

Making the grade: some principles of comparative grading

J E Stake

52 *J Legal Educ* 4, 2002, pp 583–618

ASSESSMENT METHODS

Most law teachers are concerned about their grading but spend little time thinking about how their grades fit into the context of marks given by their peers. They do not consider whether grading practices that would be benign standing alone could be problematic in that larger context. They spend little of their faculty energy overseeing the institutional process of calculating summary statistics, such as grade point averages (GPA) and class ranks.

Although GPAs and class ranks contain specific messages that are quite limited, they are read to mean more. Our grades communicate broader meanings to the students and their potential employers. The signals we send to students relate to both whether they are studying well or enough, and whether they have the aptitude for a career in the law. Grades can influence the ways students think about themselves, swelling their heads or shaking their confidence.

Some believe that one of our most useful functions as law teachers is to sort students for their employers. And whether or not it is our purpose to sort grades do have that effect. Employers act as if grades reflect aptitude for being a lawyer. Many of the most prestigious and high-paying firms limit their hiring to students above a specified grade point average or class rank, or limit their interviews to students who are on a law journal, which is usually determined in part by grades.

Because law school grades send messages, messages that open and close doors, it is obviously unfair to say students performed differently when they in fact performed the same. It is also unfair to a higher-performing student to say that she performed the same as a student who performed much less well. Of course two students rarely perform exactly the same, so there is always some unfairness in using the same grade twice, but that unfairness grows with the difference in their performances.

In addition to being unfair, inaccurate communication via grades can also be inefficient because it misleads employers and it fails to set up appropriate incentives for students. A discouraged student may drop out of school. An unduly encouraged student may gain a false sense of confidence that he can do legal work without as much preparation as others.

Grading communicates information about a student's performance relative to the performance of other students in the same school. Not all teachers employ grades to this end. Some view grades as carrots and sticks — which indeed they always are, whether or not so intended — without any regard to the assessing and sorting functions. Other teachers, probably a minority in law but a larger percentage in other fields, see grades as actual measures of achievement on criteria that are absolute, not referenced against other students' performance. Such teachers adopt standards for each grade in the grading scale and then assign grades on the basis of whether students have exceeded the announced standards. Employers and other readers of law school transcripts do not have the information or the time they would need to figure out what grades mean for individual teachers. Even when they seek shared goals, even when they are partially constrained by rules and customs, some teachers award lower grades, some higher.

Grading is never perfect. Assessment instruments will never be precisely able to measure student ability, or learning, or anything else relevant to performance as a lawyer. Subjectivity in grading will never be eliminated. Exam coverage may not be what we want it to be. Neither essay questions nor true/false exams give us an undistorted picture of the real abilities of our students to perform as legal professionals. No test is perfectly reliable or valid.

There is an argument that we should allow teachers to give higher or lower grades because some teachers teach better. Certainly teachers are not equal in their effectiveness. We could allow the more effective teachers to give higher grades as a reward for teaching better. This would add to the incentives for good teaching along with salary and other rewards. One obvious problem with this rationale for disparate average grades is that it is very hard to identify the teachers who teach better. Self-reporting will not do, as most teachers report that they are better than average, and if there is some positive payoff to being a great teacher, most of us will report that we qualify.

Law employers often want to know not how much the student has learned in law school but rather how capable the student is of learning, compared to the other students in her school. The student who learned less in the poorer teacher's course may be just as capable of learning as the student who learned more in the better teacher's course, contrary to the implication of the inferior grade. Thus, allowing the grade averages to vary according to the quality of the teaching might send the wrong signals to employers.

Teachers should spread their grades to the same degree in all classes. One measure of variability in grades, as in any other set of numbers, is the range. This is not a very useful measure, however, because it is based entirely on the top and bottom scores and ignores all the scores in between. A

Grading is never perfect. Assessment instruments will never be precisely able to measure student ability, or learning, or anything else relevant to performance as a lawyer. Subjectivity in grading will never be eliminated.

better measure of the variation is the standard variation, which is essentially a measure of how far the grades fall away from the mean.

The standard deviation should be equalised across classes. But we can justify limited exceptions to allow leeway for small classes, to account for differences in prior performances of the students, and to minimise the impact of poor evaluation instruments if students are properly warned of that possibility in advance. The argument made here for standardisation should not be read as an argument for normalisation. Some schools attempt to solve the problem of unequal spreads across courses by mandating that certain percentages of the class fall into each of the available grade intervals. If the percentage ranges for each interval are very narrow, the forced curves will have the salutary effect of keeping all teachers to the same mean and standard deviation. The forced curve has the added benefit of avoiding differences in skewness and other statistics that are used to describe distributions. Nevertheless, there are a number of potential problems with specifying the precise percentages of grades to be awarded in each of the grading intervals.

One of the common misconceptions about grading is that it is better to use a coarse grading scale with fewer and larger groupings because the grader can have more confidence in the grades. Clearly it is true that no one will get the wrong grade if one grade is given to all students. Likewise, teachers probably make few errors when they have to determine only whether the students pass or fail. But that is only half of the story, and is the less important half. There is a trade-off between, on one side, the frequency of errors and, on the other side, both the precision of information recorded for subsequent retrieval and the magnitude of the errors in that information.

Once we have established the principle, it becomes obvious that grade inflation generates a number of problems. When some teachers increase grades, they create incentives for students to take their courses rather than other courses. At the margin, the higher grades in a grade-inflated course will cause some students to take that course instead of a course that would have been better for their educational development. That in turn puts pressure on other teachers to inflate their grades because many teachers want to have students in their elective classes. With time, there is pressure on other schools to inflate their grades so that their graduates have a fair shot in the competitive education and employment markets. Thus, grade inflation spirals.

Any time a faculty imposes grading constraints on its members, it risks forcing miscommunication. If the class is abnormal, the constrained teacher is unable to send an accurate message of comparative performance because she cannot award abnormal grades. On the other hand, if the class is normal, an unconstrained teacher with a defective assessment instrument might send inaccurate messages of comparative performance by sending abnormal grades.

Because these risks are somewhat speculative, increased uniformity in grading will not necessarily lead to increased accuracy in communication. Indeed, grading constraints will undoubtedly lead to some sets of grades that are less accurate than they would have been if the teacher had been unconstrained. But the issue is not whether uniformity will improve matters in every case, but rather whether it improves communication in the long run. Forced uniformity will often increase accuracy, and overall the odds lie against complete professorial freedom in grading.

Assessment of student performance is never perfect. Choosing the types of instruments for measuring performance will always involve tradeoffs. Nevertheless, we should try to avoid exacerbating deficiencies in our assessment by making errors in the numerical methods we use for coming to summary comparative statistics. Judgments about proper grading practices must take serious account of the context in which our grades reside. If we are to give grades and class ranks, let them be fairly calculated. We teach that justice matters; let us do our best when it is our turn to hand down the decisions.

CLINICAL LEGAL EDUCATION

Lawyering for justice and the inevitability of international human rights clinics

D Hurwitz

28 *Yale J Int'l L*, 2003, pp 505–546

Globalisation may be taken for granted, but is the US legal profession prepared? A 1996 survey conducted by the American Bar Association Section on International Law and Practice found that law schools are responding to the demand for global relevance in legal studies by offering multiple and diverse courses in international and comparative law. Yet, as the ABA survey found, most students never take an international law course, and other opportunities for exposure to international law in law school are scant.

How will law schools prepare students to participate in the 'new' global society in a meaningful way? The answer is not simply offering more courses, or even making those already offered mandatory. While each of these measures would help, the better answer lies in a particular pedagogical approach — an approach that requires students to grasp and digest the inherently transnational dimension of