

---

## TESTING, HOMEWORK, AND GRADING

---

For many students, grades constitute the number-one academic priority. Tests, or any other means professors use to determine grades, are the number-two priority. Because of this concern about grades, tests and scoring of tests generate a great deal of anxiety which can translate into anxiety for the professor. It is easy to deplore students' excessive focus on grades; however, this excessive focus is at least in part the fault of the professor. In addition, a student's focus on grades and tests can be used to help the student learn the material.

Testing and homework can help the professor design a course which satisfies the learning principles discussed in Section 1.4. Homework and exams force the student to practice the material actively and provide an opportunity for the professor to give feedback. With graduated difficulty of problems, the professor can arrange the tests so that everyone has a good chance to be successful at least initially. This helps the professor approach the course with a positive attitude toward all the students, which in turn helps them succeed. The desire to achieve good grades can help motivate students to learn the material, particularly if it is clear that the tests follow the course objectives. Anxiety and excessive competition can be reduced by using cooperative study groups. Thought-provoking questions can be used both in homework and in exams to use the students' natural curiosity as a motivator. Students can be given some choice in what they do in course projects.

Although testing and homework can help the professor satisfy many learning principles, they also can serve as a barrier between students and professors which inhibits learning. It is difficult for students to truly use the professor as an ally to learn if they know he or she is evaluating and grading them (Elbow, 1986). Perhaps the ideal situation would be to completely separate the teaching and evaluation functions. One professor would teach, coach, and tutor students so that they learn as much as possible. Then a second professor would test and grade them anonymously. An alternate method with which to approach this ideal can be obtained with mastery tests and contract grading (see Section 7.4). If these alternatives are not

possible, there will always be tension between learning on the one hand and testing and grading on the other. In the remainder of this chapter we will assume that you have resolved to live with this tension.

Why does one test and how often does one test? What material should be included on the test? What types of tests can be used? How does one administer a test, particularly in large classes? These are the questions we'll consider in this chapter. Then our focus will shift to scoring tests and statistical manipulation of test scores. Homework and projects will be explored. How much weight should be placed on homework? How does the professor limit procrastination on projects? Finally, the professor's least favorite activity, grading, will be considered from several angles.

## 11.1. TESTING

Testing requires careful thought. Fair tests which cover the material can increase student motivation and satisfaction with a course. As long as a test is fair and is perceived as being fairly graded, rapport with students will not be damaged even if the test is difficult. Unfair and poorly graded exams cause student resentment, increase the likelihood of cheating, decrease student motivation, and encourage aggressive student behavior.

### 11.1.1. Reasons for and Frequency of Testing

There are many educational reasons for having students take tests. Tests motivate many students to study harder. They also aid learning since they require students to be active, provide practice in solving problems, and offer feedback. Tests also provide feedback for the professor on how well students are learning various parts of the course.

Tests are stressful since they are so closely associated with grades. Stress and pressure are part of engineering. Mild stress can actually increase student learning and performance on tests, but excessive stress is detrimental to both learning and performance for students and practicing engineers. In addition, exams can be stressful for the professor because they are so tightly coupled with grades. What can be done to harvest the benefits of tests while simultaneously reducing the stress they induce?

Give more tests!

Giving more tests reduces the stress of each one since each exam is less important in deciding the student's final grade. Courses with only a final or a comprehensive exam make the test enormously important and thus very stressful. If there are four tests during the semester, each one is significantly less important. If there are fifteen quizzes throughout the semester, then each quiz has a modest amount of stress associated with it. Having frequent tests or quizzes also allows professors to ignore an absence or discard the lowest quiz grade.

Frequent testing spreads student work throughout the semester, which increases the total amount of student effort and improves the retention of material. The more-frequent feedback to the students and to the professor is beneficial. Both the students and the professor know much earlier if the material is not being understood. The increased forced practice, repetition, and reinforcement of material aids student learning. Because stress is reduced, frequent testing serves as a better motivator for students. The net result is improved student performance (Johnson, 1988). One of the advantages of PSI and mastery courses is that they require frequent testing (see Chapter 7). Frequent exams also provide a more valid basis for a grade since one bad day has much less of an effect.

Frequent tests do have negatives. The considerable amount of class time required may reduce the amount of content that can be covered; however, the content that is covered will probably be learned better. A considerable amount of time may also be required to prepare and grade the frequent examinations. At least some of this time is available since less homework needs to be assigned when there are frequent exams. Perhaps the most important drawback of frequent tests in upper-division courses is that they do not encourage students to become independent, internally motivated learners.

We have adopted the following compromise solution to the question of how frequently to test. In graduate-level courses we give infrequent tests (two or three a semester) but usually have a course project which represents a sizable portion of the grade. In senior courses we use slightly more tests (three or four). In junior courses, despite the great deal of material to be covered, we increase the number to six or seven during the semester. In sophomore courses where there is often little new material to learn but students need to become expert at applying it, we have gone as high as two quizzes per week (and no homework). For these courses one quiz per week seems to work well. This frequency may also be appropriate for computer programming courses. Frequent quizzes ensure that students are practicing the material and are receiving frequent feedback.

What about finals? There are very mixed emotions about finals (for example, see Eble, 1988; Lowman, 1985; McKeachie, 1986). Finals do require students to review the entire semester and to integrate all the material. They can also be useful for slow learners and for those who initially have an inadequate background since they allow these students to show that they have learned the material. Finals are also useful for assigning the course grade. Unfortunately, they are very stressful for students and are almost universally disliked. In addition, feedback to the professor is too late to do any good in the current semester. To the students it is almost nonexistent. Many students look only at the final grade and do not study their mistakes on the test.

A professor choosing to give a final has several interesting options which can reduce the stress. If other tests have been reasonably frequent during the semester, students can be told that the final can only increase but not decrease their grade. When this is done, it may make sense to tell students their current earned grade and then make the final exam optional. In PSI and mastery courses an optional final can be used as one way to improve students' final grades with no risk. Another option is to give a required final but tell students that their grades will automatically be the higher of their composite grade for the entire course or their grade on the final. The reasoning behind this strategy is that it makes sense to give high grades to students who prove at the end of the semester that they have mastered the material, but having only a

final is too stressful. In this way you are also rewarding them for what they know at the end of the term instead of penalizing them for deficiencies they may have had at the start of the semester. Feedback can be made more meaningful by going over the final in a follow-up course the next semester.

Many universities have a scheduled finals period. If the professor decides not to have a final, this time may be used for other purposes. In a course with projects, the final examination period is an excellent time for student oral reports on projects. This period can also be used for a last hour examination which is not a final. One advantage of using the finals period for an hour examination is that more time is usually allotted for the final, and students taking an hour examination during this period have sufficient time to finish even if they work slowly.

One additional type of quiz is the unannounced, surprise, or “pop” quiz. Some professors like to give several of these during the semester. After answering questions the professor announces there will be a pop quiz. Once the students’ groans subside, a short quiz is administered. The advantages of pop quizzes are that they help keep students current and they reward attendance. The major disadvantage is that they increase stress. This increase in stress can be controlled by:

- 1 Noting in the syllabus that there will be unannounced quizzes.
- 2 Making the quizzes a small fraction (2 to 3 percent) of the course grade.
- 3 Giving some points for the student’s name (i.e., rewarding attendance).
- 4 Throwing out the lowest quiz grade. This helps students who miss a class which happens to have an unannounced quiz.
- 5 Making the quizzes short (five to ten minutes).

### 11.1.2. Coverage on Tests

How does a professor decide what to put on a test? If objectives have been developed for the course, the decision is relatively simple. The important objectives are tested. At what level in Bloom’s taxonomy (see Chapter 4) should the test be? If at the higher levels, then the test questions need to be evaluated for appropriateness.

An effective method for ensuring that the test covers the objectives appropriately is to develop a grid (Svinicki, 1976) as illustrated in Figure 11-1. For each objective or topic, think of a question or problem which allows you to test at appropriate levels of Bloom’s taxonomy. It may not be necessary to have any problems which are solely at the knowledge or comprehension levels since these levels are usually included in higher-level problems.

Once the preliminary grid has been developed, you can check it to see if the proposed test satisfies your goals for a particular section of the course. Since not all objectives or topics can be included at all levels of the taxonomy in a single test, you need to make some compromises. Is the coverage of topics on the test a fair representation of the coverage during lectures and of the homework? If not, the exam probably is not a fair test of the course objectives, and students are likely to think it is unfair. Although not all topics can be covered, one should try

FIGURE 11-1 EXAMPLE GRID FOR TEST PREPARATION

Objectives or Topics	Level					
	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
1	X		X			
2		X		X		
3	X	X				X
4	No problem for this objective					

to have reasonably wide coverage. If a topic is discussed in two separate parts of the course, it might be reasonable to include it in one test and not the other. The levels of the questions also need to be considered. If higher-level activities are important, they need to be included in homework and in tests. Without a conscious effort, it is highly likely that only the three lowest levels will be used since questions at these levels are the easiest to write (Stice, 1976). For the grid shown in Figure 11-1, the instructor has decided not to test for objective 4 or to include any questions at the evaluation level on the test.

Should the test be open book or closed book? The argument in favor of open book tests is that practicing engineers can use any book they want to solve a problem. Open book tests also reduce stress. One argument against them is that too many students use the book as a crutch and try to find the answer in the book instead of by thinking. Another opposing argument involves logic. The practicing engineer argument relies on a false analogy because the purpose of the open book is different: Unlike students, these engineers are not being tested on their knowledge. One problem with closed book tests is that students may be forced to memorize equations which they would always look up in practice. Closed book tests may encourage memorization of all content and not just the equations.

Some compromise arrangements are between the extremes of open book and closed book tests. The instructor can prepare a sheet of important equations for students to use during the exam and hand this sheet out to them before the test so that they know what will be available for the test. When the exam is administered, each student receives a clean set of equations. The advantage of this compromise is that the professor has control over the information each student has available during the test. Another compromise is to allow each student to bring a key relations chart (see Section 15.1) on one piece of paper or an index card. The advantage of this procedure is that students benefit from preparing the chart and often do not glance at it during the test.

### 11.1.3. Writing Test Problems and Questions

How does one write the problems or questions for tests? What style of questions is appropriate? This section discusses some general rules for writing exams and then explores specific formats for questions.

In writing examination questions, avoid trivial questions even when testing at the knowledge level. Avoid trick questions also since they do not test for the student's understanding and ability in the course. Problems should be as unambiguous as possible unless you are explicitly testing for the ability to do the *define* step of problem solving. To test for clarity have another professor or your TA read the test and outline the solutions. The time required for the exam can be estimated by taking the time you require to solve the problems and multiplying by a factor of about 4. The number of points awarded for each problem should be clearly shown on the test so that students can decide which problem to work on if time is short.

Solve the problems before handing out the test. This aids in grading and helps to prevent the disaster which will occur if an unsolvable problem is on the exam. (If you want the students to perform a degree of freedom analysis to determine if the problem is solvable, then it is reasonable to have an unsolvable problem on the test. However, warn them ahead of time that this may happen; otherwise, they will assume all problems are solvable.)

If tests are returned to students (which is a useful feedback mechanism), then you should assume that files exist on campus for all old exams. Even if you require students to return tests after they have seen their grades, you should assume that at least rudimentary files exist. Since the purpose of a test is to determine how much a student has learned and not who has the best files, you should write new tests. If exams are given frequently, this is a considerable amount of work. Once a large number of questions, particularly of the multiple-choice variety, have accumulated, you can recycle a few questions on each test. Old test questions do make good homework problems, and students appreciate the opportunity to practice on real test problems. Since some students have files, many professors provide files of old tests so that everyone has equal access to information. Most university libraries place test files on reserve. Another more drastic solution to the file problem is to periodically revise the curriculum and reorganize all the courses.

Although it may sound contrary to the previous advice, we suggest that every once in a while a homework problem should be put on a test. This rewards students who have diligently solved problems on their own and is a clear signal to students that they should work on the homework.

How does the professor generate interesting problems which test for the objectives at the correct level but are not clones of textbook or homework problems? One way is to take an existing problem and do permutations of which variables are dependent and which are independent. Changing the independent variable often changes the solution method remarkably. Brainstorm possible novel problems. Use problems from other textbooks (but if this is done consistently, some students will catch on). Set up an informal network with friends at other universities to share test problems and solutions. As part of their homework assignments have students write test problems. The occasional use of one of these will reward the student who made it up. (In our class on teaching methods the second test is based entirely on student-generated questions.) Don't wait until the last minute to start generating problems. It is often productive to generate ideas throughout the semester. Then, the details of the problem and the solution can be worked out when the exam is made up.

Test problems usually fit into one of the following categories: short-answer, long-answer, multiple-choice, true-false, and matching. Since true-false and matching have scant use in engineering, they will not be considered here but are discussed elsewhere (Canelos and Catchen, 1987; Eble, 1988; Lowman, 1985; McKeachie, 1986).

**Short-answer.** Short-answer problems include problems requiring identification of a principle, a brief essay, and short problems. In engineering, short problems are the most common. As long as complete long problems are also employed, short problems are an excellent way to determine if students have mastered certain principles. These problems are set up so that three to five lines of calculation give the desired answer. The problem is tightly defined so that the student is tested for application to a single principle.

Short-answer problems can also be used to develop students' skills as problem solvers. The problem focuses on one or two stages in the problem-solving strategy. For example, students can be asked to define the problem clearly but not solve it. Or, they can be given a "solution" to the problem and asked either to check the solution or to generalize it. Students need instruction in doing this type of short answer problem since they always want to calculate.

**Long-answer.** Long-answer problems include essay and complete long problems. In engineering, complete problems are probably the most common type of test problem. They are necessary to determine if students can find a complete solution. Unfortunately, an exam consisting entirely of a few long problems cannot test for all the objectives covered in the course. Thus, a mix of both long- and short-answer problems is often appropriate. Long-answer problems can also be difficult to score for partial credit (see Section 11.2.1).

**Multiple-choice.** With the regrettable but probably inevitable increase in class size at many engineering schools, multiple-choice examinations will become increasingly popular. They are easy to grade and, if properly constructed, can be as valid as short-answer questions (Kessler, 1988). Unfortunately, proper construction of the classical type of multiple-choice question is more time-consuming than constructing a short-answer question. Thus, the professor transfers some of her or his time from grading to test construction. This trade makes sense only with large classes.

General rules for constructing classical-style multiple-choice questions are given by Eble (1988), Lowman (1985), and McKeachie (1986), while examples for particular engineering courses are presented by Canelos and Catchen (1987) and Leuba (1986a,b). The stem, which is the question itself without the choices, should be complete, unambiguous, and understandable without reading the choices. The correct answer and the incorrect answers (the distractors) should be written as parallel as possible. Thus, all possible answers should be grammatically correct and about the same length. There should be no "cues" which allow a good test taker who is unfamiliar with the material to discard any of the distractors or to pick the right answer. Most authors suggest a total of four choices, all of which should appear reasonable. The instruction should ask the student to pick the "best" choice so that arguments with students can be minimized.

In writing a multiple-choice question, the professor usually starts with a short-answer problem. The correct answer is then obvious. Indicate that the answer is a number within a given percentage (say, 1 percent). The challenge lies in choosing distractors. If a similar short-answer question has been used in the past, look at the students' solutions to find common errors. Then construct the distractors so that the numerical answer follows from these common student mistakes. Most authors suggest that "none of the above" is an improper distractor or answer. Once the distractors have been written, randomly assign the answer and the distractors as a, b, c, and d.

When questions have numerical answers, there is a clever alternate type of multiple-choice question (Johnson, 1991). For each question, list ten numbers in numerically increasing order. Tell the students to select the choice nearest to their calculated answer. If the calculated answer is the average of two adjacent choices, tell them to select the higher choice. The effort in writing distractors is thereby reduced. Now all you have to do is to pick choices over a feasible range at reasonably narrow intervals. This procedure also reduces the probability of a guess being correct. With the usual type of multiple-choice question the student who doesn't get one of the listed answers knows that he or she has made a mistake, but this procedure does not provide this clue. In addition, if you initially make a mistake solving the problem or there is a typographical error in the problem statement, all is not lost. As long as the problem is solvable, one of the choices is correct.

One of the advantages or disadvantages of multiple-choice questions (depending upon your viewpoint) is that there is no partial credit. Students who know how to do the problem but who make an algebraic or numerical error will receive the same credit as students who have no idea how to do the problem. Since numerical and algebraic errors cause loss of all credit, we suggest that multiple-choice questions be used only to replace short-answer questions and not long problems. Both multiple-choice and one long-answer problem can be included on a test. This will significantly reduce the grading in a large class without significantly decreasing the validity of the test.

Tests are stressful for students. This stress can be reduced by providing space on the examination for student comments. Tell the students the purpose of this space and explain that the comments will not affect their grades. Then, when you read a comment which says "This problem stinks," you will realize that the student is just letting off pressure.

#### 11.1.4. ADMINISTERING THE TEST

The first part of administering a test occurs the class period before it is given. Discuss the exam with the students. Clearly state the content coverage by telling them which book chapters and which lecture periods will be covered. Explain the type of test and show a few old problems as examples. Discuss the ground rules, such as staggered seating, closed book or open book, time requirements, and so forth. Particularly for lower-division students, it is helpful to give a few hints on studying and test taking.

Many instructors find optional help sessions useful. If you plan to have an optional help session, set the rules for the session first. We hold help sessions in which students must ask questions. When the student questions stop, the help session is over. If a student asks a question which is very similar to a test problem, the best idea is to answer the question in exactly the same manner as you answer other student questions.

McKeachie (1986) suggests making up about 10 percent extra exams. It is easy for the secretary to miscount or to collate a few exams with blank pages. The extra copies allow you to rectify these problems quickly. Take reasonable precautions to safeguard the test copies, such as locking them up in a briefcase or desk in a locked office.



To students the exam is one of the most important parts of the class, so plan on being there if it is at all possible. As the professor, only you can answer student questions properly and help students understand what they are supposed to do. In addition, if a student finds a typographical error, only you can make last-minute changes to correct the problem. Professors usually have better control of the class than do TAs.

Come early and have the TAs come early. This gives you time to check the lighting, straighten up the chairs, and start to arrange the students in alternate seats. Plan to pass out the tests as quickly as possible to give everyone equal time. In very large classes put a cover sheet on the exams and tell the students not to open them until given the signal to start. Have them put their names on the test immediately. Then have them count the questions to be sure they have a complete test.

If your school does not have an honor code, it is traditional to proctor the examination. It is also helpful to have someone present to answer student questions. A circulating proctor can do wonders in reducing the desire students might have to cheat. A TA standing discretely in the back of the room can also be a major deterrent. It is much better to prevent cheating than to deal with it after it has occurred (see Chapter 12).

Periodically write on the board the time remaining. Then state, “You have two minutes, please finish your papers.” When the time is up, stop the class firmly and collect the papers. It is best to give tests where there is effectively no time limit, but this is often difficult to schedule.

As soon as the examination is over, count the tests. Then check them in against the student roster. It is best to know immediately if a student has not handed in a test or was not present. Students have been known to occasionally complain that their test was lost.

## 11.2. SCORING

We will draw a distinction between scoring tests, which has a feedback function essential for the student’s learning, and grading, which is a communication at the end of the semester of how well the student has done in the course. Grading will be discussed in Section 11.5. Unfortunately, both of these activities are often called grading.

### 11.2.1. Scoring Tests

Extra effort taken while preparing an examination is recovered when the tests are scored. Multiple-choice tests can be machine-scored or with a homemade stencil. In fact, the attractiveness of multiple-choice tests for large classes lies in the ease of scoring.

For other tests an answer sheet and a detailed scoring sheet should be prepared by you as the professor. Evaluation is difficult, and a professor can do a better job than a TA in preparing both the answer sheet and deciding the breakdown of points. The scoring sheet should be developed for the “standard solution.” The TA should be instructed to show you unique

solution paths. Occasionally, a student develops a creative solution path but makes a numerical error and gets the wrong answer. To avoid dampening creativity, it is important that you carefully consider these alternate solutions.

Whoever scores the test should do so without looking at the name. Students should receive the score that they earn, not the score that the grader thinks they should earn. Extremely important tests such as qualifying examinations should probably use a code letter for every student instead of a name. It is best to grade every test for one problem before grading a second problem. This procedure helps to ensure that grading is uniform. For a series of short-answer questions it might be feasible and faster to grade the entire sequence on each test paper before proceeding to the next. After one problem has been graded on all tests, review the scoring, particularly of the first few tests that were graded. Be sure that the scoring is uniform.

For long problems it is often useful to look at a few sample tests before grading everything or before giving the tests to the grader. The sample tests may show a common mistake that will require adjustment of the grading scheme, or they may indicate a second correct solution path. If a grader is available, sit down with him or her for a few minutes and go over both the solution and the scoring sheet. Indicate the type of feedback you want put on the tests. Give the TA or the grader a reasonable deadline for return of the exams as well as some hints on how to grade that type of test. Tell the TA to bring in any nonstandard solutions so that you can check them over.

We believe in awarding partial credit for long problems. Crittenden (1984) presents the opposite viewpoint that partial credit should either be given sparingly or not at all. Our reason in favor is that students can often demonstrate understanding of how to solve a problem and not have the correct solution because of a relatively small error in technique, an algebraic error, or a numerical error. On the other hand, students also need to realize that engineers must be accurate. Problems without partial credit can be given as short-answer or multiple-choice questions.

If partial credit is to be awarded, develop the scoring sheet for the standard solution. Do this in advance and then adjust it after looking at a few tests. You can determine partial credit by awarding points for parts of the solution that are correct or by subtracting points for parts that are wrong or missing. In long problems these two approaches often result in different scores, and if a scoring sheet is not used will certainly result in different scores. For the highest reliability use a scoring sheet and calculate a score by adding positive items and subtracting negative ones. Discrepancies in the results obtained are a signal that the scoring needs to be reconsidered.

In addition to scoring the exam, provide written feedback and marks on the test or instruct the TA to do so. Correct parts of the test can be indicated quickly with check marks, while incorrect parts can be crossed out. Be sure that there is some mark on each page, including empty pages, so that the student will be sure that every page has been seen. Both positive and negative comments should be written on the test. Comments which explicitly correct the student's work are much more useful than writing "wrong" or "incorrect" without explaining why. Positive comments such as "good" or "clever derivation" serve as motivators.

To be effective, feedback must be prompt. Ideally, feedback would be given immediately after the student has finished the test. This procedure is used in some PSI classes (see Section 7.4). In large classes it takes longer to grade tests, but there is no excuse for taking a month or longer to return tests. If possible, hand them back the next class period. If that is not possible,

be sure to return them within one week. Tell the TAs in advance which weeks there will be tests so that they can arrange to have sufficient time to grade the exams quickly. Ericksen (1984, p. 119) believes “this business of immediate feedback is overdone.” He suggests taking more time to do detailed critiques and evaluations.

If it is to be useful, students must pay attention to the feedback. There are several methods that can be used to ensure that this happens.

**1** Hand back the test and discuss it in class. A variant of this is to have small groups discuss the exam. This procedure is useful since it can reduce student aggression.

**2** Before discussing the solution, assign one of the test problems as a homework.

**3** Give one or more of the problems on a second test.

**4** Ask students who obviously do not understand the material to see you privately. Student scores on exams are private, privileged information. Write the score on the inside or fold the test over when returning the test papers. If grades are posted, use student numbers or code letters instead of names.

### 11.2.2. Data Manipulation and Critiquing the Test

After the test fix up any problems which are not quite perfect for later use as homework or in that book you will write someday. Correct any typographical errors on all copies of the test you keep and in your computer files. If some students misinterpreted the problem, reword it so that this will be less likely to occur in the future. Perhaps one of the misinterpretations will give you an idea for an alternate test problem which can be used next year. Write the idea down and put it into your test file for future use.

It is easy to determine if an exam problem discriminates between students who do well on the test and those who do poorly. Johnson (1988) suggests a simple procedure for doing this. Separate out the tests of the ten (or fifteen in large classes) students with the highest scores on the test and of the ten (or fifteen) students with the lowest scores. For problems where no partial credit is given, let  $H$  = number of top ten students who got the problem correct, and  $L$  = number of bottom ten students who got the problem correct.

The test has positive discrimination if  $H - L > 0$ , and negative discrimination if  $H - L < 0$ . If a problem has negative discrimination, the better students are having more difficulty. These problems need to be rewritten. The sum of correct scores,  $H + L$ , can also be looked at. Johnson (1988) suggests that this figure should be between 7 and 17 (except in mastery courses where 20 may be reasonable). If partial credit is given, the discrimination of each item can be determined by looking at the sum of scores for the ten best and for the ten worst students.

In large classes (more than twenty students), standard scores can be useful for comparing student scores on different tests and for deciding final grades (Cheshier, 1975). Calculate the mean test score  $\bar{X}$  for each student ( $N$  = number of students,  $x_i$  = test score),

$$\bar{X} = \frac{\sum x_i}{N} \quad (1)$$

and the standard deviation  $s$ ,

$$s = \frac{1}{N} \sqrt{N \sum (x_i^2) - (\sum x_i)^2} \quad (2)$$

Then the  $z_i$  score is

$$z_i = \frac{x_i - \bar{x}}{s} \quad (3)$$

The  $z_i$  score is a normalized score for each student which has a mean of zero and a standard deviation of 1. The  $z$  scores can be converted to  $T$  scores where the  $T$  score has a mean of 50 and a standard deviation of 10.

$$T_i = 10 z_i + 50 \quad (4)$$

The standardized scores are easily calculated with a calculator or computer. If the class follows a normal distribution, which does not always happen, then the  $z$  and  $T$  scores are shown in Figure 11-2. The  $z$  or  $T$  scores for each student can be averaged and then compared to other students' scores. Doing this for raw scores is not statistically valid since both the means and the standard deviations vary from test to test. A very simple example may help to clarify the use of standard scores. Consider Debbie who has the following scores on three tests: 60, 40, 80. Her corresponding  $z$  scores are 0, +1, and  $-1$ , while the  $T$  scores are 50, 60, and 40. Compared to the class, her lowest grade is the last one which looks highest on the basis of raw scores.

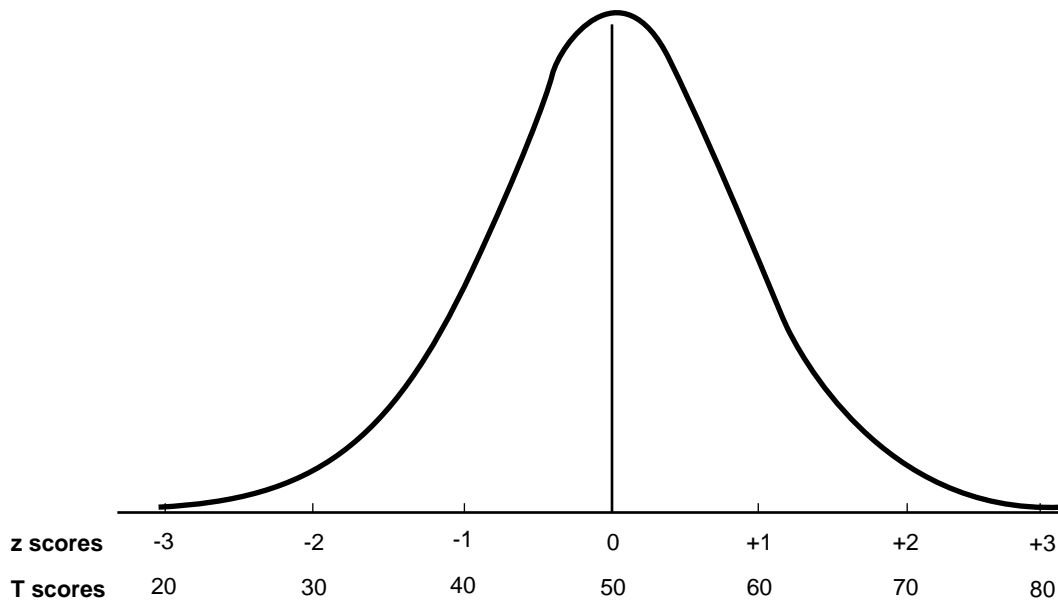
There can be problems with the use of standard scores. First, in small classes they are not statistically valid and should not be used. Second, scores of 100 or 0 do not remain 100 or 0 when translated to  $T$  scores. Extreme scores can become negative or greater than 100. Thus,  $T$  scores can be misleading for these extreme scores. Third, the usual interpretation of the meaning of one standard deviation is valid only for normal distributions.  $T$  and  $z$  scores can still be used but must be interpreted with care. Cheshier (1975) highly recommends the use of standard scores, but McKeachie (1986) does not think they are worth the effort. You get to choose. If you do use standard scores, it is important to spend a few minutes explaining them to the class. Of course, in a class which uses statistics or discusses error analysis, the use of standard scores can be a useful part of the course objectives.

### 11.2.3. REGRADES

Allow regrades! If handled properly, regrades make the professor seem fair, reduce student aggression, force some students to reexamine the test problems, and do not take much time.

In small classes regrades can be handled informally by discussions between the students and the professor. In large classes a more formal procedure is necessary (Wankat, 1983). Regardless of the method used, the regrade procedure should be discussed with the class when the first test is returned. Students are ready to listen at that time.

If the scoring error the student wishes to correct is the incorrect addition of points, then we encourage the student to see the professor immediately following the class. In large classes



**FIGURE 11-2** DISTRIBUTION OF T AND Z SCORES FOR NORMAL DISTRIBUTION OF SCORES

there will be several students clustered around the professor at this time. Thus, it is a good idea to collect the tests to allow time to check the addition.

The second type of scoring error is a mistake in the scoring where the student believes he or she deserves more points. In large classes we require a written regrade request. Students are told to make no additional marks on their tests. On a separate sheet of paper the student is asked to logically explain why he or she deserves more points. The emphasis here is on “logical,” not the plea “I deserve more points.” For example, a student who uses a different solution path than the standard solution may claim that his or her path was correct but that the answer was incorrect because of an algebraic or numerical error. The student can then rework the problem by using his or her path and show that the correct solution is obtained. Based on this type of argument, we have occasionally given a student a large increase in a test score. Quite often while trying this procedure the student finds that the path really does not work, and no regrade is requested.

Students are told that there may be an increase, no change, or a decrease in their test score. We ask for the entire test back but seldom regrade the entire exam. The advantage of getting the entire test back is that the professor can tell if extra pages have been inserted since the original pages will have additional staple holes in them. Some professors regrade the entire test (Evelt, 1980), but this policy seems designed to prevent students from asking for regrades instead of being for the educational benefit of the student.

Give students a deadline (one week is sufficient) for regrade requests. This prevents last minute “grade grubbing” by students. Once the regrade requests have all been collected, sit

**TABLE 11-1** RANGE OF HOMEWORK PROBLEMS (Adapted from Yokomoto, 1988)

Concrete	Abstract
Simple	Complex
Linear solution	Simultaneous solution
Linear solution	Trial-and-error
Short	Long
Answer given	Answer not given
Very clearly defined	Slightly ambiguous
Data given	Need literature data
Self Contained	Build on previous material
Forward solution	Backward solution
Hand calculation	Computer
Written	Visual
Logical	Brainstorm
Numerical	Symbolic

down with the TA and discuss them. The purpose of this is to ensure that grading is uniform. It is poor policy to give complainers higher scores just because they complain. Chronic complainers can be controlled if the professor carefully checks the TAs scoring before returning the tests.

### 11.3. HOMEWORK

What is the purpose of assigning homework? To keep students off the street and out of trouble? Or to help them learn the material? While doing homework the students are active and have a chance to practice the skills being taught in the course. A modest amount of drill can be useful since students learn how to perform certain operations quickly and accurately. Of course, the value of this practice depends on the timeliness of the feedback about the homework. To be effective, this feedback should consist of both positive and negative comments. Homework problems also provide students with a fair chance for success, yet some should also be challenging since both success and curiosity are motivating. The use of study groups should be encouraged since these groups are beneficial for extroverts and field-sensitive students. Homework is beneficial since there is a strong correlation between effort on the homework and test scores (Yokomoto and Ware, 1991).

Homework problems should cover all levels of Bloom's taxonomy and all levels of the problem-solving taxonomy (see Chapters 4 and 5). A gradation of problems, from easy to difficult, to cover many different aspects should be used. At least some of the homework problems should be at the same level of difficulty as the tests. Homework problems can focus on various aspects of the problem-solving strategy such as defining the problem, brainstorming possible solutions, and checking with an independent solution method. Other dimensions

which need to be considered are discussed by Yokomoto (1988) and are shown in Table 11-1. Computer problems should emphasize the use of software tools.

How often should homework be assigned and how many problems should be given? Students need an activity for every week whether it is a test, homework, or a project. These activities should be due on different days. By working around your test schedule, you can determine when homework needs to be done. With a large number of tests or quizzes students do less homework. The number of problems obviously depends upon length. Following the need for a range of problems as shown in Table 11-1, the professor can make some assignments consisting of five small problems and other assignments consisting of a single long problem.

However, assigning a significant amount of homework involves scoring the homework and providing adequate feedback. This is particularly significant if the professor does not have a TA. One solution is to score only selected problems. Tell the students ahead of time that not all problems will be scored. If the problems to be scored are randomly selected, the final homework grade will be proportional to the total amount of homework the student does during the semester. Students need to have solutions available for problems which are not scored. An alternative is to score the homework in class by having students score someone else's homework (Mafi, 1989). With a large number of quizzes it is not necessary to have students hand in homework. They will soon come to believe the professor when he or she tells them that students who do the homework do better on the quizzes. This realization comes quickly if a homework problem is occasionally used on a quiz.

What percentage of the course grade should be based on the homework? If the percentage is low, students will tend to ignore the homework unless a special effort is made to illustrate the correlation between homework effort and test results. If the percentage is high, many students will be encouraged to copy others' work or to cheat in other ways. A reasonable compromise seems to be 10 to 15 percent. This is low enough that you can encourage students to work in groups, but require each to hand in an individual homework paper.

Late submissions can be a difficulty. In industry, late work is accepted grudgingly and does not earn promotions or handsome raises. We suggest telling students this and then following industrial practice. Accept late work grudgingly and take off some percentage based on how late it is. We accept no homework after the solution has been posted.

Reading assignments pose somewhat different problems. Students will read the assignments if they see that reading leads to success in homework and tests. The professor's task is to ensure that the reading contributes directly to the student's success. And if the reading does not help the student achieve success, why is it assigned? Be sure that a good textbook or other readings have been selected. Skip certain material in lecture but make it clear to the students that it is material they are expected to learn from the readings. Refer to this material in the lecture, but do not cover it. Ask questions in class based on the readings. Reiterate to the class that it is important material. Then assign homework based on the material and include test problems based on it. It doesn't hurt if one of the homework problems is a clone of an example in the textbook.

## 11.4. PROJECTS

Design and laboratory reports can be considered special types of project reports; however, since they were discussed in Chapter 9, they will not be considered further here. Projects are most common in smaller classes such as senior electives and graduate courses. Long projects are not appropriate for first-year students since they are not ready to pick topics of special interest and need the discipline of more frequent assignments (Erickson and Strommer, 1991).

Projects can fulfill some educational objectives which are difficult to fulfill with lectures and tests. A project can allow a student to explore in depth a topic of her or his choice. This choice of project gives the student some control over her or his education which is often missing from other courses. Consequently, students sometimes become very strongly motivated and continue the project long after the class is over. Projects also provide an opportunity for students to work on communication skills through written and oral progress and final reports. In the ideal case, through the project the professor empowers the student to work on an area in which he or she is intensely interested, and the professor encourages the student to develop a meaningful project. Our experience in requiring graduate students to do projects is that approximately 20 percent comes close to this ideal.

Projects must have a deliverable. In a library research project the deliverable is a paper. In engineering it is often preferable to require the student to produce “something” such as a computer program, an integrated circuit, a teaching module, a laboratory experiment, a novel solution to a mathematical problem, and so forth. Then the deliverables are a short report describing the something in addition to that something. The more choices the student has on what to deliver, the more likely he or she is to become excited about the project.

Students naturally want to know what the professor wants. Explaining what is desired (i.e., the objectives for the project) can be surprisingly difficult. However, at some point the professor does decide what he or she wants and grades accordingly. Waiting until the projects are graded to decide what one wants leaves the professor open to student complaints: “Oh, so he does know what he wants; he just won’t say until it’s too late” (Starling, 1987). As a professor you need to explain what it is that you really want. If you can’t, then analyze the projects from the last time you taught the course to determine what you wanted when you graded them.

How big should a project be? If it is only 5 percent of the grade, students will treat it as a homework project. If it is 50 percent of the grade, students will feel very anxious about it. Projects that are about 25 percent of the grade have worked well for us.

Procrastination is the biggest problem involved in student projects. To limit but not eliminate procrastination, set up a series of deadlines. First, introduce the project in lecture relatively early in the semester. Describe the project and the evaluation procedure as clearly as possible. List the dates of all intermediate and final deadlines. In a small class, both written and oral progress reports are useful since they allow for early feedback and make students do at least some work throughout the semester. Individual meetings with each student can also help prevent procrastination.

Evaluation of projects is time-consuming, which is one reason they are most commonly used in small classes. If communication skills are included in course objectives, then projects should be evaluated on organization and writing ability. Professors who are very serious about



this can correct about a third of the report and have the student correct the entire report before it is evaluated for a grade. The written report should also be graded for content, including correctness, depth, and creativity. Although projects are usually turned in late in the semester, many students seriously study the feedback because they have become involved with the material. The best feedback method for oral reports is a videotape of the student presentations. Since the project represents a sizable portion of the course grade, instructors should use their discretion in accepting late reports.

## 11.5. GRADING

The best advice we can give before a professor decides on the final course grades is to go into an empty room and repeat out loud, “I can’t satisfy everyone. I can’t satisfy everyone. I can’t satisfy everyone.” This will help create the proper frame of mind for awarding final course grades.

### 11.5.1. Purpose of Grades

There are very diverse views in the literature about the purposes and suitability of the typical grading methods used in colleges. These range from grades being indispensable to worthy of being abolished. The writers who defend grades, at least moderately, include McKeachie (1976, 1986), Johnson (1988), and Lowman (1985). The purposes they note for grades include:

- 1 Reward or penalty for student accomplishment.
- 2 Communication to others about what the student has accomplished.
- 3 Predictor of future performance.

Grades certainly do serve as rewards or penalties, but for feedback purposes they come much too late in the semester to provide any motivation. As rewards, grades are often used to determine who will receive honors, scholarships, and so forth. As penalties, they are used to place students on probation and to drop students.

Using grades as a communication tool is often confusing since there is no generally agreed-upon definition of what a grade means. Professors who arbitrarily change the meaning of grades are not communicating well because those who see the grade interpret it differently. However, some communication exists since there is general agreement that an A or a B means the student has learned more than a student who receives a D or an F.

Grades are also used as predictors. Students use grades as a predictors of how well they will do in the rest of their college careers, and in this sense good grades may motivate a student to

continue. Professors also use grades to predict who will do well in later courses and who might do well in graduate school. Since grades are reasonably good at predicting grades in future courses (Stice, 1979), this use of grades is somewhat reasonable. Even in this case grades can be misused since they do not predict who will be good at research, which is a major part of many graduate school programs.

The most controversial use of grades is as predictors of success in life after school. Many employers use grades as part of their selection procedures. Unfortunately, many studies agree that there does not seem to be any correlation between grades and success after school (Eble, 1988; Stice, 1979). This is true regardless of how one defines success. What this means is that engineers who graduate are good enough in whatever it is that grades are measuring to be a success, and that other variables become important. These other variables can include drive, motivation, inherited wealth, common sense, communication skills, interpersonal skills, who the person knows, and luck. The supporters of grades note that since grades are part of the selection criteria, one cannot expect them to be a major predictor of success since the sample is already fairly homogeneous (McKeachie, 1986).

### 11.5.2. Grading Methods

Some intelligent and thoughtful people state that one should select a grading method that subverts the grading system (Eble, 1988; Elbow, 1986; Smith, 1986). The Keller plan for PSI is an example (see Chapter 7). We will assume that, being new to teaching, you are not ready or willing to subvert the system (yet). Thus, the remainder of this chapter will be on grading hints and methods.

Regardless of the grading method used, the more scores you have, the easier it is to give grades. When there are many scores, and the students know what these scores indicate, there are fewer conflicts with students when grades are awarded. If a student does complain about a grade, it is appropriate to listen to what he or she has to say. But unless a grading error has been made, it is unwise to change grades. Once a grade is changed, the word gets around and many students want their grades changed.

**TABLE 11-2** TYPICAL GRADING SCALES FOR DIFFERENT CLASS COMPOSITION IN TERMS OF T SCORES (© 1975, American Society for Engineering Education)

Letter Grade	Class Made Up Primarily of			
	Poor Students	Average Students	Exceptional Students	Graduate Students
A	69 and above	63 and above	57 and above	55 and above
B	59 - 68	53 - 62	47 - 56	45 - 54
C	49 - 58	43 - 52	37 - 46	35 - 44
D	39 - 48	33 - 42	27 - 36	25 - 34
F	38 and below	32 and below	26 and below	24 and below

The most appealing grading method uses an absolute standard. Although appealing, this is often difficult. Contract or mastery grading is another way to use a standard, but this method has its detractors since the grade no longer means what most people think it means (communication), and the grade is no longer a good predictor of grades in a “standard” class. However, Eble (1988), who is not fond of grading, states that the predictive capabilities of grades should not be taken too seriously.

Travers (1950) originally suggested a standard grading scheme which has been echoed by McKeachie (1986) and Johnson (1988). The major and minor objectives of the course are clearly defined. Then the grade is a communication to the student and to others of what fraction of the course objectives has been achieved. The meaning of the grades can be defined in a manner similar to the following:

- A: achieved all major and minor objectives
- A<sup>-</sup>: achieved all major and most minor objectives
- B<sup>+</sup>: achieved most major and most minor objectives
- B: achieved most major objectives and many minor objectives
- B<sup>-</sup>: achieved most major objectives and some minor objectives
- C: acceptable performance
- D: student is not prepared for advanced work requiring this material
- F: failed

Even with a system like this the professor needs to decide if the student meets an objective, and subjective decisions will have to be made. Normative grading, commonly known as grading on a curve, is often used because the professor does not have to develop and correctly test for absolute standards. Instead, students are compared to each other and the grade curve is broken up into A, B, C, D, and F. This has the unhealthy effect of increasing competition. In addition, a student performance which earns an A in one class may earn a C in another class merely because student competition is better. Because of this effect of class quality, a professor should never force grades into predetermined percentages.

Many professors grade on a curve but slant the curve to take into account student quality. Cheshier (1975) suggests the grading scales shown in Table 11-2, where the grades listed are the average T score for each student. The professor needs to decide what the average quality of the class is and then use these ranges as a guide. Since a T score of 50 is the average, Cheshier is suggesting that the average student in a poor class receive a low C while the average student in a good class or a graduate-level class receive a B.

An alternate procedure is to list all the students' total scores or average scores for the semester. Then decide first where the average grade in the course should be. Many professors believe that the average grade in upper-division courses should be higher than in freshmen courses since the poorest students have dropped or transferred out of engineering. Thus the average student might receive a B or a B<sup>-</sup>. Then look at the distribution and decide upon cutoff points. One method for assigning the cut point for F's is to see what the score of a good but not exceptionally brilliant student is (usually the second or third best student in the class). Then any student who receives less than half this number of points fails the course. If no one is that

low, then everyone passes. With the grade of the average student chosen and the F's chosen, the other grades can be selected. It is convenient to look for natural gaps in the grade distribution and put the A-B and B-C, boundaries there. When this has been done, look at the grades that are on the edges. Should they be moved up or down? Many professors prefer to give students the benefit of a doubt if their scores have been increasing throughout the semester. This can also be accomplished by giving greater weight to later tests. Try to apply wisdom at the boundaries. We have seldom been wrong when we have found reasons to be generous.

## 11.6. CHAPTER COMMENTS

Much more could be said about testing. Tests can be analyzed for validity and reliability by statistical methods. Multiple-choice tests in particular are easy to analyze. We do not think that this type of analysis is particularly valuable in engineering education and doubt that many engineering educators would use these procedures. In any case, most large universities have a testing service which machine-scores multiple-choice tests and calculates the appropriate statistics.

Students usually consider exams to be the most important part of a course since they are the main determiner of their grades. Professors can lament the students' values, but these values are difficult to change. Complaints about tests can be decreased by making them as fair as possible and by having enough tests so that one test will not completely determine a student's grade. Because they consider exams to be so crucial, some students will be tempted to cheat, and cheating is another fact of life which must be faced (see Chapter 12).

## 11.7. SUMMARY AND OBJECTIVES

After reading this chapter, you should be able to:

- Discuss the advantages and disadvantages of different types of test questions. Write test questions using each of the major test question styles.
- Develop a grid to determine the course material to be covered on a test.
- Explain to a TA how to score a test fairly. Score a test fairly.
- Determine the discrimination of test questions and calculate  $z$  and  $T$  scores for students.
- Develop a scheme for using homework and projects as part of a course which satisfies learning principles.
- Develop a personal value system for giving course grades.

## HOMEWORK

- 1 Assume you will write a test for this chapter.
  - a Develop a test grid to decide on the coverage of the test.
  - b Write at least two long-answer questions, two short-answer questions, and two multiple-choice questions for the test.
  - c Select the questions for the test so that it can be given in a regular fifty-minute class period.
  - d Write the solutions and the grading scheme for this test.
- 2 Do a project on teaching engineering material. The project must involve both content (engineering subject matter) and teaching method. You must have a deliverable such as a videotape, CAI, a self-paced module, a laboratory experiment, a National Science Foundation proposal for curriculum development, a student handbook for commercial software, or course demonstrations.

## REFERENCES

- Canelos, J., and Catchen, G. L., "Test preparation and engineering content," *Proceedings ASEE Annual Conference*, ASEE, Washington, DC, 1624, 1987.
- Cheshier, S. R., "Assigning grades more fairly," *Eng. Educ.*, 343 (Jan. 1975).
- Crittenden, J. B., "Partial credit: Not a God-given right," *Eng. Educ.*, 288 (Feb. 1984).
- Eble, K. E., *The Craft of Teaching*, 2nd ed., Jossey-Bass, San Francisco, 1988.
- Elbow, P., *Embracing Contraries: Explorations in Learning and Teaching*, Oxford University Press, New York, 1986.
- Ericksen, S. C., *The Essence of Good Teaching*, Jossey-Bass, San Francisco, 1984.
- Erickson, B. L. and Strommer, D. W., *Teaching College Freshmen*, Jossey-Bass, San Francisco, 1991.
- Evet, J. B., "Cozenage: A challenge to engineering instruction," *Eng. Educ.*, 434 (Feb. 1980).
- Johnson, B. R., "A new scheme for multiple-choice tests in lower-division mathematics," *Amer. Math. Mon.*, 427 (May 1991).
- Johnson, G. R., *Taking Teaching Seriously: A Faculty Handbook*, Texas A&M University, Center for Teaching Excellence, College Station, TX, 1988.
- Kessler, D. P., "Machine-scored versus grader-scored quizzes—An experiment," *Eng. Educ.*, 705 (April 1988)
- Leuba, R. J., "Machine scored testing, Part I: Purposes, principles and practices," *Eng. Educ.*, 89 (Nov. 1986a).
- Leuba, R. J., "Machine scored testing, Part II: Creativity and item analysis," *Eng. Educ.*, 181 (Dec. 1986b).
- Lowman, J., *Mastering the Techniques of Teaching*, Jossey-Bass, San Francisco, 1985.
- Mafi, M., "Involving students in a time-saving solution to the homework problem," *Eng. Educ.*, 444 (April 1989).
- Masih, R. Y., "Perfecting the engineering education and its exams," *Proceedings ASEE Annual Conference*, ASEE, Washington, DC, 200, 1987.
- McKeachie, W. J., "College grades: A rationale and mild defense," *AAUP Bull.*, 320 (Autumn 1976).

- McKeachie, W. J., *Teaching Tips*, 8th ed., D.C., Heath, Lexington, MA, 1986.
- Smith, K. A., "Grading and distributive justice," *Proceedings ASEE/IEEE Frontiers in Education Conference*, IEEE, New York, 421, 1986.
- Starling, R., "Professor as student: The view from the other side," *Coll. Teach.*, 35 (1), 3 (1987).
- Stice, J. E., "A first step toward improved teaching," *Eng. Educ.*, 394 (Feb. 1976).
- Stice, J. E., "Grades and test scores: Do they predict adult achievement?" *Eng. Educ.*, 390 (Feb. 1979).
- Svinicki, M. D., "The test: Uses, construction and evaluation," *Eng. Educ.*, 408 (Feb 1976).
- Travers, R. M. W., "Appraisal of the teaching of the college faculty," *J. Higher Educ.*, 21, 41 (1950).
- Wankat, P. C., "Regarding tests: A chance for students to learn," *Eng. Educ.*, 746 (April 1983).
- Yokomoto, C. F., "Writing homework assignments to evoke intellectual processes," *Proceedings ASEE Annual Conference*, ASEE, Washington, DC, 579, 1988.
- Yokomoto, C. F. and Ware, R., "The seven practices—Persuading students to do their homework," *Proceedings ASEE Annual Conference*, ASEE, Washington, DC, 1767, 1991.