



Part VIII

Testing and Grading

38. Promoting Academic Honesty
39. Quizzes, Tests, and Exams
40. Allaying Students' Anxieties about Tests
41. Multiple-Choice and Matching Tests
42. Short-Answer and Essay Tests
43. Grading Practices
44. Calculating and Assigning Grades

Promoting Academic Honesty

In surveys, one-half to nearly three-quarters of college students admit to having cheated at least once during their academic career (Cizek, 1999; Lang, 2008; Maramark and Maline, 1993; McCabe and Trevino, 1996). Students explain their dishonest behavior in various ways: they cheat in response to pressures to get good grades, or in an effort to cope with classes that seem unfair or too demanding, or because they are uncertain about the line between acceptable and unacceptable conduct, or when they feel that their instructors are uncaring or indifferent to their own teaching or students' learning. Procrastination, carelessness, panic in the face of deadlines, peer pressure to support a friend, or belief that everyone does it are other reasons students may cheat.

Some of these factors are beyond an instructor's control, but you can take several steps to promote academic honesty (adapted from Aiken, 1991; Davis et al., 1992; Roberts and Rabinowitz, 1992; Whitley, 1998):

- State the standards for scholarship and conduct. Put these standards in the syllabus and discuss them in class.
- Explain how cheating harms students and describe campus sanctions for dishonesty.
- Structure your course so that students will not be tempted to lie (for example, allow students to miss one quiz without penalty).
- Take visible precautions to detect cheating; let students see that you will not tolerate unethical conduct.
- If cheating occurs, respond swiftly.

General Strategies

Discuss standards of academic honesty at the beginning of the term. General admonitions to “avoid cheating” are relatively ineffective. Help students distinguish between acceptable and unacceptable behavior by giving examples of plagiarism, impermissible collaboration, and other practices. For example, is it

plagiarism if students incorporate ideas from peers or tutors at the student learning center? Is it okay for students to work with friends on homework assignments? Discuss university policies, procedures, and penalties for academic violations. Explain the rationale for proper citations of others' work. Ask students to think about how cheating "takes the place of and prevents learning" (Isserman, 2003). Open the floor to a discussion on questions such as Why should each of us care about academic integrity? or What type of academic environment inspires ethical conduct? (Sources: Landau et al., 2002; Lipson and Reindl, 2003)

Distribute institutional policies. Some colleges and departments produce handouts that include definitions of honest and dishonest conduct, and they require students to sign a statement that they have understood the material. Here are sample definitions from one department at the University of California at Berkeley:

Cheating means getting unauthorized help on an assignment, quiz, or examination. (1) You must not receive from any other student or give to any other student any information, answers, or help during an exam. (2) You must not use unauthorized sources for answers during an exam. You must not take notes, books, cell phones, PDAs, calculators, laptops, or other technological devices to the exam when such aids are forbidden, and you must not refer to any book, notes, or aids while you are taking the exam unless the instructor indicates it is an "open book" exam. (3) You must not obtain exam questions illegally before an exam or tamper with an exam after it has been corrected.

Plagiarism means submitting work as your own that is someone else's. For example, copying material from a book, the web, or other sources without acknowledging that the words or ideas are someone else's and not your own is plagiarism. If you copy an author's words exactly, treat the passage as a direct quotation and supply the appropriate citation. If you use someone else's ideas, even if you paraphrase the wording, appropriate credit should be given. You have committed plagiarism if you purchase a term paper online or from another student, copy a paper from someone who has previously taken the course, download a paper from the web, or submit a paper as your own that you did not write.

Include information about academic integrity in your syllabus. Some faculty distribute a brief statement about the importance of academic integrity, excerpts from their institution's policy on academic dishonesty, definitions of permissible and impermissible collaboration, and information about sanctions ("If you are found to have cheated on homework or an exam, you will receive an F on that

assignment”). Sample statements are posted on the Web site of the Center for Advancement of Teaching at Rutgers University.

You may also wish to include the following information (adapted from Whitley and Keith-Spiegel, 2002):

- a list of campus resources for students (counseling center, student learning center, library workshops, tutoring programs)
- recommendations for resources that discuss how to cite reference works
- an invitation to report concerns about academic dishonesty: “Please come to me with any concerns you have about the conduct of other students. You will be helping me and everyone else in this class, and I will hold your comments in complete confidence.”
- An invitation to discuss any questions about an assignment: “If you are having trouble with an assignment or if you are uncertain about permissible and impermissible conduct, please come to me with your questions.”

Develop a classroom climate and group norms that support honesty. Students are less apt to cheat when they feel that their instructor treats students fairly, grades consistently, encourages and praises students’ contributions, promotes group work, is accessible to students, and has good relationships with students. You might also ask your class to vote to conduct exams under the honor system (without proctors). Research shows lower levels of cheating on campuses that have honor codes or modified honor codes and strong cultures of academic integrity. (Sources: McCabe and Pavela, 2000; McCabe et al., 2001; McKeachie and Svinicki, 2006; Pulvers and Diekhoff, 1999; Whitley and Keith-Spiegel, 2002)

Ensure equal access to study materials. Establish a file (in the library or department office or on the course Web site) for old homework assignments and exam questions. Or attach a sample of past exam questions to the syllabus. (Source: Singhal and Johnson, 1983)

Before exams, revisit the topic of academic integrity. Let students know your expectations and the criteria you will use in evaluating their performance. Give them information about campus resources for help in studying and managing stress. Because students are less likely to cheat if they know they can succeed without resorting to dishonesty, give more rather than fewer tests and encourage students to discuss their difficulties with you (see Chapter 40, “Allaying Students’ Anxieties about Tests,” and Chapter 43, “Grading Practices”). Review students’ work throughout the term so that they know that you know their abilities and achievement levels. (Sources: Eble, 1988; Malehorn, 1983)

Distinguish between fraudulent, legitimate, and unacceptable excuses.

A legitimate excuse is based on events beyond a student's control; a fraudulent excuse is one fabricated to avoid an academic responsibility; an unacceptable excuse, such as forgetting when a paper was due, may be truthful but is not a justifiable reason for failure to do the assigned task. Let students know what you consider to be acceptable and unacceptable excuses, and tell them that you may request proof. But try not to become so cynical that you assume every excuse is an invention. Some faculty who give multiple exams allow students to drop the lowest score. Students who miss a test for any reason receive a zero and can just eliminate that score. Other faculty give a single makeup test that any student can take and substitute that score for an earlier exam. (Sources: Caron et al., 1992; Segal, 2000)

Assessing Students' Understanding of Academic Norms

Give a quiz or exercise on academic conduct. The simplest quiz asks students to indicate whether certain actions are acceptable or not (for example, collaborating with others on an assignment, using the views of another without proper attribution, hiding library books). Iowa State, Indiana University, Penn State, and the University of Southern California, among others, have developed formal online tests, and Roig's survey (1997) assesses students' understanding of plagiarism (reprinted in Cizek, 2003, pp. 98–99). Researchers (Landau et al., 2002) report that students who receive feedback about their performance on the survey are better able to detect plagiarism than students who don't complete the survey.

An online tutorial developed at Dalhousie University discusses plagiarism and how to avoid it. Georgetown University's online tutorial "Scholarly Research and Academic Integrity Tutorial" is required for new students. A Web search for "plagiarism tutorial" will yield other citations.

One faculty member gives students a for-credit assignment to define plagiarism. Unsigned excerpts from the responses are distributed in class, and a short class discussion focuses on discrepancies among the definitions.

Give a homework assignment on plagiarism. One faculty member gives assignments in which students are presented with a redacted page of text from a draft manuscript of a journal article (usually the first page) and asked to indicate where citations are needed, for example, "Previous research has indicated . . ." Students who completed the citation homework assignments reported a better understanding of plagiarism and had fewer citation problems than a control group. (Source: Schuetze, 2004)

Hold a class discussion on cheating. In a sociology course, a faculty member asks students to write unsigned responses to the following questions: Have you ever cheated in school or college? If yes, how would you explain your behavior? Why did you cheat? What were the circumstances? Did you believe it was wrong at the time? Students also indicate whether their responses can be shared with the class. From those granting permission, the instructor reads a few responses aloud as part of a class discussion. (Source: Brezina, 2000)

Gauge any special needs of international students. International students who are new to your institution may be unaware of American academic norms. They may have grown up with different beliefs about, for example, permissible collaboration or the value of copying or paraphrasing the words of respected authorities (for example, copying—without attribution—may be viewed as a sign of respect for and understanding of the work of an established scholar). You may need to make a special effort to help these students. Researchers recommend giving international students tips about the American classroom and the values placed on individual effort, authorship, independent thought, and achievement. (Sources: MacDonald, 2003; Whitley and Keith-Spiegel, 2002)

Defining Plagiarism

Clarify the distinctions between plagiarism, paraphrasing, and direct citation. Give students examples of correct and incorrect ways to use others' ideas and words. You might request permission to distribute the following example from *The Random House Handbook*, 6th ed., by Frederick Crews (New York: McGraw-Hill, 1992, pp. 181–183):

Consider the following source and three ways that a student might be tempted to make use of it:

Source: The joker in the European pack was Italy. For a time hopes were entertained of her as a force against Germany, but these disappeared under Mussolini. In 1935 Italy made a belated attempt to participate in the scramble for Africa by invading Ethiopia. It was clearly a breach of the covenant of the League of Nations for one of its members to attack another. France and Great Britain, as great powers, Mediterranean powers, and African colonial powers, were bound to take the lead against Italy at the league. But they did so feebly and half-heartedly because they did not want to alienate a possible ally against Germany. The result was the worst possible: the league failed to check aggression, Ethiopia lost her independence, and Italy was alienated after all. (J. M. Roberts, *History of the World* [New York: Knopf, 1976], p. 845.)

Version A: Italy, one might say, was the joker in the European deck. When she invaded Ethiopia, it was clearly a breach of the covenant of the League of Nations; yet the efforts of England and France to take the lead against her were feeble and half-hearted. It appears that those great powers had no wish to alienate a possible ally against Hitler's rearmed Germany.

Comment: Clearly plagiarism. Though the facts cited are public knowledge, the stolen phrases aren't. Note that the writer's interweaving of his own words with the source's does not render him innocent of plagiarism.

Version B: Italy was the joker in the European deck. Under Mussolini in 1935, she made a belated attempt to participate in the scramble for Africa by invading Ethiopia. As J. M. Roberts points out, this violated the covenant of the League of Nations. (J. M. Roberts, *History of the World* [New York: Knopf, 1976], p. 845.) But France and Britain, not wanting to alienate a possible ally against Germany, put up only feeble and half-hearted opposition to the Ethiopian adventure. The outcome, as Roberts observes, was "the worst possible: the league failed to check aggression, Ethiopia lost her independence, and Italy was alienated after all." (Roberts, p. 845.)

Comment: Still plagiarism. The two correct citations of Roberts serve as a kind of alibi for the appropriating of other, unacknowledged phrases. But the alibi has no force: some of Roberts' words are again being presented as the writer's.

Version C: Much has been written about German rearmament and militarism in the period 1933–1939. But Germany's dominance in Europe was by no means a foregone conclusion. The fact is that the balance of power might have been tipped against Hitler if one or two things had turned out differently. Take Italy's gravitation toward an alliance with Germany, for example. That alliance seemed so very far from inevitable that Britain and France actually muted their criticism of the Ethiopian invasion in the hope of remaining friends with Italy. They opposed the Italians in the League of Nations, as J. M. Roberts observes, "feebly and half-heartedly because they did not want to alienate a possible ally against Germany." (J. M. Roberts, *History of the World* [New York: Knopf, 1976], p. 845.) Suppose Italy, France, and Britain had retained a certain common interest. Would Hitler have been able to get away with his remarkable bluffing and bullying in the later thirties?

Comment: No plagiarism. The writer has been influenced by the public facts mentioned by Roberts, but he hasn't tried to pass off Roberts' conclusions as his own. The one clear borrowing is properly acknowledged.

Discuss "recycling" and self-plagiarism. Remind students that they cannot resubmit an old paper of theirs as a new product for your course. Ask students to check with you if they have a paper or project that they want to use as the basis

for new work. Permissible activities might include reanalyzing old data using a different method or taking the conclusions of an old paper as the springboard for a new one.

As appropriate, distribute a paper from a term-paper mill. Hundreds of Web sites exist selling thousands of papers. Consider purchasing one related to the content of your course. Have students write a critique of the paper or share with them your criteria on why the paper is not very good. Or distribute and discuss Hansen's essay (2004) that shows how poorly written downloaded papers can be. You might also warn students about term-paper mills overcharging students. If students complain, the service threatens to report them to school authorities. (Source: Campbell et al., 2000)

Assigning Papers and Written Work

Assign specific topics. Create topics that are likely to require new research, that stress thought and analysis more than recall of facts, and that are challenging but not overwhelming. Topics that are too difficult invite cheating, as do boring, trivial, and uninteresting topics (see Chapter 35, "Designing Effective Writing Assignments"). Your choice of topics will not prevent dishonest students from paying someone to write a paper, but you can make it hard for students to use a paper mill or to resubmit their own or someone else's paper if you, as appropriate, frame topics in the context of current events, local issues, or conferences or symposia held on campus. You can also discourage cheating by assigning topics that require students to conduct interviews, undertake field research, solve a problem, or compare the strengths and weaknesses of two related research papers. Consider assignments that take the form of letters to authors, fictional conversations between two authors or characters, or explanations of concepts for a specific audience, such as a friend who knows nothing about the subject or elementary school children. (Sources: Anson, 2003–04; Eble, 1988; Singhal and Johnson, 1983; Sterngold, 2004)

Change the assignments for each offering of a course. Changing the topics prevents students from resubmitting work done by former students.

Require specific references. If you require students to use particular sources (for example, certain databases) or to cite at least one source that is no more than a year old, students are unlikely to be able to meet your conditions with a purchased or cut-and-paste paper.

Give a short lecture on how to research and write a paper. Teach students the skills they need to avoid plagiarism, and help them understand that writing is a difficult process and that struggling is natural. Remind students about the materials and consultation services offered by the campus library and student learning center.

Assign one or more short papers. Short assignments—written in class or at home—help students develop their writing skills and help you assess their abilities. Also, if you assign a short paper early in the term, you will have a sample of each student’s writing, which may help you spot a term paper that isn’t the student’s own work. See Chapter 34, “Helping Students Write Better in All Courses.”

Break a major assignment into parts. Give students deadlines for each of five steps: (1) stating a topic or a preliminary thesis, (2) compiling an annotated bibliography, (3) producing research notes, (4) submitting an outline of the paper, and (5) submitting a first draft. This approach helps students write better papers and prevents them from procrastinating—or from downloading a paper. In some courses, you might ask students to share their outlines and first drafts with you, with the teaching assistant, or with each other.

Require submitted papers to be accompanied by selected cited sources. As appropriate, ask students to submit photocopies of cited sources—the first page of an article, book, or Web site—as well as copies of all pages containing passages that they quote in their paper. Or tell students that you will select one bibliographic entry and ask them to produce that item.

Have students keep a research journal or log. Ask students to list the Web searches they conducted (giving the keywords for each search), the journal databases they searched, the librarians they met with, and so on. For each entry, students indicate whether the effort was helpful or not. Or ask students to submit a log of all the people—librarians, tutors at the learning center, graduate student instructors, fellow students—who have commented on their work or drafts.

Have students write a short reflective essay about their experience in writing the paper. Ask students to describe the most important thing they learned, how they located sources, the dead ends they encountered, the sources that were most or least helpful, how their ideas evolved, or how they developed the organizational structure of the paper.

Ask students to sign a statement of authorship. Some faculty ask for a one-sentence statement of authorship, and others use a checklist for each step of the

process. Some faculty, finding such statements off-putting, use a portion of class time to develop a statement of authorship that reflects students' shared standards and common understandings.

Collect papers from students during class or have students turn in papers online. If papers are turned in to a department or faculty office, consider using locked mailboxes with slots for collection. If you ask students to submit papers online, you have the potential of checking students' work against the Web or plagiarism databases.

Detecting Plagiarism

Be alert to telltale signs of plagiarism. A plagiarized paper may have formatting inconsistent with what you requested, may contain odd sentences intermixed in an otherwise coherent passage (the result of a student's effort to customize the paper), may include out-of-date references or citations to material not available in your library, and may differ in writing style or quality from a student's past work. (Source: Lathrop and Foss, 2000)

Use a Web search to find plagiarized material. By copying a sentence or string of words into a search field, you can see if the phrase appears elsewhere. For instructions on how to use Web search engines or library databases of journal articles to detect plagiarism, consult with your librarian.

Learn about campus policies, if any, on text-matching software. Text-comparison for the detection of plagiarism is offered by commercial vendors (for example, Turnitin.com, which uses third-party servers) and packaged as part of learning management systems. Some campuses have site licenses for software; other colleges rely on the decisions of departments and individual faculty. Faculty who decide to use the software should inform their students of their intentions. McGill University offers advice on what to say to students (www.mcgill.ca/integrity/textmatching/). At some universities, students are encouraged to submit their papers through the software on their own before turning their paper in for a grade. *Caution:* Do not use free plagiarism-detection services that advertise on the Web. Some of these sites purloin the papers you submit and sell them.

Text-matching software may have a strong deterrent effect, and it can save instructors from having to do individual online searches to check for plagiarized material. However, the vendors' storage of student papers on non-university servers (for future comparisons) raises concerns about intellectual property,

copyright, and student privacy. Some faculty also believe that the use of the software introduces mistrust into the student-faculty relationship. And, of course, the software cannot detect custom-written papers. (Sources: Hansen, 2003; Johnson et al., 2004; Scanlon, 2003)

Developing and Administering Exams

Change exam questions as often as is practical. Ask students (and graduate student instructors, if you have them) to submit questions, which you can then adapt for future exams (see Chapter 39, “Quizzes, Tests, and Exams”). Be aware of online test collection Web sites where students post and can read past exams and answers. Some faculty argue that these sites could be used to cheat. Other faculty are less concerned because students have always redistributed exams. The best strategy is to make up new tests and keep your items fresh. (Source: Young, 2008)

For multiple-choice exams, use multiple forms. Scramble the order of the pages of the exam (for example, page 2 before page 1 for some copies), or scramble the order of the questions. To signify different test versions, print the first page with just the course title and date on different colored paper. However, avoid using vividly colored paper for the full exam since research shows that students perform best when tests are printed on white paper or light pastels. (Sources: Skinner, 2004; Tal et al., 2008)

Safeguard your exams. Store electronic versions on a CD that is then kept in a secure location. Never send an exam through e-mail. Keep hard copies in a locked drawer or cabinet, and destroy extra copies.

Proctor the exam. Unless your institution or class is on an honor system, monitor the test yourself or arrange for a proctor. During the exam, a proctor should walk up and down the aisles and be alert for unusual behavior: sequences of hand and feet positions or of tapping (to represent responses to multiple-choice questions), surreptitious opening of books or trading of papers, or the use of electronic devices. Cizek's examples (1999) of student misconduct include a system of color-coded M&Ms for signaling answers and the use of a gum wrapper as a crib sheet. YouTube has clips giving students advice on ways to cheat.

Proctors should spend some time in the back of the room, so that students who are thinking about cheating will have to turn around to see where the proctors are. Students are more likely to cheat when teaching assistants monitor exams than when faculty are in the room. In large classes, cheating declines as the

number of proctors increases. (Sources: Davis et al., 1992; Kerkvliet and Sigmund, 1999; Singhal and Johnson, 1983)

Seat students randomly. Impose a seating scheme that will separate friends and place students in alternate seats. Before the period begins, put the seating chart on a data projector or overhead, so that students can find their assigned seat. Have students place personal belongings on the floor rather than in empty seats. If needed, requisition a second classroom.

In large classes, check student photo IDs. Let students know in advance (in the syllabus and before the test) that you may check photo IDs against class lists to be certain that each student takes his or her own exam. Students can place their IDs on their desks while they take the exam or they can show it to the instructor when they turn in their exam. If you don't want to use photo IDs, seat students by section so that graduate student instructors can determine whether all their students are in attendance and that "ringers" are not taking tests. (Source: Whitley and Keith-Spiegel, 2002)

Keep a seating chart. Hand out blue books or exams with prerecorded seat numbers. In rooms without seat numbers, pick up the exams in the sequence of rows. (Source: Singhal and Johnson, 1983)

Ban all electronic devices, except those required for the exam. Mobile phones, PDAs, and programmable watches and pens can be used to hold, access, communicate, or disseminate information. Some faculty bring a large clock to class or write "time remaining" on the board and ask students to remove their watches and put them in their pockets. If students are allowed to use calculators, try to make certain that the memories are cleared before students begin the exam. Some faculty require students to buy a specific type of calculator so that students using unapproved devices are easy to spot. Do not post any answer keys or explanations on the Web until after the exam—dishonest students have used mobile phones to ask friends to download answers from the Web and text message them during the exam. (Source: Whitley and Keith-Spiegel, 2002)

Supply scratch paper. Some faculty do not allow students to use their own paper or pages of their blue books.

Take action if you observe inappropriate behavior. Don't let it go unchecked. If you notice "wandering eyes," whisper a warning to the student ("You may not realize this, but your behavior makes it appear as if you are looking at another

person's test"), or direct the student to another seat. If you observe cheating, position yourself near the offenders to discourage them. Or make a general public announcement: "Please do your own work." If you have suspicions about students, allow them to complete the exam, take notes on what you observe, and flag the exam for close review. (Sources: Cizek, 1999; McKeachie and Svinicki, 2006)

Ask students to sign their exams indicating that the work is their own. Some faculty have students write, sign, and date the following statement: "I have neither given nor received help on this exam."

Maintain order when students turn in their exams at the end of the period. Require students to sign an attendance sheet when they turn in their exams, or collect exams from students row by row. Count those present and make certain that the number of examinees matches the number of exams—this will prevent students who did not submit an exam from claiming that they did and that the exam was later lost or misplaced.

Using Blue Books

Have students turn in blue books before the exam. Collect blue books at a pre-exam class meeting, check them, mark them with a code, and redistribute the blue books at random on exam day. (Source: Whitley and Keith-Spiegel, 2002)

Require students to write only on one side of the blue book (left or right). Or ask students to leave a certain number of pages blank at the beginning of their blue books. This prevents students from filling the blue book with notes in advance of the exam.

Examine all the blue books before leaving the classroom. In one scam, a student pretended to take the test but submitted a blank blue book without a name on the cover. The student then completed the test at home and dropped the completed blue book in the hallway near the classroom or the professor's office. When the blue book was returned to the faculty member, he or she was supposed to assume that it slipped out from the pile.

Scoring and Returning Exams

Clearly mark incorrect answers. Use a bold X or slash mark to indicate wrong answers or blank spaces. The goal is to prevent students from changing answers

and claiming scoring errors. If you permit the regrading of exams, photocopy the exams or quizzes of students who have previously asked for regrading before returning their current test. (Source: Whitley and Keith-Spiegel, 2002)

Warn students about software that detects cheating on multiple-choice tests. Error-analysis programs compare two students' responses to determine the probability that their answers are likely to be the result of chance and not copying. This procedure, however, is not absolutely reliable; some faculty rely solely on the deterrent value of telling students that they may use the software. (Sources: Bellezza and Bellezza, 1995; Cizek, 2003; Dwyer and Hecht, 1996; Whitley and Keith-Spiegel, 2002)

Return exams and assignments in person. Maintain the security of graded exams: do not leave them in the department office or on your desk for students to pick up. For large-enrollment courses with graduate student instructors, return exams in section sessions. For large courses without graduate student instructors, use techniques described in Chapter 14, "Preparing to Teach the Large-Enrollment Course."

Handling Suspected Cases of Academic Dishonesty

Deal with the problem immediately. Addressing cases of academic dishonesty can be stressful, unpleasant, and time consuming. Some faculty hesitate to act because they fear lawsuits or retaliation, or they believe that campus policies are unresponsive of faculty or that campus policies and sanctions are too lenient, too harsh, too inflexible, or too arbitrary. But faculty who ignore cheating send the wrong message to students, and students become demoralized if they are aware of their instructors' inaction. (Sources: Cizek, 2003; Hansen, 2003; Keith-Spiegel et al., 1998; McCabe and Pavela, 2004; Tabachnick et al., 1991)

Follow your institution's policies. Some colleges allow faculty to resolve cases informally or formally, while other institutions require all cases to be reported centrally. Informal resolutions tend to be less time consuming and stressful, and they are preferred by most students. They also lead to immediate corrective action (if the student violated standards of academic integrity), and allow for simple resolution in cases where the problems were caused by misconceptions, ignorance, or mistakes. Note even if you informally sanction a student, you may still want to report the outcome centrally to your office of student judicial affairs. They keep records and can identify chronic cheaters. (Sources: Gehring and Pavela, 1994; Whitley and Keith-Spiegel, 2002)

When you suspect cheating or plagiarism, call the student into your office. Never impose a penalty without discussing the allegations with the student. Meet with students individually, if more than one student is involved in an incident. Talk with the student about your suspicions and listen carefully to his or her response. Remember, of course, that a student is innocent until proven guilty. Researchers recommend the following (adapted from Cizek, 2003; Stevens, 1996; Whitley and Keith-Spiegel, 2002):

- If you are unsure about what to say, consult with an experienced colleague or your department chair, or ask the campus student conduct office for guidelines and procedures.
- Give the student a chance to acknowledge wrongdoing by asking, “Is there anything you would like to say about this assignment or exam?”
- Explain your concerns. Treat the student with respect and fairness, but communicate the seriousness of the situation.
- Avoid using incendiary words. Instead of *cheating* or *plagiarism*, discuss *copying* or *insufficient citation of sources*.
- Ask the student an open-ended nonaccusatory question: “Tell me about this paper.” Listen carefully to the student’s explanation without interruption. In many cases you will hear a mixture of pleas, excuses, and tales of hardship and extenuating circumstances.
- If a student denies any wrongdoing, ask questions about specific aspects of the paper or exam. For example, request definitions of terms, interpretations, or restatements of points made.
- If the student admits wrongdoing, explain the consequences and take whatever actions your institution prescribes.
- If the student does not admit wrongdoing, explain that the case will be referred to the campus judicial affairs office for investigation.
- If the student becomes distraught, show some sympathy. If appropriate, suggest a referral to the counseling center.
- Avoid taking notes when you meet with the student; should there be subsequent legal action, an incomplete record of what transpired is worse than no record at all.

Decide on appropriate sanctions if a student admits guilt. Depending upon your institution’s disciplinary procedures, you may have discretion on whether to assign an F on the assignment or test, or to allow the student to write another paper or take another test, or to drop the assignment or test when calculating the student’s course grade, or devise your own sanction. For example, one faculty member, after discovering that two students had submitted identical answers to

a take-home exam, graded one of the exams and gave each student one half of the grade. Before making your decision on sanctions, you may want to ask your department chair about campus practices and options. (Source: McKeachie and Svinicki, 2006)

Do not automatically assign an F in the course. This is a severe punishment that is best made as a result of a formal hearing process. (Source: Whitley and Keith-Spiegel, 2002)

If you encounter more than one case of cheating in a course, raise your concerns in class. One instructor who found several cases of plagiarism on an assignment began class by saying, “I have read your papers. Would anyone like to talk about honesty?” During the discussion, students asked questions, confronted one another, talked about the effects of cheating on other students, and gave their opinions on sanctions (some of which were quite severe). The discussion also reinforced the notion of community standards. The instructor then returned the papers and offered the students the opportunity to revise them.

References

- Aiken, L. R. “Detecting, Understanding, and Controlling for Cheating on Tests.” *Research in Higher Education*, 1991, 32(6), 725–736.
- Anson, C. M. “Student Plagiarism: Are Teachers Part of the Solution or Part of the Problem?” *Teaching Excellence*, 2003–04, 15(1).
- Bellezza, F. S., and Bellezza, S. F. “Detection of Copying on Multiple-Choice Tests: An Update.” *Teaching of Psychology*, 1995, 22(3), 180–182.
- Brezina, T. “Are Deviants Different than the Rest of Us? Using Student Accounts of Academic Cheating to Explore a Popular Myth.” *Teaching Sociology*, 2000, 28(1), 71–78.
- Campbell, C. R., Swift, C. O., and Denton, L. “Cheating Goes Hi-Tech: Online Term Paper Mills.” *Journal of Management Education* 2000, 24(6), 726–740.
- Caron, M. D., Whitbourne, S. K., and Halgin, R. P. “Fraudulent Excuse Making Among College Students.” *Teaching of Psychology*, 1992, 19(2), 90–93.
- Cizek, G. J. *Cheating on Tests: How to Do It, Detect it, and Prevent It*. Mahwah, NJ: Erlbaum, 1999.
- Cizek, G. J. *Detecting and Preventing Classroom Cheating: Promoting Integrity in Assessment*. Thousand Oaks, CA: Corwin Press, 2003.
- Davis, S. E., Grover, C. A., Becker, A. H., and McGregor, L. N. “Academic Dishonesty: Prevalence, Determinants, Techniques, and Punishments.” *Teaching of Psychology*, 1992, 19(1), 16–20.
- Dwyer, D. J., and Hecht, J. B. “Using Statistics to Catch Cheaters: Methodological and Legal Issues for Student Personnel Administrators.” *NASPA Journal*, 1996, 33(2), 125–135.
- Eble, K. E. *The Craft of Teaching* (2nd ed.) San Francisco: Jossey-Bass, 1988.

- Gehring, D., and Pavela, G. *Issues and Perspectives on Academic Integrity*. (2nd ed.) Washington, DC: National Association of Student Personnel Administrators, 1994.
- Hansen, B. "Combating Plagiarism." *The CQ Researcher*, Sept. 19, 2003, 13(32), 773–796.
- Hansen, S. "Dear Plagiarists: You Get What You Pay For." *New York Times* Back to School section, Aug. 22, 2004, 11.
- Isserman, M. "Plagiarism: A Lie of the Mind." *Chronicle of Higher Education*, May 2, 2003, 49(34), B12.
- Johnson, D., Patton, R., Bimber, B., Almeroth, K., and Michaels, G. "Technology and Plagiarism in the University: Brief Report of a Trial in Detecting Cheating." *AACE Journal*, 2004, 12(3), 281–299.
- Keith-Spiegel, P., Tabachnick, B. G., Whitley, B. E., and Washburn, J. "Why Professors Ignore Cheating: Opinions of a National Sample of Psychology Instructors." *Ethics and Behavior*, 1998, 8(3), 215–227.
- Kerkvliet, J., and Sigmund, C. L. "Can We Control Cheating in the Classroom?" *Journal of Economic Education*, 1999, 30(4), 331–343.
- Landau, J. D., Druen, P. B., and Arcuri, J. A. "Methods for Helping Students Avoid Plagiarism." *Teaching of Psychology*, 2002, 29(2), 112–115.
- Lang, J. M. *On Course: A Week-By-Week Guide to Your First Semester of College Teaching*. Cambridge, MA: Harvard University Press, 2008.
- Lathrop, A., and Foss, K. *Student Cheating and Plagiarism in the Internet Era: A Wake-Up Call*. Englewood, CO: Libraries Unlimited, 2000.
- Lipson, A., and Reindl, S. M. "The Responsible Plagiarist: Understanding Students Who Misuse Sources." *About Campus*, July–Aug. 2003, 7–14.
- MacDonald, D. "Originality and the Paraphrasing Machine." *College Teaching*, 2003, 51(4), 166–168.
- Malehorn, H. "Term Papers for Sale and What to Do about It." *Improving College and University Teaching*, 1983, 31(3), 107–108.
- Maramark, S., and Maline, M. B. *Academic Dishonesty among College Students*. Washington, DC: Office of Research, U.S. Department of Education, 1993.
- McCabe, D. L., and Pavela, G. "Some Good News about Academic Integrity." *Change*, 2000, 32(5), 32–38.
- McCabe, D. L., and Pavela, G. "Ten (Updated) Principles of Academic Integrity." *Change*, May/June 2004, 36(3), 10–33.
- McCabe, D. L., and Trevino, L. K. "What We Know About Cheating in College." *Change*, 1996, 28(1), 29–33.
- McCabe, D. L., Trevino, L. K., and Butterfield, K. D. "Dishonesty in Academic Environments: The Influence of Peer Reporting Requirements." *Journal of Higher Education*, 2001, 72(1), 29–45.
- McKeachie, W. J., and Svinicki, M. *McKeachie's Teaching Tips*. (12th ed.) Boston: Houghton Mifflin, 2006.
- Pulvers, K., and Diekhoff, G. M. "The Relationship between Dishonesty and College Classroom Environment." *Research in Higher Education*, 1999, 40(4), 487–498.
- Roberts, D., and Rabinowitz, W. "An Investigation of Student Perceptions of Cheating in Academic Situations." *Review of Higher Education*, 1992, 15(2), 179–190.
- Roig, M. "Can Undergraduate Students Determine Whether Text Has Been Plagiarized?" *Psychological Record*, 1997, 47(1), 113–122.

- Scanlon, P. M. "Student Online Plagiarism: How Do We Respond?" *College Teaching*, 2003, 51(4), 161–165.
- Schuetz, P. "Evaluation of a Brief Homework Assignment Designed to Reduce Citation Problems." *Teaching of Psychology*, 2004, 31(4), 257–259.
- Segal, C. F. "The Dog Ate My Disk and Other Tales of Woe." *Chronicle of Higher Education*, Aug. 11, 2000, A64.
- Singhal, A., and Johnson, P. "How to Halt Student Dishonesty." *College Student Journal*, 1983, 17(1), 13–19.
- Skinner, N. F. "Differential Test Performance from Differently Colored Paper: White Paper Works Best." *Teaching of Psychology*, 2004, 31(2), 111–113.
- Sterngold, A. "Confronting Plagiarism: How Conventional Teaching Invites Cybercheating." *Change*, May/June 2004, 36(3), 16–21.
- Stevens, E. H. "Informal Resolution of Academic Misconduct Cases: A Due Process Paradigm." *College Teaching*, 1996, 44(4), 140–144.
- Tabachnick, B. G., Keith-Spiegel, P., and Pope, K. S. "Ethics of Teaching: Beliefs and Behaviors of Psychologists as Educators." *American Psychologist*, 1991, 46(5), 506–515.
- Tal, I. R., Akers, K. G., and Hodge, G. K. "Effect of Paper Color and Question Order on Exam Performance." *Teaching of Psychology*, 2008, 35(1), 26–28.
- Whitley, B. E. "Factors Associated with Cheating among College Students: A Review." *Research in Higher Education*, 1998, 39(3), 235–274.
- Whitley, B. E., and Keith-Spiegel, P. *Academic Dishonesty: An Educator's Guide*. Mahwah, NJ: Erlbaum, 2002.
- Young, J. R. "When Web Sites Post Test Answers Online, Professors Worry." *Chronicle of Higher Education*, August 1, 2008, 54(47), A8.

Quizzes, Tests, and Exams

Testing is an integral part of instruction, and well-designed tests serve four principal functions. First, tests can motivate students and help them structure their academic efforts. Researchers report that students study in ways that reflect how they think they will be tested (Martínez, 1999; McKeachie and Svinicki, 2006; Wergin, 1988). If they expect an exam focused on facts, they will memorize details; if they expect a test that will require problem solving or analysis, they will practice those skills. Second, tests give students an indication of which topics or skills they have not yet mastered and should concentrate on. Third, tests help instructors identify students' errors and misconceptions and adjust instruction to improve learning. Fourth, tests help instructors document whether students are learning what they are expected to learn. The following suggestions can enhance your ability to design tests that are effective in motivating, measuring, and reinforcing learning.

A note on terminology: Most educators use *exams* to refer to midterms and finals, in contrast to *tests*, which are more limited in scope and duration, and *quizzes*, which are more limited still. In this chapter, however, *test* and *exam* are used interchangeably, because the principles in planning, constructing, and administering them are similar.

General Strategies

Begin by focusing on learning outcomes. As you prepare a test, think about the kinds of learning you want to assess: What do you want students to remember, understand, apply, analyze, evaluate, or create? These outcomes will determine the types of items, the range of difficulty of items, the length and time limits for the test, the format and layout of the exam, and the scoring procedures.

View testing as an opportunity to understand your students' intellectual progress. In addition to determining whether students have correctly understood the course content (mastery approach to testing), tests can explore how students have mentally organized their knowledge and how students think about the concepts being studied (cognitive approach to testing). See Chapter 32, "Informally Assessing Students' Learning" and Chapter 29, "Helping Students

Learn.” (Sources: Bain, 2004; Black and Wiliam, 1998; Carver, 2006; Means, 2006; Ramsden, 2003)

Decide what to assess. In some ways, testing defines the curriculum. Assessment sends a message about the standards and amount of work required and about which aspects of the course are most important. A first step in creating a test is to identify the fundamental concepts and skills that define competence in the subject at different stages of progress. To link these fundamentals to assessment, some experts recommend creating a Table of Specifications, a grid that lists key knowledge or content down the side of the page and cognitive outcomes or competencies across the top (for example, remember, understand, apply, analyze, evaluate, create). The task then is to create test items for each cell. (Sources: Brookhart, 1999; Connor-Greene, 2000; Jacobs and Chase, 1992; Ory and Ryan, 1993; Payne, 2003; Ramsden, 2003)

Aim for validity and reliability. *Validity* refers to the soundness, trustworthiness, or legitimacy of the inferences, decisions, or actions made on the basis of the results of a test. Although people often speak about the “validity of a test,” tests themselves are not valid or invalid. Validity concerns the adequacy and appropriateness of a specific interpretation of a test’s results. For example, the scores on 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 no validity for predicting success in physics. For classroom exams, a concern is whether the content of a test adequately samples the content of the course so that an instructor can draw valid inferences about a student’s competence based on a test score. An exam that consists of only three relatively difficult problems, for example, will not yield valid inferences about what students know.

Reliability is the extent to which test scores are dependable and consistent. Are identical or nearly identical scores obtained, for example, when the test is scored by different people? Statewide and nationwide tests rely on technical procedures for determining reliability, but for general classroom tests, the best advice is to write unambiguous questions, offer clear directions, and use objective scoring criteria. One other consideration: very short tests are likely to be unreliable. (Sources: AERA, 1999; Miller et al., 2008; Moss et al., 2006)

Use a variety of testing formats. Students who have good learning skills and are confident in their academic abilities often prefer essay tests, while students with poor learning skills and those who are anxious prefer multiple-choice tests. To give all your students an opportunity to do their best, use several types of

questions. For best results, however, do not introduce a new format on the final exam: if you have given multiple-choice quizzes and midterms, don't ask students to write an all-essay final. (Sources: Bridgeman and Morgan, 1996; Jacobs and Chase, 1992; Lowman, 1995; McKeachie and Svinicki, 2006; Birenbaum and Feldman, 1998)

Create questions that test skills other than recall. Research shows that most classroom tests rely too heavily on students' recall of information (Milton et al., 1986). To measure a range of skills, write questions that use the following verbs (adapted from Anderson et al., 2001; Fuhrmann and Grasha, 1983; Montepare, 2005):

- *To measure knowledge* (of terms, facts, principles, procedures): *define, describe, identify, label, list, match, name, outline, reproduce, select, state, tabulate*. Example: "List the steps involved in determining the potential toxicity of a new drug."
- *To measure comprehension and understanding* (explaining or interpreting the meaning of material): *classify, convert, defend, distinguish, estimate, exemplify, explain, extend, generalize, give examples, infer, interpret, predict, summarize*. Example: "Summarize the basic tenets of deconstructionism."
- *To measure application* (using a concept or principle to solve a problem; applying concepts and principles to new situations): *apply, calculate, demonstrate, execute, implement, modify, make use of, operate, prepare, produce, relate, show, solve, use*. Example: "Calculate the deflection of a beam under uniform loading."
- *To measure analysis* (understanding the interrelationship of component parts; recognizing unstated assumptions or logical fallacies; distinguishing between facts and inferences): *analyze, attribute, diagram, differentiate, dissect, distinguish, illustrate, infer, organize, 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 evaluation* (judging and assessing): *appraise, assess, compare, conclude, contrast, convince, criticize, decide, describe, design, discriminate, disprove, evaluate, explain, judge, justify, interpret, prove, recommend, support, theorize*. Example: "Why is Bach's Mass in B Minor acknowledged as a classic?"
- *To measure synthesis and creating* (producing something new from component parts; integrating learning from different areas; solving problems by creative thinking): *construct, create, design, devise, explain, generate, imagine, organize, plan, produce, rearrange, reconstruct, revise, tell*. Example: "How would you restructure the school day to reflect children's development needs?"

Involve your graduate student instructors (GSIs) 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. See Chapter 38, “Promoting Academic Honesty.”

Types of Tests

Multiple-choice tests. Multiple-choice items can be used to measure both simple knowledge and complex concepts. Since students can answer multiple-choice questions quickly, you can assess their mastery of many topics on a fifty-minute exam. In addition, the items can be easily and reliably scored. Good multiple-choice questions are difficult to write, however; see Chapter 41, “Multiple-Choice and Matching Tests,” for guidance on how to develop and administer this type of test.

True-false tests. Because guessing will produce the correct answer half the time, true-false tests are likely to produce high scores. Place true-false items in a separate section, not interspersed with other types of items. Some faculty add an “explain” column in which students write a sentence or two 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 Chapter 41, “Multiple-Choice and Matching Tests,” for suggestions about developing this type of test.

Essay tests. Essay tests require students to organize, integrate, and interpret material, and to express themselves. See Chapter 42, “Short-Answer and Essay Tests,” for guidelines on creating this type of test. Research indicates that students study more efficiently for essay exams than for multiple-choice tests: students preparing for essay tests focus on broad issues, general concepts, and interrelationships rather than on specific details, and this approach results in somewhat better test performance on all types of exam questions. Essay tests also give instructors an opportunity to comment on students’ progress, the quality of their thinking, the depth of their understanding, and the difficulties they are having. However, because essay tests pose only a few questions, their content validity may be low. In addition, the reliability of essay tests can be compromised by subjectivity or inconsistencies in grading.

A variation of an essay test asks students to correct sample essay answers. One faculty member uses the following technique: Two weeks before the exam, he distributes ten to twelve essay questions, which he discusses with students in class. For the 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 sample essays contain common misunderstandings, correct but incomplete responses, and illogical inferences. (Source: McKeachie and Svinicki, 2006)

Short-answer tests. Short-answer questions can call for one or two sentences or a long paragraph. Short-answer tests are easier to write than multiple-choice tests, but they take longer to score. See Chapter 42, “Short-Answer and Essay Tests,” for guidelines.

Problem sets. In mathematics and sciences courses, tests often include problem sets. As a rule of thumb, allow students ten minutes to solve a problem you can do in two minutes. See Chapter 37, “Homework: Problem Sets,” for advice on creating and grading problem sets.

Oral exams. Oral exams are rarely used for undergraduates except in foreign language classes. Most instructors find oral tests too time-consuming, too anxiety-provoking for students, and too difficult to score, even when the exam is recorded. One math professor has experimented with individual thirty-minute oral tests in a small undergraduate seminar. Students receive the questions in advance and are allowed to drop one of their choosing, or they pick a question out of a hat. During the oral exam, the professor probes students’ understanding of the theory and underlying principles. A statistics professor administers fifteen-minute individual oral exams. Students are told to pass on questions they cannot answer (saves time), and their grade is determined not by the number of correct responses but by the extent and level of their understanding. Instructors who give oral exams recommend taking a moment to put students at ease, arranging for informal seating, being tactful when offering feedback, and letting students do most of the talking. (Sources: O’Connor, 2004; Race et al., 2005)

Performance tests. Performance tests ask students to demonstrate proficiency in conducting an experiment, executing a series of steps, 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 can be difficult to set up and hard to score, but performance tests are appropriate in classes that require students to demonstrate skills (for example,

health fields, the sciences, education). If you use performance tests, here are some tips (adapted from Race et al., 2005):

- Specify the criteria that will be used for rating or scoring (for example, the level of accuracy in performing the steps, or the time limit for completing a task).
- State the problem so that students know exactly what they are supposed to do.
- Give students a chance to perform the task more than once or to perform several task samples.
- Include a self-assessment.

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 accommodate longer and more complex questions, without sacrificing valuable class time for exams. Problem sets, short answers, and essays are the most appropriate items for take-home exams. Be wary of designing a take-home exam that is too difficult or an exam that does not include limits on the length of the response (or time spent). Also give explicit instructions on permissible and impermissible collaboration. For example, can students consult community-driven knowledge sites where users ask and answer questions of one another? A variation of a take-home test is to hand out the questions a week or more in advance but have the students write their answers in class. Some faculty hand out ten questions and announce that three of them will appear on the exam.

Open-book tests. Open-book tests simulate the workplace, where people routinely use reference books and other resources to solve problems, prepare reports, or write memos. Open-book tests are less stressful for students, but research shows that students do not necessarily perform significantly better on open-book tests. Open-book tests may reduce some students' motivation to study, and some students perform poorly because they devote more time to consulting their references than to answering the test questions. If you offer open-book tests, specify which resources are allowed and whether students will be able to use their laptops during the exam. In order to prevent the improper use of e-mail or text messaging, some faculty propose an honor code for the exam. Faculty also recommend warning students about spending too much time looking through reference materials. (Sources: Crooks, 1988; Golub, 2005; Ioannidou, 1997; Race et al., 2005; Theophilides and Koutselini, 2000; White et al., 2001)

Group exams. According to researchers, group exams offer three advantages: group work promotes deeper understanding of the material, groups outperform individuals, and group tests are less stressful.

Some faculty offer group testing as a stand-alone exam. For example, for a fifty-minute in-class exam, they administer a multiple-choice test of twenty or twenty-five items. On the first test of the semester, students are randomly divided into groups (the most effective size is three to five students). On subsequent tests, students could be assigned to groups in order to minimize differences between group scores, or to balance talkative and quiet students, or to cluster students who are performing at the same level. Each student receives the score of the group. Another variation is to give a public, oral group exam where students answer questions individually but compete for points in teams.

Other faculty have students complete a test individually and then convene as a group and submit a group exam. When the group score is higher than the individual score of any member, bonus points are added to each individual score. Or the instructor may blend the individual score (75 percent of the grade) and the group score (25 percent of the grade). A variation is to have students work on a set of test questions in groups outside of class but take the exam individually during class; in this technique, the in-class exam includes some items that appeared on the group take-home set as well as new items. Another variation is to have students take an individual exam, meet for group discussion, and then retake the exam individually. The grade might be an average of the two scores, or the first score plus bonus points based on the second score.

Faculty who use group exams offer the following tips:

- Give students practice working in groups before implementing group testing.
- Have students work individually on a problem and then compare results.
- Ask students to discuss each question fully and weigh the merits of each answer rather than simply vote on an answer.
- Take steps to minimize “social loafing,” or free-riders who do not do their share of the work.
- Ask each student to sign the group exam, verifying that it accurately reflects the work of the 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.

(Sources: Hodges, 2004; Jensen et al., 2002; Morgan, 2003; Revere et al., 2008; Shindler, 2004; Slusser and Erickson, 2006; Webb, 1997; Yuretich, 2003)

Paired testing. For paired testing, two students work on and submit a single exam. Some students may be reluctant to share a grade, but good students will most

likely earn the same grade they would have earned working alone. Researchers report that students who take an exam with a partner of choice do significantly better on later tests taken individually, compared to students who take all exams solo. Paired testing can also reduce anxiety and cheating and increase confidence and enjoyment of the course.

Pairs can be assigned in three ways: *Random assignment* breaks up cliques, avoids “last chosen,” and prevents perceptions of instructor bias. *Purposeful assignment* pairs students based on level of achievement or other characteristics. In *self-selection*, students choose their partners, which may ease some students’ anxiety as well as relieve the teacher of responsibility if students attribute poor performance to their partner. If you have an odd number of students in class, decide whether to form a triad or allow a student to work alone. For paired in-class testing, allow students enough time to confer and reach consensus. See Chapter 21, “Learning in Groups.” (Sources: Muir and Tracy, 1999; Zimbardo et al., 2003)

Portfolios. A portfolio is a selection of coursework that a student assembles in order to illustrate growth and accomplishment over the term. A portfolio might include one or more papers (drafts and revisions), journal entries, essay exams, lab reports, sketches, prototypes, or problem sets. The instructions for submitting a portfolio should state the principles for selecting the pieces, the kinds of class materials and acceptable media (hard copy, electronic files, audio, video, and so on), and the minimum and maximum amount of material to submit. Some campuses encourage students to set up electronic portfolios on the institution’s server. Some instructors grade portfolios pass/not pass. If you use letter grades, explain the grading criteria when you make the assignment. (Sources: Jacobs and Chase, 1992; Race et al., 2005; Shermis and Daniels, 2002)

Online testing. For the most part, students’ scores on online tests are similar to their scores on paper-and-pencil tests. Some students complain that online testing prevents them from highlighting key terms and marking up the questions as they think about the topic and their response. Students also dislike online tests that do not allow them to skip items and return to them later.

When an online exam is to be graded, the test must be secure and resistant to cheating. Options include using classrooms in which computers are monitored; using software that blocks access to information on a laptop, disk drive, or network; and giving open-book exams. Keeping the online test short (ten minutes or less) also minimizes the opportunities for cheating.

An excellent use for online testing is to provide practice quizzes that students take while studying. Such quizzes are especially helpful because they provide students with immediate feedback about their performance. Some learning management systems

will prepare and administer these quizzes once an instructor has taken the time to develop a bank of questions coded by topic, learning objective, and level of difficulty. That initial investment of effort is repaid when the system generates individualized online quizzes; some systems even track how much time a student spends on each question. If your learning management system does not have an online quiz feature, some commercial software packages are available. (Sources: Brooks et al., 2003; Brothen and Wambach, 2004; Daniel and Broida, 2004; Drasgow et al., 2006; LoSchiavo and Shatz, 2002; Naglieri et al., 2004)

Constructing Effective Exams

Prepare new exams each time you teach a course. Though it is time consuming to develop tests, a past exam will not reflect changes in how you presented the material and the topics you emphasized this semester. One way to make sure the exam reflects the current course is to write test questions at the end of each class session or at the end of the week. When you write a new exam, consider making copies of the old exam available to students.

Ask students to submit test questions. Faculty who solicit test questions from students limit the number of items a student can receive credit for (say, two questions per exam), specify the question format (such as multiple choice or short answer), and ask students to supply a citation from the readings or class notes for the correct answer. Tell students that their questions must involve inductive or deductive reasoning and synthesis of the material. Some faculty assemble all the student-generated questions and answers into a database and encourage students to use the database as a study tool. Others select or adapt students' items for use on exams. If you have a large-enrollment class, you can draw randomly from the pool until you have enough questions for the exam. (Sources: Carroll, 2001; Feldberg, 1999; Fellenz, 2004; Green, 1997; Hare, 1997)

Be cautious about using item banks from textbook publishers or found online. Don't take all your questions from the item bank. Some of the items may be poorly written, or focused on trivial topics or on concepts that you did not emphasize. Some test banks are filled with quickly-dashed-off items that have not been pretested. (Sources: Forsyth, 2003; Renner and Renner, 1999; Scialfa et al., 2001)

Make 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.

See Chapter 29, “Helping Students Learn.” (Sources: Bain, 2004; Halpern and Hakel, 2003)

Look at online guides on test construction. Many universities offer online guidelines for writing multiple-choice, true-false, matching, short-answer, and essay questions. If your campus does not, look at the guidelines developed by the University of Oregon or Illinois State University. The following guidelines apply to all tests (adapted from McKeachie and Svinicki, 2006; Sechrest et al., 1999):

- *Prepare clear instructions.* Test your instructions by asking a colleague or a graduate student instructor (GSI) to read them.
- *Include advice on how much time to spend on each section.*
- *Put some easy items first.* Answering easier questions helps students calm their nerves and feel more confident.
- *Challenge your best students.* Consider ending the exam with one very difficult question—but not a trick question—to challenge your best students.
- *Try out the timing.* Allow about half a minute per item for true-false questions, one minute per item for multiple-choice questions, two minutes per short-answer question 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 three or four times as long as it takes you (or a GSI) to complete the test.
- *Attend to the layout.* 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 true-false or multiple-choice questions) together. Remember that students will interpret the amount of space after a short-answer question as an indicator of the length of the answer you expect.

Responding to Subpar Test Performance by the Entire Class

Ascertain whether the exam was faulty. If all or almost all of your students performed poorly on an exam, look carefully at the exam: Were the directions clear? Were the questions thoughtfully prepared—not too tricky and not open to misinterpretation? Did the exam cover the assigned material? Did students have enough time to complete the test?

Analyze why students performed so poorly. Look carefully at students’ answers. Is there evidence that they are not reading the assigned material effectively or not understanding classroom examples? What kinds of questions are students missing?

Do not automatically raise each student's score. Simply adding points to every student's score negates the meaning of the test. Students need to earn points based on what they know and can demonstrate. Instead, you might decide to drop each student's lowest test score during the semester, or you might offer a makeup test or assignment. (Source: PsychTeacher listserv)

Talk about the disappointing results with the class. Ask your students why they think the class as a whole did poorly: shaky study habits? unrealistic expectations? time pressure?

References

- American Educational Research Association (AERA), American Psychological Association, and National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association, 1999.
- Anderson, L. W., and Krathwohl, D. R. (Eds.), *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Addison-Wesley/Longman, 2001.
- Bain, K. *What the Best College Teachers Do*. Cambridge, MA: Harvard University Press, 2004.
- Birenbaum, M., and Feldman, R. A. "Relationships between Learning Patterns and Attitudes towards Two Assessment Formats." *Educational Research*, 1998, 40(1), 90–98.
- Black, P. J., and Wiliam, D. "Assessment and Classroom Learning." *Assessment in Education: Principles, Policy and Practice*, 1998, 5(1), 7–74.
- Bridgeman, B., and Morgan, R. "Success in College for Students with Discrepancies between Performance on Multiple-Choice and Essay Tests." *Journal of Educational Psychology*, 1996, 88(2), 333–340.
- Brookhart, S. M. "The Art and Science of Classroom Assessment: The Missing Part of Pedagogy." *ASHE-ERIC Higher Education Report*, 1999, 27(1).
- Brooks, D. W., Nolan, D. E., and Gallagher, S. "Automated Testing." *Journal of Science Education and Technology*, 2003, 12(2), 183–186.
- Brothen, T., and Wambach, C. "The Value of Time Limits on Internet Quizzes." *Teaching of Psychology*, 2004, 31(1), 62–64.
- Carroll, D. W. "Using Ignorance Questions to Promote Thinking Skills." *Teaching of Psychology*, 2001, 28(2), 98–100.
- Carver, S. M. Assessing for Deep Understanding. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences*. New York: Cambridge University Press, 2006.
- Connor-Greene, P. A. "Assessing and Promoting Student Learning: Blurring the Line between Teaching and Testing." *Teaching of Psychology*, 2000, 27(2), 84–88.
- Crooks, T. J. "The Impact of Classroom Evaluation Practices on Students." *Review of Educational Research*, 1988, 58(4), 438–481.
- Daniel, D. B., and Broida, J. "Using Web-Based Quizzing to Improve Exam Performance: Lessons Learned." *Teaching of Psychology*, 2004, 31(3), 207–208.
- Drasgow, F., Luecht, R. M., and Bennett, R. E. Technology and Testing. In R. L. Brennan (Ed.), *Educational Measurement*. (4th ed.) Westport, CT: American Council on Education/Praeger, 2006.

- Feldberg, R. S. "Increasing Student Involvement in Lectures: (Very) Low Tech Innovations in a Biochemistry Lecture Class." *Biochemical Education*, 1999, 27(2), 71–73.
- Fellenz, M. R. "Using Assessment to Support Higher Level Learning: The Multiple Choice Item Development Assignment." *Assessment and Evaluation in Higher Education*, 2004, 29(6), 703–719.
- Forsyth, D. R. *The Professor's Guide to Teaching: Psychological Principles and Practices*. Washington, DC: American Psychological Association, 2003.
- Fuhrmann, B. S., and Grasha, A. F. *A Practical Handbook for College Teachers*. Boston: Little, Brown, 1983.
- Golub, E. "PCs in the Classroom and Open Book Exams." *Ubiquity*, 2005, 6(9), 1–4. <http://portal.acm.org/citation.cfm?id=1066320>
- Green, D. H. "Student-Generated Exams: Testing and Learning." *Journal of Marketing Education*, 1997, 19(2), 43–53.
- Halpern, D. F., and Hakel, M. D. "Applying the Science of Learning to the University and Beyond." *Change*, 2003, 35(4), 36–41.
- Hare, A. C. "Active Learning and Assessment in Mathematics." *College Teaching*, 1997, 45(2), 76–77.
- Hodges, L. C. "Group Exams in Science Courses." *New Directions for Teaching and Learning*, no. 100. San Francisco: Jossey-Bass, 2004, pp. 89–93.
- Ioannidou, M. K. "Testing and Life-Long Learning: Open-Book and Closed-Book Examination in a University Course." *Studies in Educational Evaluation*, 1997, 23(2), 131–139.
- Jacobs, L. C., and Chase, C. I. *Developing and Using Tests Effectively: A Guide for Faculty*. San Francisco: Jossey-Bass, 1992.
- Jensen, M., Johnson, D. W., and Johnson, R. T. "Impact of Positive Interdependence during Electronic Quizzes on Discourse and Achievement." *Journal of Educational Research*, 2002, 95(3), 161–166.
- LoSchiavo, F. M., and Shatz, M. A. "Students' Reasons for Writing on Multiple Choice Examinations." *Teaching of Psychology*, 2002, 29(2), 138–140.
- Lowman, J. *Mastering the Techniques of Teaching*. (2nd ed.) San Francisco: Jossey-Bass, 1995.
- Martínez, M. E. "Cognition and the Question of Test Item Format." *Educational Psychologist*, 1999, 34(4), 207–218.
- McKeachie, W. J., and Svinicki, M. *McKeachie's Teaching Tips*. (12th ed.) Boston: Houghton Mifflin, 2006.
- Means, B. Prospects for Transforming Schools with Technology-Supported Assessment. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences*. New York: Cambridge University Press, 2006.
- Miller, M. D., Linn, R. I., and Gronlund, N. E. *Measurement and Assessment in Teaching*. (10th ed.) Upper Saddle River, NJ: Prentice Hall, 2008.
- 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.
- Montepare, J. M. "A Self-Correcting Approach to Multiple Choice Tests." *APS Observer*, Oct. 2005, 18(10).
- Morgan, B. M. "Cooperative Learning in Higher Education: Undergraduate Student Reflections on Group Examinations for Group Grades." *College Student Journal*, 2003, 37(1), 40–49.
- Moss, P. A., Girard, B. J., and Haniford, L. C. Validity in Educational Assessment. In J. Green and A. Luke (Eds.), *Review of Research in Education*, Vol. 30. Thousand Oaks, CA: Sage, 2006.
- Muir, S. P., and Tracy, D. M. "Collaborative Essay Testing: Just Try It." *College Teaching*, 1999, 47(1), 33–35.

- Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margollis, A., and Velásquez, R. "Psychological Testing on the Internet." *American Psychologist*, 2004, 59(3), 150–162.
- O'Connor, R. J. "Using Oral Examinations in a Statistics Class." *Teaching Professor*, Jan. 2004, 5.
- Ory, J. C., and Ryan, K. E. *Tips for Improving Testing and Grading*. Newbury Park, CA: Sage, 1993.
- Payne, D. A. *Applied Educational Assessment*. (2nd ed.) Belmont, CA: Wadsworth Thomson, 2003.
- PsychTeacher Listserv: A moderated online community for instructors involved in teaching psychology; see teachpsych.org/news/psychteacher.php
- Race, P., Brown, S., and Smith, B. *500 Tips on Assessment*. (2nd ed.) London: RoutledgeFalmer, 2005.
- Ramsden, P. *Learning to Teach in Higher Education*. (2nd ed.) New York: RoutledgeFalmer, 2003.
- Renner, C. H., and Renner, M. J. How to Create a Good Exam. In B. Perlman, L. I. McCann, and S. H. McFadden (Eds.), *Lessons Learned: Practical Advice for the Teaching of Psychology*. Washington, DC: American Psychological Society, 1999.
- Revere, L., Elden, M., and Bartsch, R. "Designing Group Examinations to Decrease Social Loafing and Increase Learning." *International Journal for the Scholarship of Teaching and Learning*, 2008, 2(1). <http://academics.georgiasouthern.edu/ijstol/v2n1/articles/Revere-Elden-Bartsch/index.htm>
- Scialfa, C., Legare, C., Wenger, L., and Dingley, L. "Difficulty and Discriminability of Introductory Psychology Test Items." *Teaching of Psychology*, 2001, 28(1), 11–15.
- Sechrest, L., Kihlstrom, J. F., and Bootzin, R. How to Develop Multiple-Choice Tests. In B. Perlman, L. I. McCann, and S. H. McFadden (Eds.), *Lessons Learned: Practical Advice for the Teaching of Psychology*. Washington, DC: American Psychological Society, 1999.
- Shermis, M. D., and Daniels, K. Web Applications in Assessment. In T. W. Banta (Ed.), *Building a Scholarship of Assessment*. San Francisco: Jossey-Bass, 2002, pp. 148–166.
- Shindler, J. V. "Greater than the Sum of the Parts?" Examining the Soundness of Collaborative Exams in Teacher Education Courses." *Innovative Higher Education*, 2004, 28(4), 273–283.]
- Slusser, S. R., and Erickson, R. J. "Group Quizzes: An Extension of the Collaborative Learning Process." *Teaching Sociology*, 2006, 34(3), 249–262.
- Theophilides, C., and Koutselini, M. "Study Behavior in the Closed-Book and Open-Book Examination: A Comparative Analysis." *Educational Research and Evaluation*, 2000, 6(4), 379–393.
- Webb, N. M. "Assessing Students in Small Collaborative Groups." *Theory into Practice*, 1997, 36(4), 205–213.
- 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.
- White, B., Ceglie, R., and Puopolo, D. "Note Sheets: A Reliable Predictor of Success?" *Journal of College Science Teaching*, 2001, 31(3), 188–193.
- Yuretich, R. F. "Encouraging Critical Thinking: Measuring Skills in Large Introductory Science Classes." *Journal of College Science Teaching*, 2003, 33(3), 40–45.
- Zimbardo, P. G., Butler, L. D., and Wolfe, V. A. "Cooperative College Examinations: More Gain, Less Pain When Students Share Information and Grades." *Journal of Experimental Education*, 2003, 71(2), 101–125.

Allaying Students' Anxieties about Tests

Anxiety can interfere with students' performance on tests. You can reduce students' anxiety by taking care in how you prepare students for an exam, how you administer and return the test, and how you handle makeup tests. All students, but especially first-year students, can benefit from knowing what they will be asked to do on an exam and under what conditions. Students will also feel more relaxed and less intimidated if you provide reassurance and encouragement. The suggestions that follow are designed to help you prepare your students to do their best on tests.

General Strategies

Give students advice on how to study. Help students develop strategies for organizing and understanding the course material (adapted from Pressley et al., 1997; Roediger and Karpicke, 2006):

- Show students how to identify important ideas in the readings, and let students know the level of detail you expect them to remember.
- Describe the relative importance of the readings compared to the material addressed in class.
- Give pointers on how to take notes and review them.
- Remind students to relate new information to material they already know.
- Provide practice test items or old exams.
- Help students analyze and improve their study habits.

Your student learning center will have additional suggestions. See also Chapter 29, "Helping Students Learn."

Encourage students to study in groups. According to researchers, students who study in groups learn more than students who work alone; see Chapter 21, "Learning in Groups." (Sources: Millis and Cottell, 1998; Slavin, 1991)

Provide accommodations, as needed, for students with disabilities. Students with a documented disability are entitled to receive reasonable accommodations so that they can demonstrate their skills and accomplishments. Common accommodations include extended time to complete a test, oral presentation of a written test through assistive software or a personal assistant, a modified test setting (a quiet room), and assistive technology such as magnifying equipment or a voice-activated computer. See Chapter 6, “Students with Disabilities.” (Source: Cohen and Wollack, 2006)

Seek advice if a student asks for an accommodation because of test anxiety. Some students may exhibit symptoms of severe test anxiety. Several courts have ruled that test anxiety does not constitute a disability, but the best course is to ask staff in your campus’s disabled student services program to determine whether a student’s test anxiety warrants an accommodation. See Chapter 6, “Students with Disabilities.” (Sources: “Test Anxiety May Not Be . . .,” 2004; Zuriff, 1997)

Ask students how you can help them feel less anxious. Most often, students will ask you to describe the test format, to offer a review session, or to provide options for retesting or makeup testing if they perform poorly. (Source: Chapell et al., 2005)

Approaches to Testing

Make the first test relatively easy. Research on motivation indicates that early success in a course increases students’ motivation and confidence. (Sources: Guskey, 1988; Lucas, 1990)

Give periodic tests and quizzes. Students learn more, remember more, and do better on the final exam when they are tested throughout the term. Periodic testing also alleviates the pressure of the Big Exam, allows students to concentrate on one chunk of material at a time, and enables students and instructors to monitor academic progress. One faculty member gives frequent quizzes and a cumulative final exam that repeats items from the quizzes but with the response options scrambled. Instructors who give short quizzes at the start of every class report that students achieve significantly higher grades and retain the material longer than students who are not tested at each class session. However, such frequent testing runs the risk of students focusing more on grades than learning. In addition, time spent on testing is time taken away from other class activities. (Sources: Forsyth, 2003; Kennedy et al., 2002; Landrum, 2007; Leeming, et al., 2002; Myers and Myers, 2007; Padilla-Walker, 2006; Roediger and Karpicke, 2006; Sporer, 2001)

Decide whether to give surprise quizzes. Research shows that surprise quizzes can improve students' performance, although they may unfairly penalize the student who picks the wrong day to come unprepared. Even though students don't like pop quizzes, they recognize that the quizzes help them keep up with their studies. One faculty member gives extra credit for good performance on surprise quizzes (called "extra credit exercises") and ignores poor scores. The offer of extra credit encourages class attendance and preparation. Quizzes also help students gauge their progress and allow them to preview the kinds of questions they will see on midterms and the final exam. Another faculty member announces that there will be weekly quizzes but then flips a coin to see whether students complete the quiz as individuals or as a group and receive a group grade. (Sources: Byrnes and Byrnes, 2007; Graham, 1999; Snooks, 2004; Thorne, 2000)

Avert typical complaints about exams. Take steps to head off the four most common student complaints about classroom tests: the tests are too difficult; they do not match the content or level of the lectures or homework assignments; they are badly written (for example, questions are unclear or cannot be answered in the time available); and the test format came as a surprise. To address these concerns, give students practice exams, develop test items that reflect the course content, tell your students what you expect them to learn or be able to do, and discuss the structure of the test (number of questions, multiple-choice or essay, open or closed book). To gauge the length of the test, take it yourself and estimate that students are likely to need triple the time it took you. (Source: Hativa, 2000)

Preparing Students for an Exam

Give a diagnostic test early in the term. An early diagnostic test alerts students to the skills and knowledge they need to succeed in your class. Some faculty give a knowledge survey in which students rate (on a three-point scale) their confidence in their ability to correctly answer the question. (Sources: Nuhfer and Knipp, 2003; Ochs, 1998)

Post old exams. Reviewing past exams gives students clues about what to study. Students can analyze old exams for format (length of test, number of points for each type of question), types of questions, and level of difficulty.

Give or distribute practice exams. Practice tests help students gauge what is expected of them. But practice tests are most effective when students take

the tests, rather than read them as though they were study guides. (Source: Balch, 1998)

Hand out the final exam on the first day of class. A faculty member in the physical sciences gives students the thirty problems from which the final exam will be drawn. A faculty member in the social sciences hands out fifty essay questions. In both cases, the class discusses the items during the term, and the final exam consists of a subset of the items. Students know that if they can answer the sample items, they will do well on the final exam, and they are relieved of weeks of worrying about what will be on the final.

Schedule extra office hours before a test. Scheduling extra office hours during the week before an exam gives students a chance to ask questions and review difficult points. Some instructors especially encourage study groups to visit during these extra hours.

Schedule review sessions before the final exam. See Chapter 59, “The Last Days of Class,” for advice on how to structure a review session.

Give students a dose of commonsense advice. Remind students to avoid cramming, to get a good night’s sleep, to eat sensibly before the exam, and to arrive early. (Sources: Beilock et al., 2004; Flippo et al., 2000)

Ask students to share their tips. Ask students to offer study strategies and tips, and post these online for your class. In qualitative studies, top-performing students report the following behaviors:

- Regularly attend class.
- Pay attention in class.
- Read strategically (skim; focus on introductions and conclusions), paying special attention to readings mentioned in class.
- Try to spread out studying, as opposed to cramming.
- Vary study activities: rereading texts and notes, rewriting class notes, highlighting and memorizing information, and posing questions about the material while studying.
- Seek help from campus resources.
- Study in a quiet environment with few distractions and interruptions.
- Study with others who are well prepared and can help each other.
- Get enough sleep.

(Sources: Perlman et al., 2007; Van Etten et al., 1997)

Emphasize the importance of students rigorously monitoring their own learning. Metacognitive monitoring—which experts define as the degree to which learners are aware of how well they have or have not acquired skills and knowledge—can be as important as the actual level of skills or knowledge they have attained. People typically have little insight into their limitations and tend to overestimate their expertise and talent. Students in particular have a poorly defined metacognitive sense of how well prepared they are for an exam. See Chapter 29, “Helping Students Learn.” (Sources: Koriatic and Bjork, 2006; Peverly et al., 2003)

Help students distinguish their academic emotions. Research shows that it is okay for students to feel some anxiety before a test. In fact, anxiety is less likely to impact achievement than feelings of hopelessness and boredom. (Source: Pekrun et al., 2002)

Let students know about campus resources. Many counseling centers offer assistance to students who are feeling overwhelmed. Tell your students about the services available on your campus.

Administering Tests

Administer the test yourself. You will want to be present to announce any corrections (of typographical errors, for example) or changes in the exam. Your presence can also motivate and reassure students and signal to them the importance of the test. Arrive early on the day of the test to greet students as they enter the room and answer last-minute questions, and stay late to talk with students. (Sources: Jacobs and Chase, 1992; Lowman, 1995)

Read the instructions aloud at the beginning of class. Even if you write the clearest of instructions, it is helpful to read them aloud to the class and answer students' questions about procedures.

Plan for contingencies and emergencies. Decide how you will respond to questions such as “What if I don't finish?” or “What if I think two answers are correct?” or “What if I need to go to the restroom?”

You should also have a plan in case the exam is disrupted by a fire alarm. Obviously, everyone must leave the building. Some faculty ask students to take their exams with them and sit outside to complete them. Other faculty ask students leave their exams on their desks. Whatever you decide, share the

information on the syllabus and again before the test. Here is what one instructor includes in the syllabus:

When a final exam is interrupted by a fire alarm:

- If the alarm is pulled after the exam has been underway for at least two hours (for a three-hour final), the exam will be deemed complete, and I will adjust the grading scale accordingly.
- If the alarm is pulled after the exam has been underway for less than 15 minutes, the exam will continue after the all-clear signal and our return to the classroom.
- If the alarm is pulled at any other time, I will administer a new exam (time and place to be announced).

When a midterm exam is interrupted by a fire alarm:

- The above procedures will apply except that if the evacuation lasts more than 10–15 minutes, I will reschedule the midterm.
(Source: Listserv at UC Berkeley)

Minimize opportunities for cheating. Actively proctor exams, unless your institution is on the honor system. Be watchful, but don't hover over the class. See Chapter 38, "Promoting Academic Honesty," for advice on ways to reduce cheating during exams.

If there is no clock in the room, keep students apprised of the time. At the start of the exam write on the board the beginning time, the finishing time, and the time remaining. Update the time remaining every twenty minutes or so, and announce the last segment ("You have five minutes left"). Some faculty give prompts during the test: "If you are not yet on question 5, you need to work a little more quickly." Enforce the finishing time; it is unfair to allow some students to continue working when others must leave to go to another class.

Devote part of the session to reviewing the answers with students. One faculty member gives a thirty-minute midterm in a fifty-minute class. Students turn in their answer sheets after thirty minutes, but they keep the question sheet, and the remaining class time is devoted to discussing the correct answers. A variation on this technique is to divide the class into small groups and have them review answers and then reconvene as a class to discuss areas of disagreement or confusion. Another faculty member distributes the answer sheet to the exam as students exit the classroom.

Letting Students Show What They Know

Give students opportunities to explain their responses. Researchers report that giving students space on the test to explain their responses to multiple-choice items helps relieve students' anxiety and reduces posttest complaints. Students were directed to write a short justification for any answer they felt needed more explanation or for questions they perceived to be tricky. The researchers noted that students averaged less than one explanation per test over four tests. The instructors added a point for a "good explanation of a wrong answer" and subtracted a point for "a bad explanation of a right answer." The students' explanations, both good and bad, can also be incorporated into a class discussion on critical thinking. (Sources: Dodd and Leal, 1988; Kee, 1994; Nield and Wintre, 1986; Wallace and Williams, 2003)

Include one or more extra-credit questions. Give students the opportunity to answer additional questions for extra credit at the end of the test.

Give partial credit. For multiple-choice tests, some faculty let students know that they can select more than one answer but will have points deducted for incorrect responses. Some faculty allow students to select only one response but give partial credit for a close but incorrect choice. Another way to give partial credit is to use scratch-off technology as an alternative to optical scanning. With scratch-off forms, the correct answer is indicated by a star underneath a waxy coating. All the other alternatives are blank underneath the coating. For each question, the student picks an answer and scratches off the coating. If the student's selection is correct, a star appears and the student moves to the next question. If the student's selection is wrong, a blank space appears. The student continues scratching alternatives until the star is revealed. Students earn full credit for answering correctly on the first try. If it takes them two or three tries to find the correct answer, they earn progressively less credit. Faculty report that students seem to like scratch-off because it lets them know right away the correct answer to every question, and because it gives them partial credit. The downside is that it may take longer for students to complete the exam. (Sources: Bush, 2001; Denyer and Hancock, 2002; Di Battista et al., 2004)

Let students "buy" information during the exam. Tell students that midway through the exam they can ask you questions at the cost of losing points. For example, asking whether an answer is right or wrong might cost one point; asking for an equation or formula, two points; asking for a diagram setup, four points. A faculty member in mathematics reports that half the class usually buys information to help them "unfreeze" on difficult problems. A chemistry professor gives

every student a “test insurance page” in a lottery scratch-off format; the page contains clues to answers, and each time a student scratches off a clue, points are deducted from the test score. (Sources: Ellis, 1992; Gordon, 1988)

Decide whether to allow “crib sheets.” Some faculty allow students to prepare one five-by-eight index card or one sheet of paper that they can consult during the exam. Crib sheets alleviate pretest anxiety, and they force students to make decisions about which material is most important. The disadvantages are that some students may study less, and others may fail to answer exam questions appropriately because they simply transcribe what is on their crib sheet rather than read the question carefully. Research shows that crib sheets make students feel more confident but do not significantly facilitate learning or enhance exam performance. If you allow crib sheets, researchers recommend that you tell your students to treat crib sheets as a study aid, as a tool for reviewing and organizing the material—not as a substitute for careful preparation. (Sources: Dickson and Miller, 2005; Janick 1990; Vessey and Woodbury, 1992; White et al., 2001)

Ask students to write an unsigned evaluation of the exam. At the next class session, some faculty ask students to give a letter grade to the content, format, and fairness of the test, while others pose the following kinds of questions:

- Identify the questions you didn’t expect to see on the test.
- Were the questions clear enough that, even though you may not have known the answer, you knew what was being asked?
- What questions confused you?

Encourage students to reflect on their test performance. Ask students to think about their preparation, study habits, readiness for the exam, and changes they will make in preparing for the next exam. For example, students can ask themselves, What did I do well on and why? What did I do poorly on and why? What can I do to improve on the next exam? (Sources: Aldrich, 2001; Kher et al., 2002)

Give students a second chance to learn. The following strategies offer students an opportunity to raise their scores on multiple-choice tests:

- After students turn in their in-class, closed-book exam, they receive a second copy to take home and complete as an open-book exam, including conferring with other students. Both exams are scored, and students receive additional points for correcting their incorrect answers.
- Several days after exams are scored and returned, students take a second test that contains equivalent items. Both scores are used, with the lower score weighted at, say, 25 percent and the higher score at 75 percent.

- Students have one week after receiving their scored exam to present a written rebuttal for a specified number of incorrect answers. The rebuttal must include citations from the readings and class notes that support the response they selected. Students earn additional points for a persuasive rebuttal.
- Students have one week after receiving their scored exam to write explanations for why the correct answer to a question is better than the incorrect option they selected. Students earn additional points for thoughtful explanations.

(Sources: Deeter, 2003; Hamilton, 2003; Hare, 1997; Kottke, 2001; Montepare, 2005)

Returning Examinations

Explain that testing and assessment are part of learning. Help students view assessment as part of the learning process, not as a reward or punishment: give timely feedback, offer suggestions for improvement, and maintain a respectful tone. (Source: Wiggins, 1998)

Return tests promptly. Most students are anxious to know how they have done, and a quick turnaround also encourages relearning or corrective learning. Most experts recommend that tests be returned within five days. Laws governing the privacy and confidentiality of student records forbid the posting of grades by name, initials, or student identification numbers, although instructors may post grades by using a unique code known only by the instructor and the student. Confidentiality and concerns about security also dictate that exams not be left in a pile in the department office or circulated in a stack in class for students to help themselves. Instead divide the tests into batches and have your graduate student instructor (GSI) or reader help distribute exams in different corners of the room by calling out names. Or prepare folders or manila envelopes with students' names and place these around the room for pick up. (Source: Lowman, 1995)

Use class time to discuss the overall results. Explain how the tests were graded and the criteria you used. Indicate how the class performed as a whole, or show the distribution of scores. Note which items were missed by many students, and correct widespread misunderstandings. For essay tests, describe the elements of a good answer and mention the most common problems. Some faculty read or distribute unsigned excerpts from outstanding essays. For multiple-choice tests, some faculty have students discuss the results in small groups. Having students review exams in groups often takes less time than an instructor's own review, and students may enjoy it more. One faculty member gives students their grades prior

to going over the exam in class so that they know their overall score but not which specific items they missed. During class time, students discuss why they selected the option they did, and then the instructor lets the class know the correct answer. (Sources: McKeachie and Svinicki, 2006; Lucas, 2002; Wininger, 2005)

Give students advice on how to go over the exam on their own. Suggest they look at missed questions to determine which came from the readings and which from class notes. (Source: Weimer, 2002).

Schedule extra office hours after returning a test. Be prepared for students who are unhappy about their grade:

- Ask students who want to discuss their grade to wait twenty-four hours before coming to see you. This gives them a chance to reread the exam, cool down, and prepare specific questions.
- Let students know that if they request a review of their test, the review might result in a higher or a lower grade.
- Ask students to come with specific questions (not “Why is my grade so low?”). Some faculty request that students prepare a brief paragraph or complete a short form, which might be titled “Request to Review a Test Item,” expressing their complaint and justifying the correctness of their answer with specific reference to readings or class notes.
- Listen carefully to the student’s complaint. Do not interrupt to rebut each point.
- Try to shift the focus of the discussion from grades to problem solving. Ask, “What can we do to help you do better next time?” Help the student think less about blaming you or the test and more about how to learn and study more effectively.
- Don’t change a grade out of sympathy or compassion. Make a change only if you made a clerical error or mistakenly evaluated a response.
- Set a one-week time limit for students to ask for a review of exam grades.

(Sources: Jacobs and Chase, 1992; Kher et al., 2002; Lucas, 2002; McKeachie and Svinicki, 2006)

Arranging Makeup Tests

Decide whether to give makeup tests. The majority of students requesting a makeup exam have reasonable needs, such as illness, family emergency, religious observance, or unavoidable travel for graduate school interviews or athletic

competition. Encourage students to notify you right away if they will miss the exam. Options for makeup exams include giving the same test at a mutually agreed-on time or creating a new test in the same format.

Double the weight of tests students have taken. Makeup tests can be problematic: a new test might not be comparable to the original test, but using the same test may advantage students who talked to others. Scheduling makeup tests can also be difficult. Some faculty give two midterms and double the weight of one when a student misses the other. (Source: McKeachie and Svinicki, 2006)

Decide whether to let students drop their lowest test score. Other faculty avoid makeup tests by giving multiple exams during the term and letting students drop the lowest score. Some research suggests that students may not study as diligently when they know they can drop a test score. If you do decide to do this, keep in mind that the lowest score may not be the best one for students to drop unless all tests are worth the same number of points. Kane and Kane (2006) give the following example: Leslie scores 80 out of 100 on midterm 1, 20 out of 100 on midterm 2, and 1 out of 20 on midterm 3. If midterm 3 is dropped, her mean score on midterms 1 and 2 would be $(80 + 20) / (100 + 100) = 50$. On the other hand, if midterm 2 is dropped instead, she would receive a mean score of $(80 + 1) / (100 + 20) = 67.5$. Kane and Kane have developed an efficient algorithm to maximize the resulting average grade. If you do allow students to drop their lowest test score, be sure to give a cumulative final exam so that students study all the material in the course. (Sources: Forsyth, 2003; Kane and Kane, 2006; McKeachie and Svinicki, 2006)

Give students options on the number of tests they take. Some faculty offer students three options: (1) four multiple-choice tests, (2) four multiple-choice tests and a final, or (3) three multiple-choice tests and a final. In options 1 and 3, each test is worth 25 percent of the course grade; in option 2, each test is worth 20 percent. Students who miss one of the multiple-choice tests must elect option 3, and students who miss two tests are handled on a case-by-case basis. (Source: Buchanan and Rogers, 1990)

Give an additional exam for the entire class at the end of the semester. The grade on this extra test can replace a missed exam or replace a low grade for a student who had an off day. (Source: Shea, 1995)

Hand out essay questions in advance. If you distribute a list of essay questions from which the midterm questions will be taken, you will not have to write a makeup test. (Source: Lewis, 1990)

Give a shorter final exam and use the last hour for makeup tests. By administering makeup tests during the time block reserved for the final exam, you can avoid the complexities of special scheduling.

Give an oral exam as a substitute. In small classes and advanced courses, you can offer an oral exam. Oral exams typically cover less material but in more depth than written exams. You could also include a customized follow-up writing assignment to probe more deeply into a relevant portion of the material. (Source: Listserv at the University of California, Berkeley)

Assign a paper for a missed exam. Grading a paper may take less time than designing and scoring a new exam.

References

- Aldrich, H. E. "How to Hand Exams Back to Your Class." *College Teaching*, 2001, 49(3), 82.
- Balch, W. R. "Practice versus Review Exams and Final Exam Performance." *Teaching of Psychology*, 1998, 25(3), 181–185.
- Beilock, S. L., Kulp, C. A., Holt, L. E., and Carr, T. H. "More on the Fragility of Performance: Choking under Pressure in Mathematical Problem Solving." *Journal of Experimental Psychology: General*, 2004, 133(4), 584–600.
- Buchanan, R. W., and Rogers, M. "Innovative Assessment in Large Classes." *College Teaching*, 1990, 38(2), 69–73.
- Bush, M. "A Multiple Choice Test That Rewards Partial Knowledge." *Journal of Further and Higher Education*, 2001, 25(2), 157–163.
- Byrnes, M., and Byrnes, J. F. "Quizzes Are the Right Answer." *Teaching Professor*, March 2007.
- Chapell, M. S., Blanding, Z. B., Silverstein, M. E., Takahashi, M., Newman, B., Gubi, A., and McCann, N. "Test Anxiety and Academic Performance in Undergraduate and Graduate Students." *Journal of Educational Psychology*, 2005, 97(2), 268–274.
- Cohen, A. S., and Wollack, J. A. Test Administration, Security, Scoring and Reporting. In R. L. Brennan (Ed.), *Educational Measurement*. (4th ed.) Westport, CT: American Council on Education/Praeger, 2006.
- Deeter, L. "Incorporating Student Centered Learning Techniques into an Introductory Plant Identification Course." *North American Colleges and Teachers of Agriculture (NACTA) Journal*, June 2003, 47–52.
- Denyer, G., and Hancock, D. "Graded Multiple Choice Questions: Rewarding Understanding and Preventing Plagiarism." *Journal of Chemical Education*, 2002, 79(8), 961–964.
- Di Battista, D., Mitterer, J. O., Gosse, L. "Acceptance by Undergraduates of the Immediate Feedback Assessment Technique for Multiple-Choice Testing." *Teaching in Higher Education*, 2004, 9(1), 17–28.
- Dickson, K. L., and Miller, M. D. "Authorized Crib Cards Do Not Improve Exam Performance." *Teaching of Psychology*, 2005, 32(4), 230–233.
- Dodd, D. K., and Leal, L. "Answer Justification: Removing the 'Trick' from Multi-Choice Questions." *Teaching of Psychology*, 1988, 15(1) 37–38.

- Ellis, A. "Scratching for Grades." *National Teaching and Learning Forum*, 1992, 1(5), 4–5.
- Flippo, R. F., Becker, M. J., and Wark, D. M. Preparing for and Taking Tests. In R. F. Flippo and D. C. Caverly (Eds.), *Handbook of College Reading and Study Strategy Research*. Mahwah, NJ: Erlbaum, 2000.
- Forsyth, D. R. *The Professor's Guide to Teaching: Psychological Principles and Practices*. Washington, DC: American Psychological Association, 2003.
- Gordon, L. "Cost-Benefit Testing." *Academic Leader*, 1988, 4(4), 1–2.
- Graham, R. B. "Unannounced Quizzes Raise Test Scores Selectively for Mid-Range Students." *Teaching of Psychology*, 1999, 26(4), 271–273.
- Guskey, T. R. *Improving Student Learning in College Classrooms*. Springfield, IL: Thomas, 1988.
- Hamilton, T. M. "Everyone Deserves a Second Chance: Using the Day after the Exam as a Learning Opportunity." *College Teaching*, 2003, 51(1), 21.
- Hare, A. C. "Active Learning and Assessment in Mathematics." *College Teaching*, 1997, 45(2), 76–77.
- Hativa, N. *Teaching for Effective Learning in Higher Education*. Norwell, MA: Kluwer Academic Publishers, 2000.
- Jacobs, L. C., and Chase, C. I. *Developing and Using Tests Effectively: A Guide for Faculty*. San Francisco: Jossey-Bass, 1992.
- Janick, J. "Crib Sheets." *Teaching Professor*, 1990, 4(6), 2.
- Kane, D. M., and Kane, J. M. "Dropping Lowest Grades." *Mathematics Magazine*, June 2006, 79(3), 181–189.
- Kee, C. "Multiple-Choice Questions: A New Twist for an Old Standard." *Teaching Professor*, 1994, 8(6), 6.
- Kennedy, E. J., Lawton, L., and Plumlee, E. L. "Blissful Ignorance: The Problem of Unrecognized Incompetence and Academic Performance." *Journal of Marketing Education*, 2002, 24(3), 243–252.
- Kher, N., Juneau, G., and Molstad, S. "Test Feedback Class Sessions: Creating a Positive Learning Experience." *College Teaching*, 2002, 50(4), 148–150.
- Koriat, A., and Bjork, R. A. "Illusions of Competence during Study Can Be Remedied by Manipulations That Enhance Learners' Sensitivity to Retrieval Conditions at Test." *Memory and Cognition*, 2006, 34(5), 959–972.
- Kottke, J. L. "Students' Reactions to Written Test Item Rebuttals." *Journal of Instructional Psychology*, 2001, 28(4), 256–261.
- Landrum, R. E. "Introductory Psychology Student Performance: Weekly Quizzes Followed by a Cumulative Final Exam." *Teaching of Psychology*, 2007, 34(3), 177–180.
- Leeming, F. C. "The Exam-a-Day Procedure Improves Performance in Psychology Classes." *Teaching of Psychology*, 2002, 29(3), 210–212.
- Lewis, K. G. *Taming the Pedagogical Monster*. (3rd ed.) Austin: Center for Teaching Effectiveness, University of Texas, 1990.
- Lowman, J. *Mastering the Techniques of Teaching*. (2nd ed.) San Francisco: Jossey-Bass, 1995.
- Lucas, A. F. "Using Psychological Models to Understand Student Motivation." In M. D. Svinicki (Ed.), *The Changing Face of College Teaching*. New Directions for Teaching and Learning, no. 42. San Francisco: Jossey-Bass, 1990.
- Lucas, S. G. Returning Graded Assignments Is Part of the Learning Experience. In B. Perlman, L. I. McCann, and S. H. McFadden (Eds.), *Lessons Learned: Practical Advice for the Teaching of Psychology*. Vol. 2. Washington, DC: American Psychological Society, 2002.
- McKeachie, W. J., and Svinicki, M. *McKeachie's Teaching Tips*. (12th ed.) Boston: Houghton Mifflin, 2006.

- Millis, B., and Cottell, P. G. *Cooperative Learning for Higher Education Faculty*. American Council on Education and Oryx Press, 1998.
- Montepare, J. M. "A Self-Correcting Approach to Multiple Choice Tests." *APS Observer*, October 2005, 18(10).
- Myers, C. B., and Myers, S. M. "Assessing Assessments: The Effects of Two Exam Formats on Course Achievement and Evaluation." *Innovative Higher Education*, 2007, 31(4), 227–236.
- Nield, A. F., and Wintre, M. "Multiple Choice Questions with an Option to Comment: Students' Attitude and Use." *Teaching of Psychology*, 1986, 13(4), 196–199.
- Nuhfer, E., and Knipp, D. The Knowledge Survey: A Tool for All Reasons. In C. M. Wehlburg and S. Chadwick-Blossey (Eds.), *To Improve the Academy*. Vol. 21. San Francisco: Jossey-Bass, 2003.
- Ochs, R. S. "The First-Day Quiz as a Teaching Technique." *Journal of Chemical Education*, 1998, 75(4), 401–403.
- Padilla-Walker, L. M. "The Impact of Daily Extra Credit Quizzes on Exam Performance." *Teaching of Psychology*, 2006, 33(4), 236–239.
- Pekrun, R., Goetz, T., Titz, W., and Perry, R. P. "Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research." *Educational Psychologist*, 2002, 37(2), 91–105.
- Pearlman, B., McCann, L. I., and Prust, A. "Students' Grades and Ratings of Perceived Effectiveness of Behaviors Influencing Academic Performance." *Teaching of Psychology*, 2007, 34(4), 236–240.
- Peeverly, S. T., Brobst, K. E., Graham, M., and Shaw, R. "College Adults Are Not Good at Self-Regulation: A Study on the Relationship of Self-Regulation, Note Taking and Test Taking." *Journal of Educational Psychology*, 2003, 95(2), 335–346.
- Pressley, M., Yokoi, L., van Meter, P., van Etten, S., and Freebern, G. "Some of the Reasons Why Preparing for Exams Is So Hard: What Can Be Done to Make It Easier?" *Educational Psychology Review*, 1997, 9(1), 1–38.
- Roediger, H. L., and Karpicke, J. D. "Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention." *Psychological Science*, 2006, 17(3), 249–255.
- Shea, M. A. *Compendium of Good Ideas on Teaching and Learning*. Boulder: Faculty Teaching Excellence Program, University of Colorado, 1995.
- Slavin, R. E. "Synthesis of Research on Cooperative Learning." *Educational Leadership*, 1991, 48(5), 71–82.
- Snooks, M. K. "Using Practice Tests on a Regular Basis to Improve Student Learning." *New Directions for Teaching and Learning*, no. 100. San Francisco: Jossey-Bass, 2004, 109–113.
- Sporer, R. "The Quick Fix: The No-Fault Quiz." *College Teaching*, 2001, 49(2), 61.
- "Test Anxiety May Not Be a Disability But It Can Impair Functioning." *Disability Compliance for Higher Education*, Aug. 2004, 10(1), 3.
- Thorne, B. M. "Extra Credit Exercise: A Painless Pop Quiz." *Teaching of Psychology*, 2000, 27(3), 204–205.
- Van Etten, S., Freebern, G., and Pressley, M. "College Students' Beliefs about Exam Preparation." *Contemporary Educational Psychology*, 1997, 22(2), 192–212.
- Vessey, J. K., and Woodbury, W. "Crib Sheets: Use with Caution." *Teaching Professor*, 1992, 6(7), 6–7.
- Wallace, M. A., and Williams, R. L. "Multiple-Choice Exams: Explanations for Student Choices." *Teaching of Psychology*, 2003, 30(2), 136–138.

- Weimer, M. *Learner-Centered Teaching: Five Key Changes to Practice*. San Francisco: Jossey-Bass, 2002.
- White, B., Ceglie, R., and Puopolo, D. "Note Sheets: A Reliable Predictor of Success?" *Journal of College Science Teaching*, 2001, 31(3), 188–193.
- Wiggins, G. *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco: Jossey-Bass, 1998.
- Wininger, S. R. "Using Your Tests to Teach: Formative Summative Assessment." *Teaching of Psychology*, 2005, 32(3), 164–166.
- Zuriff, G. E. "Accommodations for Test Anxiety under ADA?" *Journal of the American Academy of Psychiatry and the Law*, 1997, 25(2), 197–206.

Multiple-Choice and Matching Tests

A multiple-choice item presents a question or incomplete statement (a *stem*) and three to five suggested answers or completions, one of which is best (the *key*); the incorrect choices are called *distractors*. A matching-test item presents two columns of information (such as terms, names, or dates) and asks students to choose from the second column a response that fits the stimulus in the first column. Multiple-choice and matching tests are the two best-known forms of selected-response tests (also called *forced choice tests*, *constructed response tests*, or *choice-type tests*).

Selected-response items are useful for testing the breadth of students' learning (a selected-response test can cover more topics than an essay test) and for testing different levels of learning (from recall of factual information to problem solving). Compared to short-answer and essay tests (called *supply-type tests*), selected-response tests require more time to prepare but less time to score. Although selected-response tests are often criticized as measuring only rote memorization, selected-response items can also be used to assess students' mastery of complex concepts and ideas.

General Strategies

Write test items throughout the term. Good test items are difficult to create, and it is impossible to compose an effective selected-response test at the last minute. Spread out the work by writing three to five items every week. (Source: Jacobs and Chase, 1992)

Give students advice on how to take a selected-response test. The following tips will help your students do their best (adapted from McKeachie and Svinicki, 2006; Sechrest et al., 1999):

- Go through the test once and answer all the questions you can.
- When you go through the test again, spend a reasonable amount of time on each question but move on if you get stuck.

- If you are stuck, try to reason the correct answer from some general concept or principle.
- Change your answer if you wish; research shows that most students gain more than they lose on changed answers (Kruger et al., 2005).
- Save some time at the end to double-check your answers and make sure you haven't made any clerical errors.

Types of Multiple-Choice Questions

Include some items that require complex thinking. For example, write a question that requires students to predict the outcome of a situation, or ask students to select examples that illustrate an abstraction or a principle. Or give examples and ask students to select the principle or theory that the examples illustrate.

The following multiple-choice items probe a range of learning outcomes (adapted from Welsh, 1978).

Question 1 requires students to make comparative generalizations.

1. Which of the following has contributed most to long-term economic growth in the United States?
 - A. Increasing personal income tax rates.
 - B. Reducing hours worked per week to spread employment among more people.
 - C. Increasing tariffs on imported goods that compete with domestically produced goods.
 - D. Increasing levels of education and technological improvement.

Question 2 asks students to apply supply-and-demand principles to a specific situation.

2. A large city is investigating the elimination of rent controls on housing at a time when the vacancy rate is extremely low—only 1 percent. Which of the following is most likely to occur if rent controls are eliminated?
 - A. An increase in the demand for housing, followed by a decrease in the supply of housing.
 - B. An increase in rents, followed by an increase in the supply of housing.
 - C. A decrease in rents and a decrease in the supply of housing.
 - D. No change in rents because price controls are usually set where supply and demand intersect.

For question 3, students must analyze the situation, select the most appropriate policy, and predict the expected effects of the policy.

3. Because of rapidly rising national defense expenditures, the country of Parador will experience price inflation unless measures are taken to restrict the growth of aggregate private demand. If Parador wishes to minimize the adverse effects of anti-inflationary policies on economic growth, it should implement
- A. A tight monetary policy because that would restrict consumption expenditures more than investment.
 - B. A tight monetary policy because that would restrict consumption expenditures.
 - C. An increase in personal income taxes because that would restrict consumption expenditures more than investment.
 - D. Either a tight monetary policy or an increase in personal income taxes because both depress investment equally.

Create “You are the teacher” questions. Some multiple-choice items require students to evaluate the response to a short-answer question (from Jensen et al., 2006, p. 69):

Pretend you are a science teacher who is correcting the following answer on a quiz. How many scientific errors does the answer contain? *Note:* There is a maximum of one error per sentence.

During the depolarization phase of an action potential, sodium gates are open and sodium diffuses from the extra-cellular fluid to the intra-cellular fluid. At the end of the depolarization phase, sodium gates close and potassium gates open. Repolarization begins when potassium moves by active transport from the intra-cellular fluid to the extra-cellular fluid of the cell. After the action potential passes, ion gradients are maintained by the sodium/potassium pump.

- A. 0 errors
- B. 1 error
- C. 2 errors
- D. 3 errors
- E. 4 errors

Jensen et al. (2006) recommend that this type of question be used only after students have done practice sets before the exam.

Experiment with assertion-reason questions (ARQs). An ARQ consists of two statements—an assertion and a reason—linked by *because*. The student selects from five response options that indicate the correctness of each statement and the validity of the reasoning. Here is an example (adapted from Williams, 2006, p. 292):

(Assertion) In a small open economy, if the prevailing world price of a good is lower than the domestic price, the quantity supplied by the domestic producer will be greater than the domestic quantity demanded, increasing domestic producer surplus.

Because

(Reason) In a small open economy any surplus in the domestic market will be absorbed by the rest of the world. This increases domestic consumer surplus.

- A. The assertion and reason are both correct, and the reason is valid.
- B. The assertion and reason are both correct, but the reason is invalid.
- C. The assertion is correct, but the reason is incorrect.
- D. The assertion is incorrect, but the reason is correct.
- E. Both the assertion and the reason are incorrect.

If you want to use ARQs, give your students time to become familiar with the format and offer examples and practice quizzes. ARQs pose special problems for students whose reading comprehension skills are below average. (Source: Williams, 2006)

Constructing Multiple-Choice Test Items

Instruct students to select the “best answer” rather than the “correct answer.” Asking for the correct answer may invite arguments from contentious students that their selections are correct as well. If you ask for the best answer, you can acknowledge that other responses have some element of truth or accuracy but that the keyed response is the best. (Source: Jacobs and Chase, 1992)

In the instructions, state the rewards or penalties for guessing. Some instructors encourage students to make their best guess, even when they are unsure about the correct answer. Other instructors penalize students for guessing. Some faculty award partial credit for reasoned but flawed answers. (Source: Baranchik and Cherkas, 2000)

Express the full problem in the stem, typically as a question. Make sure that students can understand the problem before reading the choices. Direct questions are usually clearer than sentence completions, although incomplete statements may avoid cumbersome phrasing. The stem may also include a map, diagram, picture, or graph.

In the following pair, the stem of the poorly constructed item is so brief that it fails to state the problem.

Poor: Grading is

- A. Most often used to distinguish between students.
- B. A way of reporting students' progress.
- C. The only reason students study.
- D. Something teachers put off if they can.

Better: What is the main reason most universities use a letter-grading system?

- A. Convenience in reporting students' progress.
- B. Utility in keeping permanent records.
- C. Ease in distinguishing among students.
- D. Usefulness in motivating students to learn.

When the item concerns the definition of a term, the preferred format is to present the definition in the stem and list several terms as options, rather than placing the term in the stem and listing several definitions as options. (Source: Kehoe, 1995)

Put all relevant material in the stem. Do not repeat a phrase in the options if the phrase can be stated in the stem. In the following example, students must waste time reading the repetitions of *chosen by* (adapted from Ory and Ryan, 1993):

Poor: In national elections in the United States the president is officially

- A. chosen by the people.
- B. chosen by members of Congress.
- C. chosen by the House of Representatives.
- D. chosen by the Electoral College.

Better: In national elections in the United States, who officially chooses the president?

- A. the people.
- B. members of Congress.

- C. the House of Representatives.
- D. the Electoral College.

Keep the stem short. In the following example, the unnecessary information in the stem is likely to confuse students and waste their time (adapted from Frary, 1995):

Poor: The presence and association of the male seems to have profound effects on female physiology in domestic animals. Research shows that in cattle the presence of a bull has the following effects:

Better: Research shows that the presence of a bull has which of the following effects on cows?

Write the correct response (key) first, and then create appealing distractors. Clearly formulate the best or correct answer, and then draft a set of plausible distractors. If the distractors are farfetched, students will too easily guess the correct answer. Effective distractors represent errors commonly made by students: statements that include errors in logic or interpretation, statements that are too general or too specific for the requirements of the problem, statements that are accurate but do not fully meet the requirements of the problem, and incorrect statements that will seem correct to the poorly prepared student. (Sources: Clegg and Cashin, 1986; Forsyth, 2003; Sechrest et al., 1999)

Limit the number of response alternatives. Studies show that three-choice items are more effective or about as effective as four-choice items, and yet the four-choice format is the most popular. Never give students more than five choices. You need not give the same number of choices for each test item. However, if students use optical scanning sheets or need to mark their answers on a scoring sheet, use the same number of distractors for each question to minimize the chances of transposition errors. (Sources: Green, 1997; Haladyna, 2004; Landrum et al., 1993; McKeachie and Svinicki, 2006; Sechrest et al., 1999)

Make all choices roughly equal in length. Do not signal the best choice by making it longer, more detailed, or more nuanced than the alternatives. (Sources: Green, 1997; Kehoe, 1995; Sechrest et al., 1999)

Avoid negative wording. Negative wording often confuses students, for example:

Poor: Which of the following is not a characteristic of Brutalism?

Better: Which of the following best distinguishes Brutalism from other architectural movements?

If you must use negatives, emphasize them with underlining, capital letters, or bold type. (Sources: Clegg and Cashin, 1986; Kehoe, 1995)

Refrain from using words such as *always, never, all, or none*. Savvy students know that few statements are absolute or universally true. (Source: Clegg and Cashin, 1986)

Avoid giving “all of the above” or “none of the above” as choices. These items do not discriminate well among students with differing knowledge. Students need only compare two choices: if both are acceptable, then “all of the above” is the logical answer, even if the student is unsure about a third choice. (Sources: Haladyna, 2004; Jacobs and Chase, 1992)

Begin with a few easy items. Some research shows that students do better when the sequence of the items reflects the order in which the material was presented in the readings and in class, but placing a few easy questions first will calm nervous students. (Source: Sechrest et al., 1999)

Make the choices grammatically consistent with the stem. Read the stem and each of the choices aloud to be sure that each is correct in the use of *a* or *an*, singular and plural, and subject-verb agreement. In the following example, choices B and C should be reworded (adapted from Welsh, 1978):

Poor: The functions of the Federal Reserve are to provide the nation with an elastic money supply and to

- A. help stabilize the economy.
- B. correction of national income statistics.
- C. correction of tax laws.
- D. help levy property taxes.

Vary the position of the best answer. Research shows that faculty tend to locate the best answer in the B or C position. If appropriate, list options alphabetically or in some meaningful order (for example, numerical, chronological, or conceptual). Or use a deck of cards to locate the correct responses randomly. (If the card you turn over is a heart, the correct answer goes in slot A; if it is a spade, in slot B; and so on.) For ease of reading, mark each option with a capital letter and list the options vertically. (Source: Haladyna, 2004)

Guard against overlapping items. Make test questions independent of one another so that the stem or alternative in one question doesn't give students a clue to the correct answer to another question. (Source: Haladyna, 2004)

Keep the test length manageable. Students can complete between one and two multiple-choice items per minute. (Source: Lowman, 1995)

Print the test on white paper. Though some observers recommend printing alternate forms of multiple choice exams on paper of different colors to prevent cheating, researchers report that white paper leads to better test performance. If you do scramble questions, add a cover sheet that is a different color (with just the name of the class and the date), and keep the questions themselves on white paper. (Source: Skinner, 2004)

Matching-Test Items

Write clear instructions. Let students know the basis on which items are to be paired, where to write answers, and whether a response may be used more than once. For example: “Next to each literary movement in column 1, write the letter of the work in column 2 that best exemplifies that movement. You may use each work in column 2 more than once or not at all.”

Do not mix different classes of items in a column. For example, column 1 might list events and column 2 might list dates, but do not combine events, dates, and names in a single column. (Source: Fuhrmann and Grasha, 1983)

Place the responses in order. When the items in column 2 are in order (alphabetical, chronological, or conceptual), students will be able to read the series quickly and locate answers rapidly. Limit column 2 to five to ten items; longer lists require students to spend too much time searching for responses. (Source: Ory and Ryan, 1993)

Be conscious of layout and format. Do not allow the columns to break across a page; students should not have to flip back and forth. Place answer blanks to the left of each entry in column 1. Use numbers for the items in column 1, and use capital letters for the responses in column 2. Place a heading at the top of each column. (Source: Fuhrmann and Grasha, 1983)

Scoring the Test

Use machine scoring systems. Optical scanning equipment can quickly score selected-response exams. Your learning management system may also have features for online testing and scoring. Scratch-off technology can also save time.

Have students score their own exams. Self-scoring provides students with immediate feedback on their performance. To assure that the scoring is accurate and fair, give students a separate answer sheet (“declaration sheet”) along with the test. While taking the test, students mark their answers on the exam itself and they also mark (“declare”) their answers on their signed declaration sheet. When all students have completed the test, the declaration sheets are collected, the answers are revealed, and students score their own tests and turn them in. Studies show that students are accurate in scoring and reporting their scores so faculty need not spend time marking each student’s exam. (Source: Carkenord and Laws, 2005)

Item Analysis

Perform an item analysis to evaluate the test. In classes large enough to provide statistically useful results (say, fifty students or more), you can perform an item analysis to determine which items are too easy or too hard and how well items distinguish between students at the top and bottom. Most test scoring software and learning management systems include options for these calculations; your campus testing office may also have analytic tools. The results will help you in improving future versions of the test items. (Source: Ory and Ryan, 1993)

Look at the difficulty of each item. Calculate the percentage of students who answered each item correctly. If you are administering the test to identify the best performers (norm-referenced testing), the ideal test will contain only a few items that are very difficult (answered correctly by less than 30 percent of students) or very easy (answered correctly by more than 90 percent). For norm-referenced testing, experts recommend the following difficulty levels: 5 percent of the items are answered correctly by 90 percent of the students (to boost confidence), 5 percent of the items are answered correctly by 10 percent of the students, and the remainder of the items are answered correctly by an average of 50 percent of the students.

In contrast, if you are administering the test to measure information, skills, and competencies that all students need to have acquired (standards-referenced testing), then the point of your item analysis is not to check the distribution of scores but to ascertain how well the test items represented the targeted competencies. If an item proved too difficult for many of your students, consider three possibilities: Was the item poorly written or unclear? Was the content too challenging? Were too many students insufficiently prepared? (Sources: Jacobs and Chase, 1992; Lowman, 1995; Scialfa et al., 2001; Sechrest et al., 1999; Wergin, 1988)

Look at how well each item discriminates between high and low scorers. *Item discrimination* is a statistical technique for calculating how well an individual test item differentiates between the top scorers and the bottom scorers on a test. Test-scoring software can generate a discrimination ratio—a number between -1.0 and $+1.0$ —for each item. The closer the ratio is to $+1.0$, the more effectively the item distinguishes students who know the material from those who don't. Ideally, each item will have a ratio of at least $.30$, but some items that have lower discrimination ratios are useful to ensure that a test contains a few items that everyone can answer correctly. If you have an item that few of the top-scoring students answered correctly, look for defects in the structure of the item and the distractors. (Sources: Lowman, 1995; Miller et al., 2008; Schmeiser and Welch, 2006; Sechrest et al., 1999)

Examine the reliability coefficient. If your test-scoring software calculates a reliability estimate, a coefficient of $.65$ or higher is desirable. Interpret a coefficient of less than $.60$ to mean that the test is unlikely to produce similar results on second administration. (Source: Ory and Ryan, 1993)

Use these analyses to improve your tests. Let the results of the statistical analyses suggest which items you might drop or revise for future tests. Aim for a test on which the mean percentage of correct answers is about 70 percent, with items whose difficulty level is between 30 percent and 70 percent (items within this range can be expected to have an acceptable discrimination ratio, $+.30$ or higher). When an item has a high difficulty level and a low discrimination ratio (below $+.30$), the item should be revised. Items that fall on the borderline (a discrimination ratio just under $+.30$ and a difficulty level between 30 percent and 70 percent) do not necessarily need revision. (Sources: Schmeiser and Welch, 2006; Sechrest et al., 1999)

Examine the distractors. Look at the percentage of students who selected a particular distractor, and consider replacing a distractor that few students chose. You may also want to double-check items missed by a majority of students, especially those items for which any one distractor was selected more often than the correct answer. (Source: Sechrest et al., 1999)

References

- Baranchik, A., and Cherkas, B. "Correcting Grade Deflation Caused by Multiple-Choice Scoring." *International Journal of Mathematical Education in Science and Technology*, 2000, 31(3), 371–380.
- Carkenord, D. M., and Laws, E. L. "Assuring Accuracy of Student Self-Scored Quizzes." *Teaching of Psychology*, 2005, 32(3), 175–177.

- Clegg, V. L., and Cashin, W. E. "Improving Multiple-Choice Tests." *Idea Paper*, no. 16. Manhattan: Center for Faculty Evaluation and Development in Higher Education, Kansas State University, 1986.
- Forsyth, D. R. *The Professor's Guide to Teaching: Psychological Principles and Practices*. Washington, DC: American Psychological Association, 2003.
- Frary, R. B. "More Multiple-Choice Item Writing Do's and Don'ts." *Practical Assessment, Research and Evaluation*, 1995, 4(11). <http://pareonline.net/getvn.asp?v=4&n=11>
- Fuhrmann, B. S., and Grasha, A. F. *A Practical Handbook for College Teachers*. Boston: Little, Brown, 1983.
- Green, D. H. "Student-Generated Exams: Testing and Learning." *Journal of Marketing Education*, 1997, 19(2), 43–53.
- Haladyna, T. M. *Developing and Validating Multiple-Choice Test Items*. (3rd ed.) Mahwah, NJ: Erlbaum, 2004.
- Jacobs, L. C., and Chase, C. I. *Developing and Using Tests Effectively: A Guide for Faculty*. San Francisco: Jossey-Bass, 1992.
- Jensen, M., Duranczyk, I., Staats, S., Moore, R., Hatch, J., and Somdahl, C. "Using a Reciprocal Teaching Strategy to Create Multiple Choice Exam Questions." *American Biology Teacher*, 2006, 68(6), 67–71.
- Kehoe, J. "Writing Multiple-Choice Test Items." *Practical Assessment, Research and Evaluation*, 1995, 4(9).
- Kruger, J., Wirtz, D., and Miller, D. T. "Counterfactual Thinking and First Instinct Fallacy." *Journal of Personality and Social Psychology*, 2005, 88(5), 725–735.
- Landrum, R. E., Cashin, J. R., and Theis, K. S. "More Evidence in Favor of Three-Option Multiple Choice Tests." *Educational and Psychological Measurement*, 1993, 53, 771–778.
- Lowman, J. *Mastering the Techniques of Teaching*. San Francisco: Jossey-Bass, 1995.
- McKeachie, W. J., and Svinicki, M. *McKeachie's Teaching Tips*. (12th ed.) Boston: Houghton Mifflin, 2006.
- Miller, M. D., Linn, R. I., and Gronlund, N. E. *Measurement and Assessment in Teaching*. (10th ed.) Upper Saddle River, NJ: Prentice Hall, 2008.
- Ory, J. C., and Ryan, K. E. *Tips for Improving Testing and Grading*. Newbury Park, CA: Sage, 1993.
- Schmeiser, C. B., and Welch, C. J. Test Development. In R. L. Brennan (Ed.), *Educational Measurement*. (4th ed.) Westport, CT: American Council on Education/Praeger, 2006.
- Scialfa, C., Legare, C., Wenger, L., and Dingley, L. "Difficulty and Discriminability of Introductory Psychology Test Items." *Teaching of Psychology*, 2001, 28(1), 11–15.
- Sechrest, L., Kihlstrom, J. F., and Bootzin, R. How to Develop Multiple-Choice Tests. In B. Perlman, L. I. McCann, S. H. and McFadden (Eds.), *Lessons Learned: Practical Advice for the Teaching of Psychology*. Washington, DC: American Psychological Society, 1999.
- Skinner, N. F. "Differential Test Performance from Differently Colored Paper: White Paper Works Best." *Teaching of Psychology*, 2004, 31(2), 111–113.
- Welsh, A. L. Multiple Choice Objective Tests. In P. Saunders, A. L. Welsh, and W. L. Hansen (Eds.), *Resource Manual for Teaching Training Programs in Economics*. New York: Joint Council on Economic Education, 1978.
- 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.
- Williams, J. "Assertion-Reason Multiple Choice Testing as a Tool for Deep Learning: A Qualitative Analysis." *Assessment and Evaluation in Higher Education*, 2006, 31(3), 287–301.

Short-Answer and Essay Tests

Short-answer items call for answers of about fifty words or less: Define *impedance*; List three causes of concrete failure; What happens during a solar eclipse? Essay questions, in contrast, require students to demonstrate their understanding of a topic and their ability to think critically and organize their thoughts. Although short-answer and essay questions are easier to create than multiple-choice items, they are more time-consuming to score, and maintaining consistency in grading is more difficult when evaluating essays. On the other hand, essay tests are the best measure of students' skills in higher-order thinking and written expression. In addition to the suggestions below, see Chapter 35, "Designing Effective Writing Assignments," and Chapter 36, "Evaluating Students' Written Work," for other ideas about creating topics and grading essays.

General Strategies

Do not use essay questions to evaluate understanding that could be tested with multiple-choice questions. Save essay questions for assessing reasoning and thinking skills. Appropriate tasks for essays include the following (adapted from Reiner et al., 2003):

Analyzing: "Find and correct the reasoning errors in the following passage."

Applying: "Describe a situation that illustrates the principle of . . ."

Comparing: "Identify two similarities and two differences between . . ."

Defending: "Present an argument to support your recommendation for . . ."

Designing: "Create a plan to . . ."

Evaluating: "Assess the strengths and weaknesses of . . ."

Generalizing: "State a set of principles that explains the following events."

Inferring: "How would character X respond if . . .?"

Interpreting: "Interpret the second stanza of . . ."

Justifying: "Explain why you agree or disagree with the following statement."

Predicting: “What will happen when . . . ?”

Relating cause and effect: “What are the major causes of . . . ?” “What would be the most likely effects of . . . ?”

Give students some test-taking advice. Most students will benefit from pointers like the following:

- Survey the entire test quickly, noting the directions and estimating the importance and difficulty of each question. If ideas or answers come to mind, jot them down quickly.
- Do a “brain dump”—write down anything that you fear you might forget later.
- Divide the time available among the questions. Allow more time for important or difficult questions. Allocate some time at the end to review your answers. Stick to your plan—four partially complete answers will earn a better grade than two extremely complete answers and two blanks.
- Analyze each question and its parts. The key nouns in the question should suggest the topic and subtopics, and the verbs (*compare, define, predict*) will indicate an approach to the topic. Observe any limitations (for example, “from 1900 to 1945”) expressed in the question.
- If you are completely stumped by a question, jot down *anything* you can think of that might be relevant. Free association may prompt your memory.
- Outline each answer before you begin to write. Write down your main points, arrange them in a pattern, and add details under each point. A quick outline will help you write with greater purpose, clarity, completeness, and speed.
- The first paragraph of an essay response should include a thesis statement that expresses your main point or conclusion. If the test question states a topic, create a *how* or *why* question about that topic. Your answer to that question will become your thesis. For example, if the test reads “Discuss the concept of love in D. H. Lawrence’s novel *Women in Love*,” one *how* question is “How does Lawrence portray different types of love in this novel?” The thesis statement could be “In depicting bisexual, homosexual, and familial love in *Women in Love*, Lawrence shows that hatred and isolation are present even in the closest love relationships” (based on Walvoord, 1986, p. 11).
- Follow your outline as you write. Skip every other line to leave room for additions or changes that occur to you as you reread your response.
- Support your thesis with examples and evidence.
- Include a conclusion, even if only a sentence or two that ties together the main points and states their importance.

- Reread your exam before you turn it in. Check for omissions, repetitions, and errors. Cross out and insert words as neatly as possible.
- If you are running out of time, list your main points and examples and write, “ran out of time.”

(Sources: Brooks, 1990; Forsyth, 2003; McKeachie and Svinicki, 2006; Sanders, 1966; and Walvoord, 1986)

Don't give students a choice of questions to answer. There are three drawbacks to giving students a choice. First, some students will waste time trying to decide which questions to answer. Second, you will not know whether all students are equally knowledgeable about all the topics covered on the test. Third, since some questions are likely to be harder than others, the test could be unfair. (Sources: Jacobs and Chase, 1992; Reiner et al., 2003)

Ask students to write more than one essay. Tests that ask only one question are less valid and reliable than those with a wider sampling of items. On a fifty-minute test, you could pose two or three essay questions or eight to ten short-answer questions.

Writing Effective Test Questions

State the question clearly and precisely. Avoid broad questions that invite different interpretations and that will make it hard to score the responses. Rather than asking students to “discuss” a topic, select a verb that suggests an approach to the topic. Often, adding *how* or *why* to an essay question will help students develop a thesis. The following examples illustrate ineffective and effective questions (adapted from Cashin, 1987; Jacobs and Chase, 1992; Reiner et al., 2003; Welsh, 1978):

Poor: What are three types of market organization? In what ways are they different from one another?

Better: Define *oligopoly*. How does *oligopoly* differ from both *perfect competition* and *monopoly* in terms of number of firms, control over price, conditions of entry, cost structure, and long-term profitability?

Poor: Name the principles that determined American foreign policy after September 11.

Better: Describe three principles on which American foreign policy has been based since September 11, 2001; illustrate each of the principles with two actions taken by the executive branch of government.

Poor: You are the president of the United States. What economic policies would you pursue?

Better: You are the president of the United States. State your goals for employment, price levels, and the rate of real economic growth. What fiscal and monetary policies would you implement to achieve your goals?

Poor: Why does an internal combustion engine work?

Better: Explain the functions of fuel, carburetor, distributor, and the operation of the cylinder's components in making an internal combustion engine run.

Poor: Was the above passage written by a classical or patristic Latin writer? Why do you think that?

Better: Decide whether the above passage was written by a classical or patristic Latin writer. Support your position by identifying specific phrases or other linguistic features of the author's style.

Specify the point value and criteria for each question. Giving the point value will help students allocate their time. Stating the criteria—for example, accuracy, completeness, relevance, clarity and strength of argument—may help students focus their efforts. (Source: Reiner et al., 2003)

Write out the correct answer yourself. Use your version to help you revise the question and to estimate how much time students will need. If you can answer the question in ten minutes, students will probably need twenty-five to thirty-five minutes. Some instructors give students advice on how much time to spend on each question.

Grading and Evaluating Exams

Create and use a scoring guide or rubric. Decide which facts or ideas a student must mention to earn full credit or partial credit. The following guide outlines the scoring of a 10-point essay question (adapted from Erickson et al., 2006, p. 173):

- *9 or 10 points.* The essay clearly states a position, provides support for the position, and raises at least one counterargument or alternative view and refutes it. Evidence is both persuasive and original. Counterargument is significant.
- *7 or 8 points.* The essay states a position, supports it, and raises a counterargument and refutes it. The essay contains one or more of the following ragged edges: evidence is not uniformly persuasive, counterargument is not a serious threat to the position, some points are extraneous or out of place.

- *6 points.* The essay states a position and raises a counterargument, but neither is well developed. The objection or counterargument seems rather trivial.
- *5 points.* The essay states a position and provides evidence supporting the position, but it does not raise possible objections or counterarguments.
- *3 or 4 points.* The essay states a position and provides some support, but the evidence is sparse, trivial, or general. The essay achieves its length largely through repetition of ideas and inclusion of irrelevant information.
- *1 or 2 points.* The essay does not state a position but simply restates the question and summarizes evidence discussed in class or in the reading.

Here is an example from an economics class (adapted from Ory and Ryan, 1993, p. 85). Students are given the following question: “Baseball is far less necessary than food and steel, yet professional ball players earn a lot more than farmers and steelworkers. Why? Take two or three sentences to summarize how an economist would explain the difference in salaries.” The scoring guide might allocate a total of 7 points for this question:

3 points for mentioning that salaries are based on the demand relative to the supply of a service

2 points for mentioning that excellent ball players are rare (low supply)

2 points for mentioning that many ball clubs want excellent players (high demand)

Read the exams without looking at the students’ names. Try not to bias your grading by carrying over your perceptions about individual students. Some faculty ask students to put a number or pseudonym on the exam and to place that code on an index card that is turned in with the test. Other faculty have students write their names on the last page of the blue book or on the back of the test.

Skim all the exams quickly, without assigning any grades. Before you begin grading, try to get an overview of the general level of performance and the range of students’ responses. (Source: McKeachie and Svinicki, 2006)

Grade the exams question by question rather than grading all questions for a single student. Shuffle the papers before scoring the next question to distribute your fatigue factor randomly. By shuffling the papers you also avoid ordering effects (that is, Riley’s “B” work always follows Coco’s “A” work and suffers from the comparison). (Sources: Fuhrmann and Grasha, 1983; Ory and Ryan, 1993)

Avoid judging exams on extraneous factors. Don't let handwriting, use of pen or pencil, or similar factors influence your judgment about the quality of the response.

Write comments on students' exams. Write brief notes on strengths and weaknesses to indicate what students have done well and where they need to improve. Writing comments will also keep your attention focused and will jog your memory if a student comes to talk to you about the exam. Some faculty ask students to write only on the odd-numbered pages of their blue book, leaving the even-numbered pages for the instructor's comments. Try to balance positive and critical comments, and focus on the organization of the response, not on whether you agree or disagree with a student's ideas. (Sources: Cashin, 1987; McKeachie and Svinicki, 2006; Sanders, 1966)

Read only a modest number of exams at a time. Set a time limit for each paper, and take regular short breaks. If possible, read all the responses to a single question in one session in order to minimize the effects of extraneous factors (time of day, temperature, and so forth) on your grading.

If time permits, read some of the papers twice. Wait a day or two and review a few randomly selected exams without looking at the grades you assigned. Rereading helps you increase your reliability as a grader.

Place the grade on the last page of the exam. Take care to protect students' privacy when you return or they pick up their tests.

Create standardized procedures for assistants who grade exams. The following process can promote consistency among graduate student assistants and readers (adapted from McKeachie and Svinicki, 2006):

- Meet as a group to discuss the answers to each question. Decide how many points will be given for what types of answers. Review the scoring criteria, rubric, and model answers prepared by the faculty member.
- Establish two- or three-person teams for each essay question. Give each team eight or ten exams and have each team member independently grade the team's question on each exam. Compare the grades that team members assigned and discuss the discrepancies until consensus is reached.

- If needed, have the teams grade and discuss a second batch of exams so that teams feel confident that they have arrived at common criteria.
- From this point on, each member grades independently. If any team member is unsure about a particular exam, it is passed to another team member for an opinion.

Returning Essay Exams

Return exams promptly. A quick turnaround reinforces learning and capitalizes on students' interest in the results. Try to return tests within a week or so.

Review the exam in class. Give students a copy of the scoring guide, rubric, or grading criteria. Let students know the features of a good answer and the most common errors the class made. If time permits, read examples of good and bad answers that you have created. Give students information on the distribution of scores so they know where they stand. (Source: McKeachie and Svinicki, 2006)

Convene groups to discuss test questions. Some faculty break the class into small groups to discuss answers to the test. Unresolved questions are brought up to the class as a whole. (Source: McKeachie and Svinicki, 2006)

Asks students for their opinions of the test. Ask students to tell you what was particularly difficult or unexpected. Find out how they prepared for the exam and what they wish they had done differently. Pass along the best tips to next year's class. (Source: Walvoord, 1986)

Keep a file of essay questions. Include a copy of the test with your annotations on ways to improve it, the mistakes students made in responding to various questions, the distribution of students' performance, and any comments that students made about the exam. Keep copies of good and poor exams. (Source: Cashin, 1987)

References

- Brooks, P. *Working in Subject A Courses*. Berkeley: Subject A Program, University of California, 1990.
- Cashin, W. E. "Improving Essay Tests." *Idea Paper*, no. 17. Manhattan: Center for Faculty Evaluation and Development in Higher Education, Kansas State University, 1987.

- Erickson, B. L., Peters, C. B., and Strommer, D. W. *Teaching First-Year College Students*. San Francisco: Jossey-Bass, 2006.
- Forsyth, D. R. *The Professor's Guide to Teaching: Psychological Principles and Practices*. Washington, DC: American Psychological Association, 2003.
- Fuhrmann, B. S., and Grasha, A. F. *A Practical Handbook for College Teachers*. Boston: Little, Brown, 1983.
- Jacobs, L. C., and Chase, C. I. *Developing and Using Tests Effectively: A Guide for Faculty*. San Francisco: Jossey-Bass, 1992.
- McKeachie, W. J., and Svinicki, M. *McKeachie's Teaching Tips*. (12th ed.) Boston: Houghton Mifflin, 2006.
- Ory, J. C., and Ryan, K. E. *Tips for Improving Testing and Grading*. Newbury Park, CA: Sage, 1993.
- Reiner, C. M., Bothell, T. W., and Sudweeks, R. R. *Preparing Effective Essay Questions*. Stillwater, OK: New Forums Press, 2003.
- Sanders, N. M. *Classroom Questions: What Kinds?* New York: Harper & Row, 1966.
- Walvoord, B. F. *Helping Students Write Well: A Guide for Teachers in All Disciplines*. (2nd ed.) New York: Modern Language Association, 1986.
- Welsh, A. L. Essay Questions and Tests. In P. Saunders, A. L. Welsh, and W. L. Hansen (Eds.), *Resource Manual for Teacher Training Programs in Economics*. New York: Joint Council on Economic Education, 1978.

Grading Practices

As Erickson, Peters, and Strommer (2006) point out, decisions about grading practices depend a great deal on an instructor's values, assumptions, and educational philosophy. For example, faculty who view introductory courses as “weeder” classes (courses that separate out students who seem unlikely to succeed in the field) will take a different approach to grading than faculty who view introductory courses as teaching important skills that all students need to master.

In addition to providing information on how well students are learning, grades are often used for other purposes (Scriven, 1974):

- to describe the worth, merit, or value of a piece of work
- to stimulate and encourage good work by students
- to communicate the instructor's judgment of the student's progress
- to inform the instructor about what students have and haven't learned
- to select people for rewards or continued education
- to improve students' ability to identify good work, to improve their self-evaluation or discrimination skills

The suggestions below are designed to help you develop clear and fair grading policies that emphasize learning. For tips on calculating final grades, see Chapter 44, “Calculating and Assigning Grades.”

General Strategies

Grade students on their mastery of knowledge and skills. Restrict your evaluations to academic performance. Eliminate nonacademic considerations, such as classroom behavior, punctuality, attitude, personality, need, or interest in the course material. Some faculty grade students' work without looking at the students' names to increase objectivity and guard against bias. (Sources: Guskey and Bailey, 2001; Jacobs and Chase, 1992; Scott, 1995)

Try not to overemphasize grades. Explain your grading policies at the beginning of the term, but do not repeatedly revisit the topic. Dwelling on grades is likely

to increase students' anxieties, decrease their motivation to learn for the sake of learning, and foster unhealthy, unproductive competitiveness. (Sources: Church et al., 2001; Fuhrmann and Grasha, 1983; Weimer, 2002)

Keep students informed of their progress throughout the term. For each paper, assignment, midterm, or project that you grade, give students a sense of what their score indicates about how well they are learning the material. Brief progress reports help students improve their study strategies and avoid unpleasant surprises at the end of the term. One faculty member gives each student midsemester ratings (*strong, OK, or needs improvement*) on the following dimensions: writes effectively; asks constructive questions during discussion; builds on others' ideas during discussion; takes responsibility for self and learning; displays wonder and curiosity; tolerates ambiguity; sees thinking as a way of life. (Source: Kloss, 1997)

Minimizing Students' Complaints about Grades

State your grading procedures in the course syllabus, and review this information in class. Students want to know how grades will be determined, the weights of various tests and assignments, and whether the class will be graded on a curve or by another system. Explain your policies regarding extra credit, late assignments, and revision of papers, and consistently enforce these policies throughout the semester. Because midcourse changes in policy may erode students' confidence in your impartiality, offer students a full explanation for any such changes.

Provide many opportunities for students to demonstrate what they know. By giving students many chances to show what they have learned, you can obtain a more accurate picture of their abilities and avoid penalizing a student who has an off day at the time of a test. In addition to a final exam, give several midterms and one or two short papers. For freshman and sophomore courses, Erickson, Peters, and Strommer (2006) recommend giving a quiz, short test, or written assignment every two or three weeks.

When possible, distribute grading criteria along with the assignment. For example, accompany a writing assignment with a description of the characteristics of an A or B paper; see Chapter 36, "Evaluating Students' Written Work." (Source: Walvoord and Anderson, 1998)

Remind students that grades are not judgments about people. Some students take grades as a sign of an instructor's approval or disapproval, or as a measure

of their self-worth. Remind these students that instructors grade only a piece of paper, not a person. For students who see a low grade as dooming their professional aspirations, you might mention that research studies disagree about the relationship between grades and eventual career success. (Sources: Roth et al., 1996; Waldman and Korbar, 2004)

Give encouragement to students who are performing poorly. Students receiving poor grades on exams or papers may need help in recalibrating their overly optimistic expectations; researchers report that most students overestimate the grade they will achieve in a course by an average of one full grade. Offer these students advice on how to improve their study habits and performance; don't let them blame such factors as a lack of talent ("I just don't have the knack for this"). (Sources: Forsyth, 2003; Lowman, 1995; Svanum and Bigatti, 2006; Weimer, 2002; Wendorf, 2002)

Talk to students who are angry or upset about their grade. Ask an upset student to take a day or two to reflect and to put something on paper. Some faculty have students complete a short grade-appeals form that requires citations from the course materials in support of their case. When you meet with the student in your office, have at hand a copy of the test questions, the answer key or criteria, and examples of good answers. Let the student speak first, listen with an open mind, and avoid interrupting. Try to determine which of the following reflects the student's intent (adapted from Sabee and Wilson, 2005):

- To learn more about the source of poor performance in order to be able to improve future performance. (The conversation is about instruction.)
- To persuade you to change a lower grade to a higher grade. (The conversation is about negotiation.)
- To vent anger and frustration over a lower-than-expected grade. (The conversation is about vindicating or justifying the student's feelings.)
- To impress you and repair the student's image in your eyes. (The conversation is about your opinion of the student and the student's self-image.)

Respond in a calm manner: don't allow yourself to become antagonized, and don't antagonize the student. Describe the key elements of a good answer, and point out how the student's response was incomplete or incorrect. Help the student understand your reasons for the grade you assigned. Take time to think about the student's request or to reread the exam if you need to, but resist pressures to change a grade because of a student's situation (such as the desire to maintain a high grade-point average or to avoid academic probation). If appropriate, for final

course grades, offer to write a letter to the student's adviser or to others describing the student's work and any extenuating circumstances. (Sources: Hampton, 2002; McKeachie and Svinicki, 2006; Sabee and Wilson, 2005)

Be aware of cultural variations in grading. In some countries, students are graded only on a comprehensive final examination, not on assignments and other course requirements during the semester. International students may appreciate a conversation about cultural differences and the effects of unrealistic pressures—coming either from families or from financial sponsors—to earn the highest grades. (Source: Eberly Center for Teaching Excellence and Intercultural Communication Center, n.d.)

Keep accurate records. Most departments keep copies of final grade reports, but you will want to keep a record of all grades assigned throughout the semester in case a student wishes to contest a grade, finish an incomplete, or ask for a letter of recommendation.

Minimizing Intradepartmental Complaints about Grades

Ask about your department's policies on grade distributions. Some departments have formal or informal agreements about how many high grades are desirable in an undergraduate course. Try to understand what the norms and constraints are in your department, or propose that a discussion of grading be added to the agenda of an upcoming department meeting. Are there concerns about limiting the number of As, but not the number of Bs? Or is your department more worried about courses with high failure rates (Ds and Fs)? (Source: Walvoord and Anderson, 1998)

Be prepared to show colleagues and administrators your syllabus, assignments, tests, criteria, and standards. Compile samples of student work at each grade level so that your colleagues can examine your standards. (Source: Walvoord and Anderson, 1998)

Try to dispel common misperceptions about grading practices. According to Lowman (1995), research studies contradict some common beliefs about grading:

- It is not true that the quality of education students receive is related to the difficulty of earning high grades. The quality of a college education is more a function of the quality of the faculty, the teaching, and the overall student population.

- It is not true that hard grading leads to student dissatisfaction. Students like hard graders and easy graders equally, and the majority of students seek out difficult courses rated positively over easier courses rated poorly.
- It is not true that strict grading motivates students to study. Students are more motivated by relationships (caring faculty, classmates, and friends) and by a desire for competence.

Grade Inflation

Be aware of broad national trends in grading. Grade-point averages have been rising for several decades, especially at research universities and highly selective colleges. This increase may reflect better preparation by entering students, improvements in teaching, wider use of standards-referenced grading, increased opportunities for students to revise and resubmit their work or to be retested, greater institutional attention to undergraduate teaching, and more lenient policies that allow low-achieving students to drop courses. Some campuses and departments also worry that some part of the increase is due to grade inflation—the gradual dilution of standards for A and B work. (Sources: Basinger, 1997; Boretz, 2004; Kuh and Hu, 1999; Kwon et al., 1997; Rosovsky and Hartley, 2002)

Be aware of some of the myths around grade inflation. Some people believe that grade inflation is caused by instructors watering down their courses and upping the grades they assign to secure higher ratings on students' end-of-course evaluations. The research does not bear this out (Marsh and Roche, 2000). Faculty do not get higher-than-average student ratings by offering easier courses and by giving students higher grades than they deserve. Some observers do point to a correlational (not a causal) relationship between grades and student ratings (Eiszler, 2002; Johnson, 2003), but there are competing explanations for these findings. See Chapter 60, "Student Rating Forms."

While some people believe that the biggest problem associated with grading is grade inflation, a more serious problem on some campuses is a grading disparity among different courses. This disparity, especially when it affects students' choices of courses, can cause institutional grade-point averages to rise and look like grade inflation (Hu, 2005).

Discuss the topic of grade inflation with your colleagues. Campuses and departments are experimenting with techniques for limiting or reversing the trend toward higher grades. These techniques include urging faculty to limit the number of high grades; recommending that faculty use a formula to index grades; adding supplementary information next to each grade on students' transcripts—either a

rank in class or the average course grade; and publicizing each faculty member's grade distributions. Web sites, such as pickaprof.com, routinely post grade distributions for individual courses. To date, there is no published research on the effects of the public dissemination of grade distributions. (Sources: French, 2005; Kuh and Hu, 1999; Rosovsky and Hartley, 2002)

Maintain appropriate academic standards. Campus and departmental concerns about grade inflation may be minimized if all instructors make sure that their grading reflects a considered judgment of students' achievement and performance in reference to clearly articulated learning goals or standards. (Sources: Guskey and Bailey, 2001; Kwon et al., 1997; Walvoord and Anderson, 1998)

Tactics and Policies

Return the first graded assignment or test before the add/drop deadline. An early assignment can help students decide whether they are prepared to take the class. Some faculty members give students the option of throwing out the score on the first test. Students may receive a low score because they did not know what the instructor required or because they underestimated the level of preparation needed to succeed.

Select a grading system appropriate to the assignment or test. The traditional letter grade system with pluses and minuses consists of thirteen levels (A+ through F). Other grade systems have fewer levels:

Five levels: A, B, C, D, F

Four levels: Check plus, check, check minus, no check; 4, 3, 2, 1

Three levels: Outstanding, competent, unacceptable; Exemplary, satisfactory, needs improvement; Excellent, acceptable, unacceptable; 2, 1, 0 (or 4, 2, 1)

Two levels: Pass, not pass; Credit, no credit; Satisfactory, unsatisfactory

Experts recommend using the fewest grading levels needed to suit the task and promote effective student learning. In general, the fewer categories you use, the more reliable and valid the grades will be. And the fewer the levels, the faster you can grade. Even if your campus uses a thirteen-level system for final course grades, you can use a simpler system for at least some tests and assignments and then convert to a common metric before calculating final grades. (Sources: Baker and Bates, 1999; Bressette, 2002; Chang, 1994; Guskey and Bailey, 2001; Landrum and Dietz, 2006; McClure and Spector, 2005; Walvoord and Anderson, 1998)

Set policies on extra credit. Students appreciate opportunities to earn extra credit, but some faculty worry that students might neglect required aspects of the course in pursuit of extra-credit points. Studies, however, show that the students most likely to undertake extra-credit work are those doing well in the course; struggling students are less likely to pursue extra credit. If you allow extra credit, the following are useful policies (adapted from Forsyth, 2003; Palladino et al., 1999):

- Make extra credit available to all students in the class, not just those doing poorly, with the exception of those who failed to complete key assignments or participate adequately in class.
- Announce options for extra credit at the beginning of the term, and explain your rationale and intent.
- Select extra-credit assignments that are pedagogically sound and relevant to the course content.
- Provide several choices for extra-credit opportunities.
- Limit the amount of extra credit that a student can earn.

(Sources: Forsyth, 2003; Hardy, 2002; Moore, 2005; Palladino et al., 1999)

Set policies on late assignments. Will you refuse to accept any late work? Deduct points according to how late the work is submitted? Handle late work on a case-by-case basis? Whatever policies you adopt, state them in your syllabus, and on your course Web site, and refer to these policies in class.

Some faculty refuse to accept late work and give students an F on the assignment. Other faculty argue that too much emphasis is put on penalizing late work. Other faculty impose various kinds of markdown penalties:

- a set fraction of a letter grade (from B to B-, for example) or a certain number of points
- a sliding penalty (for example, 5 points for one day late, an additional 1 point for each subsequent day late) with a cutoff date (assignments more than a week late receive a failing score)
- a grade penalty plus one of the following additional conditions:
 - New assignments cannot be submitted until the late work has been submitted.
 - Late work will not be graded until the end of the semester.
 - Late work will be returned without any comments or feedback.

Some faculty offer a small number of bonus points for assignments submitted on time (a decorum bonus). Late assignments do not receive the decorum bonus and are subject to markdown penalties.

Set policies on incompletes. Check with your department for policies on assigning Incompletes. On some campuses, incompletes are allowed only in case of situational hardship such as health issues or family emergencies. On other campuses, incompletes may be given if the students' work to date is of passing quality and the missing work is "minor." In no cases should an incomplete be given to students who want the additional time to be able to do extra work to raise their grades.

Grade fairly. Fairness is a major concern among students, and perceptions of unfairness are a source of many complaints. Fairness in grading is fostered by the following practices (adapted from Glenn, 1998):

- setting clear standards and announcing them at the beginning of the term and throughout the course
- applying the same standards consistently to all students
- resisting students' pressure to change a grade when you are in the right
- admitting when you made a mistake and correcting it
- taking your students' work seriously
- treating students impartially and respectfully

(Sources: Forsyth, 2003; Glenn, 1998; Holmes and Smith, 2003; Walvoord and Anderson, 1998)

Evaluating Your Grading Policies

Compare your grade distributions with those for similar courses in your department. Differences between your grade distributions and those of your colleagues do not necessarily mean that your methods are faulty. But glaring discrepancies should prompt you to reexamine your practices. (Source: Frisbie et al., 1979)

Include questions about grading policies on the end-of-course evaluation. On the evaluation form, ask your students questions like these (adapted from Frisbie et al., 1979, p. 22):

- Were the grading procedures for the course fair?
- Were the grading procedures for the course clearly explained?
- Did you receive adequate feedback on your performance?
- Were requests for regrading or review handled fairly?
- Did the instructor evaluate your work in a meaningful and conscientious manner?

References

- Baker, H. E., and Bates, H. L. "Student and Faculty Perceptions of the Impact of Plus/Minus Grading: A Management Department Perspective." *Journal of Excellence in College Teaching*, 1999, 10(1), 23–33.
- Basinger, D. "Fighting Grade Inflation: A Misguided Effort?" *College Teaching*, 1997, 45(3), 88–91.
- Boretz, E. "Grade Inflation and the Myth of Student Consumerism." *College Teaching*, 2004, 52(2), 42–46.
- Bressette, A. "Arguments for Plus/Minus Grading: A Case Study." *Educational Research Quarterly*, 2002, 25(3), 29–41.
- Chang, L. "A Psychometric Evaluation of 4-Point and 6-Point Likert-Type Scales in Relation to Reliability and Validity." *Applied Psychological Measurement*, 1994, 18(3), 205–215.
- Church, M. A., Elliot, A. J., and Gable, S. L. "Perceptions of Classroom Environment, Achievement Goals and Achievement Outcomes." *Journal of Educational Psychology*, 2001, 93(1), 43–54.
- Eberly Center for Teaching Excellence and Intercultural Communication Center. "Recognizing and Addressing Cultural Variations in the Classroom." Carnegie Mellon University, n.d. / <http://www.cmu.edu/teaching/resources/PublicationsArchives/InternalReports/culturalvariations.pdf>
- Eiszler, C. F. "College Students' Evaluations of Teaching and Grade Inflation." *Research in Higher Education*, 2002, 43(4), 483–501.
- Erickson, B. L., Peters, C. B., and Strommer, D. W. *Teaching First-Year College Students*. San Francisco: Jossey-Bass, 2006.
- Forsyth, D. R. *The Professor's Guide to Teaching: Psychological Principles and Practices*. Washington, DC: American Psychological Association, 2003.
- French, D. P. "Grade Inflation: Is Ranking Students the Answer?" *Journal of College Science Teaching*, 2005, 34(6), 66–67.
- Frisbie, D. A., Diamond, N. A., and Ory, J. C. *Assigning Course Grades*. Urbana: Office of Instructional Resources, University of Illinois, 1979.
- Fuhrmann, B. S., and Grasha, A. F. *A Practical Handbook for College Teachers*. Boston: Little, Brown, 1983.
- Glenn, B. J. "The Golden Rule of Grading: Be Fair." *PS: Political Science and Politics*. 1998, 31(4), 787–788.
- Guskey, T. R., and Bailey, J. M. *Developing Grading and Reporting Systems for Student Learning*. Thousand Oaks, CA: Corwin Press, 2001.
- Hampton, D. R. "The Quick Fix: Making Complaining Appealing." *College Teaching*, 2002, 50(2), 62.
- Hardy, M. S. "Extra Credit: Gifts for the Gifted?" *Teaching of Psychology*, 2002, 29(3), 233–234.
- Holmes, L. E., and Smith, L. J. "Student Evaluations of Faculty Grading Methods." *Journal of Education for Business*, 2003, 78(6), 318–323.
- Hu, S. "Beyond Grade Inflation: Grading Problems in Higher Education." *ASHE-ERIC Higher Education Report*, 2005, 30(6).
- Jacobs, L. C., and Chase, C. I. *Developing and Using Tests Effectively: A Guide for Faculty*. San Francisco: Jossey-Bass, 1992.
- Johnson, V. E. *Grade Inflation: A Crisis in College Education*. New York: Springer, 2003.

- Kloss, R. J. "The Way It Looks Right Now: A Simple Feedback Technique." *College Teaching*, 1997, 45(4), 139–142.
- Kuh, G. D., and Hu, S. "Unraveling the Complexity of the Increase in College Grades from the Mid-1980s to the Mid-1990s." *Educational Evaluation and Policy Analysis*, 1999, 21(3), 297–320.
- Kwon, I. G., Kendig, N. L., and Bae, M. "Grade Inflation from a Career Counselor's Perspective." *Journal of Employment Counseling*, 1997, 34(2), 50–54.
- Landrum, R. E., and Dietz, K. H. "Grading without Points: Does It Hurt Student Performance?" *College Teaching*, 2006, 54(4), 298–301.
- Lowman, J. *Mastering the Techniques of Teaching*. (2nd ed.) San Francisco: Jossey-Bass, 1995.
- Marsh, H. W., and Roche, L. A. "Effects of Grading Leniency and Low Workload on Students' Evaluations of Teaching: Popular Myth, Bias, Validity, or Innocent Bystanders?" *Journal of Educational Psychology*, 2000, 92(1), 202, 228.
- McClure, J. E., and Spector, L. C. "Plus/Minus Grading and Motivation: An Empirical Study of Student Choice and Performance." *Assessment and Evaluation in Higher Education* 2005, 30(6), 571–579.
- McKeachie, W. J., and Svinicki, M. *McKeachie's Teaching Tips*. (12th ed.) Boston: Houghton Mifflin, 2006.
- Moore, R. "Who Does Extra-Credit Work in Introductory Science Courses?" *Journal of College Science Teaching*, 2005, 34(7), 12–15.
- Palladino, J. J., Hill, G. W., and Norcross, J. C. Using Extra Credit. In B. Perlman, L. I. McCann, and S. H. McFadden (Eds.), *Lessons Learned: Practical Advice for the Teaching of Psychology*. Washington, DC: American Psychological Society, 1999.
- Rosovsky, H., and Hartley, M. *Evaluation and the Academy: Are We Doing the Right Thing? Grade Inflation and Letters of Recommendation*. Cambridge, MA: American Academy of Arts and Sciences, 2002.
- Roth, P. L., BeVier, C. A., Switzer, F. S., and Schippmann, J. S. "Meta-analyzing the Relationship between Grades and Job Performance." *Journal of Applied Psychology*, 1996, 81(5), 548–556.
- Sabee, C. M., and Wilson, S. R. "Students' Primary Goals, Attributions, and Facework during Conversations about Disappointing Grades." *Communication Education*, 2005, 54(3), 185–204.
- Scott, E. L. "Mokita—The Truth That Everybody Knows But Nobody Talks About: Bias in Grading." *Teaching English in the Two-Year College*, 1995, 22(3), 11–16.
- Scriven, M. "Evaluation of Students." Unpublished manuscript, University of California, Berkeley, 1974.
- Svanum, S., and Bigatti, S. "Grade Expectations: Informed or Uninformed Optimism or Both?" *Teaching of Psychology*, 2006, 33(1), 14–18.
- Waldman, D. A., and Korbar, T. "Student Assessment Center Performance in the Prediction of Early Career Success." *Academy of Management Learning and Education*, 2004, 3(2), 151–167.
- Walvoord, B. E., and Anderson, V. J. *Effective Grading: A Tool for Learning and Assessment*. San Francisco: Jossey-Bass, 1998.
- Weimer, M. *Learner-Centered Teaching: Five Key Changes to Practice*. San Francisco: Jossey-Bass, 2002.
- Wendorf, C. A. "Grade Point Average and Changes in (Great) Grade Expectations." *Teaching of Psychology*, 2002, 29(2), 136–138.

Calculating and Assigning Grades

Instructors' decisions about grading depend on the type of course they are teaching, their department's policies, and their views of the purpose of grades. In the broadest terms, the choice is between standards-referenced grading and norm-referencing grading.

Under standards-referenced grading (also called *criterion-referenced grading*, *task-based grading*, or *absolute grading*), a letter grade reflects a student's level of achievement against a specified standard or benchmark, independent of how other students in the class have performed. If all the students in a seminar give strong oral presentations, they will all receive As or Bs on that project. Conversely, if none of the students in a class scores better than 80 percent on a midterm, then no one in the class will receive higher than a B– on the exam.

Under norm-referenced grading (also called *grading on the curve*, *relative grading*, or *group-referenced grading*), in contrast, a letter grade reflects a student's level of achievement relative to other students in the class. Typically the proportion of students receiving each grade follows a pattern of a few As, lots of Bs and Cs, and a few Ds and Fs.

In large classes in which students' test scores are fairly well distributed, it may not matter which model an instructor chooses. But in smaller classes (forty students or fewer), it can matter a great deal: under norm-referenced grading, only a handful of the highest-scoring students will earn an A—no matter how many students did very well on an exam.

Educational measurement experts strongly recommend standards-referenced systems, especially for smaller classes, so that a grade reflects how well a student has mastered the material rather than how well a student performs compared to others (Brookhart, 1999; Dominowski, 2002; Gronlund and Waugh, 2008; Payne, 2003; Shepard, 2006).

The suggestions below are designed to help you understand the advantages and disadvantages of various grading strategies.

General Strategies

Familiarize yourself with department standards. Check to see how grading has been handled for the course in past semesters, and try to obtain grade distributions for earlier offerings. Ask colleagues who have taught the course before about their grading criteria and their classes' overall performance. (Source: Heppner, 2007)

Relate department standards to your conception of the course. Identify the objectives or goals you want your students to meet. What skills and knowledge are absolutely essential for students to pass the course? What would you wish from an A student?

Weight various course components in proportion to their importance. A three-hour final exam or a fifteen-page research paper should, obviously, count more than scores on two fifteen-minute quizzes. Another consideration: usually the final exam should count for no more than a third of the course grade. If the final is weighted too heavily, students will cram at the end of the term rather than work at an even pace. (Source: Lowman, 1995)

Use electronic gradebooks. Most learning management systems have an e-gradebook feature, and standalone gradebook software is available. E-gradebooks simplify record keeping, make it easy to enter and tally numerical information, and offer options for generating statistical data (such as mean scores and standard deviations).

Standards-Referenced Approaches

Grading according to absolute standards. *Absolute standards* refers to performance levels that an instructor sets at the beginning of the term. (Some faculty occasionally grade more generously than the announced standards.) For example, an instructor might decide on the following scale for a test worth 100 points: 93 or higher = A; 90–92 = A-; 86–89 = B+, and so on. The segments for each grade do not have to be of equal size.

This is the most commonly used standards-referenced model. An advantage of this approach is that it imposes no limit on how many students may earn As and Bs. Also, students know in advance how well they have to perform to earn the grade they want.

One difficulty arises in how to set rational standards. Experienced faculty can set cutoffs based on how students typically perform (a norm-referenced perspective), but new instructors may need guidance. Dominowski (2002, p. 137) proposes the following guidelines: A = 90 percent and above; B = 82–89; C = 63–81 (that is, the center of the C range is 72 percent); D = 48–62; F = 47 or below.

If many students perform very poorly, some instructors reset the standards to reflect students' performance, but such adjustments defeat the premise of creating absolute standards; see Chapter 39, "Quizzes, Tests, and Exams," for a discussion of how to respond when the entire class performs poorly. (Sources: Dominowski, 2002; McKeachie and Svinicki, 2006; Sadler, 2005)

Grading according to achievement of course objectives. This standards-referenced approach requires instructors to prepare a list of detailed objectives or learning outcomes, the measurable skills and knowledge students are expected to attain. The instructor then evaluates whether a student has or has not attained those outcomes. Under this model, most students who work hard enough and receive good instruction will obtain good grades. The difficulty is that the instructor must be able to clearly define the knowledge and skills that each grade represents as well as determine the minimum level of performance necessary to attain each grade and the best ways to measure achievement. In many college courses, the content is so extensive that an instructor cannot specify the requisite knowledge and skills with precision. (Sources: Frisbie et al., 1979; Hanna and Cashin, 1988; Ory and Ryan, 1993; Sadler, 2005)

Grading according to specified patterns of achievement. Under this model, an instructor specifies various levels of performance for each grade; for example, a student will receive an A if he or she performs at the A level on at least two-thirds of the assessment tasks and at the B level on the other third. The advantage of this approach is that different "formulas" and relative weights allow students to compensate for weak performance in some areas by superior performance in others. The disadvantages include complexity (the instructor must devise clear specifications and track performance levels for each component) and the need to familiarize students with this unusual method. (Source: Sadler, 2005)

Challenges in setting standards. All these methods require instructors to grapple with how to set standards. Numerical cutoffs seem arbitrary, but reliance on professional intuition—the expertise and "feel" that an instructor develops over time—is not transparent or objective. Describing exemplars (listing the characteristics of

typical performance at designated levels of quality or competence) is not realistic or suitable for all courses and all instructors. (Source: Sadler, 2005)

Norm-Referenced Approaches

Grading on a curve. In this approach, grades are determined by comparing a student's performance with that of other students. All the scores in a class are listed from highest to lowest, and grades are assigned according to cutoff points. Some instructors rely on a preset allocation; for example, Gronlund and Waugh (2008) recommend the following guidelines: 10 to 20 percent As, 20 to 30 percent Bs, 40 to 50 percent Cs, 10 to 20 percent Ds, 0 to 10 percent Fs. Other faculty use software to calculate the mean score and standard deviation, convert each student's score into a standard score and then convert the standard scores into a percentile based on a normal curve. A student's score is indicated as being in the 80th percentile or the 60th percentile, with the instructor determining in advance percentiles that represent letter grades.

Grading on the curve is a flexible approach that rewards students whose academic performance is outstanding in comparison to their peers. It also capitalizes on students' competitive tendencies. But it has many drawbacks:

- The grades do not indicate how much or how little students have learned—only where they stand in relationship to others.
- No matter how strong the class is, some students will receive low grades; no matter how weak the class is, some students will receive high grades. Some faculty compensate for inequities by adjusting the cutoff scores or by assigning a higher percentage of As when a class is unusually good.
- Grading standards may fluctuate from term to term. A student who earns a C+ might have received a B– a term earlier.
- Researchers note that grading on the curve may encourage exclusion, isolation, and competitiveness. It also may threaten students' sense of autonomy and fairness—their belief that their grade should depend on their efforts, not on how others perform.
- Grading on a curve assumes that every class is large enough and diverse enough to generate a full range of grades. But a small class of highly motivated students might not contain any subpar students. In difficult courses, the enrollment may be unrepresentative because struggling students dropped the course after doing poorly on the first quiz.
- Grading on the curve obscures the effects of course design and teaching (whether excellent or poor) because those factors are not scrutinized in analyzing student performance.

- Grading on the curve makes it difficult for an instructor to tell students at the beginning of the term what they must know or be able to do in order to earn an A or B.

(Sources: Dominowski, 2002; Forsyth, 2003; Guskey and Bailey, 2001; Hanna and Cashin, 1988; Sadler, 2005)

Grading according to department practices or faculty consensus. Some faculty try to have their grade distributions reflect the averages reported in their department. All faculty who teach the same course might develop a consensus on the distribution of grades suitable for a typical class (say, 20 percent As, 25 percent Bs, 30 percent Cs, 20 percent Ds, and 5 percent Fs), with adjustments for an unusually strong or unusually weak class. (Source: Hanna and Cashin, 1988)

Grading according to breaks in the distribution. In this model, the instructor lists students' scores from highest to lowest and looks for natural gaps or breaks in the distribution. For example, if six students score 80 or higher, and no one scores between 73 and 79, and two students score 72, the instructor might assign As to students who scored 80 and above, and start the Bs at 72. Significant gaps, however, are rare in large classes. Even a gap that looks meaningful might not represent a true difference in achievement as much as the vagaries of an unreliable test, good guessing, or poorly written items. Further, the grade distribution depends on judgments made after students have taken the test rather than on preestablished guidelines that can be stated prior to testing. (Sources: Forsyth, 2003; Jacobs and Chase, 1992; Ory and Ryan, 1993; Payne, 2003)

Grading on a bell curve. A bell curve is a symmetrical statistical model that is inappropriate for grading coursework. The classic bell curve results from the following distribution of grades: 7 percent As, 24 percent Bs, 38 percent Cs, 24 percent Ds, 7 percent Fs. Bell curves have their uses, but student performance is not distributed normally within a class, nor are teacher-made tests so well designed as to yield bell-shaped distributions. Even proponents of grading on the curve do not recommend using the bell curve. (Source: Gronlund and Waugh, 2008)

Hybrid Models

Grading according to highest scores earned and percentages thereof. This hybrid model (developed by Carter as reported in Fuhrmann and Grasha, 1983, p. 184) combines the advantages of standards-referenced and norm-referenced

approaches. Grades are assigned by comparing each student's score with the highest scores earned in the class:

- Compute a score for each student.
- Compute the mean score of the best-performing portion of the class:
 - For an average class, use the scores of the top 10 percent; for a superior class, the scores of the upper 15 or 20 percent; for a weak class, the upper 5 to 8 percent.
 - Add all the scores in this best-performing sample and divide by the number of scores in the sample. The result is the mean for the best performers.
- Assign grades according to a predetermined scale; for example: A = 95 percent of the mean of the best-performing sample, B = 85 percent of the mean, C = 75 percent of the mean, D = 65 percent of the mean.

In this model, class performance affects the score needed for each grade, but the number of students who can earn As and Bs is not limited. Some faculty take a shortcut by using the highest score (rather than the mean of the highest scores), but this shortcut is far too dependent on a single student's score.

Grading according to core and developmental objectives. In this model, the instructor identifies the *core objectives*, or essential content that all students must master to a satisfactory degree to receive a passing grade. The instructor also identifies those aspects of the content that students may never fully master but to which they can aspire (called *developmental objectives*). In Chemistry 1A, for example, core objectives would include the concepts, skills, and knowledge that students will need in order to be prepared for Chemistry 1B; developmental objectives might include “thinking like a scientist.” The core objectives are graded against an absolute standard (students either know or do not know the core), but developmental objectives are graded by comparisons among students. (Source: Gronlund and Waugh, 2008)

Other Approaches

Self-grading and peer grading. Some faculty let students grade themselves. The grade must be accompanied by a detailed justification that takes into account the extent and level of their learning, their performance on exams and assignments, their perceived grasp of the material, the amount of time spent on the course, and the amount of reading completed. This approach has the advantage of fostering students' abilities to evaluate their work, but it abrogates one of the faculty's chief responsibilities, the rendering of professional judgments about students' learning and performance. Self-grading, however, can be used

formatively to give students the skills they need to assess their own learning. Peer grading, where students grade one another's work, works best in classes that feature a lot of small-group work; see Chapter 21, "Learning in Groups," for suggestions. (Sources: Adams and King, 1995; Dochy et al., 1999; Jacobs and Chase, 1992; MacGregor, 1993; Strong et al., 2004)

Contract grading. In contract grading, instructors create a menu of required and elective assignments, and each student decides which assignments to do and how much weight each assignment will carry. For example, the following menu requires a student to make choices that total 200 points:

Complete two required assignments:

Two exams (worth 20 to 35 points each)

Group project (50 to 70 points)

In addition, complete two of these elective assignments:

Prepare an annotated bibliography (15 to 35 points)

Review a journal article (15 to 35 points)

Write the abstract of an unpublished manuscript (15 to 35 points)

Keep a weekly learning log (15 to 35 points)

Contract grading gives students choices in their learning goals and how they are evaluated, which may increase motivation and interest in learning. The drawback is the increased administrative burden, as instructors must develop grading alternatives, negotiate and renegotiate contracts, and keep track of each student's progress which quickly becomes impractical in medium- or large-enrollment courses. (Source: Hiller and Hietapelto, 2001)

Calculation of Final Grades for the Course

Become familiar with various methods. There are many different ways to calculate course grades, using weighted letter grades, accumulated points, median grading, and holistic grading or mastery of key elements; see Brookhart (1999), Ory and Ryan (1993), Walvoord and Anderson (1998), or Zlokovich (2004). One commonly used strategy is described below.

Convert the grades on all course components to numerical scores. Calculating final grades is usually easiest if the grades on all assignments are converted to

numerical scores. For this purpose, many faculty use $A = 95$, $A^- = 90$, $B^+ = 87$, and so on. (Source: Brookhart, 1999; Forsyth, 2003)

Decide how to handle missing scores. Some instructors ignore the missing work and base their calculations on a student's other scores; other instructors take into account the reason for the missing work. If you want to assign a number for a missed assignment or test, your choices include (1) assigning a zero—but this is likely to overpenalize the student and is not recommended; (2) assigning a score at the bottom of the F range; (3) assigning a score somewhat lower than the lowest score earned by a classmate on the assignment; (4) assigning the lowest score earned by a classmate. (Sources: Brookhart, 1999; Forsyth, 2003)

Weight each course component. Use the weights that you announced at the beginning of the semester. In general, weights should reflect a component's importance to the course goals and its complexity and difficulty; in addition, most faculty assign greater weight to tests later in the term. One professor adjusts the weighting to reflect a student's performance: in a course with two midterms and a final, the student's highest test score is weighted 50 percent, the middle score 30 percent, and the lowest 20 percent. Students appreciate having their best work weighted so heavily in their course grade, but the disadvantage is that students do not know until after the final exam how much weight each component carries. (Sources: Brookhart, 1999; Dominowksi, 2002; Forsyth, 2003; Guskey and Bailey, 2001)

Create a composite score. Your grading software will adjust the numerical scores on each component by the weights you indicate and provide a composite score for each student. (If you are doing these calculations by hand, consult with your campus testing office.) You would then assign final grades using preset cutoff points (standards referenced) or relative standing in the class (norm referenced) or a hybrid method.

If you use a standards-referenced model, you can simplify the calculations by using a total-points method of grading as described by Forsyth (2003): at the beginning of the term, state the maximum number of points students can earn for each assignment and test, with these points reflecting the importance of each component. For example, each of two tests might be worth 100 points, the homework 150, and the final exam 150—for a total of 500 points. In order to earn an A, a student would need to accumulate 450 points (90 percent of 500 points).

If you use a norm-referenced model, you may need to take the additional step of equalizing the variances—otherwise a test with a wide range of scores will more heavily influence the final grades than a test with a narrower spread. Most

testing and grading software perform these conversions; consult with your testing office for more information.

Review borderline cases. Because graders are fallible and grading is imprecise, take a careful look at students who fall right below the cutoff for a higher grade. Consider the full range of the student's achievement and trajectory of improved performance. Would dropping the first test score or the lowest test score bump up the final grade? If the final exam were cumulative, an instructor might give the higher grade if the student did particularly well.

One faculty member allows students to “purchase” grade “insurance” by completing a small number of problems any time before the final exam. If a student's course grade is not on the borderline, the insurance problems are not graded. But students within one percentage point of the cutoff will receive the higher grade if they answered the insurance problems correctly. Another faculty member includes an optional question on the final exam. The optional question is read only if a student is within five points of the next highest grade, and it is scored 0 through 5. (Sources: Brookhart, 1999; Dominowski, 2002; Peploski, 2004)

Decide how to handle improvement. Some faculty believe that course grades should take into account a measure of a student's growth and development over the semester. Otherwise, these faculty argue, a student who enters the course fairly knowledgeable will receive an A even if he or she learns very little or demonstrates little effort. But an overemphasis on improvement can also produce inequities: a student who scores 55 percent on the first two quizzes and 70 on the final has made substantial progress but knows less about the subject than the student who scored 85 on the first two quizzes and 88 on the final. And only the second student deserves a grade high enough to proceed to a more challenging upper-level course. Grading on the basis of improvement also makes it difficult for students to interpret what their grades mean: does a B mean that their work is above average or that their improvement is above average? Experts recommend that any bonus for improvement be quite modest: a few extra points for steady, significant improvement throughout the semester.

One math professor quantifies students' improvement over the semester. Say a class has two midterms worth 100 points each and a final worth 200 points. Ole's scores are 50 out of 100 on the first midterm, 80 out of 100 on the second midterm, and 190 out of 200 on the final. His unadjusted total score for the course is 320. To take into account Ole's steady improvement, the professor weights Ole's scores for the second midterm and final more heavily. The weight of the second midterm is calculated by subtracting Ole's score on the first midterm

from the total points available for both first and second midterms, or $200 - 50 = 150$. His actual score on the second midterm is then multiplied by this weighting factor. Thus his adjusted score for the second midterm is $(200 - 50)(80/100) = 120$. His first score and his second, adjusted score are then added together, $50 + 120 = 170$. His adjusted score on the final is calculated using the same process. Ole's cumulative total score (170) is subtracted from the total points available for both midterms and the final (400). His actual score on the final is then multiplied by this new weighting factor, or $(400 - 170)(190/200) = 218.5$. To calculate the total adjusted score, add his adjusted score on the midterms (170) to his adjusted score on the final (218.5) for a total of 388.5. (Sources: McKeachie and Svinicki, 2006; Ory and Ryan, 1993)

Decide how to handle effort. Studies show that undergraduate students believe that if they work hard, they should receive at least a C in the course even if they cannot demonstrate that they have met the academic goals of the course. Most faculty, however, do not believe that effort (however measured) trumps learning in the assignment of grades. Experts recommend that instructors clarify this issue at the beginning of the term. (Sources: Adams, 2005; Gaultney and Cann, 2001; Miley and Gonsalves, 2004)

References

- Adams, J. B. "What Makes the Grade? Faculty and Student Perceptions." *Teaching of Psychology*, 2005, 32(1), 21–24.
- Adams, C., and King, K. "Towards a Framework for Student Self-Assessment." *Innovations in Education and Training International*, 1995, 32(4), 336–343.
- Brookhart, S. M. "Art and Science of Classroom Assessment: The Missing Part of Pedagogy." *ASHE-ERIC Higher Education Report*, 1999, 27(1). Washington, DC: George Washington University, Graduate School of Education and Human Development.
- Dochy, F., Segers, M., and Sluijsmans, D. "The Use of Self-, Peer and Co-Assessments in Higher Education: A Review." *Studies in Higher Education*, 1999, 24(3), 331–350.
- Dominowski, R. L. *Teaching Undergraduates*. Mahwah, NJ: Erlbaum, 2002.
- Forsyth, D. R. *The Professor's Guide to Teaching: Psychological Principles and Practices*. Washington, DC: American Psychological Association, 2003.
- Frisbie, D. A., Diamond, N. A., and Ory, J. C. *Assigning Course Grades*. Urbana: Office of Instructional Resources, University of Illinois, 1979.
- Fuhrmann, B. S., and Grasha, A. F. *A Practical Handbook for College Teachers*. Boston: Little, Brown, 1983.
- Gaultney, J. F., and Cann, A. "Grade Expectations." *Teaching of Psychology*, 2001, 28(2), 84–87.
- Gronlund, N. E., and Waugh, C. K. *Assessment of Student Achievement*. (9th ed.) Boston: Allyn and Bacon, 2008.

- Guskey, T. R., and Bailey, J. M. *Developing Grading and Reporting Systems for Student Learning*. Thousand Oaks, CA: Corwin Press, 2001.
- Hanna, G. S., and Cashin, W. E. "Improving College Grading." *Idea Paper*, no. 19. Manhattan: Center for Faculty Evaluation and Development, Kansas State University, 1988.
- Hepppner, F. *Teaching the Large College Class: A Guidebook for Instructors with Multitudes*. San Francisco: Jossey-Bass, 2007.
- Hiller, T. B., and Hietapelto, A. B. "Contract Grading: Encouraging Commitment to the Learning Process through Voice in the Evaluation Process." *Journal of Management Education*, 2001, 25(6), 660–684.
- Jacobs, L. C., and Chase, C. I. *Developing and Using Tests Effectively: A Guide for Faculty*. San Francisco: Jossey-Bass, 1992.
- Lowman, J. *Mastering the Techniques of Teaching*. (2nd ed.) San Francisco: Jossey-Bass, 1995.
- MacGregor, J. (Ed.). "Student Self-Evaluation: Fostering Reflective Learning." *New Directions in Teaching and Learning*, no. 56. San Francisco: Jossey-Bass, 1993.
- McKeachie, W. J., and Svinicki, M. *McKeachie's Teaching Tips*. (12th ed.) Boston: Houghton Mifflin, 2006.
- Miley, W. M., and Gonsalves, S. "Grade Expectations: Redux." *College Student Journal*, 2004, 38(3), 327–332.
- Ory, J. C., and Ryan, K. E. *Tips for Testing and Grading*. Newbury Park, CA: Sage, 1993.
- Payne, D. A. *Applied Educational Assessment*. (2nd ed.) Belmont, CA: Wadsworth/Thomson, 2003.
- Peploski, J. "Grade 'Insurance' in Large Enrollment Courses." *Teaching Professor*, October 2004, 18(8), 6.
- Sadler, R. "Interpretations of Criteria-Based Assessment and Grading in Higher Education." *Assessment and Evaluation in Higher Education*, April 2005, 30(2), 175–194.
- Shepard, L. A. Classroom Assessment. In R. L. Brennan (Ed.), *Educational Measurement*. (4th ed.) Westport, CT: American Council on Education/Praeger, 2006.
- Strong, B., Davis, M., and Hawks, V. "Self-Grading in Large General Education Classes: A Case Study." *College Teaching*, 2004, 52(2), 52–57.
- Walvoord, B. E., and Anderson, V. J. *Effective Grading: A Tool for Learning and Assessment*. San Francisco: Jossey-Bass, 1998.
- Zlokovich, M. S. Grading for Optimal Student Learning. In B. Perlman, L. I. McCann, and S. H. McFadden (Eds.), *Lessons Learned: Practical Advice for the Teaching of Psychology*. Vol. 2. Washington, DC: American Psychological Society, 2004.