

# Evaluating Students' Work

By Betty E. Gridley and W. Holmes Finch

One of the most time-consuming and often stress-producing tasks for instructors is evaluating students' work and ultimately assigning grades. Unfortunately, many first-time teachers pay little attention to this task until it is actually time to assign grades. The evaluation of students should be among the first things instructors should consider, not among the last! As Moore (1994) points out, "Assessment and teaching/learning should be treated as two sides of the same coin." One way to minimize difficulties associated with grading is by thorough planning and preparation. For example, the time to assign criteria for grading is when the syllabus is presented to students during the first class period. The most effective and constructive way to develop tests is when planning for class. A second time for review is immediately after each class, when the material is fresh in one's mind.

## *Determining Evaluative Criteria*

When establishing evaluative criteria for a course, one should remember to check with the department to determine whether there are standards already in place that must be followed. For example, many programs and/or their accrediting bodies have performance standards that must be met. Instructors should determine the nature of the data that must be collected through their assessments. Mentors and other professors who have taught courses are good sources of information. However, it is prudent to remember that assessment are difficult to construct and as such must be evaluated continuously. Many different types of assessment may be appropriate and effective ways to evaluate students' work. Some types of assessment fit course goals and course content better than others, and instructors should consider the nature of their course and the course objectives when deciding what type of assessment will best reflect students'

progress. Instructors may also incorporate knowledge of developmental differences such as those outlined in the Perry schema (see [Characteristics of Effective Teaching](#)) as they consider evaluating learning outcomes. For the purposes of this chapter, assessment is defined as ". . .any of a variety of procedures used to obtain information about student performance. . . . Assessment answers the question, 'How well does the individual perform?'" (Linn & Gronlund, 1995).

Some useful guidelines include:

- **Determine the learning outcomes (objectives) to be assessed.** Assessments should be directly linked to plans for instruction. Instructors should consider not only the content to be tapped but also the levels of thinking. There are a number of taxonomies that might be used to develop outcomes that reflect different levels of thinking.
- **Match assessments with learning outcomes.** One of the most effective methods of matching assessments with outcomes is to develop a "test blueprint." (See below.)
- **Use multiple indicators where possible.** Students are not all alike in their ability to perform on tests. In addition, all learning outcomes are not tapped equally well with one particular type of assessment. For example, performance on a musical instrument cannot be adequately assessed using a paper-and-pencil test.
- **Consider the characteristics of the learner.** Students with disabilities are becoming a greater portion of the student population. Assessments must allow for modifications based on disabilities. All learners have distinct characteristics (see [Characteristics of Effective Teaching](#)) that make some assessment techniques more or less

appropriate. By using many sources of data, such learners are not unduly penalized.

### *Types of Assessment*

Many assessment processes may be used to gather data about students to provide appropriate evaluation. As mentioned above, the time to begin developing these assessments is when the course outline and materials are being developed. Additionally, these can be modified throughout the course depending on varying outcomes of instruction and/or other assessments.

Many textbook publishers make available "banks" of test items that have been developed for a specific text. Instructors should use these items cautiously! Many of these items have been developed by graduate assistants or other developers who have had little instruction in, or experience with, teaching and assessment. The quality of these items is often suspect and little is known about the psychometric properties of such test banks. However, all is not lost; many of these items can be made respectable with a small amount of work. Before using test bank items, instructors should read all items carefully and check them against the text. Reword the items for clarity. It is suggested that test bank items be matched against the test blueprint or table of specifications to ensure that coverage of item content is appropriate. Test bank items often focus on factual knowledge to the detriment of higher level skills such as application and integration.

### *Traditional Assessment*

Traditional assessments include those customarily called tests. In developing such instruments, it is useful first to develop a "test blueprint." This blueprint or Table of Specifications is a summary of the various learning outcomes and test items designed to tap at these outcomes. Differences in the number of items assigned to each category might reflect instructional time, material covered, or importance to later learning. An example of a

test blue print for an introductory statistics course is given below:

### *Other Bases for Evaluation*

Many instructors include performance on class assignments, class participation, term papers, etc., when determining students' final grades. Again, new instructors need to remember that it is very important to be explicit about the relative value of each component. The syllabus should contain this type of information. As is the case with other traditional forms of assessment, set standards, stick to them, and be consistent. Instructors should make only those concessions that can be extended to all students in the course.

### *Tapping Higher-Order/Critical Thinking Skills*

Examples of questions that the instructor might use to evaluate higher-order or critical thinking skills are described in [Characteristics of Effective Teaching](#). If an instructor plans to assess higher-order thinking, then this must be made explicit to students. They need ample opportunities to practice in class and to be taught to recognize the intent of such items on a test. A rationale for including these items should also be provided.

### *Developing Test Items<sup>19</sup>*

"Selected response" describes test questions that require students to choose one or more alternatives (multiple choice, true-false, matching, fill-in-the-blank), and "constructed response" refers to questions that require students to formulate their own answers (essay, short answer). The choice of selected versus constructed versus combined (selected response and constructed response items) format should depend on an instructor's purposes for the test. Constructed response questions can provide more opportunity for creative or divergent thinking than selected response items. Selected response items give little chance for demonstration of writing skills. However, it is probably a misconception that selected choice formats cannot be used to tap

higher order thinking skills. It is possible for "higher" levels of understanding to be evaluated using a selected response format. Advantages and disadvantages for each type of item are listed below:

Item Format	Advantages	Disadvantages
Selected Response	Easier to score- can use computer Easier to defend scoring Accepted methods of determining quality (reliability, validity, item analysis) Can sample over a broader area in same amounts of time Allow faster feedback time Cut down on grading time especially with large classes Easier to reuse-items can be "refined" using item analysis Different, equivalent exams can be generated from the same item bank with large numbers of items	Time consuming to construct May penalize divergent thinking Favors students who read well Cheating is easier (cheating can be minimized through use of alternative forms and/or controlled seating)

**Multiple Choice Items**

One selected choice alternative is the multiple choice item, which consists of one or more introductory sentences followed by a list of two or more suggested responses. The students choose

Item Format	Advantages	Disadvantages
Constructed Response	Less time-consuming to construct (but still needs time) Can provide more opportunity for creative or divergent thinking Can sample one area in more depth and more comprehensively Provides opportunities to improve writing Can be more revealing about specific student strengths and weaknesses Fewer opportunities for cheating during the test	Harder to score Harder to defend scoring Harder to grade reliably (one can use methods such as establishing model answers, holistic scoring, multiple raters-see below) Few accepted ways evaluate quality easily Harder to reuse-students remember a few items more easily than many

the correct answer from those responses. The part of the item that asks the questions is referred to as the *stem*. The terms *responses*, *choices*, *alternative*, and *distracters* are used to refer to the parts of the item that will be used to answer the question. "Distracters" refer to alternatives that *appear* plausible or that are often incorrectly associated with information in the stem.

An example of a multiple choice item is provided below:

**Stem**

A college instructor wishes to identify test items for which high-scoring examinees have a high probability of answering correctly and low-scoring examinees have a low probability of answering correctly. Her main goal would be to obtain which of the following?

### Responses

- a. high item difficulty indexes (distracter).
- b. high item discrimination (correct answer).
- c. high item-total correlations (distracter).
- d. both b and c (incorrect alternative).

### Constructing Multiple Choice Items

- Ask a direct question or clearly formulate a problem for the students to solve.
- Control vocabulary and sentence structure.
- Keep the stem brief, but include the important concepts so they need not be repeated in the responses.
- Avoid negatively worded stems. If negatively worded items are used, highlight the negative term.
- Avoid wording directly from the textbook.
- Keep all alternatives in a similar format (i.e., phrases, sentences, etc.).
- Make alternatives grammatically consistent with the stem.
- Make alternatives approximately the same length.
- Use misconceptions and plausible errors as basis for distracters.
- Have an equal numbers of alternatives for each item.
- Make sure that there is only one best or correct answer.
- Randomly distribute correct responses.

### True-False Items

A true-false statement consists of a proposition or statement that the student must evaluate as being either true or false. True-false statements may be the easiest type of item to construct because they

come directly from content. Some consider true-false items to be the least acceptable items because there is much more room for "guessing" than with other item types. Instructors may wish to consider alternate true-false items where two statements are given and the student must determine the trueness of each. In such cases, care must be taken so that the pairs of statements are related concepts. Such items tend to eliminate the

chances of guessing with ordinary true-false items. Another type of true-false item asks the student to correct the incorrect statements.

### Constructing True-False Items

- Create pairs of items, one true and one false, related to the same idea. Use only one this time, but save the other for another semester.
- Balance the number of "True" items and "False" items on each test.
- Keep language and wording simple.
- Avoid taking items verbatim from the text.
- Write false statements that reflect actual student misconceptions.
- Make sure the item is definitely true or definitely false.
- Avoid long statements.
- Avoid use of negatives and ambiguous or trick items.
- Avoid use of specific determiners such as "always," "never," etc.
- Focus on one idea.

### Matching Items

Matching items, which are variations of multiple choice items, are brief and used for *who*, *what*, and *when* questions. They permit efficient use of space and are easy to score. It is difficult to construct matching items that tap more than basic rote learning, but they may be the most efficient way to test memorization of definitions, dates, etc. An example of a set of matching items is provided below:

*Match each author with his/her work:*

- |                          |                            |
|--------------------------|----------------------------|
| 1. James Fenimore Cooper | a. <i>Ethan Frome</i>      |
| 2. Washington Irving     | b. <i>The Deerslayer</i>   |
| 3. Mark Twain            | c. <i>Rip Van Winkle</i>   |
| 4. Louisa May Alcott     | d. <i>Sister Carrie</i>    |
| 5. Edith Wharton         | e. <i>Huckleberry Finn</i> |
| 6. George Eliot          | f. <i>Robinson Crusoe</i>  |
|                          | g. <i>Silas Marner</i>     |
|                          | h. <i>Rose in Bloom</i>    |

### *Constructing Matching Items*

- Use only homogeneous material for each set of items. For example, do not mix names, dates, etc.
- Use more involved information in the stem and shorter answers for the responses.
- Supply clear, direct instructions for the basis of the match.
- Avoid giving grammatical cues.
- Make sure that there is only one correct response for each stem.
- Arrange items in the response column in a logical order---alphabetical, numerical, chronological.
- Provide more responses than stems to make process of guessing less likely.
- Use fewer than 15 stems per set; longer sets can be so cumbersome for students and those who are scoring the test, that they defeat the purpose of the assessment.

### *Completion/Fill-in-the-Blank Items*

These items are included under selected response items because there are a limited number of answers and the scoring is right or wrong. An instructor is well advised to have a list of acceptable responses available for scoring. This limits discussions about the appropriateness of certain answers. However, there may be acceptable answers that the instructor has not thought of. Completion items are probably the least efficient and flexible of all the selected response items. An example of a completion item follows:

1. Multiple choice, true/false, and matching are all referred to as \_\_\_\_\_ items. ,  
(Correct answer: "selected response".)

### *Constructing Completion Items*

- Word each item clearly so that the answer is a single word, brief phrase or number.
- Avoid taking questions directly from the text.
- The omitted word should be an important word.

- The blanks in all items should be approximately the same length.
- Avoid providing grammatical cues.

### *Constructed Response Items*

Many instructors use constructed response items such as short answer or essay questions to assess student learning. An advantage of essay and short answer items is in the depth of understanding that the student is asked to display. Essay/short answer items are relatively easy to construct, but may be more problematic and time consuming to score.

### *Constructing Essay/Short Answer Questions*

- Use novel problems or material when possible; make sure these problems are related to class material.
- Require more than recall of verbal information.
- Make the intent of the question clear.
- Clarify response length, purpose, time limits, evaluation criteria.
- Establish scoring criteria before students take the test. Two commonly used scoring systems focus on **analytic** and **holistic** criteria. For analytic methods, an outline or list of major elements that students should include in an ideal answer is compiled. **In** holistic scoring, the instructor makes a judgment about the overall quality of each student's response. Although analytic scoring has been shown to have greater agreement among raters, there may be some loss of qualitative interpretations if only analytic scoring is used.
- Decide before testing whether grammar, sentence structure, spelling, etc., is to be evaluated.

### *Performance Assessment*

Performance assessment is known by a number of different names, including alternative and authentic assessment. Performance tasks generally require students to demonstrate learning by producing more elaborate written or spoken answers, engaging in group activities, creating a product, or other more "authentic" methods. The

idea is that the expected performance or activity parallels more closely what might be expected in a "real-life" situation. Many areas such as the performing arts, fine arts, architecture, and laboratory and clinical classes have been using such methods for a number of years. However, such methods are relatively new to others. Data from performance assessment may be accumulated on video or audiotapes, in portfolios, etc. A complete discussion about all the advantages and disadvantages of alternative assessment as well as complete suggestions for implementing such procedures is beyond the scope of this handbook. However, the reader is urged to consult a number of sources such as Stiggins (1994) and Linn and Gronlund (1995). Some examples of alternative assessment are given in the following section.

### *Portfolios*

Portfolios may be used in many different ways. It is important to make clear the purpose of the portfolio and what goes into it. The clearer students are about learning objectives, the better they are able to design a portfolio. Although portfolios may serve multiple purposes at once, it is a mistake to expect that they will be able to do everything. An instructor may wish to limit the portfolio to a particular objective the first time this type of assessment is used. For example, a portfolio may consist of the best examples of what the student can do. If improvement or growth is an important criterion, students may be asked to include examples of their work at various stages of development. Students may also be asked to include reflective writings about their work and/or self-evaluations.

### *Evaluation by Jury and by Audition*

Both evaluation by jury and evaluation by audition are used primarily in the fine arts and architecture where such procedures are appropriate for evaluating a product such as a sculpture, landscape design, or music composition. Juries may consist of a group of instructors, a group of professionals external to

the university, peers, or a combination of persons from these groups. Similarly, auditions may involve evaluators from many different professional or academic groups. It is wise to check with departmental procedures for auditions and juried evaluations. A good source for information about juries is Anthony's (1991) book listed in the bibliography.

### *Scoring Rubrics*

Rubric evaluation involves a set of standards that is used to evaluate the level of performance in a student's work. Rubric assessment may involve numerical categories ( e.g., scores from one to three) or verbal categories ( e.g., "Exceptional," "Capable," and "Developing"). Standards for each rubric score or category are explicitly stated, often in behavioral terms. For example, "contains no grammatical errors" or "uses varied sentence structure" might be among the standards for the "Capable" category for evaluating an essay for a composition class. Because of the potential for increased subjectivity when using alternative assessments, establishing clearly defined scoring rubrics is essential. Scoring rubrics help assure more reliability in scoring. Rating scales comprise a type of scoring rubric (Linn & Gronlund, 1995), and can be very useful in evaluating performances.

### *Scoring Rubrics*

An important fact to keep in mind is that rubrics are under constant development and improvement. Precisely because they are often used to measure complex tasks, they themselves tend to be complex, and warrant continual review and revision where necessary in order to remain useful.

Below is a list of websites that might be helpful in rubric construction and testing. The reader should note that the resources listed below are merely a sampling of what is available on the web. These were chosen in large part because they are user friendly, and often provide information on several types of rubrics, as opposed to just one.

### General guides to designing rubrics

<http://pareonline.net/getvn.asp?v=7&n=3>  
<http://pareonline.net/getvn.asp?v=7&n=25>  
<http://edtech.tennessee.edu/~itce/5rubrics/rubrics.htm>  
<http://www.rubrics.com/>  
<http://learnweb.harvard.edu/alps/thinking/docs/rubricar.htm>

### Rubric generators for a variety of disciplines

<http://rubistar.4teachers.org/index.php>  
[http://www.teach-nology.com/web\\_tools/rubrics/](http://www.teach-nology.com/web_tools/rubrics/)  
<http://www.techtrekers.com/rubrics.html>

### Examples of rubrics

<http://www.bethel.edu/~tanswe/ens102/Rubric.html#oral>  
<http://www.calstate.edu/AcadAff/SLOA/links/rubrics.shtml>  
[http://www.cde.state.co.us/cdeassess/csap/rubrics/as\\_wri\\_rubrics.htm](http://www.cde.state.co.us/cdeassess/csap/rubrics/as_wri_rubrics.htm)  
[http://www.cde.state.co.us/cdeassess/csap/rubrics/as\\_math\\_rubrics.htm](http://www.cde.state.co.us/cdeassess/csap/rubrics/as_math_rubrics.htm)  
[http://www.nwrel.org/msec/science\\_ing/guides.html](http://www.nwrel.org/msec/science_ing/guides.html)  
[http://itc.utk.edu/newsletter/spring\\_04/rubricsmultimedia.shtml](http://itc.utk.edu/newsletter/spring_04/rubricsmultimedia.shtml)

## *Administering Assessments*

### *In Class Assessment*

Furthermore, a statistical report will be provided to the instructor, which includes various types of item analyses. For further information on the services available, contact the UCS scanning and multimedia group at 285-1358, or visit their website at <http://learnweb.harvard.edu/alps/thinking/docs/rubricar.htm>.

Selected response items lend themselves very well to computer scoring. Instructors can administer these items and have students respond on scan sheets. The responses can then be directly scanned into a computer and given back to the instructor in the form of electronic media such as a diskette. See *determining the Quality of Assessments* for more details.

Clegg (1994) makes the following suggestions for administering exams:

- **Monitor your own tests when possible.** This enables students to ask questions about test items that are unclear. If the instructor cannot be present, someone who has knowledge of the area should proctor the test. Often TA's are asked to monitor exams that are outside their area of expertise. It is important for the instructor to go over the test with the proctor to enable the proctor to more intelligently answer questions.
- **Remind students the class meeting before the exam** if they are required to bring certain supplies or resources (such as a #2 pencil or calculator) for the exam.
- **Try to be as fair as possible in distribution of tests.** If the class is very large, assistance may be needed to distribute papers so that no one has an unfair time advantage. It is wise to start with different distribution points for subsequent tests so that some students are not frustrated by the fact that they are always the last to begin the test.
- **Encourage legitimate questions** (no one writes perfect items). However, instructors should limit the amount of help that will be appropriate. It is O.K. to indicate that answering a particular question is beyond the scope of providing clarification.
- **If tests are thoroughly proofed and reviewed by a colleague, corrections should be kept to a minimum.** However, sometimes errors occur that have been overlooked by even the most careful proof-reader. In such cases, all students should be informed of the error as quickly and straightforwardly as possible.
- **Minimize test anxiety long before testing is to take place.** For example, sharing the types of questions to be encountered, making sure that the objectives and expectations are explicit, etc., will help. Instructors can also help by keeping the testing situation as calm as possible. Arriving at the classroom slightly earlier than usual, keeping one's voice low

and speaking distinctly, and bringing extra supplies for emergencies, all add to an overall positive atmosphere.

### *Computer Administered Assessment*

Many assessments can be administered on the computer. A nice feature available at Ball State is the inQsit program for computer assessment. The inQsit system was developed at Ball State and provides a number of different testing formats. To ensure security, students can take exams in proctored labs. Professors who use the system report many advantages. It is fairly easy to use and is available to all instructors. Instructors who are interested in finding out more can do so at <http://www.bsu.edu/inqsit/info/index.cgi>. We particularly recommend the *inQsit Quick Start Guide* that can be accessed through <http://www.bsu.edu/inqsit/info/index.cgi>.

### *Returning Exams*

Exams should be returned as promptly as possible. Delay increases anxiety and may eliminate the opportunity for students to get feedback they need to correct misunderstandings and to avoid future mistakes. There is nothing more distressing for students than to take a second exam before they have the results from the first one. Returning and discussing test results can aid instruction in a number of ways. First, students receive specific information about the instructor's tests. Second, the instructor can provide specific suggestions for improvement on certain types of items. Third, students can get information about their misunderstandings of the subject matter and what kinds of midcourse corrections they will need to make in their study habits.

Confidentiality of test results is extremely important. Names or social security numbers should never be used in posting grades. Providing an alphabetical roster of the class and merely assigning an i.d. number is also questionable practice; students are able to figure that one out fairly easily! Interestingly, even the practice of

showing a frequency distribution of scores for an entire class, where individual students are assigned only an "X," can lead to consternation on the part of some students.

### *Determining the Quality of Assessments*

After a test has been administered, it is useful to evaluate its quality, especially if the test items are to be used again. Items that have been problematic for the students can be reworded or rewritten. Once items have been refined (see following sections for suggestions on item improvement), using the same items over also allows for comparisons among groups and for the instructor to have a good understanding of the item's content and purpose.

Constructed response assessments create a dilemma for the instructor because evaluation of content can be very subjective. One of the best ways to evaluate test items is to ask other instructors to review assessments before administration. Another way to ensure more reliability in scoring is to establish scoring rubrics prior to the examination so that all students will be judged equitably. To assure anonymity of the student, develop a system whereby students cannot be identified until after grading is completed.

With selected response items, computer scoring and test analysis that provide information about the reliability of the test as well as individual items are available through Examination and Scanning Services (RE 237; 285-1506). To use the service, students must complete a machine-scorable answer sheet, and the instructor must provide a scoring key on the same kind of form. "Cover sheets" that provide the service with identifying information and special instructions are available at Examination and Scanning Services or in many departments. Turn-around time is short, usually within hours. The service is very reliable and can be useful for feedback to

students because incorrect items are noted on individual answer sheets.

### *Reliability of a Test*

Reliability refers to how well results of the test can be trusted. Instructors are often concerned with one type of reliability called internal consistency, which has to do with the consistency of items on the test itself. Internal consistency reliability can have a number of names (Cronbach's alpha, KR-20 and KR-21, and split-half); each refers to the same property but is calculated slightly differently. All estimate how well the items of the test are related to one another, i.e., whether the items of the test essentially measure one construct, such as mathematics. Keep in mind that the consistency of a test will depend, in part, upon the number of items on the test, with longer tests being more reliable than shorter tests.

### *Validity of a Test*

Perhaps the most important consideration for evaluating the quality of a test is its validity. Test validity refers to how well the test measures what it is designed to measure, i.e., whether the test items measure the content of the material covered in class. The most frequent, and often justified, criticism that students have about tests is a lack of content validity. Most often this happens when attention has not been paid to the correspondence between instructional outcomes and goals and test items. This also happens when textbooks are changed, but tests are not. Low content validity can also occur when test items are indiscriminately chosen from a test bank with little attention paid to the quality of the items and/or number of items that deal with any aspect of content.

### *Item Analysis*

Item analysis is used to examine the quality of individual test items and may be the most useful information for college instructors. However, the

usefulness of item analysis depend upon an instructor's goals. A discussion of how to use item analysis information, that takes into account the instructor's goals, is provided in Linn and Gronlund (1995) and in Gronlund (1993). Typically, two pieces of information are given: item difficulty and item discrimination.

**Item difficulty** refers to the proportion of students passing a particular item. This item statistic is probably misnamed because it actually refers to how easy an item is rather than how difficult it is. For example, if 50% of students pass an item, the item has a difficulty level of .5 (medium); if 75 percent of the students pass, .75 (easy), and if only 15% pass, item difficulty is .15 (hard). No absolute standards exist for determining what is "good" item difficulty. In fact, the desired difficulty level of items depends on the purpose of the test. For example, if the goal of the test is to evaluate mastery, higher item difficulties ( e.g., .75) might be common because most students are expected to answer most items correctly. On the other hand, greater discrimination between high and low performers can be established using items of medium (.5) difficulty. Items that are too difficult for most students may be used, but keep in mind that this difficulty should not stem from lack of instruction, poorly written questions, or other content-irrelevant factors.

**Item discrimination** refers to the ability of an item to separate good performers from poor performers. In other words, how did this item work for the top 1/3 of the students versus the bottom 1/3 of the students? One would expect that, overall, the top students would perform better than those at the bottom on an item-by-item basis. When this does not occur, negative item discrimination values are obtained. When items have negative discrimination values, it is wise to reexamine the items in question to determine why student performance was not as expected. Negative discrimination values can occur when items are were keyed wrong (we all do it!) and when distracters are misleading. If the cause is not

obvious, it can be good practice to ask high scorers why they missed a particular item.

### *Assigning Grades*

Grading systems in general can be divided into two categories: absolute and relative (Higgins, 1994). In absolute grading systems, grades are assigned in comparison with established standards or criteria. In such systems, students are not limited in their grades because of their position relative to classmates. In such criterion-referenced systems, all students are able to obtain the maximum grade if their performance reaches certain standards. For example, an instructor is using an absolute system when assigning point values or percentages to particular grades ( e.g., 75 of 100 points equals a "Pass," or 90% equals an " A "). In such systems, it is important to communicate class standards to students. Such systems have the advantage that if students work and study hard and perform well, theoretically, all can obtain good grades. Also, instructors can take advantage of using some alternative teaching methods such as cooperative learning and group projects without students being concerned about jeopardizing their own grades. However, for some instructors this may seem like a disadvantage because concerns about grade inflation may arise and the instructor may have to explain why all students did so well. New instructors may find it difficult to know what standards they should set. Consulting with instructors who have experience with the course may be helpful.

In relative grading systems, comparisons are made among students. Relative systems are often thought of as "grading on a curve" because of the reference to the normal distribution called the "bell curve." Relative grading systems are more appropriate when large numbers of unselected students are being compared with each other. Few relative grading systems at the university level of education actually follow the bell curve religiously, despite the fact that instructors often refer to curving their grades. While it is common

for instructors to adjust their final distribution of grades, it is typically not appropriate to use the strict bell curve as a basis for assigning grades in many college courses. Furthermore, relative grading is almost never justifiable for graduate courses because of the highly select nature of the group. There are some serious pitfalls associated with relative grading. First, no matter what the quality of a group of students is, there are some students who will not do well, and conversely, no matter how poor the students, there are some who will do very well. Therefore, the meaning of grades fluctuates with the composition of the group. Grades are difficult to interpret with a relative grading system based on a single class or year. Instructors can overcome some of the disadvantages of such a system by developing norms based on several years of students or by using departmental data for a large number of students.

One caution concerning development of normative comparisons from data from a number of classes and/or years: If tests and/or textbooks are changed, the basis for comparisons may be fundamentally flawed.

### *Standards for Grading*

- Have standards for grading in place before the beginning of classes. Include in the syllabus information about the types and schedule of examinations, quizzes, papers, etc.
- Relate the grading procedures to the intended learning outcomes.
- Return and review all assessment results as soon as possible. Plan to have some evaluation completed before the midterm grading period.
- Be explicit about the relative weighting of various types of information and the translation of these into final grades.
- Be fair. Avoid bias. When in doubt, review the evidence. If still in doubt, assign the higher grade.

- Document decisions carefully. Be consistent. Keep grade books secure and retain them after the course is over.
- Specify in advance if attendance and/or participation will be included in the grade. Make expectations explicit.
- Check on departmental policies for assigning grades.

### *Electronic Grade Books*

Communicating with students about grades can provide many disconcerting experiences. The best way to deal with questions is to be prepared for them. Evaluating and assigning grades can be an onerous task. However, it is very important to provide students with timely feedback. Instructors may choose to keep their grade books locally through age-old paper pencil methods or through use of spreadsheets or other programs. However, Ball State has two web-base programs that not only provide storage but also can be used to communicate to students and to submit grades to the Registrar's office. There is a grade book located within the blackboard system as well as a stand alone Electronic Gradebook. It appears that there will soon be a web-based system available that will allow evaluation of performance-based assessments using rubrics.