What Would You Do?

Teachers' Casebook

Your school requires that you give letter grades to your class.

You can use any method you want, as long as an A, B, C, D, or F appears for each of the subject areas on every student's report card, every grading period. Some teachers are using worksheets, quizzes, homework, and tests. Others are assigning group work and portfolios. A few teachers are individualizing standards by grading on progress and effort more than final achievement. Some are trying contract approaches and experimenting with longer-term projects, while others are relying almost completely on daily class work. Two teachers who use group work are considering giving credit toward grades for being a "good group member" or competitive bonus points for the top-scoring group. Others are planning to use improvement points for class rewards, but not for grades. Your only experience with grading was using written comments and a mastery approach that rated the students as making satisfactory or unsatisfactory progress toward particular objectives. You want a system that is fair and manageable, but also encourages learning, not just performance. And you want some system that gives the students feedback they can use to prepare for the proficiency tests required by NCLB.

Critical Thinking

- What would be your major graded assignments and projects?
- Would you include credit for behaviors such as group participation or effort?
- How would you put all the elements together to determine a grade for every student for every marking period?
- How would you justify your system to the principal and to the parents?
- How will these issues affect the grades you will teach?

Collaboration

With 2 or 3 other members of your class, develop a section of a class handbook that describes your grading policy. Be prepared to defend the policy.

As you read this chapter, you will examine both tests and grades, focusing not only on the effects they are likely to have on students, but also on practical ways to develop more efficient methods for testing and grading.

We begin with a consideration of the many types of tests teachers prepare each year and approaches to assessment that don't rely on traditional testing. Then, we examine the effects grades are likely to have on students. Because there are so many grading systems, we also spend some time identifying the advantages and disadvantages of one system over another. Finally, we turn to the very important topic of communication with students and parents. How will you justify the grades you give?

By the time you have completed this chapter, you should be able to answer these questions:

- How will you test students on a unit of work?
- How can you evaluate tests that accompany textbooks and teachers' manuals?
- How should you create multiple-choice and essay test items for your subject area?
- Will you use authentic assessment approaches, including portfolios, performances, exhibitions, and scoring rubrics?
- What are the potential positive and negative effects of grades on students?
- What are examples of criterion-referenced and norm-referenced grading systems?
- How will you explain your grading system to parents who do not understand their child's grades?
Formative and Summative Assessment

As a teacher, you may have little to say about the grading system for your school. Many school districts have a standard approach to grading. Still, you will have choices about how you use your district’s grading system and how you assess your students’ learning. Will you give tests? How many? What kinds? Will students do projects or keep portfolios of their work? How will homework influence grades? Will you use journals, and if so, will you “grade” them?

There are two general uses or functions for assessment: formative and summative. Formative assessment occurs before or during instruction. The purposes of formative assessment are to guide the teacher in planning and to help students identify areas that need work. In other words, formative assessment helps form instruction. Often students are given a formative test prior to instruction, a pretest that helps the teacher determine what students already know. Sometimes a test is given during instruction to see what areas of weakness remain so teaching can be directed toward the problem areas. This is generally called a diagnostic test, but should not be confused with the standardized diagnostic tests of more general learning abilities discussed in the previous chapter. A classroom diagnostic test identifies a student’s areas of achievement and weakness in a particular subject. Older students are often able to apply the information from diagnostic tests to “retro” themselves. Pretests and diagnostic tests are not graded. And because formative tests do not count toward the final grade, students who tend to be very anxious about “real” tests may find this low-pressure practice in test taking especially helpful.

Summative assessment occurs at the end of instruction. Its purpose is to let the teacher and the students know the level of accomplishment attained. Summative assessment, therefore, provides a summary of accomplishment. The final exam is a classic example.

The distinction between formative and summative assessment is based on how the results are used. Any kind of assessment—traditional, performance, project, oral, portfolio, and so on—can be used for either formative or summative purposes. If the goal is to obtain information about student learning for planning purposes, the assessment is formative. If the purpose is to determine final achievement (and help determine a course grade), the assessment is summative. In fact, the same assessment could be used as...
TABLE 15.1

Using Tests to Make Instructional Decisions

The best use of assessment is to plan, provide, and target instruction. Here are some decisions that can benefit from assessment results:

<table>
<thead>
<tr>
<th>Decision Category</th>
<th>Typical Assessment Strategy</th>
<th>Decision Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>What to teach in the first place?</td>
<td>Preassessment before instruction</td>
<td>Whether to provide instruction for specific objectives?</td>
</tr>
<tr>
<td>How long to keep teaching toward a particular instructional objective?</td>
<td>En route assessments of students' progress</td>
<td>Whether to continue or cease instruction for an objective either for an individual or for the whole class?</td>
</tr>
<tr>
<td>How effective was an instructional sequence?</td>
<td>Comparing students' posttest to pretest performances</td>
<td>Whether to retain, discard, or modify a given instructional sequence the next time it's used?</td>
</tr>
</tbody>
</table>

Source: From Classroom Assessment: What Teachers Need to Know, 4th ed. by J. W. Popham. Published by Allyn and Bacon, Boston, MA. Copyright © 2005 by Pearson Education. Adapted by permission of the publisher.

A formative evaluation at the beginning of the unit and as a summative evaluation at the end.

The formative uses of assessment are really the most important in teaching. James Popham (2005a) believes that “Unfortunately, too many teachers still think that the chief role of classroom tests is the assignment of grades. But any teacher who uses tests dominantly to determine whether students get high or low grades should receive a solid F in classroom assessment” (p. 262). Tests and all assessments should be used to help teachers make better instructional decisions. Table 15.1 gives some examples.

Getting the Most from Traditional Assessment

STOP THINK WRITE Think back to your most recent test. What was the format? Did you feel that the test results were an accurate reflection of your knowledge or skills? Have you ever had to design a test? What makes a good, fair test?

When most people think of assessment, they usually think of testing. As you will see shortly, teachers today have many other options, but testing is still a significant activity in most classrooms. Let’s consider your options for assessing students using the traditional testing approach. In this section we will examine how to plan effective tests, evaluate the tests that accompany standard curriculum materials, and write your own test questions.

Planning for Testing

One result of high-stakes testing programs in many districts is that teachers now have more classroom assessments and reviews focused on the standards behind the high-stakes tests. These assessments need to be well organized and planned. When you have a good plan, you are in a better position to judge the tests provided in teachers’ manuals and to develop class assessments yourself.

When to Test? Frank Dempster (1991) examined the research on reviews and tests and reached these useful conclusions for teachers:

Connect and Extend to PRAXIS II

Types of Assessment (II.C.1) Understand the purposes of formative and summative assessment. Explain how teachers and students can make effective use of the information generated by each type of test.
1. Frequent testing encourages the retention of information and appears to be more effective than a comparable amount of time spent reviewing and studying the material.

2. Tests are especially effective in promoting learning if you give students a test on the material soon after they learn it, then retest on the material later. The retests should be spaced farther and farther apart.

3. The use of cumulative questions on tests is a key to effective learning. Cumulative questions ask students to apply information learned in previous units to solve a new problem.

Dempster argues that students will learn more if the curriculum includes fewer topics, but explores those topics in greater depth and allows more time for review, practice, testing, and feedback (Dempster, 1993).

Judging Textbook Tests. Most elementary and secondary school tests today come complete with supplemental materials such as teaching manuals, handout masters, and ready-made tests. Using these tests can save time, but it is good teaching practice. The answer depends on your objectives for your students, the way you teach the material, and the quality of the tests provided. If the textbook test is of high quality, matches your testing plan, and fits the instruction you actually provided for your students, then it may be the right test to use. Check the reading level of the items provided and be prepared to revise/improve them (Airasian, 2002; McMillan, 2004). Table 15.2 gives key points to consider in evaluating textbook tests.

What if there are no tests available for the material you want to cover, or the tests provided in the textbook are not appropriate for your students? Then it’s time for you to create your own tests. We will consider the two major kinds of traditional tests—objective and essay.

Objective Testing

Multiple-choice questions, matching exercises, true/false statements, and short-answer or fill-in items are all types of objective testing. The word “objective” in relation to testing

| TABLE 15.2 |

| Key Points to Consider in Judging Textbook Tests |

1. The decision to use a textbook test or pre-made standard achievement test must come after a teacher identifies the objective that he or she has taught and now wants to assess.

2. Textbooks and standard tests are designed for the typical classroom, but since few classrooms are typical, most teachers deviate somewhat from the text in order to accommodate their pupils’ needs.

3. The more classroom instruction deviates from the textbook, the less valid the textbook tests are likely to be.

4. The main consideration in judging the adequacy of a textbook or standard achievement test is the match between its test questions and what pupils were taught in their classes.
   a. Are questions similar to the teacher’s objectives and instructional emphases?
   b. Do questions require pupils to perform the behaviors they were taught?
   c. Do questions cover all or most of the important objectives taught?
   d. Is the language level and terminology appropriate for pupils?
   e. Does the number of items for each objective provide a sufficient sample of pupil performance.


Objective testing Multiple-choice, matching, true/false, short-answer, and fill-in tests: scoring answers does not require interpretation.
means "not open to many interpretations," or "not subjective." The scoring of these types of items is relatively straightforward compared to the scoring of essay questions because the answers are more cut-and-dry and likely yield more consistent scores. How should you decide which item format is best for a particular test? Use the one that provides the most direct measure of the learning outcome you intended for your students (Gronlund, 2003). In other words, if you want to see how well students can write a letter, have them write a letter, don't ask multiple-choice questions about letters. But if many different item formats will work equally well, then use multiple-choice questions because they are easier to score fairly and can cover many topics. Switch to other formats if writing good multiple-choice items for the material is not possible or appropriate. For example, if related concepts such as terms and definitions need to be linked, then a matching item is a better format than multiple-choice. If it is difficult to come up with several wrong answers for a multiple-choice item, try a true/false question instead. Alternatively, ask the student to supply a short answer that completes a statement (fill-in-the-blank). Variety in objective testing can lower students' anxiety because the entire grade does not depend on one type of question that a particular student may find difficult. We will look closely at the multiple-choice format because it is the most versatile—and the most difficult to use well.

Using Multiple-Choice Tests. Even though about three-fourths of education professors reject the use of multiple-choice tests in determining students' grades, about half of public school teachers endorse these tests (Banks, 2005), so you should know how to use these tests well. In fact, many schools require teachers to give students experience answering multiple-choice tests in order to prepare them for high-stakes testing (McMillan, 2004). Of course, multiple-choice items can test facts, but these items can assess more than recall and recognition if they require the student to deal with new material by applying or analyzing the concept or principle being tested (Gronlund, 2003; McMillan, 2004). For example, the following multiple-choice item is designed to assess students' ability to recognize unstated assumptions, one of the skills involved in analyzing an idea:

An educational psychology professor states, "A z score of +1 on a test is equivalent to a percentile rank of approximately 84." Which of the following assumptions is the professor making?

1. The scores on the test range from 0 to 100.
2. The standard deviation of the test scores is equal to 3.4.
3. The distribution of scores on the test is normal. (Correct answer)
4. The test is valid and reliable.

Writing Multiple-Choice Questions. All test items require skillful construction, but good multiple-choice items are a real challenge. Some students jokingly refer to multiple-choice tests as "multiple-guess" tests—a sign that these tests are often poorly designed. Your goal in writing test items is to design them so that they measure student achievement, not test-taking and guessing skills. The stem of a multiple-choice item is the part that asks the question or poses the problem. The choices that follow are called alternatives. The wrong answers are called distractors because their purpose is to distract students who have only a partial understanding of the material. If there were no good distractors, students with only a vague understanding would have no difficulty in finding the right answer.

The Guidelines on the next page, adapted from Gronlund (2003), Popham (2005a), and Smith, Smith, & De Laist (2001), should make writing multiple-choice and other objective test questions easier.
Guidelines: Writing Multiple-Choice Test Items

The stem should be clear and simple, and present only a single problem. Unessential details should be left out.

Poor
There are several different kinds of standard or derived scores. An IQ score is especially useful because . . .

Better
Which of the following is an advantage of an IQ score?

The problem in the stem should be stated in positive terms. Negative language is confusing. If you must use words such as not, no, or except, underline them or type them in all capitals.

Poor
Which of the following is not a standard score?

Better
Which of the following is NOT a standard score?

Do not expect students to make extremely fine discrimination among answer choices.

What is the percentage of area in a normal curve falling between +1 and −1 standard deviations?

Poor
a. 66% b. 67% c. 68% d. 69%.

Better
a. 14% b. 34% c. 68% d. 95%.

As much wording as possible should be included in the stem so that phrases will not have to be repeated in each alternative.

Poor
A percentile score
a. indicates the percentage of items answered correctly.
b. indicates the percentage of correct answers divided by the percentage of wrong answers.
c. indicates the percentage of people who scored at or above a given raw score.
d. indicates the percentage of people who scored at or below a given raw score.

Better
A percentile score indicates the percentage of
a. items answered correctly.
b. correct answers divided by the percentage of wrong answers.
c. people who scored at or above a given raw score.
d. people who scored at or below a given raw score.

Each alternative answer should fit the grammatical form of the stem, so that no answers are obviously wrong.

Poor
The Stanford-Binet test yields an
a. IQ score. c. vocational preference.
b. reading level. d. mechanical aptitude.

Better
What does the Stanford-Binet assess?
(a) intelligence. c. vocational preference
(b) reading level d. mechanical aptitude

Categorical words such as always, all, only, or never should be avoided unless they can appear consistently in all the alternatives. Most smart test makers know that categorical answers are usually wrong.

Poor
A student's true score on a standardized test is
a. never equal to the obtained score.
b. always very close to the obtained score.
c. always determined by the standard error of measurement.
d. usually within a band that extends from +1 to −1 standard errors of measurement on each side of the obtained score.

Better
Which one of the statements below would most often be correct about a student's true score on a standardized test?
a. It equals the obtained score.
b. It will be very close to the obtained score.
c. It is determined by the standard error of measurement.
d. It could be above or below the obtained score.

Avoid including two distractors that have the same meaning. If only one answer can be right and if two answers are the same, then these two must both be wrong. This narrows down the choices considerably.

What is the term for the most frequently occurring score in a distribution?

Poor
a. mode b. median c. arithmetic average d. mean

Better
(a) mode b. median c. standard deviation d. mean

Avoid using the exact wording found in the textbook.

Poor students may recognize the answers without knowing what they mean.

Avoid overuse of all of the above and none of the above.
Such choices may be helpful to students who are guessing. Using all of the above may trick a quick student who sees that the first alternative is correct and does not read on.

Obvious patterns on a test also aid students who are guessing.
The position of the correct answer should be varied, as should its length.

Incorrect alternatives should reflect common student misunderstandings.
Be careful, however, that the wrong alternatives are not so obviously wrong that students don’t even consider them.

For a good summary on writing test items, see http://www.msu.edu/ depts/ome/website/Chapter15.html
Essay Testing

The best way to measure some learning objectives is to require students to create answers on their own. An essay question is appropriate in these cases. The most difficult part of essay testing is judging the quality of the answers, but writing good, clear questions is not particularly easy, either. We will look at writing, administering, and grading essay tests, with most of the specific suggestions taken from Gronlund (2003). We will also consider factors that can bias the scoring of essay questions and ways you can overcome those problems.

STOP | THINK | WRITE  Evaluate these two essay questions from Popham (2005a, pp. 172-173):

1. (High school level) You have just viewed a videotape containing three widely seen television commercials. What is the one classic propaganda technique present in all three commercials?

2. (Middle school level) Thinking back over the mathematics lesson and homework assignments you had during the past 12 weeks, what conclusions can you draw? Take no more than one page for your response. |

Constructing Essay Tests. Because answering takes time, true essay tests cover less material than objective tests. Thus, for efficiency, essay tests should be limited to the assessment of more complex learning outcomes.

An essay question should give students a clear and precise task and should indicate the elements to be covered in the answer. (Are the questions above clear and precise?) The students should know how extensive their answer should be and about how much time they should spend on each question. Question 2 above gives a page limit, but would you know what is being asked?

Students should be given ample time for answering. If more than one essay is being completed in the same class period, you may want to suggest time limits for each question. Remember, however, that time pressure increases anxiety and may prevent accurate assessment of some students. Whatever your approach, do not try to make up for the limited amount of material an essay test can cover by including a large number of essay questions. It would be better to plan on more frequent testing than to include more than two or three essay questions in a single class period. Combining an essay question with a number of objective items is one way to avoid the problem of limited sampling of course material (Gronlund, 2003).

Evaluating Essays: Dangers. In 1912, Starch and Elliot began a classic series of experiments that shocked educators into critical consideration of subjectivity in testing. These researchers wanted to find out the extent to which teachers were influenced by personal values, standards, and expectations in scoring essay tests. For their initial study, they sent copies of English examination papers written by two high-school students to English teachers in 200 high schools. Each teacher was asked to score the papers according to his or her school's standards. A percentage scale was to be used, with 75% as a passing grade.

The results? Neatness, spelling, punctuation, and communicative effectiveness were all valued to different degrees by different teachers. The scores on one of the papers ranged from 64% to 98%, with a mean of 88.2. The average score for the other paper was 80.2, with a range between 50% and 97%. The following year, Starch and Elliot (1913a, 1913b) published similar findings in a study involving history and geometry papers. The most important result of these studies was the discovery that the problem of subjectivity in grading was not confined to any particular academic area. The main difficulties were the individual standards of the grader and the unreliability of scoring procedures.

Evaluating Essays: Methods. Gronlund (2003) offers several strategies for grading essays that avoid problems of subjectivity and inaccuracy. When possible, a good first step

Connect and Extend to Your Teaching/Portfolio

Some students prefer essay tests because they can write down something even if it doesn't answer the question, and receive at least partial credit for the question. Do you think that granting partial credit is a common grading practice among teachers? Considering that essay questions sample only a limited amount of material, does the practice seem commendable or defensible?
is to construct a set of scoring criteria or a rubric (more on this later) and share it with students. Even when students are given some choice in testing, teachers can decide what type of information should be in any answer. Here is an example from TenBrink (2003, p. 326).

**Question**: Defend or refute the following statement: Civil wars are necessary to the growth of a developing country. Cite reasons for your argument, and use examples from history to help substantiate your claim.

**Scoring Rubric**: All answers, regardless of the position taken, should include (1) a clear statement of the position, (2) at least five logical reasons, (3) at least four examples from history that clearly substantiate the reasons given.

Once you have set your expectations for answers, you can assign points to the various parts of the essay. You might also give points for the organization of the answer and the internal consistency of the essay. You can then assign grades such as 1 to 5 or A, B, C, D, and F, and sort the papers into piles by grade. As a final step, skim the papers in each pile to see if they are comparable in quality. These techniques will help ensure fairness and accuracy in grading.

When grading essay tests with several questions, it makes sense to grade all responses to one question before moving on to the next. This helps prevent the quality of a student’s answer to one question from influencing your reaction to the student’s other answers. After you finish reading and scoring the first question, shuffle the papers so that no students end up having all their questions graded first (when you may be taking more time to give feedback or are applying stricter standards, for example) or last (when you may be tired of writing feedback or more lax in your standards). You may achieve greater objectivity if you ask students to put their names on the back of the paper, so that grading is anonymous. A final check on your fairness as a grader is to have another teacher who is equally familiar with your goals and subject matter grade your tests without knowing what grades you have assigned. This can give you valuable insights into areas of bias in your grading practices.

Now that we have examined both objective and essay testing, we can compare examples of the different approaches. Table 15.3 presents a summary of the advantages and disadvantages of each.

### TABLE 15.3

<table>
<thead>
<tr>
<th>Type</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>True/False</td>
<td>Tests the most facts in shortest time. Easy to score. Tests recognition. Objective.</td>
<td>Difficult to measure complex learning. Difficult to write reliable items. Subject to guessing.</td>
</tr>
<tr>
<td>Matching</td>
<td>Excellent for testing associations and recognition of facts. Although terse, can test complex learning (especially concepts). Objective.</td>
<td>Difficult to write effective items. Subject to process of elimination.</td>
</tr>
<tr>
<td>Multiple choice</td>
<td>Can assess learning at all levels of complexity. Can be highly reliable. Objective. Tests fairly large knowledge base in short time. Easy to score.</td>
<td>Difficult to write. Somewhat subject to guessing.</td>
</tr>
</tbody>
</table>

*Source: From Classroom Teaching Skills, Seventh Edition by James Cooper. Copyright © 2000 by Houghton Mifflin Company. Adapted with permission.*

560 Chapter 15: Classroom Assessment and Grading
Alternatives to Traditional Assessments

We have been considering how to make traditional testing more effective; now, let’s look at a few alternative approaches to classroom assessment. One of the main criticisms of standardized tests—that they control the curriculum, emphasizing recall of facts instead of thinking and problem solving—is a major criticism of classroom tests as well. Few teachers would dispute these criticisms. Even if you follow the guidelines we have been discussing, traditional testing can be limiting. What can be done? Should alternatives in classroom assessment make traditional testing obsolete? This Point/Counterpoint addresses this question.

Point/Counterpoint

Which Is Better—Traditional Tests or Authentic Assessments?

We have seen the advantages and disadvantages of standardized tests, but what about classroom testing? Are traditional multiple-choice and essay tests useful in classroom assessment?

**Point**

Traditional tests are a poor basis for classroom assessment.

In his article “Standards, Not Standardization: Evoking Quality Student Work,” Grant Wiggins (1991) makes a strong case for giving students standards of excellence against which they can judge their accomplishments. But these standards should not be higher scores on multiple-choice tests. When scores on traditional tests become the standard, the message to students is that only right answers matter and the thinking behind the answers is unimportant. Wiggins notes:

> We do not judge Xerox, the Boston Symphony, the Cincinnati Reds, or Dom Perignon vineyards on the basis of indirect, easy-to-test, and common indicators. Nor would the workers in those places likely produce quality if some generic, secure test served as the only measure of their success in meeting a standard. Demanding and getting quality, whether from students or adult workers, means framing standards in terms of the work that we undertake and value. And it means framing expectations about that work which make quality a necessity, not an option. Consider:

  - the English teacher who instructs peer-editors to mark the place in a student paper where they lost interest in it or found it slapdash and to hand it back for revision at that point;
  - the professor who demands that all math homework be turned in with another student having signed off on it, where one earns the grade for one’s work and the grade for the work that each person (willingly) countersigned. (p. 22)

Wiggins continues to argue for assessment that makes sense, that tests knowledge as it is applied in real-world situations. Understanding cannot be measured by tests that ask students to use skills and knowledge out of context. “In other words, we cannot be said to understand something unless we can employ our knowledge wisely, fluently, flexibly, and aptly in particular and diverse contexts.” (Wiggins, 1993, p. 200).

**Counterpoint**

Traditional tests can play an important role.

Most psychologists and educators would agree with Wiggins that setting clear, high, authentic standards is important, but many also believe that traditional tests are useful in this process. Learning may be more than knowing the right answers, but right answers are important. While schooling is about learning to think and solve problems, it is also about knowledge. Students must have something to think about—facts, ideas, concepts, principles, theories, explanations, arguments, images, opinions. Well-designed traditional tests can evaluate students’ knowledge effectively and efficiently (Arasian, 2005; Kirst, 1991b). Some educators believe that traditional testing should play an even greater role than it currently does. Educational policy analysts suggest that American students, compared to students in many other developed countries, lack essential knowledge because American schools emphasize process—critical thinking, self-esteem, problem solving—more than content. In order to teach more about content, teachers will need to determine how well their students are learning the content, and traditional testing provides useful information about content learning.

Tests are also valuable in motivating and guiding student learning. There is research evidence that frequent testing encourages learning and retention. In fact students generally learn more in classes with more rather than fewer tests (Dempster, 1991).

What do you think?

Vote online at www.mylabschool.com

Source: From “Standards, Not Standardization,” by G. Wiggins, Educational Leadership, 48(5), pp. 18–25. Copyright © 1991 by the American Association for Supervision and Curriculum Development. Reprinted with permission from ASCD. All rights reserved. The American Association for Supervision and Curriculum Development is a worldwide community of educators advocating sound policies and sharing best practices to achieve the success of each learner. To learn more, visit ASCD at www.ascd.org.
One solution that has been proposed to solve the testing dilemma is to apply the concept of authentic assessment to classroom testing.

**Authentic Classroom Assessment**

**WHAT WOULD YOU SAY?**

In your interview with the search team in an elementary school known for innovation, one of the teachers asks, "What do you know about using portfolios, performances, projects, and rubrics to assess learning?"

Authentic assessments ask students to apply skills and abilities as they would in real life. For example, they might use fractions to enlarge or reduce recipes. Grant Wiggins made this argument over 15 years ago:

If tests determine what teachers actually teach and what students will study for—and they do—then the road to reform is a straight but steep one: test those capabilities and habits we think are essential, and test them in context. Make [tests] replicate, within reason, the challenges at the heart of each academic discipline. Let them be—authentic. (1989, p. 41)

Wiggins goes on to say that if our instructional goals for students include the abilities to write, speak, listen, create, think critically, do research, solve problems, or apply knowledge, then our tests should ask students to write, speak, listen, create, think, research, solve, and apply. How can this happen?

Many educators suggest we look to the arts and sports for analogies to solve this problem. If we think of the "test" as being the recital, exhibition, game, mock court trial, or other performance, then teaching to the test is just fine. All coaches, artists, and musicians gladly "teach" to these "tests" because performing well on these tests is the whole point of instruction. Authentic assessment asks students to perform. The performances may be thinking performances, physical performances, creative performances, or other forms. So performance assessment is any form of assessment that requires students to carry out an activity or produce a product in order to demonstrate learning (Airasian, 2005).

It may seem odd to talk about thinking as a performance, but there are many parallels. Serious thinking is risky, because real-life problems are not well defined. Often, the outcomes of our thinking are public—others evaluate our ideas. Like a dancer auditioning for a Broadway show, we must cope with the consequences of being evaluated. Like a sculptor looking at a lump of clay, a student facing a difficult problem must experiment, observe, record, imagine, and test solutions, apply both basic skills and inventive techniques, make interpretations, decide how to communicate results to the intended audience, and often accept criticism and improve the initial solution (Eisner, 1999; Herman, 1997). Table 15.4 lists some characteristics of authentic tests.

**Portfolios and Exhibitions**

The concern with authentic assessment has led to the development of several approaches based on the goal of performance in context. Instead of circling answers to "factual" questions about nonexistent situations, students are required to solve real problems. Facts are used in a context where they apply—for example, the student uses grammar facts to write a persuasive letter to a software company requesting donations for the class computer center. The following example of a test of performance is taken from the Connecticut Core of Common Learning:

Many local supermarkets claim to have the lowest prices. But what does this really mean? Does it mean that every item in their store is priced lower, or just some of them? How can you really tell which supermarket will save you the most money? Your assignment is to design and carry out a study to answer this question. What items and prices will you compare and why? How will you justify the choice of your "sample"? How reliable is the sample, etc.? (Wolf, Bixby, Gleim, & Gardner, 1991, p. 61)
TABLE 15.4

Characteristics of Authentic Tests

A. Structure and Logistics
1. Are more appropriately public; involve an audience, a panel, and so on.
2. Do not rely on unrealistic and arbitrary time constraints.
3. Offer known, not secret, questions or tasks.
4. Are more like portfolios or a season of games (not one-shot).
5. Require some collaboration with others.
6. Recur—and are worth practicing for, rehearsing, and retaking.
7. Make assessment and feedback to students so central that school schedules, structures, and policies are modified to support them.

B. Intellectual Design Features
1. Are “essential”—not needlessly intrusive, arbitrary, or contrived to “shave off” a grade.
2. Are “enabling”—constructed to point the student toward more sophisticated use of the skills or knowledge.
3. Are contextualized, complex intellectual challenges, not “atomized” tasks, corresponding to isolated “outcomes.”
4. Involve the student’s own research or use of knowledge, for which “content” is a means.
5. Assess student habits and repertoires, not mere recall or plug-in skills.
6. Are representative challenges—designed to emphasize depth more than breadth.

C. Grading and Scoring Standards
1. Involve criteria that assess essentials, not easily counted (but relatively unimportant) errors.
2. Are graded not on a “curve” but in reference to performance standards (criterion-referenced, not norm-referenced).
3. Involve demystified criteria of success that appear to students as inherent in successful activity.
4. Make self-assessment a part of the assessment.
5. Use a multifaceted scoring system instead of one aggregate grade.

D. Fairness and Equity
1. Ferret out and identify (perhaps hidden) strengths.
2. Strike a constantly examined balance between honoring achievement and native skill or fortunate prior training.
3. Minimize needless, unfair, and demoralizing comparisons.
4. Allow appropriate room for student learning styles, aptitudes, and interests.
5. Can be—should be—attempted by all students, with the test “scaffolded up,” not “dumbed down,” as necessary.


Students completing this “test” will use mathematical facts and procedures in the context of solving a real-life problem. In addition, they will have to think critically and write persuasively. The Stories of Learning/Tributes to Teaching on the next page gives another example from my own education.

Portfolios and exhibitions are two approaches to assessment that require performance in context. With these approaches, it is difficult to tell where instruction stops and assessment starts because the two processes are interwoven (Smith, Smith, & DeLisi, 2001).

Portfolios. For years, photographers, artists, models, and architects have had portfolios to display their skills and often to get jobs. A portfolio is a systematic collection of work, often including work in progress, revisions, student self-analyses, and reflections on what the student has learned (Poplak, 2005a). One student’s self-reflection is presented in Figure 15.1 on the next page.

Written work or artistic pieces are common contents of portfolios, but students might also include graphs, diagrams, snapshots of displays, peer comments, audio- or videotapes, laboratory reports, and computer programs—anything that demonstrates learning in the area being taught and assessed (Belinoff & Dickson, 1991; Camp, 1990; Wolf, Bidy, Glenn, & Gardner, 1991). There is a distinction between process portfolios and final or “best work” portfolios. The distinction is similar to the difference between formative and

Connect and Extend to the Research

Portfolio. A collection of the student’s work in an area, showing growth, self-reflection, and achievement.
When I was in graduate school in Texas, the standard doctoral candidacy examination was three days worth of closed book, handwritten, blue-book-filled answers to questions. Candidates were examined in three areas of educational psychology: sociocultural, learning/cognition, and statistics/research methodology. This was a "high-stakes," make-or-break three days that determined whether you had to leave school with a "terminal Master's" (sounds deadly doesn't it!), instead of being allowed to go on to your doctoral research. By the time we took the exam, we had completed our coursework and internships. All that remained to earn the Ph.D. was to design, complete, analyze, and write up an independent research project—the dissertation. This is a complex bit of work that takes many students months or even years to finish successfully.

Now this was the 70s—a time in Texas of protest. It was a time when students were more active in asserting their rights. And I was part of that spirit. I was certain that the three-day test was not a good assessment of our learning. I asked the faculty, "When will we have to sit down for three days and answer multiple-choice questions, write essays without any resources, or solve statistics problems that are not related to our research questions?" I wanted a more "authentic" assessment—something connected to our learning and our future. I talked to other students, wrote position papers, and lobbied professors. Finally I was asked to speak at a faculty meeting about the situation.

To my amazement and gratitude, the faculty voted to allow an alternative to the three-day test. Candidates could write a research proposal that examined a significant research question from the perspectives of social/cultural and learning/cognition in educational psychology. In addition, we were required to design a study that used appropriate methodology and statistics. I chose this option (after making all that fuss, I almost had to). Some of my friends said that I was crazy because I spent months preparing my proposal while they crammed a few weeks for their traditional exam. But then I used my proposal, with some improvements, to do my dissertation and finished a year ahead of them. I got a grant to do it, based on the thinking and writing that went into completing the option of that "authentic" assessment project. At my alma mater, the alternative candidacy exam is still in place today, and that is the form I use for my graduate students at Ohio State—thanks to the willingness of a good faculty to entertain student suggestions over 35 years ago.

---

**FIGURE 15.1**

A Student Reflects on Learning: Self-Analysis of Work in a Portfolio

Not only has this student's writing improved, but the student has become a more observant and self-critical writer.

```
Today I looked at all my stories in my writing folder. I read some of my writing since September. I noticed that I've improved some stuff. Now I edit my stories and revise. Now I use periods, quotation marks. Sometimes my stories are longer. I used to misspell my words and now I look in a dictionary or ask a friend and now. I write exciting and scary stories and now I have very good endings. Now I use capital I used to leave out words and write short simple stories.
```

### TABLE 15.5

**Process and Best Works Portfolios for Individuals and Groups**

Here are a few examples of how to use portfolios in different subjects:

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>Individual Student</th>
<th>The Process Portfolio</th>
<th>Cooperative Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science</td>
<td>Documentation (running records or logs) of using the scientific method to solve a series of laboratory problems</td>
<td>Documentation (observation checklists) of using the scientific method to solve a series of laboratory problems</td>
<td></td>
</tr>
<tr>
<td>Mathematics</td>
<td>Documentation of mathematical reasoning through double-column mathematical problem solving (computations on the left side and running commentary explaining thought processes on the right side)</td>
<td>Documentation of complex problem solving and use of higher-level strategies</td>
<td></td>
</tr>
<tr>
<td>Language Arts</td>
<td>Evolution of compositions from early notes through outlines, research notes, response to others' editing, and final draft</td>
<td>Rubrics and procedures developed to ensure high-quality peer editing</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>Individual Student</th>
<th>The Best Works Portfolio</th>
<th>Cooperative Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Arts</td>
<td>The best compositions in a variety of styles—expository, humorous, satire, creative (poetry, drama, short story), journalistic (reporting, editorial column, reviewer), and advertising copy</td>
<td>The best dramatic production, video project, TV broadcast, newspaper, advertising display</td>
<td></td>
</tr>
<tr>
<td>Social Studies</td>
<td>The best historical research paper, opinion essay on historical issue, commentary on current event, original historical theory, review of a historical biography, account of academic controversy participated in</td>
<td>The best community survey, paper resulting from academic controversy, oral history compilation, multidimensional analysis of historical event, press corps interview with historical figure</td>
<td></td>
</tr>
<tr>
<td>Fine Arts</td>
<td>The best creative products such as drawings, paintings, sculptures, pottery, poems, thespian performance</td>
<td>The best creative products such as murals, plays written and performed, inventions thought of and built</td>
<td></td>
</tr>
</tbody>
</table>

*Source: From Meaningful Assessment: A Meaningful and Cooperative Process by D. W. Johnson and R. T. Johnson. Published by Allyn and Bacon, Boston, MA. Copyright © 2002 by Pearson Education. Adapted by permission of the publisher.*

summative evaluation. Process portfolios document learning and show progress. Best work portfolios showcase final accomplishments (Johnson & Johnson, 2002). Table 15.5 shows some examples for both individuals and groups.

**Exhibitions.** An exhibition is a performance test that has two additional features. First, it is public, so students preparing exhibitions must take the audience into account; communication and understanding are essential. Second, an exhibition often requires many hours of preparation, because it is the culminating experience of a whole program of study. Thomas Guskey and Jane Bailey (2001) suggest that exhibits help students understand the qualities of good work and recognize those qualities in their own productions and performances. Students also benefit when they select examples of their work to exhibit and articulate their reasons for making the selections. Being able to judge quality can encourage student motivation by setting clear goals.

The Guidelines on the next page give some ideas for using portfolios and exhibits in your teaching.

### Evaluating Portfolios and Performances

Checklists, rating scales, and scoring rubrics are helpful when you assess performances, because assessments of performances, portfolios, and exhibitions are criterion-referenced, not norm-referenced. In other words, the students’ products and performances are compared to established public standards, not ranked in relation to other students’ work.

**Connect and Extend to PRAXIS II™**

Portfolio Assessment (II, C1, 2)

For a discussion of the advantages, limitations, design, and implementation of portfolio programs, and to examine samples of portfolio checklists, go to Teachervision.com (http://www.teachervision.com/lesson-plans/lesson-4536.html).

**Exhibition** A performance test or demonstration of learning that is public and usually takes an extended time to prepare.
Guidelines: Creating Portfolios

Students should be involved in selecting the pieces that will make up the portfolio.

**Examples:**

1. During the unit or semester, ask each student to select work that fits certain criteria, such as "my most difficult problem," "my best work," "my most improved work," or "three approaches to."

2. For their final submissions, ask students to select pieces that best show how much they have learned.

A portfolio should include information that shows student self-reflection and self-criticism.

**Examples:**

1. Ask students to include a rationale for their selections.

2. Have each student write a "guide" to his or her portfolio, explaining how strengths and weaknesses are reflected in the work included.

3. Include self- and peer critiques, indicating specifically what is good and what might be improved.


The portfolio should reflect the students' activities in learning.

**Examples:**

1. Include a representative selection of projects, writings, drawings, and so forth.

2. Ask students to relate the goals of learning to the contents of their portfolios.

The portfolio can serve different functions at different times of the year.

**Examples:**

1. Early in the year, it might hold unfinished work or "problem pieces."

2. At the end of the year, it should contain only what the student is willing to make public.

Portfolios should show growth.

**Examples:**

1. Ask students to make a "history" of their progress along certain dimensions and to illustrate points in their growth with specific works.

2. Ask students to include descriptions of activities outside class that reflect the growth illustrated in the portfolio.

Teach students how to create and use portfolios.

**Examples:**

1. Keep models of very well done portfolios as examples, but stress that each portfolio is an individual statement.

2. Examine your students' portfolios frequently, especially early in the year when they are just getting used to the idea. Give constructive feedback.

For more ideas about using portfolios, see http://www.cleon.edu/students/portfolio/

---

Connect and Extend to PRAXIS II

**Scoring Rubrics** (H, C3)

- Kathy Schock's Guide for Educators (http://school.discovery. com/schrocksguide/assess.html) provides information about every aspect of the use of scoring rubrics in the classroom as well as an extensive collection of rubrics that can be used or adapted by teachers.

- **Scoring rubrics**—rules that are used to determine the quality of a student performance.

(Cambourne & Turbill, 1990; Wiggins, 1991). For example, Figure 15.2 gives three alternatives—numerical, graphic, and descriptive—for rating an oral presentation.

**Scoring Rubrics.** A checklist or rating scale gives specific feedback about elements of a performance. Scoring rubrics are rules that are used to determine the quality of a student performance (Mabry, 1999). For example, a rubric describing an excellent oral presentation might be:

- Pupil consistently faces audience, stands straight, and maintains eye contact; voice projects well and clearly; pacing and tone variation appropriate; well-organized; points logically and completely presented; brief summary at end. (Arrasian, 1996, p. 155)

Performance assessment requires careful judgment on the part of teachers and clear communication to students about what is good and what needs improving. In some ways, the approach is similar to the clinical method first introduced by Binet to assess intelligence: It is based on observing the student perform a variety of tasks and comparing his or her performance to a standard. Just as Binet never wanted to assign a single number to represent the child's intelligence, teachers who use authentic assessments do not try to assign one score to the student's performance. Even if rankings, ratings, and grades have to be given, these judgments are not the ultimate goals—improvement of...
FIGURE 15.2

Three Ways of Rating an Oral Presentation

Numerical Rating Scale

Directions:
Indicate how often the pupil performs each of these behaviors while giving an oral presentation. For each behavior circle 1 if the pupil always performs the behavior, 2 if the pupil usually performs the behavior, 3 if the pupil seldom performs the behavior, and 4 if the pupil never performs the behavior.

Physical Expression
A. Stands straight and faces audience.
   1 2 3 4
B. Changes facial expression with change in the tone of the presentation.
   1 2 3 4

Graphic Rating Scale

Directions:
Place an X on the line that shows how often the pupil did each of the behaviors listed while giving an oral presentation.

Physical Expression
A. Stands straight and faces audience.
   always usually seldom never
B. Changes facial expression with change in the tone of the presentation.
   always usually seldom never

Descriptive Rating Scale

Directions:
Place an X on the line at the place that best describes the pupil's performance of each behavior.

Physical Expression
A. Stands straight and faces audience.
   stands straight, always looks at audience weaves, fidgets, eyes roam from audience to ceiling constant, distracting movements, no eye contact with audience
B. Changes facial expression with change in the tone of the presentation.
   matches facial expressions to content and emphasis facial expressions usually appropriate, occasional lack of expression no match between tone and facial expression; expression distracts


learning is. Some of the Guidelines for developing rubrics on page 569 are taken from Goodrich (1997) and Johnson and Johnson (2002).

It is often helpful to have students join in the development of rating scales and scoring rubrics. When students participate, they are challenged to decide what quality work looks or sounds like in a particular area. They know in advance what is expected. As students gain practice in designing and applying scoring rubrics, their work and their learning often improve. Figure 15.3 on the next page is an evaluation form for self- and peer assessment of contributions to cooperative learning groups.

Reliability, Validity, Generalizability. Because judgment plays such a central role in evaluating performances, issues of reliability, validity, and generalizability are critical considerations. As we saw in the previous chapter, judges assessing the Vermont Portfolios often have not agreed on ratings, so reliability may not be adequate. When raters are experienced and scoring rubrics are well developed and refined, however, reliability may improve (Herman & Winters, 1994; LeMahieu, Gitomer, & Fresh, 1993). Some of this

Connect and Extend to Other Chapters
See Chapter 6 for a discussion of how to use praise effectively. These guidelines apply to written feedback as well.

Connect and Extend to Your Teaching/Portfolio
Adapt the Guidelines for developing rubrics for the age group you plan to teach.
FIGURE 15.3

**Self- and Peer Evaluation of Group Learning**

**STUDENT SELF- AND PEER EVALUATION FORM**

This form will be used to assess the members of your learning group. Fill one form out on yourself. Fill one form out on each member of your group. During the group discussion, give each member the form you have filled out on them. Compare the way you rated yourself with the ways your groupmates have rated you. Ask for clarification when your rating differs from the ratings given you by your groupmates. Each member should set a goal for increasing his or her contribution to the academic learning of all group members.

Person Being Rated: ____________________________

Write the number of points earned by the group member:
(4 = Excellent, 3 = Good, 2 = Poor, 1 = Inadequate)

— On time for class.  
— Arrives prepared for class.  
— Reliably completes all assigned work on time.  
— Work is of high quality.  
— Contributes to groupmates' learning daily.  
— Asks for academic help and assistance when it is needed.

— Gives careful step-by-step explanations (doesn't just tell answers).  
— Builds on others' reasoning.  
— Relates what is being learned to previous knowledge.  
— Helps draw a visual representation of what is being learned.  
— Voluntarily extends a project.


improvement in reliability occurs because a rubric focuses the raters' attention on a few dimensions of the work and gives limited scoring levels to choose from. If scorers can give only a rating of 1, 2, 3, or 4, they are more likely to agree than if they could score based on a 100-point scale. So the rubrics may achieve reliability not because they capture underlying agreement among raters, but because the rubrics limit options and thus limit variability in scoring (Mabry, 1999).

In terms of validity, there is some evidence that students who are classified as “master” writers on the basis of portfolio assessment are judged less capable using standard writing assessments. Which form of assessment is the best reflection of enduring qualities? There is so little research on this question, it is hard to say (Herren & Winters, 1994). In addition, when rubrics are developed to assess specific tasks, the results of applying the rubrics may predict performance on anything except very similar tasks, so we do not know whether the students' performance on the specific task will generalize to the larger area of study (Haertel, 1999; McMillan, 2004).

**Diversity and Equity in Performance Assessment.** Equity is an issue in all assessment and no less so with performances and portfolios. With a public performance, there could be bias effects based on a student's appearance and speech or the student's access to expensive audio, video, or graphic resources. Performance assessments have the same potential as other tests to discriminate unfairly against students who are not wealthy or who are culturally different (McDonald, 1993). And the extensive group work, peer editing, and out-of-class time devoted to portfolios means that some students may have access to more extensive networks of support and outright help. Many students in your classes will come from families that have sophisticated computer graphics and desktop publishing capabilities. Others may have little support from home. These differences can be sources of bias and inequity.
Guidelines: Developing a Rubric

1. **Look at models:** Show students examples of good and not-so-good work. Identify the characteristics that make the good ones good and the bad ones bad.
2. **List criteria:** Use the discussion of models to begin a list of what counts in quality work.
3. **Articulate gradations of quality:** Describe the best and worst levels of quality, then fill in the middle levels based on your knowledge of common problems and the discussion of not-so-good work.
4. **Practice on models:** Have students use the rubrics to evaluate the models you gave them in Step 1.
5. **Use self- and peer assessment:** Give students their task. As they work, stop them occasionally for self- and peer-assessment.
6. **Revise:** Always give students time to revise their work based on the feedback they get in Step 5.
7. **Use teacher assessment:** Use the same rubric students used to assess their work yourself.

Note: Step 1 may be necessary only when you are asking students to engage in a task with which they are unfamiliar. Steps 3 and 4 are useful but time-consuming; you can do these on your own, especially when you’ve been using rubrics for a while. A class-structured rubric-based assessment can streamline the process so that it begins with listing criteria, after which the teacher writes out the gradations of quality, checks them with the students, makes revisions, then uses the rubric for self-, peer, and teacher assessment.


The article includes several links such as [http://www.teach-nology.com/web_tools/rubrics/](http://www.teach-nology.com/web_tools/rubrics/) and [http://rubistar.4teachers.org/](http://rubistar.4teachers.org/) that allow you to create and customize rubrics for your class.

Informal Assessments

Informal assessments are ungraded (formative) assessments that gather information from multiple sources to help teachers make decisions (Banks, 2005). Early on in the unit, assessment should be formative (provide feedback, but not count toward a grade), saving the actual graded assessments for later in the unit when all students have had the chance to learn the material (Tomlinson, 2005a). Some examples of informal assessment are student observations and checklists, questioning, journals, and student self-assessment.

Journals are very flexible and widely used informal assessments. Students usually have personal or group journals and write in them on a regular basis. In their study, Michael Pressley and his colleagues (2001) found that excellent 1st-grade literacy teachers used journaling for three purposes:

- As communication tools that allowed students to express their own thoughts and ideas.
- As an opportunity to apply what they have learned.
- As an outlet to encourage fluency and creative expression in language usage.

Teachers may use journals to learn about their students in order to better connect their teaching to the students’ concerns and interests. But often journals focus on academic learning, usually through responses to prompts. For example, Banks (2005) describes one high-school physics teacher who asked his students to respond to these three questions in their journals:

1. How can you determine the coefficient of friction if you know only the angle of the inclined plane?
2. Compare and contrast some of the similarities and the differences between magnetic, electronic, and gravitational fields.
3. If you were to describe the physical concept of sound to your best friend, what music would you use to demonstrate this concept?

When he read the students’ journals, the teacher realized that many of the students’ basic assumptions about friction, acceleration, and velocity came from personal experiences and not from scientific reasoning. His approach to teaching had to change to reach the students. The teacher never would have known to make the changes without reading the journals (Banks, 2005).

Connect and Extend to the Research

There are many other kinds of informal assessments—keeping notes and observations about student performance, rating scales, and checklists. Every time teachers ask questions or watch students perform skills, the teachers are conducting informal assessments. Look at Table 15.6. It summarizes the possibilities and limitations of aligning different assessment tools with their targets. One major message in this chapter is to match the type of assessment tools used to the target—what is being assessed.

**Involving Students in Assessments**

One way to connect teaching and assessment while developing students' sense of efficacy for learning is to involve the students in the assessment process. Students can keep track of their own progress and assess their improvement. Table 15.7 gives ideas for how to help students judge their own work and participate in assessing their own learning.

**Effects of Grading on Students**

**STOP | THINK | WRITE:** Think back on your report cards and grades over the years. Did you ever receive a grade that was lower than you expected? How did you feel about yourself, the teacher, the subject, and school in general as a result of the lower grades? What could the teacher have done to help you understand and profit from the experience?

**TABLE 15.6**

<table>
<thead>
<tr>
<th>Target to Be Assessed</th>
<th>Selected Response</th>
<th>Essay</th>
<th>Performance Assessment</th>
<th>Personal Communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge Mastery</td>
<td>Multiple choice, true/false, matching, and fill-in can sample mastery of elements of knowledge</td>
<td>Essay exercises can tap understanding of relationships among elements of knowledge</td>
<td>Not a good choice for this target—three other options preferred</td>
<td>Can ask questions, evaluate answers, and infer mastery—but a time-consuming option</td>
</tr>
<tr>
<td>Reasoning Proficiency</td>
<td>Can assess understanding of basic patterns of reasoning</td>
<td>Written descriptions of complex problem solutions can provide a window into reasoning proficiency</td>
<td>Can watch students solve some problems and infer about reasoning proficiency</td>
<td>Can ask student to &quot;think aloud&quot; or can ask follow-up questions to probe reasoning</td>
</tr>
<tr>
<td>Skills</td>
<td>Can assess mastery of the prerequisites of skillful performance—but cannot tap the skill itself</td>
<td>Can assess mastery of the prerequisites of skillful performance—but cannot tap the skill itself</td>
<td>Can observe and evaluate skills as they are being performed</td>
<td>Strong match when skill is oral communication proficiency; also can assess mastery of knowledge prerequisite to skillful performance</td>
</tr>
<tr>
<td>Ability to Create Products</td>
<td>Can assess mastery of knowledge prerequisite to the ability to create quality products—but cannot assess the quality of products themselves</td>
<td>Can assess mastery of knowledge prerequisite to the ability to create quality products—but cannot assess the quality of products themselves</td>
<td>A strong match can assess: (a) proficiency in carrying out steps in product development and (b) attributes of the product itself</td>
<td>Can probe procedural knowledge and knowledge of attributes of quality products—but not product quality</td>
</tr>
</tbody>
</table>

TABLE 15.7
Involving Students in Classroom Assessment
Here are a dozen ways of using assessment to the service of student learning.

1. Engage students in reviewing strong and weak samples to determine attributes of a good performance or product.
2. Before a discussion or conference with the teacher or peer, students identify their own perceptions of strengths and weaknesses on a specific aspect of their work.
3. Students practice using criteria to evaluate anonymous strong and weak work.
4. Students work in pairs to revise an anonymous weak work sample they have just evaluated.
5. Students write a process paper, detailing the process they went through to create a product or performance, in it they reflect on problems they encountered and how they solved them.
6. Students develop practice test plans based on their understanding of the intended learning targets and essential concepts in material to be learned.
7. Students generate and answer questions they think might be on the test, based on their understanding of the content/processes/skills they were responsible for learning.
8. A few days before a test, students discuss or write answers to questions such as: “Why am I taking this test? Who will use the results? How?” “What is it testing?” “How do I think I will do?” “What do I need to study?” “With whom might I work?”
9. Teacher arranges items on a test according to specific learning targets, and prepares a “test analysis” chart for students, with three boxes: “My strengths,” “Quick review,” and “Further study.” After handing back the corrected test, students identify learning targets they have mastered and write them in the “My strengths” box. Next, students categorize their wrong answers as either “simple mistake” or “further study.” Then, students list the simple mistakes in the “Quick review” box. Last, students write the rest of the learning targets represented by wrong answers in the “Further study” box.
10. Students review a collection of their work over time and reflect on their growth: “I have become a better reader this quarter. I used to . . . , but now I . . . .”
11. Students use a collection of their self-assessments to summarize their learning and set goals for future learning: “Here is what I have learned . . . . Here is what I need to work on . . . .”
12. Students select and annotate evidence of achievement for a portfolio.


When we think of grades, we often think of competition. Highly competitive classes may be particularly hard on anxious students, students who lack self-confidence, and students who are less prepared. So, although high standards and competition do tend to be generally related to increased academic learning, it is clear that a balance must be struck between high standards and a reasonable chance to succeed. Rick Stiggins and Jan Chappuis (2005) observe:

From their very earliest school experiences, our students draw life-shaping conclusions about themselves as learners on the basis of the information we provide to them as a result of their teachers’ classroom assessments. As the evidence accumulates over time, they decide if they are capable of succeeding or not. They decide whether the learning is worth the commitment it will take to attain it. They decide . . . whether to risk investing in the schooling experience. These decisions are crucial to their academic well-being. (p. 11)

Effects of Failure
It may sound as though low grades and failure should be avoided in school. But the situation is not that simple.

The Value of Failing? After reviewing many years of research on the effects of failure from several perspectives, Margaret Clifford (1990, 1991) concluded:
It is time for educators to replace easy success with challenge. We must encourage students to reach beyond their intellectual grasp and allow them the privilege of learning from mistakes. There must be a tolerance for error-making in every classroom, and gradual success rather than continual success must become the yardstick by which learning is judged. (1990, p. 23)

Some level of failure may be helpful for most students, especially if teachers help the students see connections between hard work and improvement. Efforts to protect students from failure and to guarantee success may be counterproductive. In fact, the more able your students are, the more challenging and important it will be to help them learn to "fail successfully" (Foster, 1981). Carol Tomlinson, an expert on differentiated instruction, puts it this way: "Students whose learning histories have caused them to believe that excellence can be achieved with minimal effort do not learn to expend effort, and yet perceive that high grades are an entitlement for them." (2005b, p. 266).

Retention in Grade. So far, we have been talking about the effects of failing a test or perhaps a course. But what about the effect of failing an entire grade—that is, of being "held back"? Almost 20% of seniors have repeated at least one grade since kindergarten, usually in the earlier grades (Kelly, 1999). Retained students are more likely to be male, members of minority groups, and living in poverty (Beebe-Frankenberger, Bocian, MacMillan, & Gresham, 2004).

Most research finds that grade retention is associated with poor long-term outcomes such as dropping out of school, higher arrest rates, fewer job opportunities, lower self-esteem (Grisson & Shepard, 1989; Jimerson, Anderson, & Whipple, 2002). In their view, students generally do better academically when they are promoted. For example, in a longitudinal study of 29 retained and 50 low-achieving but promoted students, Shane Jimerson (1999) found that years later, the retained students had poorer educational and employment outcomes than the promoted students. The retained students dropped out more often, had lower-paying jobs, and received lower competence ratings from employers. In addition, the low-achieving but promoted students were comparable to a control group in all employment outcomes at age 20.

Retention assumes that the students just need more time and that they have the abilities needed to catch up. But one study in California found the students retained after second grade had the same average IQ score as those targeted for more intensive interventions under special education requirements. In addition, about 20% of the retained group had IQ scores in the range that would qualify as having mental retardation (Beebe-Frankenberger et al., 2004), so it was probably wrong to assume that these students would benefit from the same teaching. Primary-grade students who benefit from retention tend to be more emotionally immature than their peers, but have average or
above average ability (Kelly, 1999; Pierson & Connell, 1992). Even with this group, the advantage may not last. In one study that followed many students for several years, children who could have been retained, but who were promoted, did about as well as similar children who were held back, and sometimes better (Reynolds, 1992).

No matter what, students who are having trouble should get help, whether they are promoted (this is often called “social promotion”) or retained. However, just covering the same material again in the same way won’t solve the students’ academic or social problems. As Jeannie Oakes (1999) has said, “No sensible person advocates social promotion as it is currently framed—simply passing incompetent students on to the next grade” (p. 8). The best approach may be to promote the students along with their peers, but to give them special remediation during the summer or the next year (Manzoconopoulos & Morrison, 1992; Shepard & Smith, 1988). An even better approach would be to prevent the problems by differentiating instruction early (Chapter 13).

**Effects of Feedback**

The results of several studies of feedback fit well with the notion of “successful” or constructive failure. These studies have concluded that it is more helpful to tell students why they are wrong so they can learn more appropriate strategies (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991). Students often need help figuring out why their answers are incorrect. Without such feedback, they are likely to make the same mistakes again. Yet this type of feedback is rarely given. In one study, only about 5% of the teachers noticed a consistent type of error in a student’s arithmetic computation and informed the student (Bloom & Bourdon, 1980).

What are the identifying characteristics of effective written feedback? With older students (late elementary through high school), written comments are most helpful when they are personalized and when they provide constructive criticism. This means the teacher should make specific comments on errors or faulty strategies, but balance this criticism with suggestions about how to improve, as well as comments on the positive aspects of the work (Butler & Nisan, 1986; Guskey & Bailey, 2001). Working with 6th-grade teachers, Maria Flawar and Lyn Corso (1985) found that feedback was dramatically improved when the teachers used these four questions as a guide: “What is the key error? What is the probable reason the student made this error? How can I guide the student to avoid the error in the future? What did the student do well that could be noted?” (p. 166). Here are some examples of teachers’ written comments that proved helpful (Flawar & Corso, 1985, p. 164):

Juan, you know how to get a percent, but the computation is wrong in this instance. Can you see where? (Teacher has underlined the location of errors.)

You know how to solve the problem—the formula is correct—but you have not demonstrated that you understand how one fraction multiplied by another can give an answer that is smaller than either (\(\frac{1}{5} \times \frac{1}{2} = \frac{1}{10}\)).

Comments like these should help students correct errors and recognize good work, progress, and increasing skill.

**Grades and Motivation**

If you are relying on grades to motivate students, you had better think again (Smith, Smith, & De Lisi, 2001). The assessments you give should support students’ motivation to learn—not just to work for a grade. But there really is a difference between working for a grade and working to learn! The answer depends in part on how a grade is determined. As a teacher, you can use grades to motivate the kind of learning you intend students to achieve in your course. If you test only at a simple but detailed level of knowledge, you may force students to choose between complex learning and a good grade. But when a grade reflects meaningful learning, working for a grade and working to learn become the same thing. Finally, while high grades may have some value as rewards or incentives for...
meaningful engagement in learning, low grades generally do not encourage greater efforts. Students receiving low grades are more likely to withdraw, blame others, decide that the work is "dumb," or feel responsible for the low grade but helpless to make improvements. They give up on themselves or on school (Tomlinson, 2005b). Rather than give a failing grade, you might consider the work incomplete and give students support in revising or improving. Maintain high standards and give students a chance to reach them (Guskey, 1994; Guskey & Bailey, 2001).

Another effect on motivation occurs in high schools in the race for valedictorian. Sometimes, students and parents find clever ways to move ahead of the competition—but the strategies have little to do with learning. As Tom Guskey and Jane Bailey (2001) note, when a valedictorian wins by a 1/1,000 of a decimal point, how meaningful is the learning behind the difference? Some high schools now name multiple valedictorians—as many as meet the highest standards of the school—because they believe that the educators' job is "not to select talent, but, rather, to develop talent" (Guskey & Bailey, 2001, p. 39).

The Guidelines summarize the effects grades can have on students.

---

### Grading and Reporting: Nuts and Bolts

**STOP | THINK | WRITE**  Think about the other courses you are taking this term besides this one. How are your grades calculated in those courses?

In determining a final grade, the teacher must make a major decision. Should a student's grade reflect the amount of material learned and how well it has been learned, or should the grade reflect the student's status in comparison with the rest of the class? In other words, should grading be criterion-referenced or norm-referenced?

**Criterion-Referenced versus Norm-Referenced Grading**

In norm-referenced grading, the major influence on a grade is the student's standing in comparison with others who also took the course. If a student studies very hard and almost everyone else does too, the student may receive a disappointing grade, perhaps a C or D. One common type of norm-referenced grading is called grading on the curve. How you feel about this approach probably depends on where your grades generally fell along that "curve." There is good evidence that this type of grading damages the relationships among students and between teachers and students and also diminishes motivation for most students (Kramholz & Yeh, 1996). When you think about it, if the curve arbitrarily limits the number of good grades that can be given, then, in the game of grading, most students will be losers (Guskey & Bailey, 2001; Haladyna, 2002; Rohn, 1996). Over 25 years ago, Benjamin Bloom (of Bloom's taxonomy) and his colleagues (1981) pointed out the fallacy of grading on the curve:

There is nothing sacred about the normal curve. It is the distribution most appropriate to chance and random activity. Education is a purposeful activity, and we seek to have students learn what we have to teach. If we are effective in our instruction, the distribution of achievement should be very different from the normal curve. In fact, we may even insist that our educational efforts have been unsuccessful to the extent that the distribution of achievement approximates the normal distribution. (pp. 52–53)

In criterion-referenced grading, the grade represents a list of accomplishments. If clear objectives have been set for the course, the grade may represent a certain number of objectives met satisfactorily. When a criterion-referenced system is used, criteria for each grade generally are spelled out in advance. It is then up to the student to earn the grade she or he wants to receive. Theoretically, in this system, all students can achieve an A if they reach the criteria. Criterion-referenced grading has the advantage...
Guidelines: Minimizing the Detrimental Effects of Grades

Avoid reserving high grades and high praise for answers that conform to your ideas or to those in the textbook.

Examples:
1. Give extra points for correct and creative answers.
2. Withhold your opinions until all sides of an issue have been explored.
3. Reinforce students for disagreeing in a rational, productive manner.
4. Give partial credit for partially correct answers.

Make sure each student has a reasonable chance to be successful, especially at the beginning of a new task.

Examples:
1. Pretest students to make sure they have prerequisite abilities.
2. When appropriate, provide opportunities for students to retest to raise their grades, but make sure the retest is as difficult as the original.
3. Consider failing efforts as “incomplete” and encourage students to revise and improve.
4. Base grades more on work at the end of the unit, give ungraded work in the beginning of the unit.

Balance written and oral feedback.

Examples:
1. Consider giving short, lively written comments with younger students and more extensive written comments with older students.
2. When the grade on a paper is lower than the student might have expected, be sure the reason for the lower grade is clear.
3. Tailor comments to the individual student’s performance, avoid writing the same phrases over and over.
4. Note specific errors, possible reasons for errors, ideas for improvement, and work done well.

Make grades as meaningful as possible.

Examples:
1. Tie grades to the mastery of important objectives.
2. Give ungraded assignments to encourage exploration.
3. Experiment with performances and portfolios.

Base grades on more than just one criterion.

Examples:
1. Use essay questions as well as multiple-choice items on a test.
2. Grade oral reports and class participation.

For more information about the effects of grading, see http://www.naa.edu/readings/room/books/str6.html and http://www.allotkow.org/teaching/tot/kg.htm

The Point System and Percentage Grading

One popular system for combining grades from many assignments is a point system. Each test or assignment is given a certain number of total points, depending on its importance. A test worth 40% of the grade could be worth 40 points. A paper worth 20% could be worth 20 points. Points are then awarded on the test or paper based on specific criteria. An A+ paper, one that meets all the criteria, could be given the full 20 points; an average paper might be given 10 points. If tests of comparable importance are worth the same number of points, are equally difficult, and cover a similar amount of material, this system can be fair and practical. But in most schools, these points still must be converted into some form of final grade. So the teacher has to decide the standards for assigning grades.
## A Criterion-Referenced Report Card

This is one example of a criterion-referenced report card. Other forms are possible. Total criterion-referenced reports indicate student progress toward specific goals.

### LINCOLN ELEMENTARY SCHOOL
#### GRADE 5

<table>
<thead>
<tr>
<th>Student</th>
<th>Teacher</th>
<th>Principal</th>
<th>Marles Simms</th>
<th>Quarter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 3 4</td>
</tr>
</tbody>
</table>

**E** = Excellent  **S** = Satisfactory  **P** = Making Progress  **N** = Needs improvement

### READING PROGRAM
- **MATERIALS USED:**
- Reads with understanding
- Is able to write about what is read
- Completes reading group work accurately and on time
- Shows interest in reading

### READING SKILLS
- Decodes new words
- Understands new words

### INDEPENDENT READING LEVEL
- Below/At Grade Level/Above

### LANGUAGE ARTS
- Uses oral language effectively
- Listens carefully
- Masters weekly spelling

### WRITING SKILLS
- Understands writing as a process
- Creates a rough draft
- Makes meaningful revisions
- Creates edited, legible final draft

### EDITING SKILLS
- Capitalizes
- Punctuates
- Uses complete sentences
- Uses paragraphs
- Demonstrates dictionary skills

### WRITING SKILL LEVEL
- Below/At Grade Level/Above

### MATHEMATICS
- **PROBLEM SOLVING**
  - Solves teacher-generated problems
  - Solves self-generated problems
  - Can create story problems

- **INTERPRETING PROBLEMS**
  - Uses appropriate strategies
  - Can use more than one strategy
  - Can explain strategies in written form
  - Can explain strategies orally

### MATH CONCEPTS
- **UNDERSTANDS BASIC TEN**
  - Beginning/Developing/Sophisticated
- **MULTIPLICATION, BASIC FACTS**
  - Beginning/Developing/Sophisticated
- **2-DIGIT MULTIPLICATIONS**
  - Beginning/Developing/Sophisticated
- **DIVISION**
  - Beginning/Developing/Sophisticated

### OVERALL MATH SKILL LEVEL
- Beginning/Developing/Sophisticated

### ATTITUDE/WORK SKILLS
- **WELCOMES A CHALLENGE**
- **LISTENS TO DIRECTIONS**
- **CHECKS WORK**
- **PARTICIPATES IN DISCUSSION**

### IT FIGURES
- **IS WORKING ON:**
- **GOALS:**
  - Is working on achieving goal:

### SOCIAL STUDIES
- Understands subject matter
- Shows curiosity and enthusiasm
- Participates in class discussions
- Uses map skills
- Demonstrates control of reading skills by interpreting text

### SCIENCE
- Shows curiosity about scientific subject matter
- Asks good scientific questions
- Shows knowledge of scientific method
- Uses knowledge of scientific method to help set up and run experiments
- Makes good scientific observations
- Has researched scientific topic(s)

### HUMAN RELATIONS
- Shows courtesy
- Respects rights of others
- Shows self-control
- Interacts well with peers
- Shows a cooperative and positive attitude in class
- Shows a cooperative attitude when asked to work with other students
- Is willing to help other students
- Works well with other adults (teachers, student teacher, parents, etc.)

### WORKING SKILLS
- Listens carefully
- Follows directions
- Works quickly and carefully
- Checks work
- Completes work on time
- Uses time wisely
- Works well independently
- Works well in group
- Takes risks in learning
- Welcomes a challenge

### ATTENDANCE

<table>
<thead>
<tr>
<th>Attendance</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>Placement for next year:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tardy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Source:** From "Reporting Methods in Grades K-8," by K. Lake and K. Kafka. In ASCD 1986 Yearbook: Communicating Student Learning (p. 104). T. Guskay (Ed.). Copyright © 1986 by the Association for Supervision and Curriculum Development. Reprinted with permission. All rights reserved. The Association for Supervision and Curriculum Development is a worldwide community of educators advocating sound policies and sharing best practices to achieve the success of each learner. To learn more, visit ASCD at www.ascd.org.

### Percentage Grading

System of converting class performances to percentage scores and assigning grades based on predetermined cutoff points.

Using another approach, **percentage grading**, the teacher can assign grades based on how much knowledge each student has mastered—what percentage of the total knowledge he or she understands. To do this, the teacher might score tests and other class work with percentage scores (based on how much is correct: 50%, 85%, etc.) and then...
average those scores to reach a course score. These scores can then be converted into letter grades according to predetermined cutoffs. Any number of students can earn any grade. School systems often establish equivalent percentage categories for As, Bs, and so on. The percentages vary from school district to school district. An A might be 90% to 100% in one district but 94% to 100% in another.

Even though it is very common, this approach has problems. Can we really say what is the total amount of knowledge available in, for example, 8th-grade science? Are we sure we can accurately measure what percentage of this body of knowledge each student has attained? To use percentage grading appropriately, we would have to know exactly what there was to learn and exactly how much of that knowledge each student had learned (Popham, 2009a). These conditions are seldom met, even though teachers use the cutoff points to assign grades as if measurement were so accurate that a one-point difference was meaningful.

Don’t be fooled by the seeming security of points or absolute percentages. Your own grading philosophy will continue to operate, even in this system. Because there is more concern today with specifying objectives and criterion-referenced assessment, especially at the elementary-grade levels, several alternative methods for evaluating student progress against predetermined criteria have evolved.

The Contract System and Grading Rubrics

When applied to the whole class, the contract system indicates the type, quantity, and quality of work required for each number or letter grade in the system. Rubrics describe the performance expected for each level. Students agree, or “contract,” to work for particular grades by meeting the specified requirements and performing at the level specified. For example, the following standards might be established:

- F: Not coming to class regularly or not turning in the required work.
- D: Coming to class regularly and turning in the required work on time.
- C: Coming to class regularly, turning in the required work on time, and receiving a check mark on all assignments to indicate they are satisfactory.
- B: Coming to class regularly, turning in the required work on time, and receiving a check mark on all assignments except at least three that achieve a check-plus, indicating superior achievement.
- A: As above, plus a successful oral or written report on one of the books listed for supplementary reading.

This example calls for more subjective judgment than would be ideal. However, contract systems reduce student anxiety about grades. The contract system can be applied to individual students, in which case it functions much like an independent study plan.

Unfortunately, the system can lead to overemphasis on the quantity of work. Teachers may be too vague about the standards that differentiate acceptable from unacceptable work. This is where scoring rubrics for each assignment can be helpful. If clear and well-developed rubrics describe the performances expected for each assignment, and if students learn to use the rubrics to evaluate their own work, then quality, not quantity, will be at the center of grading. You can modify the contract system by including a revise option. For example, a check mark might be worth 75 points and a check-plus 90 points; a check-plus earned after revision could be worth 85 points—more than a check, but less than a check-plus earned the first time around. This system allows students to improve their work, but also rewards students for getting it right the first time. In addition, the teacher may be less reluctant to judge a project unsatisfactory because students can improve their work (King, 1979). But beware, if a school system requires a 5-point grading scale and all students contract for and achieve the highest grade (before or after revising), the teacher will wish that the principal had been consulted about the system before the

Connect and Extend to Your Teaching/Portfolio

How would you solve the following problem? You are using a contract system in one of your classes. One of the requirements for an A is “to write a book report.” However, some students are reporting on books that they read last year, and some are handing in short, superficial reports. How can you structure the contract system so that the students will do better quality work?

Connect and Extend to Other Chapters

See Chapter 6 for a discussion of behavior management contracts.

Contract system — System in which each student works for a particular grade according to agreed-upon standards.

Revised option — In a contract system, the chance to revise and improve work.
Guidelines: Using Any Grading System

Explain your grading policies to students early in the course and remind them of the policies regularly.

**Examples:**
1. Give older students a handout describing the assignments, tests, grading criteria, and schedule.
2. Explain to younger students in a low-pressure manner how their work will be evaluated.

Base grades on clearly specified, reasonable standards.

**Examples:**
1. Specify standards by developing a rubric with students—have anonymous examples of poor, good, and excellent work from previous classes.
2. Discuss workload and grading standards with more experienced teachers.
3. Give a few formative tests to get a sense of your students' abilities before you give a graded test.
4. Take tests yourself first to gauge the difficulty of the test and to estimate the time your students will need.

Base your grades on as much objective evidence as possible.

**Examples:**
1. Plan in advance how and when you will test.
2. Keep a portfolio of student work. This may be useful in student or parent conferences.

Be sure students understand test directions.

**Examples**
1. Outline the directions on the board.
2. Ask several students to explain the directions.
3. Go over a sample question first.

Correct, return, and discuss test questions as soon as possible.

**Examples:**
1. Have students who wrote good answers read their responses for the class; make sure they are not the same students each time.
2. Discuss why wrong answers, especially popular wrong choices, are incorrect.
3. As soon as students finish a test, give them the answers to questions and the page numbers where answers are discussed in the text.

As a rule, do not change a grade.

**Examples:**
1. Make sure you can defend the grade in the first place.
2. Do change any clerical or calculation errors.

Guard against bias in grading.

**Examples:**
1. Ask students to put their names on the backs of their papers.
2. Use an objective point system or model papers when grading essays.

Keep pupils informed of their standing in the class.

**Examples:**
1. Write the distribution of scores on the board after tests.
2. Schedule periodic conferences to go over work from previous weeks.

Give students the benefit of the doubt. All measurement techniques involve error.

**Examples:**
1. Unless there is a very good reason not to, give the higher grade in borderline cases.
2. If a large number of students miss the same question in the same way, revise the question for the future and consider throwing it out for that test.

For more thoughts about grading, see http://teaching.berkeley.edu/ht/guiding.html

Source: From Problems in Middle and High School Teaching: A Handbook for Student Teachers and Beginning Teachers (pp. 182–187) by A. M. Dwyer, 1979, Boston: Allyn and Bacon. Copyright © 1979 by Allyn and Bacon. Adapted by permission of the author and publisher.

Connect and Extend to Your Teaching/Portfolio

Think about your philosophy of teaching. What do you believe about testing and grading? How will you assign grades? (Consult the Guidelines for ideas.)

Other Issues in Grading

**Grading on Effort and Improvement.** Grading on effort and improvement is not really a complete grading system, but rather a theme that can run through most grading methods. Should teachers grade students based on how much they learn or on the final level of learn-
ing? One problem with using improvement as a standard for grading is that the best students improve the least, because they are already the most competent. Do you want to penalize these students because they know quite a bit initially, and the teaching and testing have limited how much learning they can demonstrate? After all, unless you assign extra work, these students will run out of things to do (Arawasz, 2003; Guskey & Bailey, 2001).

One solution is to use the individual learning expectation (ILE) system. With this system, students earn improvement points on tests or assignments for scoring above their personal base (average) score or for making a perfect score. The teacher can count these improvement points when figuring a final grade or simply use them as a basis for giving other classroom rewards. Another option is to recognize good student effort through oral or written comments, notes to parents, or other recognition.

Cautions: Being Fair. The attributions a teacher makes about the causes of student successes or failures can affect the grades that students receive. Teachers are more likely to give higher grades for effort (a controllable factor) than for ability (an uncontrollable factor). Lower grades are more likely when teachers attribute a student’s failure to lack of effort instead of lack of ability (Weiner, 1979). It is also possible that grades can be influenced by a halo effect—that is, by the tendency to view particular aspects of a student based on a general impression, either positive or negative. As a teacher, you may find it difficult to avoid being affected by positive and negative halos. A very pleasant student who seems to work hard and causes little trouble may be given the benefit of the doubt (B instead of C+), whereas a very difficult student who seems to refuse to try might be a loser at grading time (D instead of C-).

There is another aspect of fairness. If students’ grades are affected by unclean directions, difficulty reading the text or assignment questions, low grades for missing homework, a test that did not measure what was taught, time pressures, or other factors that are not related to the skills being assessed, then “grade pollution” has occurred. When flowery language or an artistic cover inflates a student’s grade, we have more grade pollution. Grades should reflect what a student knows, understands, or can do related to the learning objective specified (Tomlinson, 2005b).

Beyond Grading: Communicating with Families

**WHAT WOULD YOU SAY?**

During your interview for a job in a very diverse district, the department chair asks you, “How will you communicate with families about your homework and grading policies?”

No number or letter grade conveys the totality of a student’s experience in a class or course. Students, parents, and teachers sometimes become too focused on the end point—the grade. But communicating with families should involve more than just sending home grades. There are a number of ways to communicate with and report to families. Many teachers I know have a beginning-of-the-year newsletter or student handbook that communicates homework, behavior, and grading policies to families. Other options described by Guskey & Bailey (2001) are:

- Notes attached to report cards
- Phone calls, especially “Good News” calls
- School open houses
- Conferences, including student-led conferences
- Portfolios or exhibits of student work
- Homework hotlines
- School or class web pages
- Home visits

**Individual learning expectation (ILE)** Personal average score.

**Halo effect** The tendency for a general impression of a person to influence our perception of any aspect of that person.
Here is an example of a “Good News Call” that makes use of cell phone technology, taken from Guskey and Bailey (2001). An elementary-school principal carries her phone with her as she walks through the hallways, visits the cafeteria, supervises the playground, and observes teachers’ classes. When she sees a student performing well in class, assisting a classmate, or helping to improve the school, she immediately calls the student’s parent or guardian on her phone and announces, “Hello, this is Ms. Johnson, the principal at Judd Elementary School. I just saw Tonya...” After explaining what she observed and complimenting the child, she hands the phone to the child so that he or she can talk briefly with the parent or guardian. Everyone leaves with a big smile. One caution—if a family member didn’t answer, it would be a bad idea to leave a call back message because the message recipients might fear that their children were sick or injured.

When asked about calls to parents concerning student problems, Ms. Johnson explains, “Those I save for later school. Often, I have to think more carefully about what I’m going to say and what strategies I’m going to recommend. When I see a child doing something wonderful, however, I want to let the parents know about that right away. And I never have to weigh my words. Plus, I think it means more to the child.”

The principal’s phone calls have completely altered the culture of this school. Parent involvement and participation in school events is at an all-time high, and their regard for Ms. Johnson and the school staff is exceptionally positive. It is a small thing, but it has made a big difference.

Conferences with parents are often expected of teachers in elementary schools and can be equally important in junior high and high school. Clearly, the more skilled teachers are at communicating, the more effective they will be at conducting those conferences. Listening and problem-solving skills such as those discussed in Chapter 12 can be particularly important. When you are dealing with parents or students who are angry or upset, make sure you really hear the concerns of the participants, not just their words. The atmosphere should be friendly and unmired. Any observations about the student should be as factual as possible, based on observation or information from assignments. Infor-

---

**Family and Community Partnerships**

**Guidelines: Conferences**

Plan ahead. What are your goals?

**EXAMPLES:**
- Problem solving?
- Sharing test results?
- Asking questions that you want answered?
- Providing information you want to share?
- Describing your "next steps" in the classroom?
- Making suggestions for use at home?

Begin and end with a positive statement.

**EXAMPLES:**
- "Jacob has a great sense of humor."
- "Citelle really enjoys materials that deal with animals."
- "Yeshim is sympathetic when somebody has a problem."
- "Ashanti should be a big help in the social studies play that her group is developing."

Listen actively.

**EXAMPLES:**
- "You seem to feel frustrated when Lee doesn’t listen."

Establish a partnership.

**EXAMPLES:**
- Ask parents to follow through on class goals at home: “If you ask to see the homework checklist and go over it at home with Iris, I’ll review it and chart her progress at school.”

Plan follow-up contacts.

**EXAMPLES:**
- Write notes or make phone calls to share successes.
- Keep parents informed before problems develop.

For some dos and don’ts of conferences, see
http://www.nysut.org/newmembers/survival_conferences.html

Source: From The Successful Classroom: Management Strategies for Regular and Special Education Teachers (p. 181), by D. P. Froneberg & M. Driscoll. Published by Teachers College, Columbia University. Reprinted by permission of the publisher. All rights reserved.
TABLE 15.8
Possible Grading Accommodations for Students with Special Needs

<table>
<thead>
<tr>
<th>Adaptation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change Grading Criteria</td>
<td>Increase credit for participation in classroom group activities and decrease credit for essay examinations.</td>
</tr>
<tr>
<td>Vary grading weights assigned to different activities or products.</td>
<td>Change a C to a B if the student’s total points have increased significantly from the previous marking period.</td>
</tr>
<tr>
<td>Grade on improvement by assigning extra points.</td>
<td>Indicate in the IEP that the student will work on subtraction while the other students work on division.</td>
</tr>
<tr>
<td>Modify or individualize curriculum expectations.</td>
<td>State in the contract that student will receive a B for completing all assignments at 80% quality, quantity, timelines, accuracy, attending all classes, and completing one extra-credit report.</td>
</tr>
<tr>
<td>Use contracts and modified course requirements for quality, quantity, and timelines.</td>
<td>Write on the report card that the student’s grade reflects performance on IEP objectives and not on the regular classroom curriculum.</td>
</tr>
</tbody>
</table>

Provide Supplemental Information
| Add written comments to clarify details about the criteria used. | Note that while the student’s grade was the same this marking period, daily records show the student completed math assignments with less teacher assistance. |
| Add information from student activity log. | State that the student’s written language showed an increase in word variety, sentence length, and quality of ideas. |

Use Other Grading Options
| Add information about effort, progress, and achievement from portfolios or performance-based assignments. | Attach a checklist to the report card indicating that during the marking period, the student mastered addition facts, two-digit addition with regrouping, and counting change to one dollar. |
| Use checklists of skills and show the number or percentage of objectives met. | Students receive a “pass” for completing 80% of daily work with at least 75% accuracy, and attending 90% of class sessions. |
| Use pass/fail grades. | |


Information gained from a student or a parent should be kept confidential. The Family and Community Partnerships Guidelines offer some helpful ideas for planning and conducting conferences.

Reaching Every Student: Accommodations in Grading
In assigning and reporting grades, how do you make accommodations for students with special needs? Table 15.8 gives some options.
It may be up to you as the teacher to decide how to calculate the grades of students with special needs and students whose first language is not English, but be sure to check whether there are school or district policies already in place. Also, some schools now offer families a choice of the languages for the report card itself. The form is the same; just the language is different (Guskey & Bailey, 2001).

Diversity and Convergences in Classroom Assessment

Classroom assessment should, first and foremost, support student learning. But stereotypes and biases can interfere with assessment.

Diversity and Convergences in Classroom Assessment

Connect and Extend to the Research

Connect and Extend to Your Teaching/Portfolio
Add Table 15.2, “Key Points to Consider in Judging Textbook Tests,” Table 15.4, “Characteristics of Authentic Tests,” and Table 15.8, “Possible Grading Accommodations for Students with Special Needs” to your Teaching Resources file.
Diversity

Several studies have found that teachers may hold lower expectations for ethnic minority students and these biases can influence teaching and assessment (Banks, 2005). For example, Lipman (1997) found that teachers of ethnically diverse classes are less innovative and more traditional in their instruction. These lower expectations can show up in giving higher grades to students for lower-quality work. A national study of students in 1st, 3rd, and 7th grades compared students' grades to their standardized test scores. A grade of "A" in a high-poverty school was equivalent to a grade of "C" in a low-poverty school (Puma et al., 1997). These two kinds of schools differed in standards and curriculum, as well as expectations (McClure, 2005).

Biases can be subtle. For example, in the tests that you write, will you incorporate language and experiences that are familiar to your students? James Popham (2005a), an expert in assessment, described his first teaching job in a rural high school on an eastern Oregon town with a population of 1,500:

I had grown up in a fairly large city and had no knowledge about farming or ranching. To me, a "range" was a kitchen appliance on which one cooked meals. In retrospect, I am certain that many of my classroom test items contained "city" content that might have confused my students. I bet anything that many of my early test items unfairly penalized some of the boys and girls who had rarely, if ever, left their farms or ranches to visit more metropolitan areas. (p. 86)

Think of the range of differences in experiences of your students if they are from other countries or speak other languages. The following Guidelines, developed by several professional organizations, could guide your thinking about classroom assessment (Banks, 2005, p. 82).

Convergences

Quality teaching and quality assessment share the same basic principles and these principles hold for all students. Carol Tomlinson (2005b, pp. 265–266) suggests that good instruction and good grading both depend on a teacher who:

- Is aware of and responds to student differences.
- Specifies clear learning outcomes.
- Uses pretests and formative assessments to monitor student progress toward learning goals.
- Adapts instruction in a variety of ways to ensure, as much as possible, that each student continues to progress.
- Makes sure students know the criteria for success on summative assessments that are tightly aligned to the stated learning goals.
- Provides varied forms of assessment to ensure that students have an unobstructed opportunity to express what they have learned.

An important federal law, the Buckley Amendment, may affect you as a teacher. Also called the Family Educational Rights and Privacy Act of 1974 and the Educational Amendments Act of 1974, this law states that all educational agencies must make test results and any other information in students' records available to the students and/or their parents. If the records contain information students or parents believe is incorrect, they can challenge such entries and have the information removed if they win the challenge. This means that the information in a student's records must be based on firm, defensible evidence. Tests must be valid and reliable. Your grades must be justified by thorough assessment and observation. Comments and anecdotes about students must be accurate and fair.
## Guidelines: Standards for Educational and Psychological Testing

Assessments should be as fair as possible for students from diverse gender, race, or ethnic backgrounds.

**Examples:**
1. Avoid any slang expressions that might be offensive.
2. Eliminate items with stereotypes that might offend such as Eurocentric views of history that assume Columbus "discovered" America—there were many people ahead of him who discovered and lived in America.

Assessments should relate only to the instruction given and not to background knowledge that might penalize students from diverse backgrounds.

**Examples:**
1. Review standardized and classroom assessments to see if some questions require knowledge not available to certain groups, such as questions about sports (may penalize females), travel (may penalize students from low-income families), or driving vacations (may penalize students from families without cars).
2. Make sure certain answers don't favor particular social or political ideas.

Don't let assessments become a measure of language proficiency unless language is the focus of the test.

**Examples:**
1. Make special accommodations for English Language Learners such as extra time or the use of dictionaries.
2. Make sure math tests are not really reading proficiency tests.


### SUMMARY TABLE

**Formative and Summative Assessment (pp. 554-555)**

<table>
<thead>
<tr>
<th>Formative and Summative Assessment</th>
<th>(pp. 554-555)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What are two kinds of classroom assessment?</strong></td>
<td>In the classroom, assessment may be formative (ungraded, diagnostic) or summative (graded). Formative assessment helps form instruction, and summative assessment summarizes students’ accomplishments.</td>
</tr>
<tr>
<td><strong>Formative assessment</strong></td>
<td>Ungraded testing used before or during instruction to aid in planning and diagnosis.</td>
</tr>
<tr>
<td><strong>Pretest</strong></td>
<td>Formative test for assessing students' knowledge, readiness, and abilities.</td>
</tr>
<tr>
<td><strong>Diagnostic test</strong></td>
<td>Formative test to determine students' areas of weakness.</td>
</tr>
<tr>
<td><strong>Summative assessment</strong></td>
<td>Testing that follows instruction and assesses achievement.</td>
</tr>
</tbody>
</table>

**Alternatives to Traditional Assessments (pp. 561-570)**

<table>
<thead>
<tr>
<th>Alternatives to Traditional Assessments</th>
<th>(pp. 561-570)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What is authentic assessment?</strong></td>
<td>Critics of traditional testing believe that teachers should use authentic tests and other authentic assessment procedures. Authentic assessment requires students to perform tasks and solve problems that are similar to the real-life performances that will be expected of them outside of school.</td>
</tr>
<tr>
<td><strong>Describe portfolios and exhibitions.</strong></td>
<td>Portfolios and exhibitions are two examples of authentic assessment. A portfolio is a collection of the student's work, sometimes chosen to represent growth or improvement or sometimes featuring &quot;best work.&quot; Exhibitions are public performances of the student's understandings. With portfolios and exhibitions, there is an emphasis on performing real-life tasks in meaningful contexts.</td>
</tr>
<tr>
<td><strong>What are the issues of reliability, validity, and equity with portfolios and performance assessment?</strong></td>
<td>Using authentic assessments does not guarantee reliability, validity, and equity (absence of bias). Using rubrics is one way to make assessment more reliable and valid. But the results from assessment based on rubrics may not predict performance on related tasks. Also, rater bias based on the appearance, speech, or behavior of minority-group students or a lack of resources may place minority-group students at a disadvantage in performance assessments or projects.</td>
</tr>
<tr>
<td><strong>Authentic assessments</strong></td>
<td>Assessment procedures that test skills and abilities as they would be applied in real-life situations.</td>
</tr>
</tbody>
</table>
Performance assessments. Any form of assessment that requires students to carry out an activity or produce a product in order to demonstrate learning.

Portfolio. A collection of the student's work in an area, showing growth, self-reflection, and achievement.

Exhibition. A performance test or demonstration of learning that is public and usually takes an extended time to prepare.

Scoring rubrics. Rules that are used to determine the quality of a student performance.

Informal assessments. Ungraded (formative) assessments that gather information from multiple sources to help teachers make decisions.

Effects of Grading on Students (pp. 570-574)

How can failure support learning? Students need experience in coping with failure, so standards must be high enough to encourage effort. Occasional failure can be positive if appropriate feedback is provided. Students who never learn how to cope with failure and still persist in learning may give up quickly when their first efforts are unsuccessful.

Which is better, "social promotion" or being "held back"? Simply retarding or promoting a student who is having difficulty will not guarantee that the student will learn. Unless the student is very young or emotionally immature compared to others in the class, the best approach may be to promote, but provide extra support such as tutoring or summer school sessions. Differentiated instruction could prevent problems.

Can feedback, including grades, promote learning and motivation? Written or oral feedback that includes specific comments on errors or faulty strategies, but that balances this criticism with suggestions about how to improve, along with comments on the positive aspects of the work, increases learning. Grades can encourage students' motivation to learn if they are tied to meaningful learning.

Grading and Reporting: Nuts and Bolts (pp. 574-579)

Describe two kinds of grading. Grading can be either criterion-referenced or norm-referenced. Criterion-referenced report cards usually indicate how well the individual student has met each of several objectives. One popular norm-referenced system is grading on the curve, based on a ranking of students in relation to the average performance level. This is not recommended.

What is the point system? Tests and papers are often scored on a point system. Many schools use percentage grading systems, but the difficulty of the tests and the scoring criteria often influence the results. The difference between a B and a C may only be a matter of one or two points on paper, but the effect of the difference can be large for a student.

Describe some alternatives to traditional grading. Alternatives to traditional grading are the contract and individual learning expectation (ILE). Whatever system you use, you will have to decide whether you want to grade on effort, improvement, or some combination and whether you want to limit the number of good grades available.

What are some sources of bias in grading? Many factors besides quality of work can influence grades: the teacher's beliefs about the student's ability or effort or the student's general classroom behavior, for example. There are many appropriate accommodations that teachers can make to ensure that grading is fair.

Criterion-referenced grading. Assessment of each student's mastery of course objectives.

Norm-referenced grading. Assessment of students' achievement in relation to one another.

Grading on the curve. Norm-referenced grading that compares students' performance to an average level.

Percentage grading. System of converting class performances to percentage scores and assigning grades based on predetermined cutoff points.

Contract system. System in which each student works for a particular grade according to agreed-upon standards.

Revise option. In a contract system, the chance to revise and improve work.

Halo effect. The tendency for a general impression of a person to influence our perception of any aspect of that person.

Beyond Grading: Communication with Families (pp. 579-581)

How can communications with families support learning? Not every communication from the teacher needs to be tied to a grade. Communication with students and parents can be important in helping a teacher understand students and present effective instruction by creating a consistent learning environment. Students and parents have a legal right to see all the information in the students' records, so the contents of files must be appropriate, accurate, and supported by evidence.
Teachers’ Casebook: Connections to PRAXIS II™

Just as you have learned that there has been much research-based innovation when it comes to instruction, there has been much research-based innovation when it comes to classroom assessment.

What Would They Do?

Here is how some practicing teachers responded to the teaching situation presented at the beginning of this chapter about setting up a system to give letter grades.

**Katie Churchill** Third Grade Teacher, Oriole Park Elementary School, Chicago, Illinois

I would use a combination of assessment tools to evaluate my students. Using a rubric that students and parents alike are familiar with provides an easy-to-follow and understand grading system. The rubric needs to remain in a focal area in the classroom as a constant reminder to the students of what their expectations are.

By differentiating instruction consistently to cover all learning styles and modalities, the students would hopefully become more involved and invested in their own learning and, as a result, produce better-quality work and exceed expectations.

Several factors would play a part in obtaining a particular letter grade. The letter grade would be earned through a combination of group work, completing objectives, and following the rubric guidelines for quality work.

**MacyAyala** High School Teacher of Preparatoria Eugenio Garza Lagüera, Campus Garza Sada, Monterey, N. L. Mexico

I think it is important to assess a cross-section of student work. First, portfolios can be a useful way to gather various types of work throughout the year. Using a portfolio, a teacher can then attach a letter grade to student progress and achievement. It is important to grade children not only on progress, but also on their understanding of material. I would use meaningful, written assessments to test for retention and understanding of my students' knowledge. Finally, I would grade various projects and experiments, so that the students who are better project-based learners will be graded fairly. I also like the idea of using a rubric system to grade students on writing or projects. Under a rubric system, a teacher allocates a certain number of points to each content area. It is then easy to attach a letter grade based on the number of points received.

**Katie Piel** Kindergarten–Sixth Grade Teacher, West Park School, Moscow, Idaho

Students should be given the latitude to express achievement in different ways like group projects, daily class work, tests, and individual projects. All students would be held accountable for demonstrating their own learning. With each teacher grading on a different standard, the teachers must also take on the responsibility of collaborating with their peers. Communicating to other teachers the skills a student can be expected to bring with him or her to the next level is crucial.

**Aimee Fredette** Second Grade Teacher, Fisher Elementary School, Waypole, Massachusetts

I believe that students are not all smart in the same ways, I give the students a variety of ways to demonstrate their knowledge. I also focus on the students' ability to take their knowledge and integrate it into other subject areas across the curriculum.

I use a student portfolio for each child, compiled throughout the year and used to show growth and development. Each time I correct papers I choose a couple of pieces of work that each child has done. I put these papers in the portfolio folder. I try to choose a variety of work, not necessarily their "best work." At the end of the year the children receive the entire folder to keep.

**Allan Osborne** Assistant Principal, Snug Harbor Community School, Quincy, Massachusetts

Any grading system should consider a student's progress and effort. Grading systems also should be individualized to account for a student's unique strengths and weaknesses. Thus, a mainstreamed special education student should not be held to the same expectations as a gifted student.

The most critical aspect of any successful grading system is that it is fair. Fairness dictates that students and their parents be given information in advance about class requirements and expectations, along with a description of grading criteria. A system that is fair can be easily justified. It is also important to keep accurate and detailed records of student progress. In addition to recording grades on tests, quizzes, and projects, anecdotal records describing a student's typical performance should be kept. These records can be valuable if a report card grade is questioned.

Although group assignments can be an important learning experience, I would be reluctant to place too much emphasis on a group project grade. As we all know, each member of the group does not participate equally, and thus, a group grade does not reflect the contribution of each individual member.