3. Providing feedback is more important than assigning a grade. You can use nongraded evaluation as well as evaluation for assigning grades.

4. Try to assess the attainment of all your objectives, even if some objectives (such as increased motivation for learning) are not appropriate criteria for grades.

5. Avoid evaluation devices that increase anxiety and competition.

Supplementary Reading

Paul Ramsden's chapter "Assessing for Understanding," in his book *Learning to Teach in Higher Education* (London: Routledge, 1992), presents a wise perspective on assessment and gives examples from chemistry, anatomy, materials technology, engineering, history of art, statistics, medicine, and physics.

Tom Angelo and Pat Cross's *Classroom Assessment Techniques* (1993) has become almost a bible for faculty interested in incorporating more ongoing assessment in their classes.

500 Tips on Assessment by Sally Brown, Phil Race, and Brenda Smith (London: Kogan Page, 1996) is a marvelous compendium of useful suggestions on all types of assessment, ranging from self-assessment through group assessment, multiple-choice tests, and assessment of performance, lab work, and dissertations.

Graham Gibbs discusses modern methods of assessing learner-centered courses in his book *Assessing Student-Centered Courses* (Oxford: Oxford Centre for Staff Development, 1995). Chapters give case studies illustrating assessment of group work, projects, journals, skills, and portfolios.

Assessment Matters in Higher Education, edited by Sally Brown and Angela Glasner (Buckingham, UK, and Philadelphia: Society for Research into Higher Education and Open University, 1999), describes innovative approaches to assessment and current United Kingdom practices in a variety of disciplines. There is an entire section on peer assessment and self-assessment. (I suspect that the pun in the title was intentional.)

Assessment Essentials: Planning, Implementing and Improving Assessment in Higher Education by Catherine Palomba and Trudy Banta (San Francisco: Jossey-Bass, 1999) is a fine resource on all manner of assessment strategies and the rules that guide their use.

Chapter 8

Testing: The Details

If your assessment plans call for the use of in-class testing (and they probably will), you can do a lot to make sure that the test you design serves the assessment purposes you had in mind. In this chapter, I'm going to get down to the nitty-gritty details of writing a test. Not all the details will fit every testing situation, but the planning and execution of most tests will follow this decision process.

WHEN TO TEST

Because tests are so important in making the goals of a course concrete and influencing student methods of learning, I give an ungraded quiz during the first week and a graded test after the third or fourth week of a 14-week semester. To reduce the stress I weight early tests very little in determining the final grade. An early test gets students started—they don't delay their studying until the conventional midterm examination—and it will help you to identify problems early while they are still remediable. Thus, early tests should demand the style of learning you expect, and they need to be constructed carefully even though their purpose is more motivational and diagnostic than evaluative.

The amount and frequency of tests should depend on the background of your students and the nature of the content. In a first-year course in an area new to students, frequent short tests early in the term facilitate learning, as demonstrated in the Personalized System of Instruction (Keller, 1968). Generally, however, I want to wean students from studying for tests, so that they will become lifelong learners able to evaluate their own learning. This implies less frequent testing as learners become more experienced. It probably also implies questions requiring broader integration and more detailed analysis as learners advance.

CONSTRUCTING THE TEST

In planning your tests you may want to use a mix of different types of questions in order to balance measurements of the varied goals of education. The following sections describe the strengths and weaknesses of each type of question, as well as offer tips on constructing items.

Choosing the Type of Question

The instructor who is about to give an examination is in a conflict situation. The administration of an examination consists of two time-consuming procedures: (1) construction of the examination and (2) grading. Unfortunately, it appears to be generally true that the examinations that are easiest to construct are the most difficult to grade and vice versa.

Teachers often base their choice of question types solely on class size, using multiple-choice tests for large classes, short-answer questions for medium-sized classes, and essay questions for small classes. Class size is certainly an important factor, but your educational goals should take precedence. Your goals almost always will require the use of some essay questions, problems, or other items requiring analysis, integration, or application.

Problems. In mathematics, science, and some other disciplines, a test typically consists of problems. The value of problems depends on the degree to which they elicit the sort of problem-solving skills that are your goals. Some problems are too trite and stereotypic to have much value as indicators of whether students understand the steps they are following. In other cases the answer depends to such a large extent on tedious calculations that only a small sample of problems can be tested. In such cases you might provide calculations leading up to a certain point and ask students to complete the problem, or you might use a multiple-choice question about the proper procedure—for example, "Which of the following problems can be solved by procedure *x*?" Or you might have students set up the problem without actually calculating the final answer. Many instructors who have problem solving as their goal say that setting the problem up correctly is more than half the battle, yet students often jump right to a formulaic response. If the grade is based solely on problem setup, students will pay more attention to it. Many teachers use problems that may be solved in more than one way or that have more than one satisfactory answer. In this case special emphasis in both teaching and grading should be on justifying the solution strategy rather than on the specific answer. This has the advantage of focusing students' attention on the process rather than on the product.

Short-Answer Items. Here is an example of a short-answer item: "Give one example from your own experience of the concept of elaboration." In responding, a student might describe an experience of explaining a concept to another student or of thinking about the relationship of a fact to a general principle. Such a question is restricted enough that it is not often difficult to judge whether the expected answer is there. Furthermore, such questions can be presented in a format that allows only a small amount of space for the answer. The student tendency to employ the "shotgun" approach to the examination is thus inhibited.

Short-answer questions permit coverage of assigned materials without asking for petty details. Unfortunately, many short-answer questions test only recall of specific facts. Short-answer questions, however, can do more than testing recall. If you are trying to develop skill in analysis or diagnosis, for example, you may present case material or a description of an experiment and ask students what questions they would ask. You can then provide additional information that they can use in an analysis. Or a short-answer question can ask students to solve a problem or propose a hypothesis relevant to information learned earlier. An example is the following question from a course on the psychology of aging:

1. Given the *differences* in ways in which men and women experience middle age, and the fact that depression rises as a psychiatric symptom in middle age, how might the *causes* of the depression differ for men and women at this time in life?

Essay Items. Although the short-answer examination is very useful in certain situations, I recommend that, if possible, you include at least one

essay question on examinations in most college courses. Experiments indicate that students study more efficiently for essay-type examinations than for objective tests (d'Ydewalle, Swerts, & de Corte, 1983; McCluskey, 1934; Monaco, 1977). Thus in addition to the values of essay tests as evaluation devices, you should take into consideration their potential educational value as stimuli to students' reflection about conceptual relationships, possible applications, or aspects of thinking. One strategy is to pass out several questions the week before the test and tell students that these are the sorts of questions you will use—that, in fact, you may even use one of these very questions.

Where the tests can be returned with comments, essay examinations may give students practice in organized, creative thinking about a subject and an opportunity to check their thinking against the standards of someone with more experience and ability in the field. Johnson (1975) demonstrated that when marginal comments on earlier tests emphasized creativity, creativity on the final exam was improved.

In large classes where time is limited and in classes where the writing itself is not the point of the question, you can format the answer sheet to break a long, complex answer into its critical components, each of which has a space for an answer. For example, in my class the last problem on every exam is a case to which the students must apply whatever theory we have been studying. So at the top of the sheet, there is a short description of the scenario. Then there is a space headed "In five sentences or fewer, describe your proposed solution to this scenario based on theory X." About two inches farther down the sheet there is another instruction: "In the spaces below connect the components of your solution to three aspects of theory X that are relevant and explain their relevance." That is followed by three spaces, each headed like this:

aspect one: (space)

connection to your solution and why: (space)

This considerably speeds up my grading time because rather than searching through a long essay organized (I hope) by the student, I can at a glance see if the student has provided a reasonable solution and tied it to the theory. I'm not "giving away the answer" because the prompts are fairly broad; I'm simply imposing a little organization on the answer to make my grading easier. And maybe students learn something about structuring an answer efficiently, too.

Finally, if you read the examinations yourself (or at least some of them), you get some excellent information on what students are learning.

True-False Items. Although true-false examinations are rather easy to make up, I don't ordinarily advocate their use. Toppino and Brochin (1989) showed that students tend after the test to remember the false items as being true—an outcome not conducive to achieving your objectives. If you do use true-false items, ask students to explain their answers. This will encourage reflection and help you understand why there were some common misunderstandings.

Multiple-Choice and Matching Items. It is improbable that most teachers can adequately measure all their objectives with a test made up entirely of multiple-choice questions. Matching questions are similar to multiple-choice in that the student must discriminate between the correct answer and other choices. Nonetheless, for some purposes multiple-choice items are useful. They can measure both simple knowledge and precise discrimination. They can measure ability to apply concepts or principles; they can assess elements of problem solving. But they are not likely to assess organization of ideas, conceptual relationships, or many of the skills involved in higher-order thinking.

Good multiple-choice questions are difficult to construct. (The greater your experience in their construction, the more you realize how long it takes per item to construct a reasonably fair, accurate, and inclusive question.) Because of this difficulty, the construction of such items is probably not worthwhile unless they will be administered to several hundred students, either in a single year or in successive years. Some books that can help you write high-quality items, if you are so inclined, are referenced at the end of this chapter.

Even if you don't pretest the items on students, it is worthwhile to have someone take the test before it is in its final form. If you can persuade a skilled test taker who doesn't know the subject matter to take the test, you will probably be surprised at how many he or she gets right simply from cues that you provided in the questions.

How Many Questions Should You Use?

Obviously, the number of questions depends on the type and difficulty of each question. I prefer to give tests without a time limit, but the constraints of class scheduling usually require that you clear the classroom so that the next class can begin. Thus, you must plan the length of the exam so that the slowest students have time to finish before the end of the period. As a rule of thumb I allow about 1 minute per item for multiple-choice or fill-in-the-blank items, 2 minutes per short-answer question

Constructing Multiple-Choice Items

1. Teachers' manuals that accompany many textbooks contain multiple-choice items. You should not rely on a manual as the source of all your questions, because the manual probably will not contain many good questions and may cover only textbook material. You need to assess what students have learned in class as well as their understanding of what they have read.
2. A second source of multiple-choice items is the students themselves. They are not a particularly satisfactory source of test questions, because only about 10 percent of the items thus written will be usable. However, this technique is a useful pedagogical device because it gets the students to read their assignments more analytically. It also gives the instructor a good index of what the students are getting out of the various sections of their reading, and it gives you a chance to remind them of the goals of the course going beyond recall of details.
3. There are statistical methods for evaluating questions, but the best suggestions for improvement come from students themselves in their discussion of the test. It seems almost criminal to waste this experience with items; therefore I recommend a permanent file.
4. If you have a problem but no good distractor (incorrect alternative), give the item in short-answer or essay form and use the students' own responses for alternatives for a later use of the item in multiple-choice form.
5. Multiple-choice questions typically have four or five alternatives. Rather than wasting your and your students' time with extra alternatives that don't test a discrimination that is important, use only as many alternatives as make meaningful discriminations. Costin (1972) showed that three-choice items are about as effective as four-choice.
6. For measuring understanding, I like questions that require the student to predict the outcome of a situation, rather than questions that simply ask the student to label the phenomenon.
7. Multiple-choice items need not stand alone. You can use a sequence of related items to measure more complex thinking.
8. Grouping items under headings will improve student performance (Marcinkiewicz & Clariana, 1997).

requiring more than a sentence answer, 10 or 15 minutes for a limited essay question, and a half-hour to an hour for a broader question requiring more than a page or two to answer. You can get a rough estimate of time requirements by simply timing how long it takes to actually read the items without answering them. That can serve as a minimum time requirement. If you ask someone else to take the test as suggested above, time that person, too.

TESTS FROM THE STUDENT PERSPECTIVE

It's not surprising that our students get so concerned about tests and other assessments. More is riding on their performance than just a grade in the class. All kinds of things depend on a student's grade point average, many of them with no apparent relationship to scholarly achievement. For example, in some areas, students with good grades get lower car insurance rates! Maybe the insurance companies figure that you must be home studying all the time to get such good grades, so you're not as likely to have an accident!

On a more serious note, no one is totally comfortable with being assessed, and, rightly or wrongly, students often equate grades with self-worth. We owe it to them to help them maximize their potential for good performance by dealing with some of the things that might get in the way.

REDUCING STUDENT FRUSTRATION AND AGGRESSION

Most beginning teachers find the aggression that students direct against them after a test very disturbing. It is likely to impair the instructor's rapport with the class and may actually be a block to learning. Thus, strategies for reducing the aggression seem to be worthwhile.

The most obvious solution to the problem is to reduce students' frustration when taking tests. You can do this by emphasizing the contribution the course can make to students' long-range goals. Explaining how and why you test as you do will also help. A nongraded practice test will provide guidance. Periodic assessments of learning (not necessarily graded) to help students assess their own progress and to help you identify problems, as well as frequent explanations of why and how you test and assess learning, should reduce students' anxiety and frustration about testing.

Yet no matter how much you emphasize long-range goals, the tests will in large measure determine what students do. Do you want them to memorize details? Then give the usual memory-of-details test. But if you want more, make your objectives clear, and make sure that your tests measure the attainment of those objectives. If you used the Bloom Taxonomy of Educational Objectives or Biggs's SOLO Taxonomy as suggested in the chapter "Countdown for Course Preparation," remind the students of these levels before each test.

Test instructions should indicate whether students are to guess, what the time limit is, and any other directions that define the nature of the expected responses. For the typical classroom examination, there is no point in a correction for guessing. Emphasizing in the multiple-choice test introduction that students should choose the *best* answer may help prevent lengthy discussion with students who dream up a remote instance in which the correct alternative might be wrong.

Research I did with my colleagues (1955) and research by Smith and Rockett (1958) demonstrated that on multiple-choice tests the instruction "Feel free to write comments," with blank space by each question for the comments, results in higher scores, especially for anxious students. A problem with this strategy is that students these days have been taught to make notes to themselves on the test, so you may find yourself reading a lot of stuff not really written to you. Here is how I have solved this problem and the problem of students who want to explain every item. I allow students to explain their choices for up to three questions. They star the question that they want to elaborate on; then on the last page of the test, called the "explanations page," they write their thoughts and indicate why they answered the way they did. I only read the explanations to questions that they have missed. This process greatly reduces their anxiety and saves me grading time. It also forces them to pick their battles; they can't simply write everything they know for every question in hopes that the correct answer is in there somewhere.

HELPING STUDENTS BECOME TEST-WISE

Particularly in the case of multiple-choice examinations, I have found that a good morale builder is spending 15 minutes or so the day before the first test telling students how to take a test of this sort and familiarizing them with the format. Some of the points that I make in such a lecture follow.

Taking Multiple-Choice Tests

The student taking a multiple-choice examination is essentially in the same position as a poker player. The object is to get into a position where you are betting on a sure thing. If this is impossible, at least make your bet on the choice where the odds are in your favor. In poker, you are in the strongest position if you know exactly what your opponent has; in the examination situation, you are in the strongest position if you know the material. There is no substitute for study. Nevertheless, you are not likely to be absolutely certain of all the right answers, and when you are not, certain techniques may help.

What I recommend to the student is this. First go through the examination and answer all of the items you know. In addition to getting a certain amount of the examination done without spending a lot of time on single, difficult items, you probably will find that going through the complete test once in this way will suggest the answers to questions that might have been difficult had they been answered in numerical order. When you have gone through the test once in this fashion, go through it again and answer any questions whose answers are now obvious. Usually there will still be a few unanswered questions. It is in connection with these that certain tricks may be useful.

If the item is multiple choice, don't simply guess at this stage of the game. See whether it is possible to eliminate some of the choices as incorrect. In a four-choice multiple-choice item, the probability of getting the answer right by pure guesswork is one in four; if you can eliminate two of the choices, your chances improve to 50–50. So take advantage of the mathematics of the situation.

After completing the examination, go through the whole thing again to check your choices to make sure that you still regard them as correct and to make sure that you made no clerical errors when recording them. In this connection, it is worthwhile to point out to students the common misconception that when you change your answers, you usually change from right answers to wrong ones. Mueller and Wasser (1977) reviewed 18 studies demonstrating that most students gain more than they lose on changed answers.

Taking Essay Tests

My instructions for essay exams are simpler.

1. Outline your answer before writing it. This provides a check against the common error of omitting one part of the answer.

2. If a question completely baffles you, start writing on the back of your paper anything you know that could possibly be relevant. This starts your memory functioning, and usually you'll soon find that you have some relevant ideas.

3. If you are still at a loss, admit it, write a *related* question that you can answer, and answer it. Most instructors will give you at least a few points more than if you wrote nothing.

4. Write as well as you can. Even if I intend not to grade on writing ability, my judgment is negatively influenced when I have to struggle to read poor handwriting or surmount poor grammar and sentence structure. Moreover, because I believe that every course is responsible for teaching writing, writing always enters into my grading.

Why Teach Test Taking?

Is it wise to give students these tips? The answer to this question depends on your purposes in giving an examination. If you want to test for "test-taking" ability, you will not want to give students such hints. At any rate, this orientation seems to have the effect of conveying to students the notion that you are not trying to "outsmart" them and are instead interested in helping them get as high a grade as their learning warrants.

Coping with Test Anxiety

Many students struggle with test anxiety because of the high-stakes testing they experienced in the past and the emphasis on grades they're experiencing now. A student may know the material but blank out during the test and be unable to show what he or she knows. If my students are having such problems, I can do several things to help:

1. I can lower the stakes of any given test. By having several assessments of learning, I can lower the overall importance of any one test and thereby lower students' anxiety about their performance on it.
2. I can offer "second chances" to students who experience difficulties while taking a test. This means allowing them after the test to earn back some of the points they missed. I describe this process later in this chapter. This is a good learning strategy, and more important for test-anxious students, it relieves some of the pressure and therefore some of the anxiety.

3. I already mentioned the strategy of allowing students to explain their answers more thoroughly on the test itself. This also removes some of the pressure that comes with uncertainty about a particular answer.
4. Prior to the test day I familiarize students with what the test will actually look like, the kinds of questions, any special procedures they'll need to follow, and how I'll grade the test. This removes a lot of the unknowns associated with the test, which are a big source of anxiety.
5. I offer ideas about studying and about getting physically ready for the test, relaxation strategies (taking deep breaths, putting down the pencil and flexing your fingers, and so on). Sometimes I even coach students to think about what they're saying to themselves that contributes to their anxiety—for example, saying, "I've *got* to get an A," instead of, "I'm going to do OK," is more likely to produce anxiety.

ADMINISTERING THE TEST

Handing out a test should be a simple matter. Usually it is, but in large classes, simple administrative matters can become disasters. It is hard to imagine how angry and upset students can become while waiting only 10 minutes for the proctors to finish distributing the test forms. And if this doesn't move you, imagine your feelings when you find that you don't have enough tests for all of the students. (It has happened to me twice—deserving a place among my worst moments in teaching!)

How can you avoid such problems?

1. If you are having tests duplicated, ask for at least 10 percent extra—more if the test is administered in several rooms. (Some proctor always walks off with too many.) This gives you insurance against miscounting and against omitted or blank pages on some copies.

2. Unless there is some compelling reason to distribute the tests later, have your proctors pass out the tests as students come into the room. This protects students from mounting waves of panic while they wait for the tests to be distributed.

3. Minimize interruptions. Tell students before the exam that you will write announcements, instructions, or corrections on the board. Some exam periods are less a measure of achievement than a test of the students' ability to work despite the instructor's interruptions.

ALTERNATIVE TESTING MODELS GAINING FAVOR

Group Testing

Given the prevalence of group work in classes these days, some instructors have begun to administer group tests as well. Since the students have been encouraged and actually required to study and work in groups while learning, the logic is that asking them to perform in an individual situation on the test contradicts what they have learned about peer support. Although I may not agree with that logic, I do agree that taking a test in a group situation is a good learning experience for the same reasons that collaborative learning is a good teaching method: students learn a lot from one another and from having to explain their own answers.

The most common method for this strategy is to have the students initially take the test on their own. Then after turning their copy in, they get into a group (usually the one they've been working with all semester) and go through the test again to come up with a group response to the test. It is amazing how much energy there is during this activity! It has the advantage of giving the students immediate feedback on their test performance by comparing their responses to their groupmates', and it also corrects any misconceptions right away—something that we can't do very easily in a regular test situation. Grades are a combination of individual test performance and group test performance.

There are many concerns about this strategy, most of them having to do with grading and with difficulties posed by room configurations. There is also the possibility of one student dominating the group's responses to the test. These are the same problems that arise whenever group work is suggested, and they must be at least acknowledged. Achacoso and Svinicki's (2005) descriptions of group testing by a couple of different instructors in different settings may inform your understanding of this trend.

Online Testing

Another new trend in testing is the use of testing online. In this model students take their tests on a computer, either their own or at a testing center. There are almost as many varieties of this strategy as there are instructors. Achacoso and Svinicki (2005) provide examples of different online testing strategies.

The advantage of online testing is that it can allow an instructor to give a customized test to each student through the miracles of

technology and a large database of questions. Given what you know about computers, I'm sure you can imagine all the clever ways that the technology can modify, randomize, customize, and evaluate a student's test. For example, there is one format that calibrates the difficulty of each subsequent item based on whether the current item was answered correctly. This particular mode is being used with the large standardized placement tests, such as the GRE or LSAT. That's probably a little too fancy for a regular classroom test, but future developments in software may make it possible for individual instructors to design such systems just as we can now design online tutorials much more easily.

Another advantage of the online testing idea is that the instructor can include simulations that are interactive. Such questions would provide a much better test of student understanding than the static problems that can be included in paper-and-pencil tests.

The difficulty with such testing is maintaining testing integrity. Unless the test is administered under secure conditions—for example, in a computer lab or testing facility—the instructor may not be able to ensure that the person submitting the test is really the designated student or whether the student is making inappropriate use of support materials during the test. Many institutions are considering the feasibility of providing large computer-based testing centers, and it will be interesting to see whether such efforts are scalable to the kinds of large classes in which they might be the most useful.

WHAT TO DO ABOUT CHEATING

It may be hard for you to believe that your students would ever cheat—"Maybe other students cheat, but not mine!" Unfortunately, studies of cheating behavior over several decades invariably find that a majority of students report that they have cheated at some time (McCabe & Trevino, 1996). A recent Google search on "cheating in college" turned up over 400,000 pages! Most students would rather not cheat, but the pressures for good grades are so intense that many students feel that they, too, must cheat if they believe that other students are cheating. In my experience the most common excuse given by a student caught cheating is that other students were cheating and the teacher didn't seem to care, at least not enough to do anything to prevent or stop cheating. Many students thus feel less stress when an examination is well managed and well proctored.

Why Do Students Cheat?

The research on this question is alarmingly consistent. The most significant factor in a student's decision to cheat is peer influence (McCabe, Trevino, & Butterfield, 2001). McCabe and Trevino (1996) report that students don't believe they'll get caught because instructors are indifferent to their activities. Gerdeman (2000) reports students' belief that if they do get caught they won't be punished severely even if the institution has policies for dealing with such misconduct. In today's high-stakes testing environment, where there is such a strong emphasis on grades, students believe there is a large reward for success at any cost (Whitley, 1998). Certainly they see on the news successful cheaters in the real world constantly getting away without severe penalties.

How Do Students Cheat?

1. Students pass information to a neighbor; for example, they may loan a neighbor an eraser with the answer on the eraser.

2. Students use notes written on clothing, skin, or small note cards.

3. Students store answers in calculators or cassette recorders used during the exam.

4. Students peek at a knowledgeable neighbor's exam (sometimes seated in groups around the best student in the fraternity).

5. Students use tapping, hand code, cell phones, instant messaging, or other communication.

6. Students accuse the teacher of losing an exam (that they never turned in).

7. Students pay someone else to take an exam or write a paper for them.

8. Students copy or paraphrase material for a paper without acknowledging the source.

Preventing Cheating

"OK, so we want to prevent cheating. What can we do?"

If it's true that cheating comes from some of the causes just mentioned, then there is a lot of proactive action that you can take to prevent it or discourage it from happening. Researchers are fairly consistent in many of their recommendations. Here are a few that I've gleaned from the now extensive literature on cheating in college (Gerdeman, 2000; McMurtry, 2001; Pulvers & Diekhoff, 1999; plus Websites from teaching and learning sites at many of the major universities around the country such as the University of Illinois and the University of California at Santa Barbara). They're fairly consistent with my own practices.

An obvious first answer is to reduce the pressure. While you can't affect the general academic atmosphere that puts heavy emphasis on grades, you can influence the pressure in your own course, for example, by providing a number of opportunities for students to demonstrate achievement of course goals, rather than relying on a single examination. A second answer is to address the issue in your syllabus or have a discussion on the topic early in your course.

A third answer is to make reasonable demands and write a reasonable and interesting test. Some cheating is simply the result of frustration and desperation arising from assignments too long to be covered adequately or tests requiring memorization of trivial details. In some cases cheating is simply a way of getting back at an unreasonable, hostile teacher.

A fourth answer is to develop group norms supporting honesty. I frequently give my classes a chance to vote on whether or not we will conduct the tests on the honor system. I announce that we will not use the honor system unless the vote is unanimous, since it will not work unless everyone feels committed to it. If the vote is unanimous, I remind the students of it on the day of the exam and ask whether they still wish to have the test under the honor system. I haven't collected data on the success of this approach, but I've never had a complaint about it. Although only a minority of classes vote for the honor system, a discussion of academic dishonesty is itself useful in helping students recognize why cheating is bad. I've taken to having the students sign a pledge of academic integrity prior to each exam. I think it reminds them of my expectations and reinforces the impression that I care.

Fifth, if some students are not doing well in the course, talk to them and find out what has gone wrong and what they can do to improve. Try to reduce the stress that leads to cheating. If there are stresses originating beyond your course, suggest counseling.

What else can be done?

One principle is to preserve each student's sense that he or she is an individual with a personal relationship both with the instructor and with other students. Students are not as likely to cheat in situations in which they are known as in situations in which they are anonymous members of a crowd. Thus, if a large course has regular meetings in small discussion or laboratory sections, there is likely to be less cheating if the test is administered in these groups than if the test is administered

en masse. Moreover, if the test is given in their regular classroom, they may perform better because of the cues to their original learning (Metzger et al., 1979).

Even in small groups, cheating will occur if the instructor seems unconcerned. Graduate student teaching assistants often feel that any show of active proctoring will indicate that they do not trust the students. There is certainly a danger that the teacher will appear to be so poised to spring at a miscreant that the atmosphere becomes tense, but it is possible to convey a sense of alert helpfulness while strolling down the aisles or watching for questions.

The most common form of cheating is copying from another student's paper. To reduce this I usually ask to have a large enough exam room to enable students to sit in alternate seats. I write on the board before students arrive, "Take alternate seats." Some students fail to see the sign, so in large exams you not only need two proctors at each door passing out exams but at least one more to supervise seating.

In the event that you can't get rooms large enough to permit alternate seating, you probably should use two or more alternate forms of the test. Houston (1983) found that scrambling the order of items alone did not reduce cheating. Since I prefer to have items on a test follow the same order as the order in which the material has been discussed in the course, I scramble the order of items only within topics and also scramble the order of alternatives. I typically write separate sets of essay questions for the two tests. It is difficult to make two tests equally difficult, so you probably will want to tabulate separate distributions of scores on each form of the test.

Whether you use one form or two, don't leave copies lying around your office or the typist's office. One of our students was nearly killed by a fall from a third-floor ledge outside the office where he hoped to steal the examination, and janitors have been bribed to turn over the contents of wastebaskets thought to contain discarded drafts of the test.

Handling Cheating

Despite preventive measures, almost every instructor must at some time or another face the problem of what to do about a student who is cheating. For example, as you are administering an examination you note that a student's eyes are on his neighbor's rather than on his own paper. Typically you do nothing at this time, for you don't want to embarrass an innocent student. But when the eyes again stray, you are faced with a decision about what to do.

Most colleges have rules about the procedures to be followed in case of cheating. Yet instructors are often reluctant to begin the procedure. The reasons for instructor reluctance vary. Sometimes it is simply uncertainty about whether or not cheating really occurred. Students' eyes do wander without cheating. Answers may be similar simply because two students have studied together. "If the student denies the charge, what evidence do I have to support my accusation?"

Again, unwillingness to invoke the regulations concerning cheating may be based on distrust of the justice of the eventual disposition of the case. Cheating is common in colleges; many teachers have been guilty themselves at some stage in their academic careers. Thus, most of us are understandably reluctant to subject the unfortunate one who gets caught to the drastic possible punishments that more skillful cheaters avoid. Such conflicts as these make the problem of handling a cheater one of the most disturbing of those a new teacher faces.

Unfortunately I've never been completely satisfied that I handle the problem adequately; so my "advice" should, like the rest of the advice in this book, be regarded simply as some ideas for your consideration rather than as dicta to be accepted verbatim. However, much of what I'm going to say is backed up by most writers in this field.

First, let me support the value of following your college's procedures. Find out what they are and what legal precedents may affect what you should do. Even though it may not have been long since you were taking examinations yourself, your role as a teacher requires that you represent established authority rather than the schoolboy code that rejects "tattlers." Moreover, your memories of student days may help you recall your own feelings when you saw someone cheating and the instructor took no action.

Further, student or faculty committees dealing with cheating are not as arbitrary and impersonal as you might expect. Typically, they attempt to get at the cause of the cheating and to help students solve their underlying problems. Being apprehended for cheating may, therefore, actually be of real long-term value to the students.

Finally, following college policies protects you in the rare case in which a student initiates legal action against you for an arbitrary punishment.

There still remain cases where the evidence is weak and you're not quite sure whether or not cheating actually occurred. Even here I advise against such individual action as reducing a grade. If you're wrong, the solution is unjust. If you're right, you've failed to give the student feedback that is likely to change his behavior. In such cases I advise talking to the student and calling the head of the committee handling cheating cases or the student's counselor. It's surprising to find how often your

suspicions fit in with other evidence about the student's behavior. Even when they don't, advice from someone who has additional information about the student will frequently be helpful.

Finally, let's return to the case of the straying eyes. Here you haven't time for a phone call to get advice; your decision has to be made now. Rather than arousing the whole class by snatching away the student's paper with a loud denunciation, I simply ask the student unobtrusively to move to a seat where he'll be less crowded. If he says he's not crowded, I simply whisper that I'd prefer that he move. So far no one's refused.

AFTER THE TEST

Grading Objective Tests

Of course the most wonderful thing about objective tests is they are easier to grade. Or are they? The important point to remember is to get the scoring key right! There's nothing more disconcerting to students than to find that the test was scored incorrectly. I strongly recommend that you check and double check the keys to be sure that the marks are correct. Then, before you give the tests back, it **really** pays to do a short analysis of overall student performance on each item. This is called an item analysis. It consists of figuring what percentage of the students missed each question and how did the performance of the top third of the class compare to the bottom third as measured by their overall score. You can short-circuit a lot of student complaints by identifying items that were troublesome and knowing why. For example, if an item is missed by more than half of the class, I always reread the item to see whether there was something unclear. Or if a large number of the students in the top group miss an item, I consider which answer they gave to see if for some reasons the question misled those who actually knew a lot, maybe too much. You still have time to make the necessary adjustments in the scoring to allow for poorly worded questions or a distractor that turns out to be correct after all. By making all these adjustments before you give the papers back, you avoid a lot of confusion about which items were right and which were wrong.

Once the students recognize that you are making a good-faith effort to identify or remediate poorly worded items, they are more likely to give you the benefit of the doubt. You also have the advantage of having at your fingertips solid data on each question so that if a student challenges a question after the test, you will know whether there is any merit to that challenge and be able to respond immediately and authoritatively.

Grading Essay Questions

I recommend that you use some essay questions because of their powerful effect on the way students study, but there is a drawback. Instructors don't grade essay tests very reliably. One problem is that standards vary. First papers are graded differently than later papers. A paper graded immediately after several poor papers is graded differently than one graded after several good papers.

There are seven procedures you can initiate to improve your evaluation of essay examinations—but they entail work.

1. Establish a rubric or set of criteria—not just a list of facts to be included. Are you looking for integration, for analysis, for rational arguments for and against a conclusion? Be prepared to modify your criteria as you find student responses that you hadn't thought of. Learning to create a good grading rubric is worth the effort because it can help you write good questions, maintain reliable grading of answers, and, if shared with the students, help students understand how their answer was graded. Walvoord (1998) has an excellent book on how to create rubrics based on "primary trait analysis."

Creating a good rubric through primary trait analysis involves laying out the key aspects of the response that figure into the grade. For example, on a given essay question, the analysis might list four main points that must be included in the answer, plus criteria for a clean argument and criteria for good writing itself. Then each "trait" is described along a scale of acceptability. Here is an example of a scale for the trait of "solid argumentation":

> Best answer (100 percent credit)—An answer at this level provides clear statements of the thesis or theses being asserted in a logical order that builds to the final conclusion. Each thesis is accompanied by sufficient reasonable evidence to support it. Each thesis also considers and counters reasonable arguments against it. The theses stand together and are internally consistent with one another.
>
> Acceptable answer (80 percent credit)—An answer at this level provides fewer theses but still provides reasonable and primary ones in light of the conclusion. There is evidence offered for each thesis, although possibly overlooking some minor supporting assertions. Several of the more obvious counterarguments are raised and refuted. The order is logical and builds to the conclusion. Transitions between theses are present but ordinary.
>
> Unacceptable answer (no credit)—Any two or more of the following characteristics constitute an unacceptable answer. The answer contains many errors of assertion and omission. No evidence is given or the evidence

given is incorrect or unrelated to the assertion. No attempt or a weak attempt is made to introduce and refute counterarguments. The order of presentation is not logical or convincing. The conclusion is not justified by the arguments.

Creating this type of rubric helps you clarify for yourself what you want in an answer. It also increases the reliability of grading across graders and across time within a single grader's work.

2. Read exams without knowledge of the name of the writer.

3. If you're unsure of what to expect, first read briefly through a random sample of answers. Then, having identified papers of differing levels of excellence, compare them to determine what the distinguishing features are. You will find some characteristics that were not in your original criteria. Now set up the criteria you will use, but don't be rigid. Give students credit when they come up with creative answers that don't fit the rubric.

4. Write specific comments on the papers. One of the problems in using essay exams and in assigning term papers is that students feel that the grading represents some mysterious, unfathomable bias. The more helpful comments you can write on the paper, the more students will learn.

I am finding that computer technology is a big help in my grading of papers (but not essay exams unless they're also done on the computer). I use the editing software available in common word processing programs to read and mark the papers that my students submit in electronic format. I can give a lot more feedback because I'm not limited by how much I can squeeze into the margins, and I can type a lot faster than I can write by hand. In addition, the students can probably read my typing better than my handwriting. (There is more about this in the chapter "How to Enhance Learning by Using High-Stakes and Low-Stakes Writing.")

5. Develop a code for common comments. For example, you might want to use a vertical line alongside paragraphs that are particularly good or "NFD" for "needs further development." Or you can identify the most commonly occurring errors with numbers. When you grade, you can put the number next to the error on the paper and give students the numbered list of errors for reference. They may learn something from reading the whole list even if they didn't make any of those errors.

6. Don't simply give points for each concept or fact mentioned. Doing that just converts the essay into a recall test rather than a measure of higher-level goals of integration and evaluation. Developing rubrics like those described earlier can be helpful in increasing reliability of grading. However, don't use them mechanically. Your overall impression may be as valid.

7. If possible, do your grading in teams. My teaching assistants and I gather after administering a test. We bring in draft model answers for each question. We discuss what we expect as answers for each question. We then establish two- or three-person teams for each essay question. Each team picks eight to 12 test papers, which are circulated among the team members. Each team member notes privately his or her grade for the question. Team members then compare grades and discuss discrepancies until they reach consensus. A second group of tests is then graded in the same way, with grades compared and discrepancies discussed. This procedure continues until the team is confident that it has arrived at common criteria. From this point on, each member grades independently. When a team member is not sure how to grade a paper, it is passed to another team member for an opinion.

We stay with the grading until all the papers are done, but we make a party of it to alleviate fatigue and boredom. Funny answers are read aloud. Sandwiches are brought in from a delicatessen. Teams help other teams for a change of pace or to balance the workload.

If you don't have a team, try to develop your own strategies for maintaining motivation. If you begin to be bored, irritated, or tired, take a break. Or before beginning, pull out the answers of some of your most interesting students and read those when you begin to feel dispirited. Take notes to use in discussing the papers in class. Also take separate notes for yourself on what seem to be common problems that you need to correct in your teaching in the future.

Grading papers is still time consuming but does not become the sort of aversive task that makes for procrastination and long delays in providing feedback to students.

Helping Yourself Learn from the Test

Often we get so wrapped up in the pure mechanics of correcting and grading tests that we overlook the fact that measures of student performance not only can diagnose student weaknesses but also can reveal areas in which our teaching has failed to achieve its purposes. The item analysis process described earlier is especially helpful with this. Once you've achieved some ease with the grading process, look back at the papers to see what they reveal about problems in student understanding. There may be some things about which the entire class seems a bit shaky; in addition, there may be areas of difficulty experienced by certain

subgroups of students—perhaps those with background knowledge or experience different from that of the rest of the class. In short, think about what *you* need to do as well as about what the *students* need to do.

Returning Test Papers

Remember that tests are important tools for learning and that discussion of a test can be a worthwhile use of class time. However, it's also a pretty emotional time for some of the students, and it might pay to delay the discussion until that emotion settles down. In fact, in my own class, I give students the opportunity to challenge the answer to a question in writing before the next class period. I've found that often once the student has had a chance to look over the items and tried to justify their incorrect answers, they realize what they did wrong much more readily than if I just tell them. Sometimes they are actually able to make a good case for their choice, in which case I'll give them credit for their answer. I don't think you should discuss every question in class, but when there are common errors, try to find out why the error occurred and suggest strategies for avoiding such problems in the future. Although you should avoid spending class time quibbling over some individual items, you should make known your willingness to discuss the test individually with students who have further questions.

Helping Students Learn from a Test

The most important function of testing is *not* to provide a basis for grading. Rather, tests are an important educational tool. They not only direct students' studying but also can provide important corrective feedback. The comments that you write on essay tests are far more important than the grade. Students do learn from their corrected papers (McCluskey, 1934). I recommend looking at the suggestions for giving feedback that are included in Chapter 9 of this book. They apply equally to essays, papers, and objectively scorable tests like multiple choice. If you have the time and the temperament, you can increase the probability that students will take that opportunity to learn if you give them a chance to redo an assignment based on your feedback, as described in the previous section.

Dealing with an Aggrieved Student

What about the student who comes to your office in great anger or with a desperate appeal for sympathy but no educationally valid reason for changing the test grade? First of all, listen. Engaging in a debate will simply prolong the unpleasantness.

Ask the student to think aloud about what he or she was thinking when answering the questions that he or she is unhappy about. Once you have heard the student out, if you have decided not to change the grade, try to convert the discussion from one of stonewall resistance to problem solving. Try to help the student find alternative modes of study that will produce better results: "What can we do to help you do better next time?" Encourage the student to shift from blaming you toward motivation to work more effectively. Ask the student to summarize what he or she plans to do before the next test. Although these suggestions may save the instructor some bitter moments, they cannot substitute for the time (and it takes lots) devoted to the construction of good tests.

What Do You Do About the Student Who Missed the Test?

In any large class some students are absent from the test. Their excuses range from very legitimate to very suspicious, but making that discrimination is not always easy.

Makeup tests can involve a good deal of extra work for the instructor. If you devise a new test, you may have trouble assigning a norm with which to grade the makeup comparable to grades on the original test. If you use the same test that the student missed, you cannot tell how much the student has learned about the test from students who took it at the scheduled time. I simply average marks from the tests the student did take to determine the grade, counting the missed test neither for nor against the student.

Another strategy is to drop the lowest score or missed test out of all the tests a student takes. (This, of course, presumes you have enough exams during the semester that one can be dropped.) This also lowers test anxiety because the stakes on any one test are lower. Depending on how strongly you feel about final exams, you could allow students to use the final as the test they drop if they've taken all the other exams and are satisfied with their grade. You'd be surprised what an incentive that is for working diligently during the semester.

IN CONCLUSION

1. Consider using both graded and ungraded tests and moving from less frequent tests to more frequent, where each test can count less.

2. Select question types that target your educational goals.

3. Prepare your students to take the test.

4. Create a class atmosphere that values academic honesty and support and discourages cheating.

5. Develop grading strategies for essay questions so that you won't shy away from using them.

6. Be prepared to address students' complaints about test scores in a way that helps them learn.

7. Learn from the test yourself and show your students how to learn from it as well.

Supplementary Reading

Effective Grading: A Tool for Learning and Assessment by Barbara E. Walvoord and Virginia Johnson Anderson (San Francisco: Jossey-Bass, 1998) does a good job of describing how to create grading rubrics for all manner of written assessments.

Constructing Test Items: Multiple-Choice, Constructed-Response, Performance, and Other Formats, 2nd ed., by Steven J. Osterlind (Boston: Kluwer Academic Publishers, 1998) is a fairly complete discussion of the process of writing different types of test items. It may be a bit long on detail, but the guidelines for item construction are solid and fairly straightforward.

The following resources are drawn from the ERIC Digest series. This is a series of short summaries of research and best practices provided online for educators in a searchable database.

- www.ericfacility.net/databases/ERIC_Digests/index.
- Childs, R. (1989). *Constructing Classroom Achievement Tests*. ERIC Digest. ERIC Clearinghouse on Tests Measurement and Evaluation. ED315426.
- Grist, S., and others (1989). *Computerized Adaptive Tests*. ERIC Digest No. 107. ERIC Clearinghouse on Tests Measurement and Evaluation. ED315425.
- Kehoe, J. (1995). *Basic Item Analysis for Multiple-Choice Tests*. ERIC/AE Digest. ERIC Clearinghouse on Assessment and Evaluation. ED398237.

- Kehoe, J. (1995). *Writing Multiple Choice Test Items*. ERIC/AE Digest. ERIC Clearinghouse on Assessment and Evaluation. ED398236.

Readings about helping students:

- C. E. Weinstein and L. Hume, *Study Strategies for Lifelong Learning* (Washington: American Psychological Association, 1998). The American Psychological Association Division 15 has a whole series of publications on helping students improve their learning. Access them through the APA Publications site.
- D. Sadker and K. Zittleman, "Test Anxiety: Are Students Failing Tests—Or Are Tests Failing Students?" *Phi Delta Kappan*, 2004, *85*(10), 740.

The entire September 2004 issue of *Anxiety, Stress, and Coping* is devoted to test anxiety and research on it, including how to cope with it.

Readings about cheating:

- S. F. Davis, C. A. Grover, A. H. Becker, and L. N. McGregor, "Academic Dishonesty: Prevalence, Determinants, Techniques, and Punishments," *Teaching of Psychology*, 1992, *19*(1), 16–20.
- J. McBurney, "Cheating: Preventing and Dealing with Academic Dishonesty," *APS Observer*, January 1996, 32–35.

One might assume that it would be un-British to cheat. But Stephen Newstead, Arlyne Franklyn-Stokes, and Penny Armstrong found that British students are not much different from Americans in this respect. Their article "Individual Differences in Student Cheating," *Journal of Educational Psychology*, 1996, *88*, 229–241, is consistent with American data.

A particularly interesting set of recommendations comes from the Website "On the Cutting Edge" of the National Association of Geoscience Teachers, which provides workshops for faculty in the geological sciences (serc.carleton.edu/NAGTWorkshops/index.html).

The Center for Academic Integrity at Duke University, Durham, North Carolina (www.academicintegrity.org/cai_research.asp), can provide a lot of information and sponsors workshops and research on academic integrity. They also have a searchable database of 700-plus articles on this topic.