Random Thoughts . . .

FAQS. V. DESIGNING FAIR TESTS[1]

RICHARD M. FELDER AND REBECCA BRENT
North Carolina State University • Raleigh, NC 27695

The subject that sets off the most heated discussions in our workshops is testing. When we suggest giving tests that can be finished in the allotted time by most of the students, contain only material covered in lectures or assignments, involve no unfamiliar or tricky solution methods, and have average grades in the 70–75 range, a few participants always leap up to raise objections:

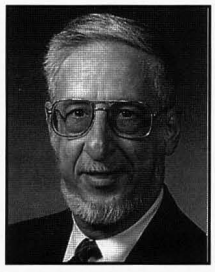
1. *What's wrong with tests that only the best students have time to finish?*
   Engineers constantly have to face deadlines; besides, if you really understand course material you should be able to solve problems quickly.
2. *Why do I have to teach everything on the test?*
   We shouldn't spoon-feed the students—they need to learn to think for themselves!
3. *If I curve grades, what difference does it make if my averages are in the 50's?*

Let's consider these questions, starting with the first one. One problem with long tests is that students have different learning and test-taking styles.[2] Some ("intuitors") tend to work quickly and are not inclined to check their calculations, even if they have enough time. Fortunately for them, their style doesn't hurt them too badly on tests: they are usually fast enough to finish and their careless mistakes only lead to minor point deductions. Others ("sensors") are characteristically methodical and tend to go over their calculations exhaustively. They may understand the material just as well as the intuitors do, but their painstaking way of working often leads to their failing exams they could have passed with flying colors if they had more time.

Being methodical and careful is not exactly a liability in an engineer, and sensors are every bit as likely as intuitors to succeed in engineering careers. (Frankly, we would prefer them to design the bridges we drive across and the planes we fly in, even if their insistence on checking their results repeatedly slows them down compared to the intuitors.) Studies have shown, however, that sensors tend to get significantly lower grades than intuitors in engineering courses[2] and that minimizing speed as a factor in test performance may help level the playing field.[3]

Tests that are too long thus discriminate against some students on the basis of an attribute that has little to do with conceptual understanding or aptitude for engineering. (True, engineers have deadlines, but not on a time scale of minutes for the types of problems on most engineering exams.) Moreover, while overlong tests inevitably frustrate and demoralize students, there is not a scrap of research evidence that they either predict professional success or help students to become better or faster problem solvers.

Richard M. Felder *is Hoechst Celanese Professor Emeritus of Chemical Engineering at North Carolina State University. He received his BChE from City College of CUNY and his PhD from Princeton. He is coauthor of the text* Elementary Principles of Chemical Processes *(Wiley, 2000) and codirector of the ASEE National Effective Teaching Institute*



Rebecca Brent *is an education consultant specializing in faculty development for effective university teaching, classroom and computer-based simulations in teacher education, and K-12 staff development in language arts and classroom management. She co-directs the SUCCEED Coalition faculty development program and has published articles on a variety of topics including writing in undergraduate courses, cooperative learning, public school reform, and effective university teaching.*

© *Copyright ChE Division of ASEE 2002*

How long is too long? Unless problems are trivial, students need time to stop and think about how to solve them while the author of the problems does not. A well-known rule-of-thumb is that *if a test involves quantitative problem solving, the author should be able to work out the test in less than one-third of the time the students have to do it* (and less than one-fourth or one-fifth if particularly complex or computation-heavy problems are included). If a test fails to meet this criterion, it should be shortened by eliminating some questions, giving some formulas instead of requiring their derivations, or asking for some solution outlines rather than requiring all the algebra and arithmetic to be worked out in detail.

How about those problems with unfamiliar twists that supposedly show whether the students can think independently? The logic here is questionable, to say the least. Figuring out a new way to tackle a quantitative problem on a time-limited test reflects puzzle-solving ability as much as anything else. If tricky problems count for more than about 10–15% of a test, the good puzzle-solvers will get high grades and the poor ones will get low grades, even if they understand the course content quite well. This outcome is unfair.

But (a workshop participant protests) shouldn't engineering students learn to think for themselves? Of course, but people learn through practice and feedback, period; no one has ever demonstrated that testing unpracticed skills teaches anyone anything.Therefore, *there should be no surprises on tests:* no content should appear that the students could not have anticipated, no skill tested that has not been taught and repeatedly practiced. To equip students to solve problems that require, say, critical or creative thinking, try working through one or two such problems in class, then put several more on homework assignments, and *then* put one on the test. If for some reason you want students to be faster problem solvers, give speed drills in class and on assignments and *then* give longer tests. The test grades will be higher—not because you're lowering standards, but because you're teaching the students the skills you want them to have (which is, after all, what teachers are supposed to do).

Finally, what's wrong with a test on which the average grade is 55, especially if the grades are curved? It is that given the hurdles students have to jump over to matriculate in engineering and survive the freshman year, an entire engineering class is unlikely to be incompetent enough to deserve a failing average grade on a fair test. If most students in a class can only work out half of a test correctly, it is probably because the test was poorly designed (too long, too tricky) or the instructor didn't do a good job of teaching the necessary skills. Either way, there's a problem.

One way to make tests fair without sacrificing their rigor is to post a detailed study guide before each one. The guide should include statements of every type of question that might show up on the test, especially the types that require high-level thinking skills.[4] The statements should begin with observable action words (*explain, identify, calculate, derive, design, formulate, evaluate,...*) 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.) A typical study guide for a mid-semester test might be between one and two pages long, single-spaced. Drawing from the study guides when planning lectures and assignments and constructing tests makes the course both coherent and effective.

Peter Elbow observes that faculty members have two conflicting functions—gatekeeper and coach.[5] As gatekeepers, we set high standards to assure that our students are qualified for professional practice by the time they graduate, and as coaches we do everything we can to help them meet and surpass those standards. Tests are at the heart of both functions. We fulfill the gatekeeper role by making our tests comprehensive and rigorous, and we satisfy our mission as coaches by ensuring that the tests are fair and doing our best to prepare our students for them. The suggestions given in this paper and its predecessor[1] address both sets of goals. Adopting them may take some effort, but it is hard to imagine an effort more important for both our students and the professions they will serve.

REFERENCES

1. This column is based on R.M. Felder, "Designing Tests to Maximize Learning," *J. Prof. Issues in Engr. Education & Practice, 128*(1), 1–3 (2002). Available at *<http://www.ncsu.edu/felder-public/Papers/TestingTips.htm>*.
2. R.M. Felder, "Reaching the Second Tier: Learning and Teaching Styles in College Science Education," *J. College Science Teaching, 23*(5), 286-290 (1993). Available at *<http://www.ncsu.edu/felder-public/Learning_Styles.html>*.
3. R.M. Felder, G.N. Felder, and E.J. Dietz, "The Effects of Personality Type on Engineering Student Performance and Attitudes," *J. Engr. Education, 91*(1), 3–17 (2002). Available at *<http://www.ncsu.edu/felder-public/Learning_Styles.html>*.
4. R.M. Felder and R. Brent, "Objectively Speaking," *Chemical Engineering Education, 31*(3), 178–179 (1997). Available at *<http://www.ncsu.edu/felder-public/Columns/Objectives.html>*. Illustrative study guides may be found at *<http://www.ncsu.edu/felder-public/che205site/guides.html>*
5. P. Elbow, *Embracing Contraries: Explorations in Learning and Teaching*, New York, Oxford University Press, 1986.

All of the *Random Thoughts* columns are now available on the World Wide Web at http://www.ncsu.edu/effective_teaching and at http://che.ufl.edu/~cee/