

DESIGNING TESTS TO MAXIMIZE LEARNING

Richard M. Felder
North Carolina State University

It's the middle of December. A colleague of yours who teaches mechanics has just gotten the tabulations of his end-of-course student evaluations and he's steaming! His students clearly hated his course, giving him the lowest ratings received by any instructor in the department. He consoles himself by grumbling that student evaluations are just popularity contests and that even though his students don't appreciate him now, in a few years they'll realize that he really did them a favor by maintaining high standards.

He's probably kidding himself. Although bashing student ratings is a popular faculty sport, several thousand research studies have shown that student ratings are remarkably consistent with retrospective senior and alumni ratings, peer ratings, and every other form of teaching evaluation used in higher education.^{1,2} Although there are always exceptions, teaching rated by most students as excellent usually is excellent, and teaching rated as atrocious usually is atrocious.

If your colleague decided to take a hard objective look at those evaluations instead of dismissing them out of hand, there is a good chance that he would find that his examinations play a major role in the students' complaints. Not the difficulty of the exams *per se*: the research also shows that the highest evaluations tend to go to some of the more demanding teachers, not the ones who hand out A's for mediocre work.² What students hate more than anything else except outright sadistic behavior are examinations that they perceive as unfair. Tests that fall into this category have any of the following features: (1) problems on content not covered in lectures or homework assignments; (2) problems the students consider tricky, with unfamiliar twists that must be worked out on the spur of the moment; (3) excessive length, so that only the best students can finish in the allotted time; (4) excessively harsh grading, with little distinction being made between major conceptual errors and minor calculation mistakes; (5) inconsistent grading, so that two students who make the identical mistake lose different points. Most students can deal with tests that they fail because they don't understand the material or didn't study hard enough; however, if they understand but do poorly anyway for any of those five reasons, they feel cheated. Their feeling is not unjustified.

If you teach a course in a quantitative discipline, there are several specific things you can do to minimize your students' perception that you are dealing with them unfairly on examinations.

- *Test on what you teach.* A common and unfortunate practice is to give fairly straightforward examples in lectures and homework and then to put high-level analysis problems or problems with unexpected twists on the test, with the argument being that "we need to teach students to think for themselves."

The logic of this argument is questionable, to say the least. People acquire skills through practice and feedback, period: no one has ever presented evidence that testing students on

unpracticed skills teaches them anything. Moreover, engineers and scientists are never presented with brand new varieties of quantitative problems and told that they have to solve them on the spot without consulting anyone. A student's ability to solve hard puzzles quickly should not be the main determinant of whether he or she should be certified to practice engineering or science. The way to equip students to solve open-ended or poorly defined problems or problems that call for critical or creative thinking is to work out several such problems in class, then put several more on successive homework assignments and provide constructive feedback, and *then* put similar problems on tests.

- Consider handing out a study guide one to two weeks before each test. It makes no sense to tell students “Here’s the 574-page text—you’re responsible for all of it. Guess what I’m going to put on the exam!” In the words of Jim Stice, teaching is not a mystery religion. *There should be no surprises on tests:* nothing should appear that the students could not have anticipated, no skill tested that has not been explicitly taught and repeatedly practiced.

Suggestions such as this and the previous one are often equated with lowering standards or “spoon-feeding” students. They are nothing of the sort. Taking the guesswork out of expectations is not equivalent to lowering them: on the contrary, I advocate *raising* expectations to the highest level appropriate for the course being taught, knowing that only the best of the students will be capable of meeting all of them. The point is that the more clearly the students understand those expectations and the more explicit training they are given in the skills needed to meet them, the more likely those with the aptitude to perform at the highest level will acquire the ability to do so.

A study guide is an effective way to communicate your expectations—among other reasons, because students are likely to pay attention to it. The guide should be thorough and detailed, with statements of every type of question you might include on the test—calculations, estimations, definitions and explanations, derivations, troubleshooting exercises, etc. The statements should begin with observable action words and not vague terms such as *know, learn, understand, or appreciate*. (You wouldn’t ask students to understand something on a test—you would ask them to do something to demonstrate their understanding.) Draw from the study guide when planning lectures and assignments and constructing the test. No surprises!

A number of benefits follow from the formulation of such *instructional objectives* for courses.^{3,4} A well-written set of objectives helps the instructor make the lectures, assignments, and tests coherent, gives other faculty members a good idea of what they can expect students who pass the course to know, and gives new instructors an invaluable head start when they are preparing to teach the course for the first time. An additional benefit in engineering is that the objectives provide accreditation visitors with an excellent summary of the knowledge and skills being imparted to the students in the course (particularly those having to do with Outcomes 3a–3k of Engineering Criteria 2000).⁵

- Minimize speed as a factor in performance on tests. Unless problems are trivial, students need time to stop and think about how to solve them while the author of the problems does not. *If your test involves quantitative problem solving, you should be able to work out the test in less than one-third of the time your students will have to do it* (and less than one-fourth or one-fifth if particularly complex or computation-heavy problems are included). If you

can't, cut it down by eliminating questions, presenting some formulas instead of requiring derivations, or asking for solution outlines rather than complete calculations.

In my courses, the problems get quite long: by the end of the course, a single problem might take two or three hours to solve completely. There's no way I can put one of those problems on a 50-minute test, but I still have to assess my students' ability to solve them. I do it with the following generic problem:

Given...(describe the process or system to be analyzed and state the values of known quantities), write in order the equations you would solve to calculate...(state the quantities to be determined). Just write the equations—don't attempt to simplify or solve them. In each equation, circle the variable for which you would solve, or the set of variables if several equations must be solved simultaneously.

The students who understand the material can do that relatively quickly—it's the calculus and algebra and number-crunching that take most of the solution time. Moreover, I know that if they can write equations that can be solved sequentially for the variables of interest, given sufficient time they could grind through the detailed calculations.

One cautionary note, however. If students have never worked on a problem framed in this manner and one suddenly appears on a test, many of them will be confused and may do worse than they would have if the problem had called for all the calculations to be done. Once again, the rule is no surprises on tests. If you plan to use this device, be sure you work similar problems in class and then put some on homework, and *then* do it on the test.

- *Always work out a test from scratch when you have what you think is the final version, then revise it to eliminate the flaws you discover and try it again.* Consider giving the test to a colleague or teaching assistant to review.

Professors don't want to do this—I certainly don't! There are only two choices, however. One is to write the test on Sunday night, give it a quick once-through to make sure there are no glaring errors, and administer it Monday morning. You'll usually find that the test is too long—only a handful of students have time to finish it, and some who really understand the material fail miserably because the only way they're capable of working is slowly and methodically. (Incidentally, people who work like that are the ones I want designing the bridges I drive across.) It may also happen—and frequently does—that 15 to 30 minutes into the test a puzzled student asks if something is missing from the statement of Problem 2, and you realize that you forgot to include an important piece of data. Telling the class at that point that they've been beating their heads against an impossible problem and then figuring out how to grade the test is not an experience you want to have.

The only alternative is to do what I have suggested. I make up my test, think it's perfect, and then sit down with my stopwatch and take it. That's when the problems invariably surface. First, it's too long—in 32 years of teaching, I have yet to make up a test that wasn't too long on the first round. And there are underspecified problems and overspecified problems and poorly worded problems and problems that call for time-consuming but relatively pointless number-crunching. Then I revise—cleaning up some questions, eliminating busywork in others, eliminating others completely—and take the test again, and

sometimes the revised version is acceptable and other times I have to go back and make still more changes.

- Set up multiple-part problems so that the parts are independent. For example, in Part (b) of a problem, say something like “Assume the answer to Part (a) was 42.8 cm/s, regardless of what you actually got.” This technique provides two benefits. First, it decouples the parts of the problem, so that even if students can’t get Part (a) they can show you whether or not they’re able to do Part (b). Second, all the students will have the same starting point for Part (b), which will greatly simplify the grading. (This is a particularly important consideration in large classes.) The usual caution applies, however: give practice on problems like this before they show up on the exam.
- Design 10–15% of the test to discriminate between A-level and B-level performance. Much less and your better students will have little incentive to go for the highest levels of understanding they are capable of achieving; much more and the test loses discriminatory ability (the A students will do well and the B, C, and D students will be clustered together in the failing range). If you have included high level questions on your study guide—e.g., explanations of physical phenomena in terms of course concepts, troubleshooting exercises, or problems involving conceptual design or critical evaluation—use them to constitute this 10–15%.
- Be generous with partial credit on time-limited tests for work that clearly demonstrates understanding and penalize heavily for mistakes on homework, where students have time to check their work carefully. Instructors often get it backwards. They collect the homework, grade it superficially and check it as correct if it looks remotely like what they had in mind, and then take the test grading seriously and penalize students for making the same mistakes they got away with in the homework. If you have graders, tell them to count off enough for careless errors so that it stings. When the students come to you complaining about the harshness of the grading, say something like “Look, when you’re a civil engineer, if you design 10 buildings and one of them collapses, they’re not going to pat you on the back and give you a 90. Small mistakes can cost you a lot in the real world—this would be a good time for you to start learning to avoid them.” In the artificial environment of an in-class exam, however, cut them some slack.
- Don’t deliberately design tests to make the average grade 60 or less. Tests on which most grades are very low serve no useful purpose. While low grades in engineering, science, and math courses may on rare occasions reflect the students’ wholesale laziness or incompetence (the default interpretation of instructors whose test averages are consistently low), they are much more likely to indicate that either the tests were poorly designed or the instructor did a poor job of teaching the students what they needed to know to do well. Low test grades are also demoralizing and can lead students who would be excellent professionals to conclude that they are in the wrong field.
- If you give a test on which the grades are much lower than you anticipated and you believe some of the responsibility is yours, consider making adjustments. The simplest method is to add the same number of points to everyone’s grade so that the top grade is 100 or the average is 70. Another method is applicable when the grades are low because virtually everyone missed a particular problem, a situation almost certain to be the instructor’s responsibility.

When that happens in my class, I announce that I will give a variation of that problem as a quiz and that the students may add their quiz grade to their test score. By the time of the quiz, that class knows how to do that problem.

- *If you are teaching a large class and use teaching assistants to grade tests, take precautions to assure that the grading is consistent and fair.* Write out a detailed solution key and breakdown of the point values for every part of every problem (3 points for this, 1 point for that, etc.) and go over it carefully with the graders. Make sure that each problem is only graded by one person. Sit with the graders for the first hour or so and help them with difficult decisions about partial credit, tell them to consult with one another thereafter if they're not sure about something, and encourage them to contact you if they can't reach agreement among themselves. Glance through the graded tests to make sure that nothing strange has happened.
- *Institute a formal procedure for students to complain about test grades.* (This one is more to protect your time than to help the students.) Announce in your course guidelines that students have one week to register complaints about how a test was graded, after which complaints will not be heard. If the complaint is that the points were incorrectly totaled, the students simply have to show you (or the student grader) the graded exam, but if they think they deserve more points on one or more problems they must make their case in writing. Give such requests serious consideration and make the grade adjustments you believe are justified. The volume of complaints you have to deal with should drop by an order of magnitude, few or none of those you get will be frivolous, and never again will you have to deal with a flood of complaints on the last day of class about grading on a test given 12 weeks earlier.

In *Embracing Contraries*,⁶ Peter Elbow notes that faculty members have two conflicting functions—gatekeeper and coach. As gatekeepers, we set and maintain high standards to assure that our students are qualified to enter the community of professional practice by the time they graduate, and as coaches we do everything in our power to help them meet and surpass those standards. Examinations are at the heart of both functions. By making our tests comprehensive and rigorous we fulfill the gatekeeper role, and by doing our best to prepare our students for them and ensuring that they are fairly graded, we satisfy our mission as coaches. The suggestions given in this paper are intended to help us serve well in both capacities. Clearly, adopting them can take time, but it is hard to imagine an expenditure of time more important to our students, their future employers, and the professions they will serve.

References

1. W.E. Cashin, (a) “Student Ratings of Teaching: A Summary of the Research,” *IDEA Paper No. 20* (1988); (b) “Student Ratings of Teaching: The Research Revisited,” *IDEA Paper No. 32* (1995). IDEA Center, Kansas State University. Available on-line at <http://www.idea.ksu.edu/products/Papers.html>.
2. R.M. Felder, “What Do They Know, Anyway?” *Chemical Engineering Education*, 26(3), 134–135 (1992). Available on-line at <http://www.ncsu.edu/felder-public/Columns/Eval.html>.
3. R.M. Felder and R. Brent, “Objectively Speaking,” *Chemical Engineering Education*, 31(3), 178–179 (1997). Available on-line at <http://www.ncsu.edu/felder-public/Columns/Objectives.html>.
4. J.E. Stice, “A First Step Toward Improved Teaching.” *Engineering Education*, 66, 394-398 (1976).
5. Accreditation Board for Engineering and Technology, “Criteria for Accrediting Engineering Programs: Effective for Evaluations during the 2001–2002 Accreditation Cycle,” http://www.abet.org/downloads/2001-02_Engineering_Criteria.pdf, November 2000.
6. P. Elbow, *Embracing Contraries: Explorations in Learning and Teaching*, New York, Oxford University Press, 1986.