

## QUIZZES, TESTS, AND EXAMS

By Barbara Gross Davis, University of California, Berkeley.  
From *Tools for Teaching*, copyright by Jossey-Bass. For purchase or reprint information,  
contact Jossey-Bass. Reprinted here with permission, September 1, 1999.

Many teachers dislike preparing and grading exams, and most students dread taking them. Yet tests are powerful educational tools that serve at least four functions. First, tests help you evaluate students and assess whether they are learning what you are expecting them to learn. Second, well-designed tests serve to motivate and help students structure their academic efforts. Crooks (1988), McKeachie (1986), and Wergin (1988) report that students study in ways that reflect how they think they will be tested. If they expect an exam focused on facts, they will memorize details; if they expect a test that will require problem solving or integrating knowledge, they will work toward understanding and applying information. Third, tests can help you understand how successfully you are presenting the material. Finally, tests can reinforce learning by providing students with indicators of what topics or skills they have not yet mastered and should concentrate on. Despite these benefits, testing is also emotionally charged and anxiety producing. The following suggestions can enhance your ability to design tests that are effective in motivating, measuring, and reinforcing learning.

A note on terminology: instructors often use the terms *tests*, *exams*, and even *quizzes* interchangeably. Test experts Jacobs and Chase (1992), however, make distinctions among them based on the scope of content covered and their weight or importance in calculating the final grade for the course. An examination is the most comprehensive form of testing, typically given at the end of the term (as a final) and one or two times during the semester (as midterms). A test is more limited in scope, focusing on particular aspects of the course material. A course might have three or four tests. A quiz is even more limited and usually is administered in fifteen minutes or less. Though these distinctions are useful, the terms *test* and *exam* will be used interchangeably throughout the rest of this section because the principles in planning, constructing, and administering them are similar.

### GENERAL STRATEGIES

**Spend adequate amounts of time developing your tests.** As you prepare a test, think carefully about the learning outcomes you wish to measure, the type of items best suited to those outcomes, the range of difficulty of items, the length and time limits for the test, the format and layout of the exam, and your scoring procedures.

**Match your tests to the content you are teaching.** Ideally, the tests you give will measure students' achievement of your educational goals for the course. Test items should be based on the content and skills that are most important for your students to learn. To keep track of how well your tests reflect your objectives, you can construct a grid, listing your course objectives along the side of the page and content areas along the top. For each test item, check off the objective and content it covers. (Sources: Ericksen, 1969; Jacobs and Chase,

1992; Svinicki and Woodward, 1982)

**Try to make your tests valid, reliable, and balanced.** A test is *valid* if its results are appropriate and useful for making decisions about an aspect of students' achievement (Gronlund and Linn, 1990). Technically, validity refers to the appropriateness of the interpretation of the results and not to the test itself, though colloquially we speak about a test being valid. Validity is a matter of degree and considered in relation to specific use or interpretation (Gronlund and Linn, 1990). For example, the results of a writing test may have a high degree of validity for indicating the level of a student's composition skills, a moderate degree of validity for predicting success in later composition courses, and essentially no validity for predicting success in mathematics or physics. Validity can be difficult to determine. A practical approach is to focus on *content validity*, the extent to which the content of the test represents an adequate sampling of the knowledge and skills taught in the course. If you design the test to cover information in lectures and readings in proportion to their importance in the course, then the interpretations of test scores are likely to have greater validity. An exam that consists of only a few difficult items, however, will not yield valid interpretations of what students know.

A test is *reliable* if it accurately and consistently evaluates a student's performance. The purest measure of reliability would entail having a group of students take the same test twice and get the same scores (assuming that we could erase their memories of test items from the first administration). This is impractical, of course, but there are technical procedures for determining reliability. In general, ambiguous questions, unclear directions, and vague scoring criteria threaten reliability. Very short tests are also unlikely to be highly reliable. It is also important for a test to be balanced: to cover most of the main ideas and important concepts in proportion to the emphasis they received in class.

If you are interested in learning more about psychometric concepts and the technical properties of tests, here are some books you might review:

Ebel, R. L., and Frisbie, D. A. *Essentials of Educational Measurement*. (5th ed.) Englewood Cliffs, N.J.: Prentice-Hall, 1990.

Gronlund, N. E., and Linn, R. *Measurement and Evaluation in Teaching*. (6th ed.) New York: Macmillan, 1990.

Mehrens, W. A., and Lehmann, I. J. *Measurement and Evaluation in Education and Psychology*. (4th ed.) New York: Holt, Rinehart & Winston, 1991.

**Use a variety of testing methods.** Research shows that students vary in their preferences for different formats, so using a variety of methods will help students do their best (Jacobs and Chase, 1992). Multiple-choice or short-answer questions are appropriate for assessing students' mastery of details and specific knowledge, while essay questions assess comprehension, the ability to integrate and synthesize, and the ability to apply information to new situations. A single test can have several formats. Try to avoid introducing a new format on the final exam: if you have given all multiple-choice quizzes or midterms, don't

ask students to write an all-essay final. (Sources: Jacobs and Chase, 1992; Lowman, 1984; McKeachie, 1986; Svinicki, 1987)

**Write questions that test skills other than recall.** Research shows that most tests administered by faculty rely too heavily on students' recall of information (Milton, Pollio, and Eison, 1986). Bloom (1956) argues that it is important for tests to measure higher-learning as well. Fuhrmann and Grasha (1983, p. 170) have adapted Bloom's taxonomy for test development. Here is a condensation of their list:

To measure *knowledge* (common terms, facts, principles, procedures), ask these kinds of questions: Define, Describe, Identify, Label, List, Match, Name, Outline, Reproduce, Select, State. Example: "List the steps involved in titration."

To measure *comprehension* (understanding of facts and principles, interpretation of material), ask these kinds of questions: Convert, Defend, Distinguish, Estimate, Explain, Extend, Generalize, Give examples, Infer, Predict, Summarize. Example: "Summarize the basic tenets of deconstructionism."

To measure *application* (solving problems, applying concepts and principles to new situations), ask these kinds of questions: Demonstrate, Modify, Operate, Prepare, Produce, Relate, Show, Solve, Use. Example: "Calculate the deflection of a beam under uniform loading."

To measure *analysis* (recognition of unstated assumptions or logical fallacies, ability to distinguish between facts and inferences), ask these kinds of questions: Diagram, Differentiate, Distinguish, Illustrate, Infer, Point out, Relate, Select, Separate, Subdivide. Example: "In the president's State of the Union Address, which statements are based on facts and which are based on assumptions?"

To measure *synthesis* (integrate learning from different areas or solve problems by creative thinking), ask these kinds of questions: Categorize, Combine, Compile, Devise, Design, Explain, Generate, Organize, Plan, Rearrange, Reconstruct, Revise, Tell. Example: "How would you restructure the school day to reflect children's developmental needs?"

To measure *evaluation* (judging and assessing), ask these kinds of questions: Appraise, Compare, Conclude, Contrast, Criticize, Describe, Discriminate, Explain, Justify, Interpret, Support. Example: "Why is Bach's Mass in B Minor acknowledged as a classic?"

Many faculty members have found it difficult to apply this six-level taxonomy, and some educators have simplified and collapsed the taxonomy into three general levels (Crooks, 1988): The first category knowledge (recall or recognition of specific information). The second category combines comprehension and application. The third category is described as "problem solving," transferring existing knowledge and skills to new situations.

**If your course has graduate student instructors (GSIs), involve them in designing exams.** At the least, ask your GSIs to read your draft of the exam and comment on it. Better

still, involve them in creating the exam. Not only will they have useful suggestions, but their participation in designing an exam will help them grade the exam.

**Take precautions to avoid cheating.**

## TYPES OF TESTS

**Multiple-choice tests.** Multiple-choice items can be used to measure both simple knowledge and complex concepts. Since multiple-choice questions can be answered quickly, you can assess students' mastery of many topics on an hour exam. In addition, the items can be easily and reliably scored. Good multiple-choice questions are difficult to write-see "Multiple-Choice and Matching Tests" for guidance on how to develop and administer this type of test.

**True-false tests.** Because random guessing will produce the correct answer half the time, true-false tests are less reliable than other types of exams. However, these items are appropriate for occasional use. Some faculty who use true-false questions add an "explain" column in which students write one or two sentences justifying their response.

**Matching tests.** The matching format is an effective way to test students' recognition of the relationships between words and definitions, events and dates, categories and examples, and so on. See "Multiple-Choice and Matching Tests" for suggestions about developing this type of test.

**Essay tests.** Essay tests enable you to judge students' abilities to organize, integrate, interpret material, and express themselves in their own words. Research indicates that students study more efficiently for essay-type examinations than for selection (multiple-choice) tests: students preparing for essay tests focus on broad issues, general concepts, and interrelationships rather than on specific details, and this studying results in somewhat better student performance regardless of the type of exam they are given (McKeachie, 1986). Essay tests also give you an opportunity to comment on students' progress, the quality of their thinking, the depth of their understanding, and the difficulties they may be having. However, because essay tests pose only a few questions, their content validity may be low. In addition, the reliability of essay tests is compromised by subjectivity or inconsistencies in grading. For specific advice, see "Short-Answer and Essay Tests." (Sources: Ericksen, 1969, McKeachie, 1986)

A variation of an essay test asks students to correct mock answers. One faculty member prepares a test that requires students to correct, expand, or refute mock essays. Two weeks before the exam date, he distributes ten to twelve essay questions, which he discusses with students in class. For the actual exam, he selects four of the questions and prepares well-written but intellectually flawed answers for the students to edit, correct, expand, and refute. The mock essays contain common misunderstandings, correct but incomplete responses, or absurd notions; in some cases the answer has only one or two flaws. He reports that students

seem to enjoy this type of test more than traditional examinations.

**Short-answer tests.** Depending on your objectives, short-answer questions can call for one or two sentences or a long paragraph. Short-answer tests are easier to write, though they take longer to score, than multiple-choice tests.

They also give you some opportunity to see how well students can express their thoughts, though they are not as useful as longer essay responses for this purpose. See "Short-Answer and Essay Tests" for detailed guidelines.

**Problem sets.** In courses in mathematics and the sciences, your tests can include problem sets. As a rule of thumb, allow students ten minutes to solve a problem you can do in two minutes. See "Homework: Problem Sets" for advice on creating and grading problem sets.

**Oral exams.** Though common at the graduate level, oral exams are rarely used for undergraduates except in foreign language classes. In other classes they are usually time-consuming, too anxiety provoking for students, and difficult to score unless the instructor tape-records the answers. However, a math professor has experimented with individual thirty-minute oral tests in a small seminar class. Students receive the questions in advance and are allowed to drop one of their choosing. During the oral exam, the professor probes students' level of understanding of the theory and principles behind the theorems. He reports that about eight students per day can be tested.

**Performance tests.** Performance tests ask students to demonstrate proficiency in conducting an experiment, executing a series of steps in a reasonable amount of time, following instructions, creating drawings, manipulating materials or equipment, or reacting to real or simulated situations. Performance tests can be administered individually or in groups. They are seldom used in colleges and universities because they are logistically difficult to set up, hard to score and the content of most courses does not necessarily lend itself to this type of testing. However, performance tests can be useful in classes that require students to demonstrate their skills (for example, health fields, the sciences, education). If you use performance tests, Anderson (1987, p. 43) recommends that you do the following (I have slightly modified her list):

- Specify the criteria to be used for rating or scoring (for example, the level of accuracy in performing the steps in sequence or completing the task within a specified time limit).
- State the problem so that students know exactly what they are supposed to do (if possible, conditions of a performance test should mirror a real-life situation).
- Give students a chance to perform the task more than once or to perform several task samples.

**"Create-a-game" exams.** For one midterm, ask students to create either a board game, word game, or trivia game that covers the range of information relevant to your course. Students must include the rules, game board, game pieces, and whatever else is needed to play. For example, students in a history of psychology class created "Freud's Inner Circle,"

in which students move tokens such as small cigars and toilet seats around a board each time they answer a question correctly, and "Psychogories," a card game in which players select and discard cards until they have a full hand of theoretically compatible psychological theories, beliefs, or assumptions. (Source: Berrenberg and Prosser, 1991)

### ALTERNATIVE TESTING MODES

**Take-home tests.** Take-home tests allow students to work at their own pace with access to books and materials. Take-home tests also permit longer and more involved questions, without sacrificing valuable class time for exams. Problem sets, short answers, and essays are the most appropriate kinds of take-home exams. Be wary, though, of designing a take-home exam that is too difficult or an exam that does not include limits on the number of words or time spent (Jedrey, 1984). Also, be sure to give students explicit instructions on what they can and cannot do: for example, are they allowed to talk to other students about their answers? A variation of a take-home test is to give the topics in advance but ask the students to write their answers in class. Some faculty hand out ten or twelve questions the week before an exam and announce that three of those questions will appear on the exam.

**Open-book tests.** Open-book tests simulate the situations professionals face every day, when they use resources to solve problems, prepare reports, or write memos. Open-book tests tend to be inappropriate in introductory courses in which facts must be learned or skills thoroughly mastered if the student is to progress to more complicated concepts and techniques in advanced courses. On an open-book test, students who are lacking basic knowledge may waste too much of their time consulting their references rather than writing. Open-book tests appear to reduce stress (Boniface, 1985; Liska and Simonson, 1991), but research shows that students do not necessarily perform significantly better on open-book tests (Clift and Imrie, 1981; Crooks, 1988). Further, open-book tests seem to reduce students' motivation to study. A compromise between open- and closed-book testing is to let students bring an index card or one page of notes to the exam or to distribute appropriate reference material such as equations or formulas as part of the test.

**Group exams.** Some faculty have successfully experimented with group exams, either in class or as take-home projects. Faculty report that groups outperform individuals and that students respond positively to group exams (Geiger, 1991; Hendrickson, 1990; Keyworth, 1989; Toppins 1989). For example, for a fifty-minute in-class exam, use a multiple-choice test of about twenty to twenty-five items. For the first test, the groups can be randomly divided. Groups of three to five students seem to work best. For subsequent tests, you may want to assign students to groups in ways that minimize differences between group scores and balance talkative and quiet students. Or you might want to group students who are performing at or near the same level (based on students' performance on individual tests). Some faculty have students complete the test individually before meeting as a group. Others just let the groups discuss the test, item by item. In the first case, if the group score is higher than the individual score of any member, bonus points are added to each individual's score. In the second case, each student receives the score of the group. Faculty who use group

exams offer the following tips:

- Ask students to discuss each question fully and weigh the merits of each answer rather than simply vote on an answer.
- If you assign problems, have each student work a problem and then compare results.
- If you want students to take the exam individually first, consider devoting two class periods to tests; one for individual work and the other for group.
- Show students the distribution of their scores as individuals and as groups; in most cases group scores will be higher than any single individual score.

A variation of this idea is to have students first work on an exam in groups outside of class. Students then complete the exam individually during class time and receive their own score. Some portion of the test items are derived from the group exam. The rest are new questions. Or let students know in advance you will be asking them to justify a few of their responses; this will keep students from blithely relying on their work group for all the answers. (Sources: Geiger, 1991; Hendrickson, 1990; Keyworth, 1989; Murray, 1990; Toppins, 1989)

**Paired testing.** For paired exams, pairs of students work on a single essay exam, and the two students turn in one paper. Some students may be reluctant to share a grade, but good students will most likely earn the same grade they would have working alone. Pairs can be self-selected or assigned. For example, pairing a student who is doing well in the course with one not doing well allows for some peer teaching. A variation is to have students work in teams but submit individual answer sheets. (Source: Murray, 1990)

**Portfolios.** A portfolio is not a specific test but rather a cumulative collection of a student's work. Students decide what examples to include that characterize their growth and accomplishment over the term. While most common in composition classes, portfolios are beginning to be used in other disciplines to provide a fuller picture of students' achievements. A student's portfolio might include sample papers (first drafts and revisions), journal entries, essay exams, and other work representative of the student's progress. You can assign portfolios a letter grade or a pass/not pass. If you do grade portfolios, you will need to establish clear criteria. (Source: Jacobs and Chase, 1992)

## CONSTRUCTION OF EFFECTIVE EXAMS

**Prepare new exams each time you teach a course.** Though it is time consuming to develop tests, a past exam may not reflect changes in how you have presented the material or which topics you have emphasized in the course. If you do write a new exam, you can make copies of the old exam available to students.

**Make up test items throughout the term.** Don't wait until a week or so before the exam. One way to make sure the exam reflects the topics emphasized in the course is to write test questions at the end of each class session and place them on index cards or computer files for later sorting. Software that allows you to create test banks of items and generate exams

from the pool is now available.

**Ask students to submit test questions.** Faculty who use this technique limit the number of items a student can submit and receive credit for. Here is an example (adapted from Buchanan and Rogers, 1990, p. 72):

You can submit up to two questions per exam. Each question must be typed or legibly printed on a separate 5" x 8" card. The correct answer and the source (that is, page of the text, date of lecture, and so on) must be provided for each question. Questions can be of the short-answer, multiple-choice, or essay type.

Students receive a few points of additional credit for each question they submit that is judged appropriate. Not all students will take advantage of this opportunity. You can select or adapt student's test items for the exam. If you have a large lecture class, tell your students that you might not review all items but will draw randomly from the pool until you have enough questions for the exam. (Sources: Buchanan and Rogers, 1990; Fuhrmann and Grasha, 1983)

**Cull items from colleagues' exams.** Ask colleagues at other institutions for copies of their exams. Be careful, though, about using items from tests given by colleagues on your own campus. Some of your students may have previously seen those tests.

**Consider making your tests cumulative.** Cumulative tests require students to review material they have already studied, thus reinforcing what they have learned. Cumulative tests also give students a chance to integrate and synthesize course content. (Sources: Crooks, 1988; Jacobs and Chase, 1992; Svinicki, 1987)

**Prepare clear instructions.** Test your instructions by asking a colleague (or one of your graduate student instructors) to read them.

**Include a few words of advice and encouragement on the exam.** For example, give students advice on how much time to spend on each section or offer a hint at the beginning of an essay question or wish students good luck. (Source: "Exams: Alternative Ideas and Approaches," 1989)

**Put some easy items first.** Place several questions all your students can answer near the beginning of the exam. Answering easier questions helps students overcome their nervousness and may help them feel confident that they can succeed on the exam. You can also use the first few questions to identify students in serious academic difficulty. (Source: Savitz, 1985)

**Challenge your best students.** Some instructors like to include at least one very difficult question -- though not a trick question or a trivial one -- to challenge the interest of the best students. They place that question at or near the end of the exam.

**Try out the timing.** No purpose is served by creating a test too long for even well-prepared

students to finish and review before turning it in. As a rule of thumb, allow about one-half minute per item for true-false tests, one minute per item for multiple-choice tests, two minutes per short-answer requiring a few sentences, ten or fifteen minutes for a limited essay question, and about thirty minutes for a broader essay question. Allow another five or ten minutes for students to review their work, and factor in time to distribute and collect the tests. Another rule of thumb is to allow students about four times as long as it takes you (or a graduate student instructor) to complete the test. (Source: McKeachie, 1986)

**Give some thought to the layout of the test.** Use margins and line spacing that make the test easy to read. If items are worth different numbers of points, indicate the point value next to each item. Group similar types of items, such as all true-false questions, together. Keep in mind that the amount of space you leave for short-answer questions often signifies to the students the length of the answer expected of them. If students are to write on the exam rather than in a blue book, leave space at the top of each page for the student's name (and section, if appropriate). If each page is identified, the exams can be separated so that each graduate student instructor can grade the same questions on every test paper, for courses that have GSIs.

#### REFERENCES

Anderson, S. B. "The Role of the Teacher-Made Test in Higher Education." In D. Bray and M. J. Blecher (eds.), *Issues in Student Assessment*. New Directions for Community Colleges, no. 59. San Francisco: Jossey-Bass, 1987.

Berrenberg, J. L., and Prosser, A. "The Create-a-Game Exam: A Method to Facilitate Student Interest and Learning." *Teaching of Psychology*, 1991, 18(3), 167-169.

Bloom, B. S. (ed.). *Taxonomy of Educational Objectives. Vol. 1: Cognitive Domain*. New York: McKay, 1956.

Boniface, D. "Candidates' Use of Notes and Textbooks During an Open Book Examination." *Educational Research*, 1985, 27(3), 201-209.

Brown, I. W. "To Learn Is to Teach Is to Create the Final Exam." *College Teaching*, 1991, 39(4), 150-153.

Buchanan, R. W., and Rogers, M. "Innovative Assessment in Large Classes." *College Teaching*, 1990, 38(2), 69-73.

Clift, J. C., and Imrie, B. W. *Assessing Students, Appraising Teaching*. New York: Wiley, 1981.

Crooks, T. J. "The Impact of Classroom Evaluation Practices on Students." *Review of Educational Research*, 1988, 58(4), 438-481.

Ericksen, S. C. "The Teacher-Made Test." *Memo to the Faculty*, no. 35. Ann Arbor: Center for Research on Learning and Teaching, University of Michigan, 1969.

"Exams: Alternative Ideas and Approaches." *Teaching Professor*, 1989, 3(8), 3-4.

Fuhrmann, B. S., and Grasha, A. F. *A Practical Handbook for College Teachers*. Boston: Little, Brown, 1983.

Geiger, T. "Test Partners: A Formula for Success." *Innovation Abstracts*, 1991, 13 (11). (Newsletter published by College of Education, University of Texas at Austin)

Gronlund, N. E., and Linn, R. *Measurement and Evaluation in Teaching*. (6th ed.) New York: Macmillan, 1990.

Hendrickson, A. D. "Cooperative Group Test-Taking." *Focus*, 1990, 5(2), 6. (Publication of the Office of Educational Development Programs, University of Minnesota)

Jacobs, L. C., and Chase, C. I. *Developing and Using Tests Effectively: A Guide for Faculty*. San Francisco: Jossey-Bass, 1992.

Jedrey, C. M. "Grading and Evaluation." In M. M. Gullette (ed.), *The Art and Craft of Teaching*. Cambridge, Mass.: Harvard University Press, 1984.

Keyworth, D. R. "The Group Exam." *Teaching Professor*, 1989, 3(8), 5.

Liska, T., and Simonson, J. "Open-Text and Open-Note Exams." *Teaching Professor*, 1991, 5(5), 1-2.

Lowman, J. *Mastering the Techniques of Teaching*. San Francisco: Jossey-Bass, 1984.

McKeachie, W. J. *Teaching Tips*. (8th ed.) Lexington, Mass.: Heath, 1986.

Milton, O., Pollio, H. R., and Eison, J. A. *Making Sense of College Grades: Why the Grading System Does Not Work and What Can Be Done About It*. San Francisco: Jossey-Bass, 1986.

Murray, J. P. "Better Testing for Better Learning." *College Teaching*, 1990, 38(4), 148-152.

Savitz, F. "Effects of Easy Examination Questions Placed at the Beginning of Science Multiple-Choice Examinations." *Journal of Instructional Psychology*, 1985, 12(1), 6-10.

Svinicki, M. D. "Comprehensive Finals." *Newsletter*, 1987, 9(2), 1-2. (Publication of the Center for Teaching Effectiveness, University of Texas at Austin)

Svinicki, M. D., and Woodward, P. J. "Writing Higher-Level Objective Test Items." In K. G. Lewis (ed.), *Taming the Pedagogical Monster*. Austin: Center for Teaching

Effectiveness, University of Texas, 1982.

Toppins, A. D. "Teaching by Testing: A Group Consensus Approach." *College Teaching*, 1989, 37(3), 96-99.

Wergin, J. F. "Basic Issues and Principles in Classroom Assessment." In J. H. McMillan (ed.), *Assessing Students' Learning*. New Directions for Teaching and Learning, no. 34. San Francisco: Jossey-Bass, 1988.

*Used with permission from Honolulu Community College.*