

Section VI

Assessing the Journey: Assignments, Grading, and Course Closure

The continual improvement of teaching and learning requires that college teachers have some idea of how things are working. Traditionally, exams and papers have been the mainstay of measuring academic performance. But as with the formal, nonparticipatory lecture, these tried and perhaps not-so-true methods have come under fire as applicable to all times, places, class sizes, and learning goals. Furthermore, college teachers have debated, time and time again, how to evaluate performance in the classroom as mentors and as public purveyors of knowledge. Some argue that research standards are clear; others, us among them, suggest that the same could be (and is increasingly becoming) so for teaching. It is a matter of effort and will, not intrinsic possibility. In this section, then, we also address the issues of how to evaluate teachers.

General Strategies for Evaluation and Grading

Faculty hold notions of grading from upholding straight bell curves to giving all students A's for participating in a course. Again, these issues are the source of a great deal of controversy. In this subsection, we ask users of the *Fieldguide* to stand back and assess how they think about the grading process.

Rather than have students play a “guessing game” about what we expect of them, Hull suggests that we provide a handout, and here we reprint his.

How to Succeed: A Handout for Students

Jay Hull

Over the years, a number of students have come to me and said, “I don’t understand why I am not doing better in your class.” I think students often approach science courses with expectations about tests (and hence study methods) that are a bit different [from] their expectations regarding tests in the social sciences and humanities. Science and math courses are often seen to involve facts and theories, whereas social science and humanities courses are often seen to emphasize general themes and perspectives. Social psychology is a blend: a body of research facts, some of which are consistent with prevalent theories, others of which are not. The exams demand a knowledge of research facts and relevant theories together with an understanding of how those facts and theories relate to a common theme and how they apply to everyday life.

My approach to testing follows from a general teaching philosophy that you have to know the material before you can apply it in useful and interesting ways. As a consequence, all of my courses (even my seminars) involve written examinations in which you have to demonstrate an adequate knowledge of the material. Beyond that, the higher the level of the

course, the more insistent I am that you be able to compare and contrast different theories, apply them to new situations, integrate distinct research literatures, provide your own insights as to directions that research literatures should take, etc. Even at the advanced level, however, I insist that students demonstrate knowledge of the factual material.

To do well, you should attend all of the lectures and read all of the material. A lecture missed or a book chapter skimmed is guaranteed to lose you points on the exam. Nevertheless, reading the material and attending the lectures is not sufficient for doing well on the tests. If you read all of the material, attend all of the lectures, and study the night before the exam, you should be satisfied with a C. In order to make a B or higher, you are going to have to know details and show an ability to integrate and apply what you know. In order to make an A, you are going to have to know everything. Part of the reason for this is the way I choose test questions. Although I do think there are major theories and findings in social psychology (and these theories and findings will almost certainly find their way onto my exams), I also believe that if I went over something in class or

[if] the book spent a paragraph discussing a topic, then that material is fair game. In my way of thinking, A students should get the majority of these “randomly chosen” items correct and B students should get a healthy number of them. Given this approach to testing, it makes little sense for students to try to second guess what material will be on the exam. Nor is it really possible on these exams to “study the wrong material.” If I talked about it in class, if it was discussed in the readings, it is fair game for the test. The longer it was talked about in class, or the more it was discussed in the book, the fairer game it is.

Given the nature of my exams, how should you study? There is no pat answer to this question. My typical advice for a person who is not doing as well as he or she wants is as follows:

1. You should read all of the material and attend all of the lectures. Skip a chapter or miss a lecture, and you will lose points. This is because my tests are balanced to include questions covering material from every lecture and every chapter of the primary text (a little over half of the material from secondary texts is covered on an exam).
2. Second, when you are in class, you should take notes. I notice when people don't take notes, and over the course of the term I keep track of how they do. They rarely make better than a B-minus in the course, and they frequently complain that they don't understand why they are doing poorly. Take as detailed notes as possible without interfering with your ability to understand what I am saying. If you don't understand what I said or if I am going too fast, raise your hand and ask me to repeat it. Never assume that you will understand it later in the comfort of your dorm—you won't. If you ask me to repeat something, I will try to explain it in a slightly different way that may be easier to understand.
3. Read the assigned book material before class. In doing this, pay attention to the format of the material. Most books and articles are written with explicit headings and subheadings. If you stripped away the sentences and paragraphs, you would still be left with a fairly detailed outline. Actively try to decode this outline before reading the material. Then use the outline to break the material into meaningful chunks. While reading, try to think about the material in terms of its relationship to this outline.
4. Reread the book material within two days following the relevant lecture(s), this time using a ballpoint pen to underline between one-quarter and one-third of the material. The purpose of this underlining is to boil down the chapter in a way that you can reread just your underlining and get all of the facts (while cutting out padded elaborations and examples). Now go back, reread just your underlining, and use a felt-tipped highlighter to highlight key words or phrases. You have now created an even more fine-grained outline and done so in a way that is meaningful to you.
5. Before the exam, review the material in a way that is consistent with the test format. Try playing Jeopardy with the material! Make up the questions to fit the answers in your text or your notes. Try making them up using the format you will face (essay, short answer, multiple choice) and see how well you answer them. Share them with others in

the class and test yourself on their questions as well.

6. Within [two] days following a lecture (and before the next lecture occurs), sit down with your lecture notes. They are probably scribbled, include unlabeled and unelaborated diagrams, and [include] chunks of phrases and terms. Get out a clean notebook and transfer your scribbled class notes into a text with complete sentences, in paragraph form, with labeled diagrams. You will be amazed at the extent to which this process will (a) cause you to remember details that you did not have time to write down in class and (b) see the coherence and overall structure of the lecture—thus improving your ability to integrate the material and thus improving your ability to remember it. Make sure to do this for one lecture before the next takes place. The longer you wait, the worse your memory and the less time you have to devote to this technique. Now apply the underlining-highlighting technique explained above to your transferred lecture notes.
7. Start preparing for the exam several days in advance. Do not spend this time transferring your notes or underlining or highlighting the book or lecture material. All of that preparation should already be done. Spend this time testing

yourself on the material (looking at the highlighting and recreating the text in your mind, making sure you understand the material, finding your weak spots and eliminating them).

8. Many people feel that they benefit from study groups. Be careful, though. Although study groups can help (e.g., by quizzing each other and filling in gaps in understanding), they can also waste time and yet convince themselves that they did a good job. If you study in groups, try to keep the group focused on its goal.
9. Get a good eight hours of sleep the night before the exam.
10. Review all of the highlighted material (not the underlining) the morning of the exam.

I realize that there are many study methods, and the above plan of attack may be overkill for you. However, keep in mind that this advice was originally formulated for people who came to me complaining that they were doing poorly in my class and did not know what to do about it. In addition, since I started making this sheet available to students at the beginning of my courses, the average grade on the first exam is higher. Whether the improvement is due to the specific advice, better anticipation of the nature of my exams, or the motivating influence of fear is unclear.

Svinicki lays out four criteria by which to evaluate our own methods for testing student learning.

Four R's of Effective Evaluation

Marilla D. Svinicki

Because so much depends upon the evaluation of a student's learning and the resulting grade, it is in everyone's interest to try to make the evaluation system as free from irrelevant errors as possible. Borrowing from the evaluation literature, I propose the four R's of evaluation—Relevant, Reliable, Recognizable, Realistic—as ways to ensure the quality of our evaluation systems.

Relevant

In the jargon, this is known as the validity of an evaluation method. This means that any activity used to evaluate a student's learning must be an accurate reflection of the skill or concept which is being tested. What are the characteristics of a relevant evaluation?

Oddly enough, one characteristic that might seem very mundane is that the evaluation activity must appear related to the course content (known in the jargon as face validity). A common student complaint is that tests are not related to the course content or what was presented in class. Although we know that what we assign is directly related to the course, the students often don't see the connection. And, student impressions aside, the more obvious the

connection, the higher the probability that we really have a valid evaluation activity.

A second characteristic of relevant evaluations is that they are derived directly from the objectives (known in the jargon as content validity). The most obvious way to achieve this is to follow the objectives as closely as possible in selecting activities.

If your objective is that the students will be able to select the appropriate statistic for analyzing a given set of data, the evaluation should provide them with a data set and have them select the analysis. It could take many forms:

- an in-class exam where no actual calculations are done;
- an out-of-class exam where no actual calculations are done;
- an out-of-class homework assignment involving extensive calculations;
- a component of a large-scale semester-long project; [or]
- an in-class exercise done in groups with class-generated data.

All of these alternatives represent relevant tests of that objective.

Another characteristic of a relevant evaluation is how well performance on that evalu-

ation predicts performance on other closely related skills, either at the same time (concurrent validity) or in the future (predictive validity). If the skill you are supposedly testing should be highly correlated with some other skill which you are also testing, chart the students' performances on each and see if they follow the same pattern.

To use a simplified example, we can say that the ability to add two single-digit numbers is a precursor to, and therefore highly correlated with, the ability to add two two-digit numbers. Therefore, students who do poorly on the former should not be able to do well on the latter. If they do, then one of the two tests is not measuring what it is supposed to be measuring and is therefore not relevant to the additional skill we are trying to evaluate.

Reliable

The second aspect of an evaluation activity is how reliably or consistently it measures whatever it measures without being affected too much by the situation. A student's grade should not hang on a single performance or on the mood of the person making the judgment. Of course, no system is perfectly reliable and will produce exactly the same evaluation of performance each time, but the goal here is to eliminate as many sources of error as possible.

The three biggest sources of error in reliably evaluating a student are:

- poor communication of expectations;
- lack of consistent criteria for judgment; and
- lack of sufficient information about performance.

Poor communication of expectations means that poor student performance may be the result of the student's failure to correctly interpret

the task requirements. In written exams, this usually is caused by ambiguous questions, unclear instructions, corrections given verbally during the test, and so on. In each case, a bad grade is the result of the student not understanding the question. The student may in fact know the material.

Lack of consistent criteria for judgment means that, if the same performance were to be judged a second time by the same grader or if another grader evaluated it, it might not receive the same grade because the basis for judging was unclear. The clearer the criterion for judging a student's performance, the more reliable the evaluation becomes.

For example, one real strength of multiple-choice tests is that the grading is very reliable. Either the students marked the correct answer or they didn't; very little is left to the judgment of the grader. On the other hand, essay tests are notoriously unreliable unless the instructor takes pains to make the criteria explicit and keeps checking to make sure he or she is not straying too far from the preset criteria.

Lack of sufficient information is the third source of error in evaluating students, not just in terms of the amount of information but also in terms of variety of information sources. Not everyone excels in every format. Using only one format may introduce a source of bias for or against some students and lower the reliability of an evaluation.

Recognizable

Our third R is the need for the evaluation system to be recognizable to the students. By this, we mean that students should be aware of how they will be evaluated, and their class activities should prepare them for those evaluations. Testing should not be a game of "Guess what I'm going to ask you."

Students don't mind "hard" tests as long as there are no surprises and they can recognize the relationship of the test to the course. Some instructors may criticize this as "teaching the test," but in reality the test should be the best statement of the course expectations and therefore should mirror the teaching. Furthermore, few courses are taught at such a low level that tests are verbatim transcripts of the class or text; rather, they are interpretations or new examples of the class or text material.

Realistic

All of the above activities require work on the part of either the students or the teacher. So, to

avoid burning out either, the final R is that the evaluation system should be realistic; the amount of information obtained is balanced by the amount of work required. Too often, we forget that our students are taking three to four other courses along with ours.

What is a realistic [evaluation system]? Unfortunately, no one can give a blanket answer to that question. I can say that several smaller assignments tend to be more valuable than one large assignment. Alternatively, if a large assignment is called for, spreading it out across the semester and requiring components to be handed in periodically is a good technique, both pedagogically and administratively.

Trying to eliminate the misconception that a grade is a “gift,” Clack and Poledink provide a method to increase student involvement in monitoring their own progress throughout a course.

A Grade Is Not a Gift

Donna Clack and Alice Poledink

“**W**hat grade are you going to give me on this paper?”

All teachers have heard the question. The idea that a grade is a gift seems embedded in our language and our students’ perceptions of the learning process.

“How are you doing in your classes?” we ask students. “Fine,” they tell us, with the word describing performances ranging from exemplary to disastrous.

Too often, these questions reflect misperceptions, especially for our at-risk students, who often have trouble assessing their performance or understanding the implication of behaviors like missing class, assignments, and quizzes.

To change the misperceptions of the words “gift” and “fine,” we have adopted the policy of having students either continuously monitor their grades or periodically average their overall course grade and submit it to the instructor for confirmation. By doing this, we’re helping our students adjust to the idea that a grade is continuously earned and that they have to frequently assess their performance rather than assume that the grade [is] a “gift” and they’re doing “fine.”

To create a *grade-monitoring sheet*, the instructor develops a guide for students or simply lists assignments. Following this assignment list is a column for recording the total points for each assignment. Next to it is a column totaling the points earned. This last column can be used to keep a running total of the grade. Also, instructors may want to include an example of how to calculate the grade since some students have difficulty computing a percentage.

The *grade-averaging sheet* is an informal worksheet created by the instructor for periodic use during the course. All assignments and other factors contributing to the course grade are listed, with directions for calculating the overall average where needed. From this sheet, students learn early in the semester if all is not going well and have time to make adjustments.

Also, using either of these systems, the instructor can compare the student’s calculation with what [is] in the grade book. After having students calculate their grade, we sometimes ask them to answer in writing the following questions:

- Are you satisfied with your grade?
- Were you surprised by your grade?

- Why do you feel you have earned this grade?
- What plans do you have to improve or maintain your grade?

These questions encourage students to reflect on their learning behavior and attitudes, thereby developing the kind of metacognitive skills that help them assume responsibility and improve performance.

We have found that we can use these sheets for several purposes. Knowing their progress gives students a sense of control and, as a result, enhances their motivation. Seeing a grade attached to each assignment enables students to more easily see what kind of help they need. They can identify their strengths and weaknesses and, perhaps with the help of the instructor, pick out patterns that can help them set more realistic goals. They may determine that they need a tutor for content skills if exam or quiz scores are low or assistance with time

management if incomplete or missing assignments are lowering their grades.

The sheets also encourage dialogue between students and instructors. Some students avoid examining their role in the learning process; the sheets give us a reason to approach them on the subject. Or, the process can provide a stress-free way of making students aware of their performance without confrontations.

We have used these sheets for a number of years and are very happy with the system. It is fast—the students do most of the listing and calculating, and explaining how to use the sheets takes very little time. Neither of us has had a student complain about a course grade since we started this system.

We believe the best time to find out that everything isn't "fine" is when there's still time to rectify the situation. And we think that the best "gift" instructors can give their students is this opportunity to regularly and realistically assess their own academic performance.

Tests and Exams

As with many other issues that we have dealt with in this volume, college teachers are often forced to rely on their own experiences *taking* tests and exams rather than being introduced to the existing body of research and thinking on these topics. In this subsection, we provide some of the basic references to improving tests and exams.

Clegg and Cashin offer a basic introduction to the multiple-choice items, their strengths and weaknesses, and recommendations for how and when to use them.

Improving Multiple-Choice Tests

Victoria L. Clegg and William E. Cashin

The tendency in course examinations is to pose the question “How much do you remember of what has been covered?” rather than “What can you do with what you have learned?”

—Dressel (1976, p. 208)

What Is a Multiple-Choice Item?

The multiple-choice item requires that students select the correct answer to a question from an array of alternative responses that are written by the instructor. All multiple-choice items have the same three elements: (1) an **item stem** that presents the problem, (2) the **correct or keyed option**, and (3) several **distractor** options, incorrect alternatives that are likely to be plausible to the student who has not completely mastered the learning being tested. Several variations of the standard multiple-choice item have been used in classroom tests. Some of these will be described later. Typically, multiple-choice items present the problem in one of two formats: the **complete question**, e.g., “What is the most frequently used type of test item in college-level examinations?” or the **incomplete statement**, e.g., “The most frequent type of test

item used in college-level examinations is ____.” The students are directed to select either the correct answer or the best answer from the list of options provided. In the **correct answer** form, the answer is correct beyond question or doubt while the others are definitely incorrect. In the **best answer** version, more than one option may be appropriate in varying degrees; however, it is essential that the keyed or “best” response be the one that competent experts would agree upon.

It may appear to be fairly simple to construct items in the multiple-choice format. Actually, the formatting is simple; it is constructing a meaningful and worthwhile item that is so difficult and time-consuming. “An ingenious and talented item writer can construct multiple-choice items that require not only the recall of knowledge but also the use of skills of comprehension, interpretation, application, analysis, or synthesis to arrive at the keyed answer” (Thorndike & Hagen, 1969, p. 103). How many of us who teach at colleges and universities would describe ourselves as “ingenious and talented” while we struggle to write effective multiple-choice items? Wilbert J. McKeachie (1986, p. 91) has said that “the greater your experience in their construction, the longer it takes per [multiple-choice] item to construct a

4. Multiple-choice items provide an **excellent basis for posttest discussion**, especially if the discussion includes why the distractors are wrong as well as why the correct answers are right.

Multiple-choice items also share many of the strengths of other selected response items, [e.g.], true-false, matching, etc.

5. Multiple-choice items can provide a **more comprehensive sample of subject material** because more questions can be asked.
6. Multiple-choice items adapt to a **wide range of content and difficulty levels**.
7. Multiple-choice items require relatively **less student time to answer**.
8. Multiple-choice items can be **easily and accurately scored** by a person or machine.

Limitations of Multiple-Choice Tests

Of course, multiple-choice items also have disadvantages.

1. Multiple-choice items are **open to misinterpretation** by students who read more into questions than was intended.
2. Multiple-choice items may **appear too picky** to students, especially when the options are well constructed.
3. Multiple-choice items, when written to assess higher levels of learning, require significant intellectual effort both in reading and in answering, **causing some students to be anxious**.

In addition, multiple-choice items share the limitations of other selected response items.

4. Multiple-choice items **deny demonstration of knowledge beyond the range of options provided**.
5. Multiple-choice items are **difficult to phrase** so that all students will have the same interpretation.
6. Multiple-choice items **take time and skill** to construct effectively.
7. Multiple-choice items are so easily constructed to assess basic factual knowledge that instructors **often fail to test higher levels of thinking**.
8. Multiple-choice items are ill suited to **assess affective or attitudinal learning** because they are easily “faked.”
9. Multiple-choice items **encourage guessing**—after all, one option is correct.

Recommendations

When Should Multiple-Choice Items Be Used?

Knowing the strengths and limitations of multiple-choice items can help instructors make better decisions about whether or not to use these items in particular testing situations. Use multiple-choice items for the following instructional goals:

1. When you wish to test the **breadth of student learning**. Multiple-choice items offer the opportunity to sample a greater breadth of learning than do questions that require a lot of student writing. Because they take considerably less time to answer, many more questions can be asked and so more content [can be] tested.
2. When you want to test a **variety of levels of learning**. Multiple-choice items are extraordinarily flexible in that they can

be used to assess the full range of Bloom et al.'s (1956) taxonomy. Do not discount multiple-choice when you want to evaluate abilities to think critically and solve problems effectively

3. When you have **many students who will be taking the test**, then multiple-choice tests are very efficient. If the class is very small in size, it usually is *not* worth the time it will take to construct an effective set of multiple-choice items. Carefully consider whether other item types will serve your testing purposes.
4. When you have **time to construct the test items**. Remember that effective multiple-choice items, which assess more than basic factual knowledge, require a great deal of time and effort to construct. If you do not have the time, another type of test will be a wiser choice.
5. **When time is limited for scoring**, then selected-response items are often the better choice. While it may have taken an hour to construct a multiple-choice item, it will take less than a second to score it.
6. When it is *not* important to determine how well the student can **formulate a correct or acceptable answer**. The answers are definitely provided in multiple-choice items. Even if the question requires critical thinking skills, it may be possible for a student to get the answer right because of clues in the options or by guessing. When it is important for students to formulate their own answers, multiple-choice will *not* do.

Required Preconditions

Before considering specific suggestions for writing multiple-choice items, there are a combination of abilities that, according to Alexander G. Wesman (1971), are necessary to write successful test items.

7. You must **have a thorough mastery of the subject matter** being tested. You must not only understand the implications of the facts and principles of a particular field, but you must also be aware of common fallacies and misconceptions.
8. You must **develop and use a set of educational objectives** to clearly guide your efforts to help students learn. Unless you have carefully considered what you want students to learn, you will not be able to evaluate their progress with any accuracy. This means that you must develop a test plan or table of specifications to guide your item writing. For the vast majority of tests, a two-dimensional table is sufficient. On one dimension, list the areas and subunits of the content you wish to test. On the second dimension, list the various levels of learning you wish to test, for example, understanding, application, and higher order cognitive objectives. You must also decide what proportion of the test you want to devote to each area of content and each level of learning. Finally, as you write the test items, you should keep a tally of how many items fall into each cell of your total plan to ensure that your test actually covers the learning as you originally intended. (See Box FG6.2.)

Box FG6.2: Levels of Learning

Topic	Understanding	Application	Higher Order
A	5%	10%	10%
B	5%	20%	10%
C	10%	20%	10%

According to this table of specifications (Box FG6.2), approximately 40 percent of the instruction time was spent on topics “B” and “C” at the application level and 20 percent at the higher order level. The test should reflect that proportion. (See Gronlund, 1985a, or Mehrens and Lehman, 1984, for further treatment of tables of specification.)

9. **Know the students** who will be taking the test in order to appropriately adjust the complexity and difficulty of the items. Sophomores in Introduction 101 may look the same semester after semester, but there are likely to be many differences in the educational backgrounds and intellectual abilities of the groups. Design your test so that the students can demonstrate their learning.
10. You must be a **master of written communication**, able to communicate with precision and simplicity, and you must use language that the students understand.

Constructing Multiple-Choice Items

The following recommendations for constructing multiple-choice items reflect the collective experience and wisdom of many authors. These recommendations are written in chronological order. Several works are listed in

the References and Further Readings section for those of you who wish to read more extensively.

11. **Spread the work across time.** It is unwise to wait until the night before an exam is scheduled to construct the test items. It is impossible to construct effective multiple-choice items in such a limited time. Not only do you need time to construct the items, you need an opportunity to review and revise. If you write a question or two after each class or on a weekly basis, the collection is more likely to be representative of your instruction.
12. **Use note cards for writing the items.** This makes it much easier to file according to your test plan, rearrange, rewrite, and discard items. Better yet, if you have access to a personal computer, use it.
13. Really concentrate on writing items to **evaluate higher levels of thinking.** Avoid the pitfall of writing items that test only memorization of basic factual knowledge. Many instructors (especially those who are writing the test questions just before the test) fall into this trap and pull their students in with them.
14. **Write the stem first.** The stem should present a single, definite problem as a question or an incomplete statement. The problem should be one of significance in the course.
15. **Concentrate on evaluating student ability to understand, apply, analyze, synthesize, and evaluate.** It is difficult to write questions that evaluate these higher cognitive levels; but if critical thinking is what you want students to do, you will have to test for it. Students have a tendency to study “what will be

on the test” and to study *only* what will be on the test.

16. **State the problem concisely but completely.** What the student is to answer must be obvious, and the student should be able to discern the problem *without* reading all of the options. A direct question usually does this more clearly than an incomplete statement. There are times, however, when the question is just too convoluted or confusing for easy interpretation; then the incomplete statement may be preferable, or perhaps an item type other than multiple-choice is more appropriate.
17. **Write the stem to include all the information essential to determining the problem** but omitting irrelevant material that merely serves as padding, unless the student’s determination of what is relevant is part of what you want to test.
18. **Avoid unnecessary repetition in the options** by including as much of the item as possible in the stem. This is especially important when using the incomplete statement format. Forcing students to reread a phrase several times wastes time they could put to better use when taking a test.
19. **State the problem or ask the question in a positive form.** The use of negatives can be confusing to even the most intelligent reader, and anxious students often completely miss little words like “not.” On those rare occasions when you decide that you must use negatives, use **boldface**, underlining, or CAPITAL letters. Do *not* use double negatives, e.g., negatives in both the stem and the options.
20. **Write the correct or best response after writing the stem.** Be certain that the best response is indeed *best*, that is, would be acknowledged as best by authorities in the field. State this response as *briefly as possible* and without ambiguities so that all knowledgeable students will read it with the same interpretation. Having colleagues or former students critique your questions for clarity before using them on a test can help to avoid such difficulties.
21. **Avoid making the correct option longer than the distractors.** Test-wise students are very aware of this fault and use this clue to choose the correct answer without knowing the correct answer. The emphasis on the keyed response being absolutely correct sometimes leads to wordiness, and instructors tend to spend much less time developing the distractors, which then tend to be shorter. Write the correct response and the distractors, and then compare the lengths. If correct answers are consistently longer (or shorter) as you write multiple-choice items, edit as necessary.
22. **Write the distractors after writing the correct option.** The effectiveness of multiple-choice items can be undermined by the sloppy preparation of the incorrect options. Designing distractors is actually quite challenging because these options must be wrong yet be plausible enough to attract the attention of students who do not know the material as well as they should.
23. **Make all distractors plausible responses.** Avoid writing poor alternatives just for the sake of having more options; they simply become throwaway options. The criterion is whether or not

the distractors test a discrimination that is important—if not, do not use it. Once in a while, a ridiculous option can relieve some of the tension that pervades a testing situation, but only once in a great while.

24. **Be sure that the distractors use words that ought to be familiar to the students.** Using highly technical language or the vocabulary of experts, terms that have not been used in class, forces students to choose correct answers without knowing the meaning of one or more of the options. If the students were not expected to learn the terms, do not include them in the options.
25. **Write distractors that are distinct from each other.** If all the distractors are too much alike, the test-wise student will use this clue to eliminate the group of look-alikes in favor of the dissimilar, correct response. Similar distractors may also indicate that the question should not be presented in the multiple-choice format. Avoid alternatives that overlap or include each other. This error is likely to be distracting to students who read carefully and know the material well, which can result in the more knowledgeable student being penalized by the instructor's lack of item-writing skills.
26. **Critique for general errors in style and format.** Delete any irrelevant clues that could lead a student to select the correct answer or eliminate one or more of the wrong options without knowing the material. Measuring the test-wiseness of the students is not the intent of the test.
27. **Be careful in using specific determiners** such as “all,” “never,” “always,” or other all-inclusive terms that are more likely to be found in incorrect options. Similarly, qualifiers such as “usually,” “sometimes,” and “maybe” are more likely to be found in the keyed item. However, sometimes the content permits using absolute specific determiners correctly and so can keep the test-wise student “honest”; e.g., “The president of the United States must always be at least 35 years old” is correct.
28. **Avoid grammatical inconsistencies** between the stem and the options. These are very useful clues for the student who is competent in syntax.
29. **Use “none of the above” as an option with caution.** Some faculty believe that the option “none of the above” should never be used in a multiple-choice item. This belief is correct for a “best answer” type item. (Nor should options like “all of the above” or “both A and B” be used in “best answer” items.) However, for “correct answer” items, where there definitely is a correct answer, the option “none of the above” may serve a useful purpose, especially for items requiring mathematical calculations or perhaps correct spelling or grammar in a language. Using “none of the above” can prevent correct answers because of guessing or [can] save students from spending an inordinate amount of time on a problem they cannot solve. To be effective, the option must occasionally be the keyed response; otherwise, the students will see it simply as a throw-away option.
30. **Check once more** to be certain that the correct options are not consistently longer than the alternatives.
31. **Arrange options in a logical order,** if one exists. Numerical answers should be

placed in numerical order and dates put in chronological order. Sometimes alphabetizing the options is appropriate.

Organizing the Layout of the Entire Test

Once the individual multiple-choice items are written, you must decide how to organize the groups of items on the test. If you are using several types of items on your exam, be sure to group all of the multiple-choice items together, etc.

32. **List options on separate lines**, arranged in a vertical column to clearly distinguish each option from the others. Printing the responses in tandem or arranging them across the page may save paper, but the result is difficult to read. You should not be testing reading skills.
33. **Use capital letters for the response options** if the student is to write the letter to indicate the selected answer. The handwritten, lowercase letters “a and d” and “c and e” can be difficult to distinguish when scoring.
34. **Check to see that the correct answers are distributed randomly among the possible option positions.** If you have had a tendency to choose one position over others, for example, “B,” it may become apparent to the test-wise student who seeks out such clues. If necessary, it is easy to rearrange the order of the options to correct this problem.

References and Further Readings

Those references below which are followed by an asterisk are standard texts on educational

measurement. Each has one or more chapters on multiple-choice and other selection items as well as chapters on other aspects of testing and grading.

- Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of education objectives, Handbook I: The cognitive domain*. New York: David McKay.
- Campbell, J. R. (1972). *In touch with students*. Columbia, SC: Kelly Press.
- Clegg, V. L., & Owens, R. E. (1984). *Tips for writing tests*. Manhattan, KS: Graduate Services and Publications.
- Dressel, R. L. (1976). *Handbook of academic evaluation*. San Francisco: Jossey-Bass.
- Ebel, R. L., & Frisbie, D. A. (1986). *Essentials of educational measurement* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Eble, K. E. (1976). *The craft of teaching: A guide to mastering the professor's art*. San Francisco: Jossey-Bass.
- Educational Testing Service. (1973). *Multiple choice questions: A closer look*. Princeton, NJ: Author.
- Fuhrmann, B. S., & Grasha, A. F. (1983). *A practical handbook for college teachers*. Boston: Little, Brown.
- Gronlund, N. E. (1985a). *Measurement and evaluation in teaching* (5th ed.). New York: Macmillan.
- Gronlund, N. E. (1985b). *Stating objectives for classroom instruction* (3rd ed.). New York: Macmillan.
- Lowman, J. (1984). *Mastering the techniques of teaching*. San Francisco: Jossey-Bass.
- McKeachie, W. J. (1986). *Teaching tips: A guidebook for the beginning teacher* (8th ed.). Lexington, MA: D. C. Heath.
- Mehrens, W. A., & Lehmann, I. J. (1984). *Measurement and evaluation in education and*

- psychology* (3rd ed.). New York: Holt, Rinehart & Winston.
- Milton, O. (1982). *Will that be on the final?* Springfield, IL: Charles C. Thomas.
- Milton, O., & Associates. (1978). *On college teaching: A guide to contemporary practices*. San Francisco: Jossey-Bass.
- Roid, G. H., & Haladyna, T. M. (1982). *A technology for test-item writing*. New York: Academic Press.
- Thorndike, R. L., & Hagen, E. (1969). *Measurement and evaluation in psychology and education*. New York: John Wiley.
- Wesman, A. G. (1971). Writing the test item. In R. L. Thorndike (Ed.), *Educational measurement*. Washington, DC: American Council on Education.

This piece parallels the information provided on multiple-choice questions but, here, focuses on essay tests—definitions of types, strengths, weaknesses, and applicability.

Improving Essay Tests

William E. Cashin

What Is an Essay Test?

Coffman (1971, p. 271) describes an essay test as “one or more essay questions administered to a group of students under standard conditions for the primary purpose of collecting evaluation data.” Their scoring requires expert judgment rather than the application of a clerical key. Administration under standard conditions distinguishes the essay test from the term paper or project report. However, I suggest that many of the recommendations concerning essay tests can also be applied, with appropriate adaptations, to term papers, project reports, oral exams, and other student products or processes used in assessing student achievement, including mathematical problems and artistic productions.

Essay questions are often divided into two types: extended response questions and restricted response questions.

Extended Response Questions. Other than stating the topic, extended response questions leave students free to determine the content and to organize the format of their answers. The students decide which facts are pertinent and

how to organize, synthesize, and evaluate them. Perhaps the classic example is “How I Spent My Summer Vacation.” Such questions are most appropriate when our objective is to test writing (composition) skills including conceptualization, organization, analysis, synthesis, and evaluation, giving the student maximum choice regarding topic.

Restricted Response Questions. These limit both the content and the form (e.g., describe vs. compare and contrast) that the student’s answer may take. Most writers agree that restricted response questions are the appropriate form when we wish to test content. All of the examples of essay questions which follow will be examples of restricted response questions unless otherwise stated.

Strengths of Essay Tests

Essay tests have a legitimate place in higher education because of the following strengths:

1. [They] can test complex learning outcomes not measurable by other means. An obvious example is the ability to express oneself in writing.

2. **[They] can test thought processes**, the students' ability to select, organize, and evaluate facts, ideas, etc., and their ability to apply, integrate, think critically, and solve problems. (Note: all of these can also be tested by appropriately designed multiple-choice items—see IDEA Paper No. 16, *Improving Multiple-Choice Tests*, Clegg and Cashin, 1986.)
3. **[They] require that students use their own writing skills**; the students must select the words, compose the sentences and paragraphs, organize the sequence of exposition, decide upon correct grammar and spelling, etc.
4. **[They] pose a more realistic task** than multiple-choice and other “objective” items. Most of life's questions and problems do not come in a multiple-choice format, and almost every occupation, including engineering, business, technical, and service jobs, requires people to communicate in sentences and paragraphs, if not in writing, at least orally.
5. **[They] cannot be answered correctly by simply recognizing the correct answer**; it is not possible to guess. (Students can bluff, however.)
6. **[They] can be constructed relatively quickly**. This advantage is short-lived because any time saved in constructing the test is lost when scoring it. All well-constructed tests require time and effort; the only choice is in when these will be expended.

Limitations of Essay Tests

The focus in this [piece] is on using essays for assessment. When essays are used as a learning

experience to provide the students an opportunity to exercise a skill and then to give them feedback about their achievement, the limitations described below are of less concern. However, as an assessment technique, essay tests have the following serious limitations.

1. **Only limited content can be sampled.** Therefore, essay tests are unreliable in assessing content. Because answering essay questions takes more time than answering “objective” items, less content can be tested. Most exams only sample a very small portion of the domain of content and skills to be learned. Therefore, when we rely solely on essay tests, differences in students' scores will to some extent reflect the “luck of the draw”—the questions you happened to include on the test—as well as reflect differences in the students' command of the entire domain of what you were trying to teach.
2. **[They] yield unreliable scores.** Not only have studies found differences in the grades assigned to essay questions by different scorers, but they have found differences for the same scorer grading the same question at different times. Thus, differences in student grades on an essay test may be due to *who* scored the question, or *when* it was scored, in addition to *what* the student knew or wrote.
3. **Scores can be influenced by the scorer's impression of the student**, e.g., general impression of the student: halo effect or test-to-test or item-to-item carryover—knowing how well the student did on the previous test or item. Obviously, multiple-choice tests do not have this limitation.

4. **Scores may be influenced by factors extraneous to the content being tested**, e.g., handwriting, writing skills, spelling, and grammar.
5. **Essay tests often provide the students with an opportunity to exercise *poor* writing skills.** When one considers the time pressure and anxiety connected with the typical essay test, it is surprising that the students do as well as they do. Most of the students' time in an essay test is spent physically writing. There is limited time to think, to organize creatively, to write a second draft, or [to] proofread.
6. **Essay tests are time-consuming to score.** Anyone who has ever graded essays needs no proof of this beyond his/her own experience.

Recommendations

These recommendations are divided into three sections: when should essay questions be used, constructing the test, and scoring the test.

When Should Essay Questions Be Used?

These recommendations are adapted from Ebel and Frisbie (1986).

1. **To test writing skills.** Obviously, the most appropriate way to test the students' ability to express themselves in writing is to have them write something (remembering that essay tests are less representative of day-to-day writing tasks than are papers, project reports, keeping a journal, etc.).
2. **To test a small group.** Despite all of the advantages of multiple-choice and other "objective" type items, when testing small groups of students, developing such items is not worth the effort. Short-answer questions, e.g., one to a few sentences identifying or defining questions, can be useful to serve in place of multiple-choice items.
3. **When the time to construct the test is more limited than the time to score it.** Testing is a teaching responsibility, so we have a professional obligation to plan ahead for it. However, constructing a makeup exam for one or a few students who were legitimately unable to take the regular exam would be an instance where the instructor would have limited time.
4. **When the instructor has more confidence in his or her ability as a critical reader than as an "objective" test constructor.** Granted that college teachers, like the rest of humanity, differ as individuals, nevertheless, I would suggest that college teachers should have in their repertoire the basic skills of their profession including the ability to construct reliable and valid tests, both "objective" and essay tests.
5. **To encourage students to explore attitudes** more than testing for cognitive achievement. This suggestion focuses more on teaching (helping the students learn) than on testing but fits into our broader approach suggesting that readers consider these recommendations not just for essay tests in the narrow sense but also for papers, reports, journals, etc. Furthermore, in a later IDEA Paper, it will be urged that testing, and all of our assessment techniques, should

be an integral part of our instructional design, not just something added on for evaluation, i.e., determining grades.

Constructing the Test

6. **Allow adequate time to construct essay questions.** Although a five-question essay test can be constructed faster than a 50-item multiple-choice test, writing an effective essay question takes thought and therefore time. One poorly designed essay question would have an effect similar to ten poor multiple-choice items.
7. **Limit the use of essay questions to learning outcomes that cannot be satisfactorily measured by “objective” items.** Given the serious limitations of essay tests, especially with respect to reliability, the recommendation is to use essay questions for assessment only when you have to. Especially, do not use essay questions to test facts or learning at the lower levels of Bloom’s taxonomy (Bloom et al., 1956). (For brief discussions of Bloom’s taxonomy, see Clegg and Cashin, 1986, and Gronlund, 1985b.)
8. **Design the essay question to test only one or a few specific instructional objectives per question.** This seems fairly clear; you must make explicit what you want to test (a necessary corollary is that you had to be clear about what you were trying to teach, i.e., expected the students to learn).

For example, the following is a poor essay question: “Why do animals migrate?” It might be better to ask: “Describe three hypotheses which might

explain why animals migrate south in the fall of the year.” This second version, however, points out that what is being tested basically is the students’ memory of what was in the lecture or text—not a recommended use of essay questions. (See Clarence H. Nelson’s chapter, “Evaluation in the Natural Sciences,” in Dressel and Associates, 1961, for ways in which the students’ understanding of these theories might be tested using “objective” items.)

A more appropriate example of a question testing a specific instructional objective, in this case a foreign language (Latin), is:

Read the above passage and decide whether it was written by a classical or patristic Latin writer. Support your position by identifying and explaining specific phrases or passages which illustrate the characteristic writing style. Also identify phrases, etc., which might support the opposing position.

The objective was to assess not simply students’ passive understanding of the elements which characterize the two different styles but also their ability to apply that knowledge in their reading. Of course, to do this the students also needed to have a certain proficiency in translating Latin. The instructor chose a passage unfamiliar to the students by an author whose writing contained elements of both styles (and a passage where the content did not serve as a clue), so it was possible for the students to make a case for either side. The primary point of the question was not whether the students correctly classified the author but how good an argument they could make for their positions and

how aware they were of the contrary evidence. I consider this question to be at least at the application level of Bloom's taxonomy.

9. **Give preference to focused questions that can be answered briefly.** When it fits your instructional objectives, several short essays will yield a more reliable score than fewer long questions. On the other hand, a short-answer question is less likely to permit the students to demonstrate complex mental processes. Also, if an instructional objective can be tested by a short essay, perhaps it can also be tested by a multiple-choice item.
10. **The question should clearly indicate the task(s) the students are to address** with respect to both content and process. On one history of philosophy exam, the students were given the following topic: "Locke: the key to Hume." While I applaud the creativity of the instructor, the question can be improved. For example:

Locke: the key to Hume. Discuss the influence of the philosophy of Locke on Hume's theory of knowledge.

or

Locke: the key to Hume. Discuss the similarities and differences in the philosophies of John Locke and David Hume with respect to: the origin and relation ideas, the nature of belief, (etc.)

Gronlund (1985a, p. 220) provides a list of 12 types of thought questions and sample item stems, e.g.:

Synthesizing: Describe a plan for _____.

Evaluating: Describe the strengths and weaknesses _____.

Hopkins and Stanley (1981, pp. 214-216) list 21 types of essay questions, e.g.:

Inferential thinking: Discuss whether the authors of this text are likely to use essay tests frequently in their measurement classes. Support your opinion with principles and recommendations given in the text.

One very helpful way to determine whether you have clearly specified the task is to give your essay questions to colleagues and see if they understand the questions. Also, ask them what instructional objective(s) they think you are trying to test with each question.

11. **Make explicit the approximate time or length for each question and/or the number of points.** This is especially important if the questions are not weighted equally. Therefore, we might add to the "Locke" question above:

(50 points, spend about 30 minutes, five pages on this question).

12. **Provide sufficient time** for the students to write the answer. See how long it takes you or a colleague to write an answer, then allow the student several times that amount of time. You do not want your tests basically to assess writing speed.
13. **Use novel questions;** otherwise, you are testing memory. Novelty can provide interest, and therefore motivation, for the students. One psychology professor teaching a Systems of Psychology course asked the students to imagine that they

were a rat in the lab of specific psychologists and then describe what might happen to them with respect to a number of experimental variables.

14. **Avoid optional questions**, i.e., letting the students choose which questions they will answer. The only advantage is student morale, and the reasons against providing the students with a choice are persuasive:

A. **The students are taking different tests.** Thus, there is no common basis for comparison, and the scoring becomes unreliable. It is almost impossible to write several essay questions which are of equal difficulty; the result is that different students are taking tests of varying difficulty, but you will grade these the same. This may penalize the “better” students because they may choose the more difficult (challenging) questions and so will not score as well as students who choose the easier questions.

B. **In real life, we usually cannot pick our problems.** In the world of work, at home, and in society, we are expected to address all of the major issues facing us, not the four out of five we feel most competent to handle.

Exception: There is one notable exception to the recommendation to avoid optional questions, and that is with extended response questions where you wish to test a skill. The most common example is assessing writing skills. Often students are given many topics and told to choose one to write on. The hope is that the list of topics will be broad enough to enable every student

to find a topic he or she knows about. Thus, differences in the final essays will not reflect differences in their knowledge of the content but will only reflect differences in writing skills. The same argument applies to a variety of other processes or skills: critical thinking, public speaking, artistic expression, etc. However, when your primary purpose is to test command of content, providing optional questions is not advised.

15. **Do not give the students a short list of essay questions to prepare before the test.** Although the intent in doing this is usually to help the students, the results are often undesirable. Such essay tests may simply test the students’ ability to memorize someone else’s thinking. If the list is short, it may encourage the students not to study all of the content. This can be exacerbated if the students know they will only have to answer, say, two out of four questions. In such cases, they may simply omit studying two of the questions.
16. **Prepare the students to take the test.** Consider whether part or all of a class session might profitably be spent letting the students respond to a typical sample of your essay questions and then discussing what you look for when scoring them. Using a question from a previous year where you kept samples of “A” papers, “B” papers, etc., could be even more helpful to the students.

Scoring Essay Tests

The following recommendations are made to enhance the reliability and validity of scoring essay tests. The goal is to ensure as much as

possible that differences in students' essay scores reflect differences in their respective achievement and nothing else.

17. **Fit the scoring approach to the type of essay question.** Two approaches are described in the literature: analytical (point-score) and global (holistic). (See Mehrens and Lehmann 1984, pp. 114-116, for a longer discussion.)

Analytic (point-score) Method. This method is recommended for restricted response questions. The ideal or model answer is broken down into several specific points regarding content. A specific subtotal point value is assigned to each. When reading the exam, you need to decide how much of each maximum subtotal you judge that student's answer to have earned.

Global (holistic) Method. This is recommended for extended response questions. The rater reads the entire essay and makes an overall judgment about how successfully the student has covered everything that was expected in the answer and assigns the paper to a category (grade). Generally, five to nine categories are sufficient. Ideally, all of the essays should be read quickly and sorted into five to nine piles, then each pile reread to check that every essay has been accurately (fairly) assigned to that pile which will be given a specific score or letter grade.

18. When using analytical scoring for restricted response questions, **outline the model (ideal or acceptable) answer before you begin to read the essays.** The specificity of the answer, however, may

vary with the question. It is recommended that you read a sample of the actual essays before you begin to assign scores. Ideally, you should read all of the essays quickly to check (and perhaps modify) your model answer, then reread all of them to assign scores. In practice, this often is not feasible, but remember that your goal is to have the students' scores reflect achievement on a common task. Also, you want to **use realistic standards**, [and] reading several or all of the essays before assigning scores helps achieve this; having a colleague read your model answer would also help.

Ebel and Frisbie (1986, p. 127) suggest that some of the common reasons that students do not obtain maximum credit are:

- a. answer includes incorrect statements;
- b. relevant material is omitted;
- c. irrelevant material is included;
- d. student commits errors in logic, reaches unsound conclusion(s);
- e. student writes unclear answer, often because of poor writing skills for poor handwriting); and
- f. student commits flagrant errors in grammar, spelling, etc.

Your model answer should provide guidelines for you to make an accurate assessment of at least the first three.

19. **Keep the identity of the student anonymous.** Your score should reflect your assessment of the adequacy of the answer and not anything else you know about the student.
20. **Score one answer at a time** (in a single, uninterrupted session if feasible). Experience suggests that most of us do not

read all of the essays first before rereading them to assign scores. Reading the answers to a single question (or mathematical problem) in a single sitting is at least a partial help for ensuring that the students are being scored based upon a common set of criteria.

21. **Shuffle the exams after scoring each question when the test consists of several items.** This is suggested for a couple of reasons. First, if you have not followed some of the recommendations suggested above about clearly defining our model answer before assigning grades, there is a tendency to change expectations while reading the answers; e.g., you may become more and more depressed as you read the students' answers because no one has gotten it "right" and so tend to score essays read later more leniently. Second, if you keep the exams in the same order, you are more likely to be influenced by how well (or poorly) that student did on the previous question. Related to this, it is helpful to have the students start each answer on a new page so that you cannot see the score on the previous question.
22. **Decide beforehand how you will handle grammar, spelling, handwriting, etc.** Obviously, if you cannot read the student's writing, you cannot score his or her answer. If the writing (composition) is so unclear that you do not understand, you should not give credit. However, there are many instances where the substance of an answer is understandable but there are obvious errors in spelling, grammar, etc. Sometimes students will take the position that these things should only influence the grade in English courses. However, college-educated people should be able to write clear, grammatical, correctly spelled English. Even engineers and businesspeople write reports and letters. You, however, must decide how much these things will count, especially in light of the time pressures typical in essay testing. You should also inform the students of your grading criteria well before they take any test.
23. **When feasible, use multiple readings and/or readers.** For really important essays like undergraduate theses, two separate readings and scores are desirable, with the grade being the average of the two scores. Better yet, having two separate readers score the test is desirable. This is often done with essays used for placement in English composition, senior comprehensive exams, and the like. It is an effective, but time-consuming, way to improve reliability.
24. **Provide extensive comments.** Although one of the purposes of an essay test, paper, etc., is to assess the students' past learning, they can and should also be used to help students continue to learn. Providing extensive comments, not just a grade, is a effective way to do this. One practice is to have the students write on every other page of the blue-book, leaving the opposite page for instructor comments. Since we can talk faster than we can write, some instructors, e.g., those teaching writing, have each student purchase an audiotape and the instructor records extensive comments on that tape. Because of time pressures, comments are likely to be brief, but be sure that they are at least clear. Again, occasionally checking with

a colleague can be very informative. If you use a code, tell the students; better yet, put it in writing. One student commented after an entire semester of receiving essays with check marks that he did not know whether the check meant good, or bad, or important, or what.

25. **Consider keeping a test file on your essay questions.** Over time, you can develop a collection of essay questions (math problems, paper assignments, etc.) for specific instructional objectives. Your file should include your instructional objective(s), the question (problem) itself, any improvements that seem appropriate based upon past use, *and* a record of how well the students performed on the question and any comments you have. Ideally, keep samples of “A” answers, “B” answers, etc. There is no sense in reinventing the wheel every year, especially if it is your own “wheel.”

References

- Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of education objectives*, Handbook I: *The cognitive domain*. New York: David McKay.
- Clegg, V. L., & Cashin, W. E. (1986). *Improving multiple-choice tests* (IDEA Paper No. 16). Manhattan: Kansas State University, Center for Faculty Evaluation and Development.
- Coffman, W. E. (1971). Essay examinations. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 271-302). Washington, DC: American Council on Education.
- Dressel, P. E., & Associates. (1961). *Evaluation in higher education*. Boston: Houghton Mifflin. (Despite its 1961 publication date, this is one of a few books that give extensive treatment to testing approaches for several different academic fields at the college level, e.g., natural sciences, social sciences, humanities)
- Ebel, R. L., & Frisbie, D. A. (1986). *Essentials of educational measurement* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Gronlund, N. E. (1985a). *Measurement and evaluation in teaching* (5th ed.). New York: Macmillan.
- Gronlund, E. E. (1985b). *Stating objectives for classroom instruction* (3rd ed.). New York: Macmillan.
- Hopkins, K. D., & Stanley, J. C. (1981). *Educational and psychological measurement and evaluation* (6th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Mehrens, W. A., & Lehmann, I. J. (1984). *Measurement and evaluation in education and psychology* (3rd ed.). New York: Holt, Rinehart & Winston.

This brief list from the faculty at California State University, San Bernardino, suggests a list of considerations in grading essays written by students.

A Guide to Evaluating Essays

*Developed by the Faculty at
California State University, San Bernardino*

I. Substance

- A. Assignment should show comprehension of the subject matter.
- B. Response should demonstrate an awareness of the significance of issues or problems posed by the assignment.
- C. Response should include development of important implications of the assignment.
- D. Treatment should be comprehensive, developing all parts of the assignment.

II. Focus

- A. Answers should address the questions or problems posed.
- B. Superfluous materials should be avoided.

III. Evidence

- A. Information should be accurate.
- B. Opinions should be adequately supported; theory should be backed by specific and relevant examples or illustrations.
- C. Sources used should be identified and documented.
- D. Evidence should lead to conclusions.

IV. Organization

- A. Papers should have a recognizable pattern of organization.
- B. Each paragraph should be coherent.
- C. Important points should be emphasized.
- D. Transitions should be clear.

V. Style

- A. Treatment should be appropriate to subject and audience.
- B. Language should be accurate and precise.
- C. Sentences should be clear and effective.
- D. Sentence structure should be varied.
- E. Tone should be consistent.

VI. Mechanics

- A. Sentences should be complete.
- B. Pronoun reference should be clear.
- C. Verb forms should be correct.
- D. Punctuation should be accurate.
- E. Spelling should be correct.
- F. Documentation should follow a prescribed style manual.

In this excerpt, Takata offers a different way in which to think about giving essay questions that rely less on guesswork and test more of whether the students understand basic concepts and logic.

How the Guided Essay Works

Susan R. Takata

The guided essay provides students with an organizational structure for their essay responses. One of the major problems of the traditional essay examination is that it encourages a “wandering,” unfocused response. The guided essay form, however, prompts students to tell “what’s next”; it asks them to develop their ideas by elaborating on the related subtopic statements, offering reasons and providing examples to illustrate points. Because the students have received the basic outline, it is easier to identify the logic and reasoning they use. The guided essay form provides space for an introduction, for five topic statements, and for a conclusion. The guided essay is basically an outline, providing students with the organizational structure for their essays. Jones and Faulkner (1971:111) explain that parallel items (or ideas) are indented equally and that the divisions of the outline are mutually exclusive.

The introduction is an explanation of the subject. What will be the focus of this essay? The essay should contain a clear indication of purpose, not too broad and not too narrow. Topic Statements 1 through 5 are the major points in the student’s essay: topic and parallel topics, statement and related actions, and generalizations followed by support. Each topic

statement may be regarded as a paragraph. Subtopic statements are in-depth explanations or further developments of parallel ideas relating to the topic statement. These statements might be considered expansions or amplifications of the topic statement (i.e., explaining why this point is important to the argument and providing illustrations from authoritative materials from the course or outside the course [that] strengthen the point). The conclusion consists of a summary, a final generalization, and concluding remarks to end the essay. (To illustrate how the guided essay form works, see the Appendix. The blank form given to students includes only what is printed in bold type in the Appendix.) The guided essay form can be adapted to meet a variety of writing needs, not [just] essay examinations.

Appendix: Guided Essay Form with a Sample Essay

Question: How would each sociological perspective—the functionalist, conflict theorist, and interactionist—examine the problem of “career floundering?” Which

perspective do you agree with the most? Explain why. Provide examples.

1. Introduction [explanation of the subject].

Each of the three theoretical perspectives—the functionalist, conflict theorist, and interactionist—would examine the problem of career floundering by utilizing different research methods. The following will show the three perspectives examining career floundering and our agreement with the interactionist research method.

2. Topic Statement #1. Career floundering is the time between college graduation and initial job satisfaction.

A. Subtopic Statement A

Career floundering is a natural part of the job search process. You accept different positions to progress to job satisfaction professionally and economically.

1. Development of Subtopic A [include explanations of why and supporting examples]. You accept short-term careers. For example, you are a paralegal and discover you enjoy research. You transfer to the criminological research department. You eventually move to the DA's [district attorney's] office and become a prosecuting attorney.

B. Subtopic Statement B

Career floundering ends when a career is targeted.

1. Development of Subtopic B [include explanations of why and supporting examples]. You must know the what element—accounting, teaching, selling—and then target the “where” element—bank, school, store. You use self-analysis, past job experiences, and goal setting to prevent prolonged career floundering.

3. Topic Statement #2. Functionalists would examine career floundering scientifically, much as a biologist examines and explains body organs and their function in the body. The functionalist focuses on major interdependent parts—the family, religion, education, the economy.

A. Subtopic Statement A

The functionalist uses generalization, scientific objectification, quantification, and experimentation.

1. Development of Subtopic A [include explanations of why and supporting examples]. To maintain social order through consensus, a functionalist generalizes that all college graduates must be employed to support the economy. The functionalist would use quantification to target specific job markets with openings.

B. Subtopic Statement B

The ultimate goal is to predict the job market and control the economy.

1. Development of Subtopic B [include explanations of why and supporting examples]. The functionalist would use data gathered to predict job openings and encourage people to limit degrees in saturated fields. The functionalist intends to match the supply of students with job market demands.

4. Topic Statement #3. The conflict theorist uses the medical approach to examine the problem of career floundering through the use of diagnosis, prescriptive treatment, and evaluation.

A. Subtopic Statement A

The conflict theorist places the problem of career floundering in a battlefield or arena and approaches the problem from

the point of view of dissensus or disagreement.

1. **Development of Subtopic A [include explanations of why and supporting examples].** The conflict theorist theorizes that the problem is a power struggle, with the “haves” limiting job options for the “have-nots.” An example would be the trend to temporary job services replacing permanent job openings.

B. Subtopic Statement B

The ultimate goal of the conflict theorist is to provide everyone with a job and equal pay to achieve social justice.

1. **Development of Subtopic B [include explanations of why and supporting examples].** The conflict theorist insists that the powerful exploit the masses. The conflict theorist contends that career floundering is not a personal problem but a political one. An example is the migration of garment workers’ jobs to foreign countries for substandard wages.
5. **Topic Statement #4.** The interactionist views career floundering as a natural process to job satisfaction. The interactionist uses field and participant observation and interviews to research career floundering.

A Subtopic Statement A

The interactionist views career floundering from a micro or small-scale aspect.

1. **Development of Subtopic A [include explanations of why and supporting examples].** The interactionist strives to appreciate and understand each phase of the quest for job satisfaction, as an art critic appreciates different artists’

concepts. An interactionist would appreciate each teacher’s approaches to teaching different levels of education.

B. Subtopic Statement B

The interactionist would track the career changes, possibly participate in career floundering, and appreciate the trial-and-error procedure.

1. **Development of Subtopic B [include explanations of why and supporting examples].** The interactionist is interested in “how people make sense of their world.” The interactionist values the process of achieving satisfaction, such as the stages a sales clerk would advance through to own her own business: sales clerk, assistant manager, manager, buyer, owner/operator.
6. **Topic Statement #5.** We agree with the interactionist view as that of an art critic in examining career floundering because it is an individual’s decision.

A. Subtopic Statement A

Job satisfaction and stability are difficult to predict at a macro level.

1. **Development of Subtopic A [include explanations of why and supporting examples].** An individual’s intersubjectivity—the day-to-day social interactions—determines job satisfaction. An example is an executive who sacrifices a promotion to spend more time with his family.

B. Subtopic Statement B

The method of examining career floundering is best studied by an interactionist because it is a micro issue.

1. **Development of Subtopic B [include explanations of why and supporting examples].** To understand and appreci-

ate job change, the best tools to use are observation and interview. We interview a nurse who accepts a part-time position because she has a new baby and her husband will provide evening child care. She achieves job satisfaction by remaining in her field, and she enjoys home life.

7. **Conclusion [summary and closing remarks].** In conclusion, the functionalist and the conflict theorist examine career floundering at the macro level—how it

affects society. The interactionist examines career floundering at the micro level—how it affects an individual's daily life.

Reference

- Jones, Alexander E. and Claude W. Faulkner. 1971. *Writing Good Prose*. New York: Scribner.

In another innovative method of testing, Stearns draws from research on small groups to develop a collaborative method of testing.

Collaborative Exams as Learning Tools

Susan A. Stearns

Based on small group theory and practice, the following are suggestions for successful group exams. Because groups should not be given simple tasks to do (Shaw 1981), exam questions should be complex enough to engage students in healthy discussion. Use the same exam at the individual and group levels, but it should be difficult for an individual to do. Between the typical high score on the group exam and low score on the individual [exam], overall exam scores balance out.

Because groups take longer than individuals to complete a task (Beebe and Masterson 1994), there must be adequate time to complete both the individual and group exams. In practical terms for 80- to 90-minute classes, I have found twenty multiple choice questions to be enough for all groups to finish both the individual and group exams without rushing any particular student or group. Further, I suggest that you encourage heterogeneity in terms of gender, nationality, ethnicity, ages (nontraditional and traditional students), and ability (if known to the instructor) because the vast majority of the research supports the benefits of heterogeneous over homogeneous groups (Shaw 1981).

Place five to six students in a group. Shaw (1981) discusses group size by emphasizing the following:

As the size of the group increases, the amount of time available for each member to participate in the group's activities decreases; the larger the group, the less opportunity each person has to participate in discussion, to express his or her opinions, etc. (169)

I have found the above to be true when one additional aspect is remembered—heterogeneity. If there are too few students in the group, it typically fails to capture the full potential of heterogeneity.

Groups of three have occurred due to students dropping the course. These groups have not been as successful on their group exam scores. This is due to the fact that these groups have a minimal level of heterogeneity. With groups of four, I have had mixed results; thus I do not recommend their use. And with groups of seven or more, while they achieve high scores on the exam, they are not as satisfied with the process. From their comments, this appears to be because each student participates less than in smaller groups. Thus, placing five to six students per group usually allows for hetero-

geneity and enough time for each student to participate.

Have the students take the individual exam before taking the same exam with their group. When I began this technique, I used only a group exam. Students initially liked the concept, but after the first exam there were a number of complaints about students who came to the group unprepared. Comments suggested that some students, realizing that others desiring high grades would come prepared enough for the entire group, felt unmotivated. There was a need for an individual level of responsibility. Therefore, I began having students take the same exam at both an individual and a group level. Balance the grading between individual and group scores. I allot two-thirds to individual exams and one-third to group exams (e.g., 100 points for their individual score and 50 points for their group [score]). When students are individually responsible for the majority of their exam scores, they come better prepared for the exam.

Finally, place students in a specific group for the entire quarter. They learn each other's strengths and weaknesses and therefore how to study better by compensating for their own and each other's weaknesses. They have a group within the class whom they learn to rely on for

additional information (also noted by Johnson and Johnson 1982) and social support (Lazar 1995). And they do not waste class time with introductions each time the group gets together. Because they are in one group for the entire quarter, I allow students to remove unproductive members from their groups after first attempting to negotiate with them over their level of work (see Steams 1995 for a thorough explanation of this process). Also, I use these same groups for other group activities.

References

- Beebe, S. A., and J. T. Masterson. 1994. *Communicating in small groups: Principles and practices*. New York: HarperCollins.
- Johnson, D. W., and R. Johnson. 1982. Healthy peer relationships: A necessity, not a luxury. In *Structuring cooperative learning: The 1982 handbook*, ed. P. Roy, 45-82. Minneapolis, MN: Interaction Book.
- Lazar, A. M. 1995. Who is studying in groups and why? Peer collaboration outside the classroom. *College Teaching* 43 (2): 61-65.
- Shaw, M. E. 1981. *Group dynamics: The psychology of small group behavior*. 3rd ed. New York: McGraw-Hill.

This list of “never’s” on testing and grading draws from McKeachie’s Teaching Tips but is fleshed out by Lowman in his book, Mastering the Techniques of Teaching.

A List of “Never’s” in Testing and Grading

Wilbert McKeachie

Evaluating Student Performance: Testing and Grading

1. Never announce that one type of exam will be given, change your mind, and distribute something different.
2. Never spend the first fifteen minutes of a one-hour exam making oral changes in the printed questions or instructions,
3. Never include two right answers on multiple-choice items.
4. Never fail to finish reading essays and cover up your failure by making a brief comment on the last page of each paper to make it appear to have been read.
5. Never give everyone the same grade on essays whether you have read them or not.
6. Never decide after the first exam (or paper) which students are “A” students, “B” students, or “C” students and sort subsequent work into appropriate piles before reading it to verify that initial impression.
7. Never fail to give an indication during a course of the grade that a student is likely to receive.
8. When a student asks about a grade, never say that he or she actually earned a grade higher than the one given but that the grade was lowered because the supervisor’s, the department’s, or the school’s guidelines indicated that there were too many “A’s,” “B’s,” or whatever.
9. Never tell students that grade distribution was really quite arbitrary and that you could have just as easily called the “C’s” “B’s” or the “B’s” “A’s.”
10. Never give a student suspected of cheating a lower grade or an “F” without telling the student or the honor council.

Writing Assignments and Research Papers

Like exams, writing assignments have been cast as both the “solution” to testing and as yet another rote exercise unless sufficient attention is given to continual revision and rewriting. On this topic, those in composition have compiled an impressive body of work on techniques and methods. In this subsection, we offer a brief introduction to this literature.

Seventeen Suggestions for Making and Presenting Writing Assignments

Andrew Moss and Carol Holder

Announcing the Assignment

1. Announce the due date of major writing assignments in the course syllabus.
2. Prepare a handout for each assignment and distribute it when you feel students should begin work on the paper.
3. In the handout, describe the topic and specify the type of paper you expect (e.g., report, term paper with certain format for documenting sources, essay, one paragraph), the due date, and the approximate length.
4. Break up a large assignment into smaller steps, especially if your students have had little experience with this type of writing.
5. Specify the audience for each assignment. Doing so will help the students make decisions about the development of ideas, the language of the paper, and matters of organization and paragraphing.
6. Provide students with information as to how the paper will be graded and to what extent the final course grade will reflect their performance on this assignment. Advise the students (preferably in the handout) that good writing will be rewarded

and poor writing penalized—that organization, paragraphing, clarity of sentence structure and wording, spelling, punctuation, documentation format, etc., will affect the grade. If you have developed a system for grading papers, share it with the students when you distribute the assignment.

Improving the Assignment

7. Give the finished assignment to another to read to check for clarity in your instructions.
8. Occasionally do an assignment yourself. You may discover changes that need to be made, and because you will have a clearer notion of the steps required in the completion of the assignment, you will be better able to break up the assignment into smaller steps or to advise the students on the *process* of doing the assignment.
9. Keep notes on the success and pitfalls of each assignment you develop and use. Such comments will help you modify the assignment the next time you want to use it. Save a good paper to distribute as a model in the future.

Types of Writing for Different Purposes

10. Experiment with different types of writing assignments suitable for your course. In addition to term papers and long reports, try assigning essays, critical summaries of assigned readings, letters, journals, and other short, more frequent, not necessarily graded writings. (Writing is most often used to *test* students' knowledge of a subject. Try instead to develop assignments that help students *learn* more about a subject.)
11. Use five-minute writings in class in place of multiple-choice type quizzes to focus attention on concepts from readings or previous lectures.
12. Giving several shorter assignments, in or out of class, early in the quarter will allow you to identify students with serious problems and direct them to other courses or a tutoring center for help with their writing. Also, early writing assignments give all students an opportunity to benefit from your comments on and evaluation of their writing before they tackle a longer paper due toward the end of the quarter. Comments made on papers after final exams often go unread [and] unseen and thus do not help the students write better on future assignments.

Presenting the Assignment to the Class

13. Take ten or fifteen minutes of class time to discuss the assignment and answer questions when you distribute the handout. Speak to problems the students can expect to encounter.
14. Early in the quarter or with each assignment, give the students a checklist to re-

inforce what you consider important in completing the assignment successfully. Another option would be to recommend a handbook, term paper manual, or other guide to writing in your department or discipline.

15. Students appreciate seeing a model of good writing of the type you will expect. Save papers from your students or set up a department file of model papers of various forms for this purpose.
16. After presenting the assignment, take some class time for small group activity to encourage discussion and generation of ideas for the paper. Also, students could present rough and final drafts to a small group for feedback from other students. Experiment with group writing exercises so that the weaker students have a chance to learn from the stronger students in a cooperative instead of competitive environment.

Collaboration of First-Class Minds

17. Find others in your department or school who also are trying to use writing more effectively in their courses. Share your experiences with writing assignments you have developed. Faculty who teach sections of the same course can pool ideas about using writing to help students learn more about the subject matter of the course.

Checklist for Evaluating Assignments

1. What is the purpose of the assignment?
2. Is the task clearly and succinctly described?

3. What verbal or conceptual abilities does the assignment ask the student to use or develop?
4. Does the assignment involve the whole communicating person (listening, speaking, reading, writing)? If not, could the assignment be revised to call all of these abilities into play?
5. Will the students have a clear idea of how their performance will be evaluated?
6. In what way does the assignment relate to preceding and ensuing course assignments in developing students' skills sequentially?
7. Can the assignment be revised to reflect more fully the course's aims in promoting the mastery of particular knowledge or the development of specific skills?

Like much contemporary research on the discrepant attitudes among different stakeholders, Sheirer stakes out the differing perspectives of students and teachers regarding research papers.

Research Papers Revisited

John Sheirer

What most teachers think when they say “research paper”:

- A meaningful essay that explores or explains an exciting, specific topic
- A terrific learning experience that reinforces classroom instruction about a topic
- Proper documentation that follows a simple set of guidelines
- Discipline-specific language and thinking rendered in written form
- Analysis and synthesis of various readily available sources

What most students think when they hear “research paper”:

- Boring topics, dull readings that only interest teachers and are hard to write
- Probably won’t do well based on previous experiences
- Arbitrary documentation systems that are impossible to get right
- Good writing equals only proper documentation, punctuation, and grammar

- Hours of digging in the library for obscure, incomprehensible information

Research paper alternatives that accomplish what teachers want but don’t intimidate students:

- Research reviews/responses: written analyses of specific outside sources that are either provided by the instructor or found by the student
- Annotated bibliographies: summaries of outside sources with bibliographic references
- Fact sheets: forms designed by the instructor calling for specific written information about a topic
- Internet searches: overviews of the results of Internet searches on specific topics
- Interviews: questions and answers from exploratory conversations between the student and an expert in the field
- Surveys: reports on answers to questions asked of groups of people (other stu-

- dents, people at the student's home or job or in the community)
- Research collages: collections of cut-and-pasted text, photos, artwork, and graphics from outside sources arranged by the student in a meaningful way
 - Research dialogues: created “conversations” between two aspects or figures important in the course, with reference to the sources of the ideas used in the dialogues

Course Closure

Often, the end of the course receives very little attention as last-minute details, attempts to wrap up or catch up, and student panic often take precedence. However, in this subsection, suggestions on the end of the course are given proper attention.

Below, Maier and Panitz summarize a set of techniques and tips used by faculty at the end of a course.

End on a High Note

Mark H. Maier and Ted Panitz

In the frenzy of final exams, the end of a course often is frustrating (Lowman, 1984, 40-44). Professors and students alike are aware primarily of what has *not* been learned. It is the infrequent student who turns in the final examination with a feeling of pride, and it is usually the teacher's role to explain deficiencies in the final product to justify a grade. More effective course endings will help students avoid a sense of inadequacy and realize how much has been accomplished.

There are several ways in which we can be creative with final examinations so that students document explicitly how they have changed their thinking during the course. Duffy and Jones (1995, 204-205) recommend the use of portfolios to document students' progress; they help students see learning as a process of rethinking and revising. They (210) also suggest that instructors use the syllabus as a tool for review. Did the course fulfill the stated goals and objectives?

Saleh A. Ebrahim (1995) asks students to create their own flow chart in the final exam, graphing the relations between concepts learned in the course. This task encourages students to review and synthesize the material. As part of the final evaluation, Karen McComas (1996) asks students to annotate their portfo-

lios, describing how their achievement compares with the course objectives.

Students might also assess goals they set for themselves at the beginning compared with what they have accomplished. Gary Wagenheim (1993) requires students to write these goals early in the course on the reverse side of a name card that they keep in front of them throughout the course. In addition to aiding the learning of names, the cards remind students of the goals they set. At the end of a course, Ted Panitz (1996) asks students, "Has your approach to math changed during this course compared to previous courses?" "What would you do differently if you had a chance to do this all over again?" His goal is to help students think about how they will approach their subsequent courses and to encourage them to make needed changes.

Dean Mancina (1995) uses the ideas of David B. Ellis (1994) in which, at the end of a course, students write a letter on three-part paper completing the sentence, "I am becoming a student who _____" in ten different ways. The student keeps one copy, the instructor keeps one, and the third is mailed to the student so that it arrives just before the start of the next semester. In this way, students receive a "reminder of the skills they developed and com-

mitments to change they made the prior semester.”

Every course of study leaves some questions unanswered. Thus, self-awareness may help students. Randal Parker (1995) models for students the self-aware learner who makes choices about what to study—we can’t do everything—and what we sacrifice in not learning a particular concept. He then asks students to reflect on the benefits and costs of *not* understanding a particular part of the course. We can help students celebrate what they have mastered and specify what it is they will learn later.

On the other hand, if students have worked in small groups for substantial periods of time, a more formal structure may be needed. Teachers could use one of the following activities:

1. Wagenheim and Gemmill (1994) ask students to write a separate letter to each group member, using sentence completion stems as a guide: “The way I experienced your behavior in the group _____” and “What I personally feel you contributed to the group _____.” The authors recommend that the letter writers “circle those things that you feel are true about you” so that students recognize the importance of projection. They point out, “When we own our feedback (e.g., ‘This is true about me and maybe it is true about you’), we usually decrease interpersonal defensiveness and increase our learning.”
2. George Jacobs (1996) adds a twist to such letters, recommending that they be put in the form of a “letter of reference” for other group members to take to their next groups.
3. Ted Panitz (1996) makes the final exam a teaching tool by asking students to take it in two stages, first individually and

then with help from other class members. He has seen students make breakthroughs in understanding material they missed during the semester. This method of examination assesses how well students work together to solve problems, a course goal omitted in a traditional format.

4. Henry Maier (1995) suggests a technique in which the instructor describes what he or she learned, including evaluation of teaching and also new insights that he or she gained about the subject. In our experience, a semester never goes by that we do not better understand something we thought we had mastered.

Afterward

In our last formal meeting with students, we might discuss explicitly our future contacts with them. Teachers extending an invitation for students to stop for a visit later could explain that usually we do not receive enough feedback from students. Innovative strategies are available for maintaining contact with students.

George M. Jacobs (1996) asks students to write a letter at least three months after the course—but no more than nine months later—telling him “one thing they learned which they have used.”

Carol Haussermann (1995) creates a connection between students by asking them to write a letter to a student who will take the course the next semester. It provides a general introduction to the course as well as student strategies that were worthwhile and those that caused problems. The letters are sealed and then delivered at random and individually to students at the start of the next semester.

Haussermann reports that students take the activity seriously and [that] the recipients appreciate receiving a student's point of view. Our experience with these letters suggests that students need our prompting to include a discussion of course content in addition to tips on how to obtain a good grade.

The techniques described here have not entirely overcome the challenges we encounter at the end of our classes. Experience has socialized us and our students into teacher-focused endings—we are still trying to cover content—while students remain passive, knowing that they will soon be leaving. It takes time for students to feel comfortable enough to tell a teacher what they do not understand.

We recommend that teachers experiment first with classroom assessment and think-pair-share, the easiest activities. However, these are successful only to the extent that teachers take them seriously. For example, in classroom assessment, we need to be willing to read student input and to respond to it at the next class meeting. In think-pair-share, we need to keep students accountable, perhaps by calling on random pairs or by asking students to write individual answers after pair work. More ambitious activities such as letters to next term's students require significant class time and considerable trust in one's students. However, the benefit is that students reflect on how they learned, which demonstrably improves retention.

The varied techniques in this [piece] show that others have struggled with the problem of class and course endings. Nonetheless, the issue remains a surprising lacuna in our knowledge about classroom teaching. We need additional thought and research to define precisely our goals in endings and to identify which activities

work well. We invite readers to share their suggestions with us.

References

- Duffy, D. K., and J. W. Jones. 1995. *Teaching within the rhythms of the semester*. San Francisco: Jossey-Bass.
- Ebrahim, S. A. 1995. altlearn posting, Pennsylvania State University, December 16.
- Ellis, D. B. 1994. *Becoming a master student*. Boston: Houghton Mifflin.
- Haussermann, C. 1995. altlearn posting, physical education. Dana College, December 11.
- Jacobs, G. M. 1996. CL posting, SEAMO Regional Language Centre, Singapore, January 1.
- Lowman, J. 1984. *Mastering the techniques of teaching*. San Francisco: Jossey-Bass.
- McComas, K. 1996. altlearn posting, communications, Marhsall University, January 3.
- Maier, H. 1995. Personal communication, University of Washington.
- Mancina, D. 1995. CL posting, GoldenWest College, December 9.
- Panitz, T. 1996. altlearn posting, mathematics, Cape Cod Community College, January 1.
- Parker, R. 1995. Personal communication, Glendale College.
- Wagenheim, G. 1993. *Feedback exchange: Managing group closure*. Ogden, UT: Weber State College, International Society for Exploring Teaching Alternatives.
- Wagenheim, G., and G. Gemmill. 1994. Feedback exchange: Managing group closure. *Journal of Management Education* 18: 265-69.

Like the article before it, this piece represents a collection of examples collected by the editors of The Teaching Professor on how to end a course in an interesting way.

How to End Courses with a Bang

Editors' Article, The Teaching Professor

**Sandra Bassendowski, Wascana
Institute (Saskatchewan):**

I use a strategy I call “From Kaleidoscopes to Collages.” I ask learners to bring several magazines (that they are willing to cut up), glue or tape, and scissors on the last day of class.

When I introduce the strategy, I encourage learners to think about the course content as pieces within a kaleidoscope. The kaleidoscope represents the views and beliefs of the learners, and depending on the position of the pieces, aspects of the content can be thought of as transforming and emerging. Each learner gets the same pieces (course content), and how the pieces are put together is important to summarizing the content. Before they start, learners are encouraged to do a quick review of the text and their notes. Then they create a collage using only pictures to summarize the content of the class.

The collages are taped to the classroom walls. Frequently, I have students look at each other's collages and offer feedback. Sometimes, individuals explain their collages. Either way, by the time the last one has been discussed, the course has been summarized better than I could ever hope to [summarize it].

**James K. Wangberg,
University of Wyoming:**

I try to end the semester on a high note and bring closure . . . with some fun, a sense of celebration, and total class involvement. I do so with a game of charades—one that I believe helps students review, synthesize, and prepare for the final. The course is beginning-level Insect Biology, and so the game is “insect charades,” but it could be adapted to many subjects. A few days before, I have students research and compile a list of challenging entomological words, phrases, or concepts that could be depicted in a charade. Any course content is legitimate. I divide the class into teams, and each team comes up with the most challenging list possible. During the last class, they take turns challenging the other teams to guess their insect term or concept. Students may act out the charade. If they seem reluctant or bashful, I conduct the charade. The opposing team has two minutes to come up with the correct response. The team with the lowest combined times wins.

“Insect charades” transforms a potentially dull review session into an energetic and engaging exercise. By preparing the lists beforehand,

students have had to review their notes, class handouts, and readings. That starts them studying. The preparation also reinforces their learning, as does the charade experience itself. The exercise is memorable, fun, and a great way to punctuate the end of a semester.

Gary Wagenheim, Purdue University (IN)

[Wagenheim] shared some strategies that he and Gary Gemmill reported in the *Journal of Management Education* (May 1994). We highlight one (p. 268):

The goal of this activity is to help students synthesize their learning. It begins in groups (it's an organizational behavior class), with members completing four sentence stems:

- I'm left feeling _____.
- Something I learned about myself _____.
- Something I learned about others _____.
- Something I learned about groups _____.

They complete the stems using various colored markers. They may draw pictures or symbols or make comments. The instructor may join the activity.

These sheets are posted around the classroom, and students read the responses of others. They should be encouraged to look for themes as well as differences.

Struggling with asking and giving feedback for clarification is a part of the learning process, so students should be encouraged not to leave until they feel comfortable with what they said and what they received as feedback.

E. R. Klein, University of North Florida:

On the last day of the course, I always do the same thing—I ask, “What is philosophy?” This

is something that one of my favorite professors in graduate school did at the end of each course, and I have found it to be so helpful, intriguing, and delightful that I have kept the tradition.

Because I have asked this question on the first day of each class as well, the responses given at the end are irrefutable proof that the students have really learned how to “do” philosophy—the *way* they answer the question is now incredibly sophisticated.

Although by now I know just what each student has learned, this process lets them know as well. (Even my majors, who have been through this process many times in different courses, look forward to this question and claim to have achieved new insights with each fresh evaluation.) Closure is achieved as the students conclude the discussion feeling justifiably satisfied with their newfound strength in critical reasoning.

Paul J. Berghoff, Pembroke State University (NC):

Endings are very difficult for me, so I frequently use a story or parable. Here's one that I often tell at the end of our last class meeting:

While walking along the beach after a storm, a man came upon a young boy. The boy was picking up clams that had been washed up on the beach. . . . The man asked the boy what he was doing. He replied, “If I don't throw these creatures back into the sea, they'll die.” “I see,” said the man, “but there must be millions and millions of clams. You can't get to all of them. What you're doing can't possibly make a difference.” The boy smiled, bent down, and picked up another clam. As he threw it back into the sea, he replied, “Made a difference to that one!”

A version of this story, “One at a Time,” appears in Jack Canfield and Mark Victor Hansen's book, *Chicken Soup for the Soul: 101 Stories to Open the Heart and Rekindle the Spirit* (1993, p. 22). The book has many in-

spirational stories that might be of value to teachers.

**Kevin Shannon, Orange
Coast College (CA):**

I have found a simple but effective way of giving students a satisfying sense of closure at the end of a course. I shake their hand and say “thank

you” when they turn in their final exam. It’s honest, it’s polite, and it makes a difference.

Reference

Canfield, Jack, and Mark Victor Hansen, comp. 1993. *Chicken soup for the soul: 101 stories to open the heart and rekindle the spirit*. Deerfield Beach, FL: Health Communications.

Given the flight of students to office hours in the last third of a course, Pescosolido suggests a “quick hit” to handle student angst without changing any of the contours of the course.

The Contract: A Solution to the End-of-the-Semester Student Panic

Bernice A. Pescosolido

Because I (and most faculty) clearly explain the procedure for totaling final grades on the syllabus, any student could derive the final grade or come close to it without last-minute office hour visits. However, after two or three semesters of addressing these questions, it was clear to me that students are really there for a number of other reasons and, by accident, I hit on “The Contract.” This solution is based on the idea that what students really want in these panicked meetings is twofold: motivation and social support. In fact, it takes very little to provide these.

Toward the end of each semester, I get a stack of brightly colored paper that I put on a corner of my desk. When a student comes in with the question, “What am I going to get in this course?,” I take out my grade book and one of the sheets of paper. I then write the student’s name in large letters at the top of the paper and underneath write their grades to date. Then, I ask what they think is the best they can offer me, and what they think they might get, etc., on the final (depending on the nature of the conversation to that point). I write that down and

simply figure out the grade. In my classes, I offer some incentive for improvement over the semester and vary the assignments, but to show a simple case where there are three equally weighted exams:

<i>Exam I</i>	<i>Exam II</i>	<i>Expect Exam III</i>	<i>Final Grade</i>
B (85)	B+ (88)	A (95)	89.3 = A-

Given the improvement, I tell them that I would jump them to an A-. I then sign the contract and ask them to do the same. I tell them that this is a binding contract, and if they do as they say they will, they will get the A-. If that does not happen, they can show me their contract, which I will keep on the same corner of my desk for one semester. This is no different from what I would have done under the stated grading policy, but it seems to work quite well in allaying student anxiety.

Often, I will work through a number of scenarios with students if they wish (e.g., offering different final exam possible outcomes) or

start from a different point. If a student comes in and says, “What do I have to do in this class to pass, get an A (etc.)?”, I simply sit down and figure it out with them, write a contract, and

sign it. Again, it does not change the outcome of the course grade but does seem to influence their feelings about attaining that higher final exam grade.

Evaluation and Documentation of Teaching

With issues of tenure reform for junior and senior faculty on the table, this final subsection of the *Fieldguide* focuses on the documentation of teaching for both formative and summative purposes.

Felder presents a candid assessment of the evaluation of teachers by students at the end of a course and what these ratings might mean.

What Do They Know, Anyway?

Richard M. Felder

Sooner or later, the conversation at the committee meeting or in the faculty lounge turns to student articles' ratings of instructors. It's a sure bet that within six seconds, someone will announce that ratings are meaningless—students don't know enough to evaluate the quality of their instruction. Others agree: one grumbles that the high ratings always go to the easy graders and entertainers; another adds with complete assurance that the rigorous instructors who are really the best teachers may get low ratings now, but in later years their students will come to appreciate them.

What is interesting is that these assertions are invariably offered without a scrap of evidence by individuals with well-deserved reputations for analytical thinking. If someone offered such unsupported arguments in a research seminar, most of us would dismiss both the arguments and the arguer out of hand. In discussions of teaching, however, we routinely suspend the rules of logical inference without a second thought.

It's not as if data on student ratings are lacking. Cashin (1988) notes the existence of 1,300 articles and books dealing with research on the subject; Feldman (1990) sees Cashin and raises him to 2,000! So, for the record and in case you happen to find yourself on a commit-

tee where student ratings come up, here are some facts to throw into the conversation.

Myth: Students lack the wisdom and experience to evaluate the effectiveness of their current instructors. Those who give instructors low ratings at the end of a course will in future years appreciate those instructors.

Fact: High correlations exist between course end ratings and ratings by those who presumably *have* the required wisdom and experience—peers (Marsh, 1987), administrators (Kulik & McKeachie, 1975), alumni (Centra, 1974; Drucker & Remmers, 1951; Overall & Marsh, 1980), and graduating seniors (Aleamoni, 1981; Marsh, 1984). If professors in your department who know how you teach rated your effectiveness, the results would probably not differ all that much from your student ratings. If students rate you highly now, they'll probably still do so when they look back in future years; if they dislike you now, the chances are that in their later wisdom they won't decide you were really a gem.

Myth: Student evaluations are just popularity contests. Easy teachers/easy graders get the highest ratings.

Fact: Teachers who assign more work and more difficult work tend to be rated as most effective (Marsh, 1984, 1986, 1987). Some studies show no effect of grading practices on overall student ratings (McKeachie, 1986; Palmer, Carliner, & Romer, 1978); others find tendencies for teachers giving higher grades to get higher ratings. The latter result does not invalidate the ratings; however, as McKeachie (1986) observes, if students learn more from a teacher (see previous Fact), one would expect both their grades and their ratings to be higher.

Myth: Even if student evaluations have some validity, there's no value in the multiple-choice forms used to collect most of them. You've got to interview students and ask open-ended questions for the results to mean anything.

Fact: Comparisons have been run on student ratings collected in three different ways: objective questionnaire items, written responses to open-ended questions, and group interviews. The average correlation among the rating methods was 0.86 (Marsh & Solomon, 1958).

Myth: Teachers who get high ratings aren't really doing a better job of teaching.

Fact: Teachers rated as effective by students tend to be those whose students perform best on achievement tests (Marsh, 1987). Classes in which students give instructors higher ratings when multiple sections are taught tend to be those in which the students score higher on common external exams (Cashin, 1988). Good teaching also motivates interest and desire to learn; students in courses taught by highly-rated teachers are subsequently more likely to elect advanced courses in the same subjects (Marsh & Solomon, 1958) and to major in those subjects (Sullivan & Skanes, 1974).

Myth: Student evaluations don't improve teaching.

Fact: Students of instructors who got student feedback scored higher on achievement tests and assessments of motivation for learning than [did] students of instructors who got no feedback (Overall & Marsh, 1979).

In short, the research shows that student evaluations of an instructor provide a reliable, valid assessment of that instructor's teaching effectiveness, especially if they reflect the views of many students in several different course offerings. So, next time someone says that there's no good way to evaluate teaching, quietly mention that one or two thousand research studies on the topic suggest otherwise. You may not change anyone's mind on the spot, but it might raise the discussion to a higher level than it usually occupies. It remains to consider how evaluations can be structured to have the maximum impact on teaching effectiveness. That's another column.

References

- Aleamoni, L. M. (1981). Student ratings of instruction. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp. 110-145). Beverly Hills, CA: Sage.
- Cashin, W. E. (1988). *Student ratings of teaching: A summary of the research*. Center for Faculty Evaluation and Development, Kansas State University.
- Centra, J. A. (1974). The relationship between student and alumni ratings of teachers. *Educational and Psychological Measurement*, 34(2), 321-326.
- Drucker, A. J., & Remmers, H. H. (1951). Do alumni and students differ in their attitudes

- toward instructors? *Journal of Educational Psychology*, 42, 129-143.
- Feldman, K. (1990, December). Quoted in *The Teaching Professor*, 5.
- Kulik, J. A., & McKeachie, W. J. (1975). The evaluation of teachers in higher education. In F. N. Kerlinger (Ed.), *Review of research in education* (Vol. 3, pp. 210-240). Itasca, IL: F. E. Peacock.
- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, potential biases, and utility. *Journal of Educational Psychology*, 76(5), 707-754.
- Marsh, H. W. (1986). Applicability paradigm: Students' evaluations of teaching effectiveness in different countries. *Journal of Educational Psychology*, 78(6), 465-473.
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. Elmsford, NY: Pergamon.
- Marsh, H. W., & Solomon, D. (1958). Student ratings of instructors: A validity study. *Journal of Educational Research*, 51, 379-382.
- McKeachie, W. J. (1986). *Teaching tips* (8th ed.). Lexington, MA: D. C. Heath.
- Ory, J. C., Braskamp, L. A., & Pieper, D. M. (1980). Congruency of student evaluative information collected by three methods. *Journal of Educational Psychology*, 22(1), 181-185.
- Overall, J. U., & Marsh, H. W. (1979). Midterm feedback from students: Its relationship to instructional improvement and students' cognitive and affective outcomes. *Journal of Educational Psychology*, 21(6), 856-865.
- Overall, J. U., & Marsh, H. W. (1980). Students' evaluations of instruction: A longitudinal study of their stability. *Journal of Educational Psychology*, 22(2), 321-325.
- Palmer, J., Carliner, G., & Romer, T. (1978). Leniency, learning and evaluation. *Journal of Educational Psychology*, 70(5), 855-863.
- Sullivan, A. M., & Skanes, G. R. (1974). Validity of student evaluation of teaching and the characteristics of successful instructors. *Journal of Educational Psychology*, 66(4), 584-590.

Perhaps one of the most confusing issues to college teachers beginning to document their teaching is the difference between dossiers and portfolios. In this reproduction of an overhead from her presentation at a Preparing Future Faculty Program conference at Minnowbrook, New York, Hutchings lays out the basic contrast.

A List of What Teaching Portfolios Are and What They Are Not

Pat Hutchings

What They Are . . .	And Are Not . . .
A coherent set of materials	A miscellany
Selective	All-inclusive
Work samples (including student work)	An extended resume (a listing of courses taught)
Evidence from multiple sources	Only student ratings
A movie	A snapshot
Authentic assessment (actual performance)	High science
Compiled and owned by faculty	Evaluation <i>done to</i> faculty
Tools for improvement	More record keeping
Collaborately developed/renewed, conversation prompting	Private, confidential files
Professional accountability	Bureaucratic accountability
Culture changing	Conversations through portfolios

EDITORS' NOTE: For discussion of the evaluation of faculty teaching, see Sorcinelli's "The Evaluation of Teaching"; Hutchings' "Behind Outcomes"; and Tierney's "Academic Community and Post-Tenure Review" in the *Handbook*.

Providing more detail on the assessment of teaching, Hutchings' excerpt from the AAHE Bulletin offers suggestions on variations in portfolios.

The Teaching Portfolio: Possible Models

Pat Hutchings

Model I: Individual Choice

Entries of choice (any number)

Model II: Focus on Development

Statement of teaching philosophy: its origins and evolution

Two entries organized around changes made over time—

e.g., two syllabi for same course, with commentary on rationale

e.g., two laboratory protocols/writing assignments

Report on teaching improvement workshops and conferences attended/participated in

Entries exploring unresolved tensions, problems, issues related to student learning—

e.g., samples of student's work, feedback on [work] with reflection

Summary and comment on student ratings over time

Model III: Departmental Key Tasks of Teaching

Two entries in each of three categories: categories reflect "key tasks of teaching on X campus in X department"—

- Course planning with one entry organized around syllabus
- Evaluating student work with one entry organized around student work and evaluation of/feedback on it
- Teaching the class with one entry including a videotape

EDITORS' NOTE: For discussion of the evaluation of faculty teaching, see Sorcinelli's "The Evaluation of Teaching"; Hutchings' "Behind Outcomes"; and Tierney's "Academic Community and Post-Tenure Review" in the *Handbook*.

Model IV: The Scholarly/ Professional Portfolio

An introductory framing statement indicating faculty's roles and responsibilities in context of [departmental/institutional] goals

Four entries organized around work samples that show how new developments in the discipline are being incorporated into classroom teaching, advising, curricular work, outreach [etc.]

Scholarship about teaching: published articles, presentations

In this first of two pieces, Sorcinelli distinguishes between self-evaluation of teaching for employment purposes and that for monitoring one's own progress, providing a template for a form.

Self-Evaluation of Teaching

Mary Deane Sorcinelli

Self-evaluation of teaching can range from personal reflection to formal assessment intended for promotion, tenure, or salary committees. As a source of information used to evaluate teaching in colleges and universities, self-evaluation has gained popularity in recent years (Seldin, 1984). Self-assessment should play a significant part in a multifaceted evaluation process. Despite its increasing use, however, self-evaluation of teaching for personnel decisions remains controversial. The advantages and limits are almost obvious. More than anyone else, the faculty member knows the amount of time, effort, and energy spent on teaching. Both individuals and review processes benefit from self-appraisal of teaching strengths, weaknesses, accomplishments, and future goals. Still, asking a teacher to rate overall teaching effectiveness for promotion or tenure is problematic, particularly in terms of credibility.

EDITORS' NOTE: For discussion of the evaluation of faculty teaching, see Sorcinelli's "The Evaluation of Teaching"; Hutchings' "Behind Outcomes"; and Tierney's "Academic Community and Post-Tenure Review" in the *Handbook*.

Research on self-evaluation is limited but indicates that self-evaluation should not be used as the basis for personnel decisions. Some studies (Centra, 1973; Clark and Blackburn, 1973) conclude that faculty consistently overrate their teaching when using self-evaluation (e.g., general rating of teaching, annual faculty report). Research findings indicate stronger reliability when students and instructors rate a specific class and use the same instrument. They generally show agreement both on overall ratings of the instructor and on such dimensions of teaching as organization and stimulation of interest (Marsh, 1984). The value of self-evaluation, then, depends on purpose. It is an important addition to a teaching file and can set other evidence about instruction in a broader context but should never be used as the focus for personnel decisions. It can be an important means of improving teaching.

Methods used for personnel decisions include the "personal statement" for tenure and promotion decisions and faculty self-assessment forms for annual review. Self-evaluation techniques available to teachers for improvement purposes include audio- and videotape analysis, student rating and self-rating, and faculty development contracts.

Self-Evaluation for Personnel Decisions

Personal Statement. Self-evaluation of teaching plays an important role in the “personal statement” prepared for tenure and promotion dossiers at Indiana University (IU). While self-rating of overall teaching effectiveness lacks credibility for personnel decisions, a narrative that provides evidence of teaching competence is valuable. In such statements, candidates describe what is taught (course content, level, range); outline a philosophy of teaching; discuss their contribution to the teaching, advising, and curricular needs of the department and university; explain teaching preferences, strengths, and weaknesses; detail teaching awards, accomplishments, and future goals; describe efforts to develop teaching competence; and interpret student ratings and colleague judgments. The *Tenure and Promotion Handbook* prepared by and available through the Dean of Faculties Office, IU Bloomington, has examples of personal statements from several academic units (Bodnar, Royce, & Snygg, 1985).

Self-Assessment Form. Some campuses use a standard teaching self-evaluation form for annual reviews. Unlike a faculty activity sheet, such forms ask not only what individuals did but [also] how well they think they did it. Such forms contain both general and summary items and are judgmental in character. Guiding questions for an “essay” form developed by Seldin (1984) include:

- In which area of your discipline do you consider yourself strongest?
 - What is your greatest strength as a teacher? Your greatest weakness?
 - If you could change one thing, what would you most like to change about your teaching?
- Compared to others in your department, how do you assess your teaching performance?
 - What was your most important accomplishment as a faculty member last year?
 - Compared to others in your department, how do you assess that accomplishment?
 - Compared to others in your department, how do you assess your research and publication activity? Your contribution to the institution? To the community?
 - Considering your answers to the previous questions, how do you assess your overall performance as a faculty member in your department?

Other forms are more structured and also seek information on student advising, research, community and institutional service (see Exhibit FG6.1).

Self-Evaluation for Improvement

Videotape and Audio Recording. An effective way to improve teaching is to be video- or audiotaped in the classroom. On many IU campuses, this can be arranged by calling a learning resources or faculty members division office and setting up a time; the service is free and confidential. Watching or listening to a tape can help instructors identify teaching strengths and weaknesses. Many faculty members, however, find the assistance of a colleague or teaching improvement consultant more useful. Studies indicate that change is far more likely if a knowledgeable colleague or consultant can help analyze the performance, provide encouragement, and suggest teaching improvement strategies (Seldin, 1984).

Exhibit FG6.1: Faculty Self-Evaluation

For purposes of promotion, tenure, contract renewal, and increments

Name: _____

Date: _____ Appraisal for period from _____ to _____

Please answer the following questions. For questions involving rating scales, rate yourself 1 to 5 or NA for Not Applicable

1 = Never 2 = Seldom 3 = Average 4=Usually 5 = Always

1. The following are skills used in classroom instruction:
 - A. I am well organized and present material clearly. NA 1 2 3 4 5
 - B. I am readily available for consultation with students. NA 1 2 3 4 5
 - C. I speak clearly, use illustrations to clarify the material, and summarize major points as well. NA 1 2 3 4 5
 - D. I encourage an open atmosphere where students feel free to ask questions and seek help if needed. NA 1 2 3 4 5
 - E. I give examinations that reflect the important aspect of the courses taught. NA 1 2 3 4 5
 - F. I am objective and able to substantiate grades given. NA 1 2 3 4 5
 - G. I use a variety of teaching materials. NA 1 2 3 4 5
 - H. What have you done differently from last year to improve your instructional capability and/or student learning? Were the changes successful? Why or why not? NA 1 2 3 4 5

2. In intellectual breadth and professional activities, I:
Related to Academic Discipline and Departmental or Division Responsibilities
 - A. Am well read beyond the subject I teach. NA 1 2 3 4 5
 - B. Can suggest reading in any area of my general field. NA 1 2 3 4 5
 - C. Make a positive contribution to the progress of my academic unit through committee participation. NA 1 2 3 4 5
 - D. What have you done to maintain and/or improve competence in your academic discipline? NA 1 2 3 4 5

(Continued)

Exhibit FG6.1: Continued

Related to College Duties and Professional Responsibilities Outside the Classroom

E. Discharge my college duties in an effective manner outside of the classroom.	NA 1 2 3 4 5
F. Meet deadlines.	NA 1 2 3 4 5
G. Cooperate with others.	NA 1 2 3 4 5
H. Work well as a member of a committee.	NA 1 2 3 4 5
I. Follow through on committee work by appropriate actions and communications.	NA 1 2 3 4 5
J. Professional activities: List organizations in which you hold membership and any leadership roles.	NA 1 2 3 4 5

Self-Rating and Student Rating. Some units encourage faculty members to assess their teaching by using the department or school student evaluation form. Instructors simply complete a self-evaluation while students complete the same form. The extent to which student information questions or confirms the instructor's perceptions can provide avenues for further consideration. In fact, the less favorable students' ratings are in comparison to the instructor's self-rating, the more likely the exercise will lead to teaching improvement. Again, greater change is possible when individual consultation is combined with rating results (Centra, 1980, 1986).

Faculty Development Contracts. Self-evaluation is an important part of a contracting process used by teaching consultants in campus offices for learning resources to help an individual define goals for faculty development. The exchange between a faculty member and consultant challenges the teacher to examine and

articulate goals in particular courses. Faculty members often begin to look more closely at their ideas about teaching and learning. With these concerns before them, the consultant and individual draw up a plan by which the faculty member will try to accomplish a goal. This goal may be as broad as working to build a course and improving teaching style or as specific as designing a computer-based instructional program. As part of this plan, individuals are encouraged to examine their professional goals and plan the direction of their careers. The substance of the plan focuses on teaching improvement but touches on other such aspects of careers as research, institutional, and community responsibilities.

References

Bodnar, J., Royce, A. P., & Snygg, F. (1985). *Tenure and promotion handbook*. Bloomington: Indiana University Press.

- Centra, J. A. (1973). Self-ratings of college teachers: A comparison with student ratings. *Journal of Educational Measurement, 10*, 287-295.
- Centra, J. A. (1980). *Determining faculty effectiveness*. San Francisco: Jossey-Bass.
- Centra, J. (1986, April). *New directions for research on student ratings of instruction*. Paper presented at a meeting of American Educational Research Association, San Francisco.
- Clark, M. J., & Blackburn, R. T. (1973). Faculty performance under stress. In A. Sockloff (Ed.), *Faculty effectiveness as evaluated by students*. Philadelphia: Temple University Press.
- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76*, 707-754.
- Seldin, P. (1984). *Changing practices in faculty evaluation*. San Francisco: Jossey-Bass.

In this second piece, Sorcinelli shifts to peer evaluation methods and rationales and gives suggested forms to use.

Faculty Evaluation by Colleagues

Mary Deane Sorcinelli

Surveys of how college teaching is evaluated demonstrate a dramatic increase in use of faculty peers as sources of information both for improvement and personnel decisions (Centra, 1980; Seldin, 1980, 1984). Colleagues are in an excellent position to assess the instructor's content knowledge as reflected by course materials (syllabus, readings, assignments, tests) and contributions to the department's teaching mission (associate instructors [A.I.'s], honors, theses, curriculum work). Colleagues are in a less advantageous position to judge teaching through classroom visits. The disadvantages of colleague evaluation (particularly classroom visits) are that it may affect collegiality, overburden faculty members in terms of time and effort, suffer from problems of sampling and rater unreliability, and duplicate information available from students.

Evaluation of Course Materials

Colleagues who have expertise in the discipline of a faculty member can evaluate course materials (syllabus, reading list, textbooks, handouts, assignments, graded exams). Although such examination is neither practiced much nor

[done] systematically, it offers several advantages. It properly uses faculty expertise, [it] is relatively nonthreatening, [it] can be done in a reasonable period of time, and peers do not need to visit classrooms to gain insight into teaching effectiveness. It is also appealing because it can be used for both personnel decisions and improvement. Finally, sharing of materials between junior and senior faculty members might help clarify standards and stimulate the exchange of ideas.

Despite the potential for peer review of course materials, there are few systematic procedures and certainly no standardized forms. Some institutions do have forms that suggest areas for examination. One particularly comprehensive form on course organization, readings, projects, and assignments and grading is included in Exhibit FG6.2.

Evaluation of Instructional Contributions

Colleagues may be in the most advantageous position to observe such teaching-related activities as curriculum development, undergraduate and dissertation advising, A.I. supervision, and

Exhibit FG6.2: Peer Evaluation of Teaching Materials

Listed below are items concerned with teaching materials. They are categorized into three major areas. For each item, indicate on a five-point scale (1-5, with 5 being high) the extent to which the materials meet the criteria as represented by each item.

Course Organization

- _____ The syllabus adequately outlines the sequence of topics to be covered.
- _____ The stated course objectives are clear.
- _____ The outline and sequence of topics are logical.
- _____ The difficulty level is appropriate for the enrolled students.
- _____ The course integrates recent developments in the field.
- _____ Time given to each of the major course topics is appropriate.
- _____ The course is responsive to the needs of the enrolled students.
- _____ The course is an adequate prerequisite for other courses.
- _____ The course objectives are congruent with the department curricula.

Readings, Projects, and Laboratory Assignments

- _____ The reading list (required/recommended) is up to date and represents the work of recognized authorities.
- _____ Readings are appropriate for [the] level of [the] course.
- _____ The texts used in the course are well selected.
- _____ Students are given ample time to complete the assignments/take-home exams.
- _____ The amount of homework and assignments is appropriate.
- _____ The written assignments and projects are carefully chosen to reflect course goals.
- _____ A variety of assignments is available to meet individual student needs.
- _____ Laboratory work is integrated into the course.
- _____ Students are given the course requirements in writing at the beginning of the course.
- _____ The assignments are intellectually challenging to the students.

Exams and Grading

- _____ The exam content is representative of the course content and objectives.
- _____ The exam items are clear and well written.
- _____ The exams are graded in a fair manner.
- _____ The grade distribution is appropriate for the level of [the] course and type of students enrolled.
- _____ The standards used for grading are communicated to the students.

participation in teaching development. Promotion and tenure criteria acknowledge the importance of such activities (see Gros Louis, 1985), and units might develop a way to look at an instructor's out-of-class contributions to teaching. Exhibit FG6.3 suggests areas to include in such a report.

Evaluation of Classroom Teaching

Colleagues who have expertise in the discipline and training in what to observe can provide evaluative information through classroom visits. In particular, a colleague's observation of such aspects of teaching as appropriateness of materials and methods, breadth and depth of material covered, relation of such material to the syllabus and goals of the course, and incorporation of recent developments in the discipline could offer a more informed appraisal of the instructor's mastery of content than could students' perceptions. Of all ways to collect information about teaching for personnel decisions, however, classroom observation by colleagues is by far the most controversial method and the most inconsistently used.

Literature on colleague evaluation of classroom teaching is inconclusive. Some researchers report that the value of observing teaching for promotion and tenure is dubious (Cohen & McKeachie, 1981; Centra, 1980), while others (French-Lazovik, 1981, Seldin, 1980) conclude that, if carefully gathered and judiciously interpreted, observation is capable of solid judgment on merit increases, promotion, and tenure. There is consensus that peer observation has enjoyed more success as a strategy for improvement, and a system must be carefully crafted if used for personnel decisions.

Departments should consider several issues when developing a system of evaluation by colleagues for promotion, tenure, or salary decisions. They include:

- Develop[ing] materials and policies and implementing an observation program places considerable demands on the time of faculty members. The department and institution will need to recognize the investment of faculty time and effort as worthwhile.
- Using evaluations by colleagues in promotion and tenure decisions could affect collegiality. Since a mandatory system, no matter how fair, may undermine relations among faculty, it is not surprising that institutions report more faculty support when a program is voluntary, [is] used primarily to improve teaching, and gives the individual faculty member the option to include data in a teaching file.
- The department must determine which faculty members (tenured senior colleagues, respected teachers, untenured colleagues) will participate, how observers will be chosen (chair, individual faculty members, peer review committee), and how often colleagues will visit (one course per year, every five years, only for tenure and promotion review).
- It is important to have explicit criteria by which colleagues make evaluations to guide the classroom observations and summarize impressions developed over several visits. Colleague judgment about such issues as the instructor's ability to present knowledge to students; the currency, appropriateness, level of difficulty, and amount of content; [and] the

Exhibit FG6.3: Suggested Form for Peer Review of Undergraduate Teaching Based on Dossier Materials

Dossier Materials

Suggested Focus in Examining Dossier Materials

1. What is the quality of materials used in teaching?

– Course outline	– Are these materials current?
– Syllabus	– Do they represent the best work in the field?
– Reading list	– Are they adequate and appropriate [for] course goals?
– Test used	– Do they represent superficial or thorough coverage of course content?
– Study guide	
– Description of non-print materials	
– Handouts	
– Problem sets	
– Assignments	

Peer Reviewer's Rating: Low ___/___/___/___/___/___/___ Very High

Comments: _____

2. What kind of intellectual tasks were set by the teacher for the students (or did the teacher succeed in getting students to set for themselves), and how did the students perform?

– Copies of graded examinations	– What was the level of intellectual performance achieved by the students?
– Examples of graded research papers	– What kind of work was given an A? a B? a C?
– Examples of teacher's feedback to students on written work	– Did the students learn what the department curriculum expected for this course?
– Grade distribution	– How adequately do the tests or assignments represent the kinds of student performance specified in the course objectives?
– Descriptions of student performances, e.g., class presentation, etc.	
– Examples of completed assignments	

Peer Reviewer's Rating: Low ___/___/___/___/___/___/___ Very High

Comments: _____

(Continued)

Exhibit FG6.3: Continued

3. How knowledgeable is this faculty member in subjects taught?

– Has the instructor kept in thoughtful contact with developments in his or her field?	– Record of colloquia or lectures given
– Evidence in teaching materials	– Is there evidence of acquaintance with the ideas and findings of other scholars? (This question addresses the scholarship necessary to good teaching; it is not concerned with scholarly research publication)
– Record of attendance at regional or national meetings	

Peer Reviewer's Rating: Low ___/___/___/___/___/___/___/___ Very High

Comments: _____

4. Has this faculty member assumed department's or university's teaching?

– Record of service on department curriculum committee, honor program, advising board of teaching support service, special committees (e.g., to examine grading policies admission standards, etc.)	– Evidence of design of new courses
– Description of activities in supervising graduate students learning to teach	

Peer Reviewer's Rating: Low ___/___/___/___/___/___/___/___ Very High

Comments: _____

5. To what extent is this faculty member teaching?

– Factual statement of what activities the faculty member has engaged in to improve his or her teaching	– Examples of questionnaires used for formative purposes
	– Examples of changes made on the basis of feedback

Peer Reviewer's Rating: Low ___/___/___/___/___/___/___/___ Very High

Comments: _____

structure and integration of lectures and topics seems most appropriate.

- A standardized observation form will yield systematic and comparable data. It should include specific questions that reflect dimensions of effective teaching on which colleagues are best able to provide information (e.g., content knowledge) and can observe in a classroom setting. It might ask for comparative information (“Would you recommend this instructor to students?” “Overall, how would you rate this instructor?”). There also should be open-ended questions and space for extensive comments.
- Training of participating faculty in what and how to observe is helpful. Often emphasized is a three-step consultation that includes a pre-observation conference in which the observer meets with the instructor to discuss the visit, observation, and another conference to discuss tentative conclusions and recommendations.
- Colleagues should observe at least three classes over a semester. Observation by several colleagues ensures adequate representation (one or more colleagues could observe an entire course). A single visit is highly suspect for personnel decision making.
- The department needs to decide how to report such observations. The instructor might comment on a colleague’s recommendation prior to submission to the department chair or summarize differ-

ent colleague evaluations to point out similarities and interpret differences in judgment.

If the purpose of visitation is improvement of teaching, greater flexibility of sources and methods is possible. The simplest procedure is for faculty members to ask a colleague to observe and review their teaching. Even more helpful are observations by experienced colleagues from similar academic disciplines or teaching improvement consultants.

References

- Centra, J. A. (1980). *Determining faculty effectiveness*. San Francisco: Jossey-Bass.
- Cohen, P. A., & McKeachie, W. J. (1981). The role of colleagues in the evaluation of college teaching. *Improving College and University Teaching*, 28, 147-154.
- French-Lazovik, G. (1981). Documentary evidence in the evaluation of teaching. In J. Millman (Ed.), *Handbook of teacher evaluation*. Beverly Hills, CA: Sage.
- Gros Louis, K. (1985). *Preparing a teaching dossier for promotion and tenure: Teaching and learning at Indiana University*. (Available from University Office for Learning Resources).
- Seldin, P. (1980). *Successful faculty evaluation programs*. New York: Coventry Press.
- Sorcinelli, M. D. (1984). An approach to colleague evaluation of classroom instruction. *Journal of Instructional Development*, 7, 11-17.

