the types of cognitive capabilities that students might demonstrate.

This practice forces you to think about both the number and the relative importance of your objectives before you start teaching or writing test items. Thus, if some objectives are more important to you than others, you will have a way of ensuring that these are tested more thoroughly. If, however, a test is going to be brief and emphasize all objectives more or less equally, you may wish to put a check mark in each box as you write

FIGURE 14.3 Example of a Table of Specifications for Material Covered in This Chapter

TOPIC	OBJECTIVES					
	Knows	Comprehends	Applies	Analyzes	Synthesizes	Evaluates
Nature of measurement and evaluation						
Purposes of measurement and evaluation						
Types of written tests						
Nature of performance tests						

Photo © Brian Weed/Shutterstock Images

questions. The important point is that by taking steps to ensure that your tests cover what you want your students to know, you will be increasing the tests' validity.

In preparing assessments for your own students, you may choose to use outcomes from the affective or psychomotor domains or you may choose not to list all of the categories in the taxonomy for all subjects or at all grade levels.

3 Use multiple assessments to tell the story of student learning.

Suppose you asked one multiple-choice question on one test to determine if your students had attained a particular objective. Suppose further that some of your students got that question correct and some did not. In telling the story of a student who got the item correct, how confident would you be in claiming that the student had learned? What about the story of one who missed the item?

Our suggestion here is that you assess the objectives you have specified a number of times, using a variety of the methods of assessment. If you found that students do well on several multiple-choice items, on

several short-answer items, on an extended essay, and on a presentation—all assessing a single objective—you would be able to make a more convincing case that those students had sufficiently demonstrated the learning required to meet the objective. The stories of learning—whether they are stories of success or of postponed completion—must be convincing if they are to be used.

4 Synthesize the information from across multiple measures.

For the same reason that it is important to use multiple assessments to collect information about what students know and are able to do, it is important that the story of each student's learning be based on the evidence of learning and that the assessments and evidence be synthesized. Simply listing test results, scores on homework, and grades on term papers provides the raw data gathered across assessments, but a list of scores does not qualify as evidence. Evidence is what is used to support a claim that a student has or has not learned. Remember that justifying your assessment and evaluation practices requires you to render an account. The

accountant for a large company does not simply list assets and liabilities; an account based on the available information is rendered. Following this advice becomes very important when communicating with parents.

5 Avoid creating "high stakes" assessment when possible.

In our current era of accountability in education, testing—especially standardized testing of the type discussed in the next chapter—has created an environment of high stakes. Think back to the emphasis placed on doing well on the state tests of achievement when you were in high school. Reflect also on what we know about academic cheating behavior: If the stakes are high enough, cheating will occur. This can happen in classrooms as well. Imagine that you took a class that had only one final exam, that your entire grade for the course rested on your performance at a particular time, on a particular day, in a particular place.

In order to gather accurate information about your students' learning, so that you can make sound evaluative judgments, do not create assessments that have high stakes.

continued

Suggestions for Teaching—*continued*

Rather than give one unit exam, give several or—following the mastery approach to assessment—provide multiple opportunities for students to demonstrate on each assessment what they know and are able to do. The next suggestion can also help avoid creating high stakes environments.

6 Make sure assessment is embedded in instruction.

In order to provide a useful and compelling account of a student's learning, you need to know not just where the student ended up but how the student got there. The story comes from the journey. We offer this suggestion not only because it will help you communicate effectively and help you justify your assessment and evaluation decisions, but because it will help you teach more effectively and will increase the probability that your students' story of learning will end where it should.

This suggestion boils down to following the guidelines for formative assessment (see Table 14.1). By assessing consistently and persistently, you align instruction with your objectives and with data about those objectives, you generate multiple measures, you provide more sources of information to help you synthesize across those measures, and you minimize high stakes environments. Embedding assessment in instruction will facilitate following the five previous suggestions.

7 Actively share and collaborate on the design and development of assessment and evaluation strategies.

Assessment and evaluation are integral to teaching; they are critical teaching functions. As promulgated by the National Board for Professional Teaching Standards, assessment and evaluation must not be treated as afterthoughts in teaching.

One way to develop your professional practice on a trajectory toward teaching expertise is to share and work with others on designing, using, and improving classroom assessments and evaluative strategies. If you develop a quiz that you intend to use formatively, share it with a colleague who teaches the same class, invite her or his critique, and offer to compare results (while maintaining appropriate confidentiality). Sharing your assessment and evaluation practices and experiences is one way to develop your professional knowledge and your skills as an author of learning accounts. Highly qualified professionals are seen as authorities on their practice. Developing your skills as an author of assessments will help establish your authority.

Having considered the Suggestions for Teaching, we close this chapter with one last thought and one last suggestion.

Aspiring teachers are, understandably, excited and eager to present what they know to their students. Aspiring teachers are enthused to find and implement classroom activities that capture students' attention, that engage students deeply, that motivate them, and that leave them thinking. But it is important to remember that teaching is more than presenting what one knows to others. As we have learned in this chapter, assessment is not something that is done after one has taught. Rather, assessment

is integral to teaching and to learning: Recall our discussions of assessment *of* learning, assessment *for* learning, and assessment *as* learning.

Our suggestion is that including your students in the assessment of their own learning captures their attention, engages them deeply in what and how they are learning, increases motivation, and helps them think and view themselves as capable of exerting influence on their own learning. Our point is this: Of all the outcomes of learning that can be assessed, the one that might have the greatest long-term impact on your students is that they are capable of assessing and influencing their own learning.



**WHAT
ELSE? RIP
& REVIEW
CARDS IN
THE BACK**