

anatomy, materials technology, engineering, history of art, statistics, medicine, and physics.

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.

McKeachie, W.J. (2006) *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers* 12th ed. Boston: Houghton Mifflin.

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 operationalizing goals 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.

I usually also give midterm and final examinations, but the amount and frequency of tests should depend on the background of your students. 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. For this reason my tests are all cumulative; that is, they cover material previously tested as well as material learned since the last test. Give students a chance to comment on the dates of tests or other assessments. They may know of potential conflicting events that could influence due dates.

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 et al., 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. Of course the teacher can also learn from students' responses to objective tests, but the impact on the teacher of what students are learning seems to be greater and more vivid in reading essay tests.

True-False Items. Although true-false examinations are rather easy to make up, I don't ordinarily advocate their use. Toppino & 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. The box that follows contains hints for their construction that you won't find in an official measurement book.

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 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

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).

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.

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 ten 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.

AFTER THE TEST

Grading Essay Questions

I recommend that you use 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 nine 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% 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% 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. Read all or several of the examinations in a preliminary fashion to establish some notion of the general level of performance.
4. 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.
5. 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. 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.")
6. 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.

7. 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.
8. 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 8 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.

9. Before you get ready to give the tests back, it really pays to do a short analysis of overall student performance on each item. You can short-circuit a lot of student complaints by identifying items that were troublesome and knowing why. 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.

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.

Assigning a Grade

The papers have been corrected, errors noted, comments written. Now you have to worry about grading. I say more about grading in the chapter "The ABC's of Assigning Grades," but for the moment let's consider grades given on a test when you are expected to convert a number of points into a letter grade such as A, B, C, D, or F.

You could grade on a "curve"—a practice also called "norm-referenced grading" because each student is compared to the norm of the rest of the class, the "normal" curve. For example, in grading on the curve one might give the top 10 percent of the

scores A's, the next 25 percent B's, the next 35 percent C's, the next 20 percent D's, and the bottom 10 percent F's. However, grading based on *relative* achievement in a given group may encourage an undesirably high degree of competition. Grading on the curve stacks the cards against cooperative learning because helping classmates may lower one's own grade.

Grading on the curve also discourages the teacher from helping less able students because any student who moves up a grade moves another student down. It increases student anxiety because a student can't interpret his or her grade without information about the overall performance of the class. Even so, many teachers prefer to grade on a curve—some because they are not sure what standards are appropriate, others because they fear that they will be accused of contributing to grade inflation if most students reach high standards of achievement. The latter is a bugaboo that persists despite evidence that grades have not gone up in the last two decades (Adelman, 1999).

A modified curving system is what I call the "gap" system. Like the curve, gap grades compare students to one another by setting grade cutoffs where there are gaps in the frequency distribution of the students' grades. So, for example, my students' grade distribution might be as shown in Table 8.1.

Notice that there is a "gap" in the grade distribution between 88 points and 82 points. If I put the cutoff for an A at 88, then there is a visible difference between the performance of the top group and the second group. (When you don't have big gaps in the distribution, you can look for low points where there is a drop in the number of students at successive levels.) You can even establish some general levels where you'd expect the grade cutoffs to be and then locate the gap nearest to those levels.

In the scenario I just described, students getting 82 points can see that they really performed differently from students getting 89, in contrast with systems where only one point separates an A from a B. Students also are much less likely to argue for more points because the number of points that they would have to talk you out of to get the next highest grade is so large.

Alternatively, and in my opinion preferably, you can grade in terms of percentage of a possible score. This is called "criterion-referenced grading" because each student is compared to a set

Score	Number at That Score	Score	Number at That Score
100	1	89	1
99	0	88	1
98	1	87	0
97	2	86	0
96	3	85	0
95	3	84	0
94	4	83	0
93	4	82	2
92	3	81	3
91	2	etc.	
90	1		

TABLE 8.1 Sample Distribution of Grades

criterion. Thus if a test has 150 possible points, I would tell the students:

If you make 140 or over (93%+), I'll guarantee an A.

135 to 139 (90%+), A-

131 to 134 (87%+), B+

125 to 130 (83%+), B

120 to 124 (80%+), B-, etc.

If everyone gets over 140 points, everyone will get an A, and I'll be very pleased if you all do well.

I tell the students that I may grade more generously than the standards I have announced, but I promise not to be tougher than announced. As it turns out, my distribution of grades has not turned out to be more generous than that of my colleagues—which may indicate that I'm not teaching as effectively as I'd like.

My "percentage of possible points" system is fairly easy to apply but lacks the educational value of criteria or standards tied more directly to course goals. Royce Sadler (1987) describes the use of exemplars and verbal descriptions of quality to set standards for grading.

Returning Test Papers

Remember that tests are important tools for learning and that discussion of a test is a worthwhile use of class time. You might begin by asking students what they learned from the test.

Were students accurate in their assessment of how well they did? (Helping students learn to assess their own learning is a worthy objective.) You don't need to discuss every question, but when there are common errors, try to find out why the error occurred and suggest strategies for avoiding such problems in the future (see Schutz & Weinstein, 1990).

Students do learn from their corrected papers (McCluskey, 1934). Although you may not wish to spend class time quibbling over some individual items, you should make known your willingness to discuss the test individually with students who have further questions.

On multiple-choice questions that many students missed, I recommend this sort of procedure: Read the stem of the item and each of the choices. For each of the incorrect choices give your reasons for regarding it as incorrect. This procedure gives you the jump on the chronic criticizer. It is more difficult to maintain that a given choice is right under these circumstances than it would be if you said nothing about the various alternatives and students could argue that the correct alternative was not completely correct.

There will still be cases in which a legitimate argument arises. If some ambiguities have gotten through the screening process and an item really is capable of two equally correct interpretations, admit it and change scores. But remember that you can't escape aggression simply by changing scores, because every time you admit a new right answer, the students who originally had the question right are likely to feel injured. That's the advantage of having done the item analysis mentioned earlier. You'll head off problems that might arise during class discussion.

For essay tests I try to describe what I expected in a good answer and the most common inadequacies. I may read an example of a good answer (without identifying the student), pointing out how it met the criteria, and I might construct a synthetic poor answer to contrast with the good one.

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.

My colleague Deborah Keller-Cohen asks students coming to see her with complaints about grades to write a paragraph describing their complaint or point of view. She declares her willingness to go over the test of anyone who brings in such a paragraph, noting that she may change the grade either positively or negatively. She reports that this technique has a calming effect, resulting in fewer unfounded complaints and more rational discussion with students who do come in.

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 more frequent tests to less frequent.
2. Select question types that target your educational goals.
3. Develop grading strategies for essay questions so that you won't shy away from using them.
4. Be prepared to address students' complaints about test scores in a way that helps them learn.
5. Learn from the test yourself.

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.

Tests from the Students' 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.