TEN MYTHS ABOUT LAW SCHOOL GRADING

DANIEL KEATING[*]

There is probably no subject more misunderstood and more clouded in myths than law school grading. I am not certain why this is so, but I do have a couple of theories. The first is that law school grades are very important and therefore are the subject of much discussion by law students. Although law students talk to each other a lot about grading, law professors do not tell students that much about the subject. Thus, students are left to concoct their own theories and folklore about how grading really works or should work.

My second proposed reason for the prevalence of grading myths might be called the Lawyer As Math Phobe Theory. That theory starts with the assumption that most law students had non-mathematical majors in college such as Political Science, English and History. Perhaps if I were the associate dean at an engineering school, I would encounter these grading myths less often. In any event, this second theory postulates that many of the myths about law school grading are due simply to a lack of understanding about basic math concepts. Indeed, many of the myths described below are held by law faculty as well as students, and I know from personal experience in helping new colleagues do grading conversions that an aversion to math is by no means confined to students.

At the outset, I realize that most readers are going to disagree with me concerning some or all of the points that are presented in this essay. Therefore, when I characterize a particular proposition as a "myth," I do not intend a put-down of those who subscribe to it. I only mean that, from my perspective, the stated proposition is erroneous or exaggerated for the reasons that I will give.

The myths that I describe below are held by students, faculty, employers, or some combination of the three. With each myth, I will endeavor to describe its content, its likely genesis, and then the countervailing truth about the same subject. I will begin with the myths that are primarily student-held and then proceed to describe faculty-held myths about law school grading.

MYTH NO. 1: LAW SCHOOL GRADES ARE A GOOD PREDICTOR OF WHO
WILL BECOME A GOOD LAWYER

This myth is held to some extent by students, faculty, and employers alike. I lump it here with "student-held myths" only because I think that a belief in this myth causes more damage to students than to faculty or employers. There is perhaps some truth to this myth, but only if we are willing to define "good predictor" as "the most convenient and easiest to quantify."

I have little doubt that there is, on the whole, some positive correlation between law school grades and success in law practice. However, to say that there is some positive correlation says nothing about the *strength* of that correlation. To use an analogy, the LSAT is currently the best known quantifiable predictor we have of a prospective law student's performance on first-year law exams. The LSAT score, however, ends up accounting for less than fifty percent of the variance among law students in their first-year law school performance. To put this another way, factors that the LSAT *cannot* measure end up accounting for more of the variance in first-year grades than do factors that the LSAT *can* measure.

The connection between law school grades and success in law practice is even more tenuous than is the connection between the LSAT test and first-year law exams. At least the LSAT is one timed test that is geared to predicting success on a different series of timed tests. While law school tests attempt to measure issue-spotting and legal analysis—two skills that are certainly important to the practice of law—real law practice generally allows a lawyer the luxury to ruminate on a client's problem for more than just three hours. Furthermore, factors such as interpersonal skills, perseverance, rain-making, and attention to detail—all of which are crucial to the success of any lawyer—are either not measured effectively or at all by law school exams.

If all of the above is true, then why do employers place so much weight on law school grades in their hiring decisions? Probably for the same reason that law schools continue to place so much weight in admissions decisions on an admittedly imprecise predictor such as the LSAT: it is fast and it is presently the best quantifiable predictor we have. The thing that students need to remember is that the "best" predictor may still not be very good, and where you start your practice is almost never where you end up.

MYTH NO. 2: LAW SCHOOL GRADING IS NOT REALLY ANONYMOUS

This fairly prevalent student-held myth is truly one of my favorites. The reason it is one of my favorites is that it makes the wild assumption that law

professors *want* to know the identity of the person whose test they are grading. The truth is, that is the last thing that I or any other grader generally wants to know.

There are a number of reasons that law professors cherish anonymous grading as much as law students do. First, most of us don't really want to have the power to decide which students have better job opportunities than others. The next best thing to not having that power at all is being able to exercise that power in the most impersonal way possible. Second, most of us professors would not trust ourselves, despite our best intentions, to be completely objective if we knew the identity of the person whose exam we were assessing.[1] Given that some exams will inevitably deserve low grades, it would be hard in a non-anonymous system to give low grades knowingly to identifiable students we like. Anonymous grading assures that our subconscious biases will play no part in the score that we give.

The final reason that most law professors like anonymous grading is that it is good for faculty-student relations. If a student with whom I have a good out-of-class relationship gets a terrible score on my exam, that student and I can generally still have a good rapport despite the bad grade. One reason for this, I am convinced, is that the student knows that there was nothing personal in the grade he received. This is one reason that grading papers in seminars or legal writing courses is a more delicate proposition: both the student and the professor know that the professor is grading the overtly personal product of that particular student.

Besides the fact that professors value anonymous grading as much as students, there is a second reason why I find the "no anonymous grading" myth such an unlikely one: it assumes a Great Conspiracy. If there really were this Great Conspiracy going on out there by which the law school told students that grading was anonymous even though it really wasn't, how likely would it be that not a single disgruntled professor would ever blow the whistle?

Any student who holds to this Great Conspiracy theory is making at least two far-fetched assumptions: first, that an entire law faculty could ever uniformly agree about any one issue, including the desirability of perpetuating the anonymous grading myth; and second, assuming we could agree, that somewhere along the way one of us wouldn't inadvertently let the cat out of the bag. Like so many great conspiracy theories, this one gives way

---

1. For this reason, several years ago I changed my seminar's grading procedure so that my co-teachers and I could grade the students' papers according to the students' exam numbers rather than their names. This anonymous grading approach would not be viable in writing seminars where, unlike in our seminar, each student is writing on a completely different topic.

too much credit to those who are thought to be the architects of the conspiracy.

Given all of the reasons above why this particular myth is so improbable, why does it persist? The answer is one that will repeat itself as a justification for many of the myths described below: because it helps students sleep through the night. Lest one get the impression that I take a cynical view of students, let me quickly clarify that my real point here is merely that students are people, and people like to rationalize. Professors are also people, and we, too, like to rationalize. We just rationalize about different things, like why our path-breaking article did not get the law review placement that it deserved.

The student rationalization that is behind the "grading is not anonymous" myth is something to the effect that, "I really knew the material, so the reason I didn't do well on the test must be that the professor didn't like me." Perhaps no student ever put it in such stark terms, but the purpose of the myth is that it serves to further a more general proposition that exam-grading is not objective and therefore is not "accurate." Whatever the truth of the general proposition that grades are not "accurate," it is a proposition that will continue to be popular as long as students continue to be human. For if students accepted that exam-grading were truly an accurate measure of a student's knowledge of the course material relative to other students, then ninety percent of the students would have to attach some legitimacy to their place outside of the top ten percent of the class.

MYTH NO. 3: "MY GRADES WOULD BE HIGHER, IF ONLY I COULD LEARN 'EXAM TECHNIQUE'"

This very prevalent student myth is probably heard most frequently early in the second semester of law school, after the students have received their very first set of law school grades. This belief is not entirely myth; there are, indeed, some students whose initial exam grades in law school are artificially low because of some problem in exam-taking technique rather than because of deficiencies in their substantive knowledge. Furthermore, students as a group probably become more adept at taking exams as they take more of them. However, how many students can legitimately expect that their exam technique will significantly improve over time *relative to their peers*, whose exam techniques are also presumably improving with experience?

I am convinced that such "technique-deficient" students are the exception rather than the rule. When students come to me after receiving their exam grade to ask how they can write a better exam, I first re-read their exam closely to see if I can discern any problems of style or approach. Typically,

however, my response to the student after re-reading the exam is the same: Know the subject matter better and be able to apply the law to the facts of the exam.

This, of course, is not the answer that students want to hear. Instead, they want to be told some secret about exam "technique" that somehow they have been missing. After all, these students are *positive* that they knew the material much better than some classmates who somehow (presumably because they knew the secret) got better scores on the final exam. It is the prevalence of this myth that keeps alive a whole industry of nationwide "exam technique" seminars that offer the dubious guarantee that if the student's grades do not increase, the student can take the course again "for free"!

This myth is perhaps the most understandable of all the student grading myths, because it is so directly a function of the student's need to "sleep through the night" (Myth No. 2). As human beings, we all have a need to explain away our failure to perform at the level where we thought we should be. I generally do not try hard to dispel this myth directly when confronted with it. Instead, I will have students read either my model answer or another student's higher-graded answer so that the students can see all of the substantive ways in which their answers were lacking.

In identifying this myth, I should quickly add that I am not thereby making the further suggestion that the grades received on exams are a reliable measure of the value of the course to the student, let alone an accurate predictor of a student's ultimate success as a lawyer. As noted in Myth No. 1, while exams provide a rough measure of some of the skills relevant to the practice of law, they do not and cannot measure all of the necessary skills. I am suggesting merely that students tend to overplay the importance of exam "technique" as an explanation for any unflattering disparities in performance between themselves and their peers.

MYTH NO. 4: LAW SCHOOL ADMINISTRATORS SERVE AS A GRADE "APPEALS COURT" TO REVERSE PARTICULARLY UNFAIR GRADES OF PROFESSORS

This student-held myth is not nearly as prevalent as many of the others, but it is a myth that nevertheless has enough adherents that there is a longstanding file in the Associate Dean's Office under the heading, "Grading Disputes." For a couple of reasons, however, confronting a student who holds this myth is actually one of the easier problems that I face as Associate Dean. First, disposing of this problem is usually very quick: "Dismissed for lack of jurisdiction." Second, there is a long history and some relevant

precedent to draw from in explaining to students why administrators almost never tinker with the grades given by faculty members.

In my most recent letter to a student on this subject, I wrote the following:

> The law school has a longstanding policy regarding faculty autonomy over final grades. In a January 9, 1985 memo concerning a grading dispute, then-Acting Dean Philip Shelton wrote: "Grades given by faculty members are final. . . Grading is an exclusive responsibility of the teacher and no further review process exists."

I can imagine, I suppose, instances in which the administration would have to get involved in a grading dispute, but these would almost have to involve allegations of clear faculty misconduct. For example, suppose a student alleged that her refusal to respond to a professor's sexual advances caused the professor to give the student a low grade in a writing seminar in which the papers were not anonymous. Without a doubt, that would be a case for administrative intervention of some kind.

When, however, the only allegation is that the professor made some kind of substantive mistake in the professor's assessment of the student's exam or paper, then it would seem inadvisable for a non-expert to attempt to second-guess the professor's professional judgment. Furthermore, once the teacher learns the identity of the exam taker who is now appealing the original grade, the principle of anonymous grading has been destroyed.

Having dispelled the myth of the Associate Dean as a grading court of appeals, let me add that I can certainly understand the student's wish that it would be so. After all, these students are law students, and just like any good lawyer, they feel an almost inherent right to *some* avenue of appeal when they've been wronged. Indeed, occasionally law students (though never at my school) have taken their grading disputes to the real court system,[2] but alas, without ever achieving any better result there than they did at the doorstep of the Associate Dean's Office.

MYTH NO. 5: STUDENTS GENERALLY BENEFIT FROM A SYSTEM IN WHICH EMPLOYERS ARE GIVEN AS LITTLE INFORMATION AS POSSIBLE ABOUT RELATIVE CLASS STANDING

About five years ago, the law school where I teach went from a grading system that included a student's exact class rank on the transcript to a system in which, except for the top ten students in the class, no student knows his or

---

2. *See, e.g., Ex-Student Sues Law School Over Grades*, HARRISBURG PATRIOT, Aug. 20, 1997, at B5.

her precise class standing. Instead, the Registrar's Office publishes a grid which shows the grade averages at various percentile cutoffs. The theory, I guess, was that by giving employers less specific information about precise class rank, employers would be less grade-focused and would consider other, non-grade factors in their selection of new lawyers.

My sense at the time was that we were following a national trend among law schools, but I have my doubts both about the effectiveness of and the theory behind this change. One justification for the change, in particular, really had me scratching my head. This argument said that by giving employers less information about the meaning of grades, we were really doing them a favor since employers tended to attach "too much weight" to differences in class ranking among the large middle segment of the class. My first response to this argument is that if I were an employer, I would want to make my own judgment about the extent to which differences in rank really mattered, since it would be my firm's livelihood that was riding on whether I attached the appropriate measure of weight to these differences.

Second, if there was a problem of employer misperception here, it would seem that the more appropriate response would be to give the employers more information, not less. Indeed, one of my colleagues made an alternative proposal that every employer receive with each transcript not only the individual student's rank, but a grid which demonstrated to the employer just how clustered the middle of the class really was. This alternative proposal was not adopted.

Beyond questioning the theory supposedly supporting the change, which seemed to be motivated by some combination of self-interest and paternalism for those apparently hapless employers, I wondered then and still wonder now whether this change really helps students. When I was a member of the recruiting committee for a major bank's in-house law department, I would always assume the worst whenever there was an ambiguity about an applicant's class standing. If a student said nothing about grades or class standing, I would assume that the student was near the bottom of the class. If a student said "top fifty percent," I would assume that the student was just inside the top fifty percent line.

What I don't think our students fully appreciated when they pushed for the abolition of class rank information is that no matter where you draw the line, some student is going to get hurt. Suppose, for example, that you had a system that indicated grade cutoffs at the tenth-, twenty-fifth-, and fiftieth-percentile levels. If you were a student in the top twenty-seven percent, the most you could say about your class standing is that you were in the top fifty

percent.[3] If you were a student in the top fifty-seven percent, the most you could say about your class standing is either nothing or that you were somewhere in the bottom half of the class. I am not sure that these students are any better off—and arguably, they are worse off—than they were in a system that provided more information on class standing to employers.

The other interesting phenomenon about class rank information is that even with schools that claim to give no information on that score, certain prospective employers will nevertheless develop their own system for discovering it. Yale Law School, for example, purports to have a "pass/fail" system and in fact provides no information about relative class rank. However, it turns out that Yale really has four grades: "high pass," "pass," "low pass," and "fail." Furthermore, although most law firms may be happy to hire any Yalies, no matter where they fall in the class, law schools are more discriminating in their new faculty hiring even with respect to Yale graduates.

Thus, you will hear at faculty and personnel committee meetings discussions of Yale-educated teaching candidates such as: "Well, she got 9 high passes out of 19 graded upper-class courses, but Susie Smith from last year had *14* high passes out of 18 graded upper-class courses." In effect, we professors yearn to know as much as we can about the very information that we have helped to obscure from the law-firm employers that come to our campuses.

Having expressed my skepticism about the common trend to abolish or diminish class ranking information, I should note for the record here that I ended up voting for our students' proposal. My feeling was that if the students and I disagreed about the efficacy of this effort to enhance their job opportunities, I would defer to them since it was their job prospects that were on the line rather than mine.

MYTH NO. 6: LAW SCHOOL GRADES HAVE OBJECTIVE, NON-RELATIVE MEANING

This myth of "non-relativity" is one that is held, to one extent or another, by students, employers, and faculty. Non-relativity says that there is some absolute meaning that can be attached to an "A" or a "B" or a "C," wholly apart from where that grade places a student within a particular class. The fact that some employers buy into the non-relativity myth helps explain why

---

3. I suppose this student could also indicate what the 25% cutoff was, and that her grades put her close to it. It does seem a little strange, however, for the student to draw the employer's specific attention to a standard that she did not even meet.

there has been such a trend toward grade inflation in law schools nationally.[4] One simple, if somewhat crude, justification for grade inflation is that if some employers think that an "A" is a good thing in the abstract, then by golly, let's give out more "A's."[5]

Indeed, a colleague and I unsuccessfully tried a few years ago to inflate our relatively stingy grading scale with a proposal that began as follows:

> If potential employers were wholly rational, they would care about only two facts in assessing a student's grades: where a student job applicant ranked within his or her class, and the overall quality of the students at the applicant's school compared to those at other schools. The nominal figure assigned to express a student's grade average would be irrelevant except insofar as it gave information about the student's relative position within the class.
>
> In conversations that we have had with employers and career service professionals, we are convinced that whereas most employers do not attach independent significance to a student's nominal grade average, some employers do attach significance to the grade average number for its own sake. Our sense is that such irrationality is less prevalent with large firms and more common with employers who are less frequent players in the market.

Perhaps we fall prey to the non-relativity myth because we are thinking back to our third-grade math class, where you either could do your multiplication tables correctly in the allotted time or you couldn't. In that setting there really was an absolute standard, and all the relevant parties could agree on what the appropriate minimum competence level was and how to measure it. Law school is not like that. Even if the faculty could agree on which skills were necessary for minimum competence, we could probably not agree on how to objectively assess them. And even if we could agree on how to assess them, there would be no way to ensure that different graders would be consistent in applying those assessments.

Somewhat to my surprise, there are some law school professors who buy into the myth of grade non-relativity.[6] There are at least two ways in which I

---

4. *See, e.g., High-Grade Fever,* RICHMOND-TIMES DISPATCH, June 16, 1997, at A8; Jonathan Yardley, *The High Cost of Grade Inflation,* WASH. POST, June 16, 1997, at C2.

5. A more "defensive" and less crude justification for grade inflation is that it responds to a collective action problem: if lots of other schools are inflating their grades, we had better follow suit or our students will be relatively disadvantaged in a marketplace where some employers attach weight to absolute grade letters or numbers.

6. One who clearly does not is Nicholas L. Georgakopoulos, author of the insightful article, *Relative Rank: A Remedy for Subjective Absolute Grades,* 29 CONN. L. REV. 445 (1996).

have seen this. First, most law schools rely on the LSDAS index score as a major factor in placing prospective students in the categories of "certain admit," "go to committee," and "certain reject." These LSDAS index scores, however, use an undergraduate GPA figure that has not been adjusted for either the relative grade distribution at a particular undergraduate institution or for the relative strength of the student body at the institution. Both of these pieces of information are available to law schools on a student's LSDAS report, but many if not most law schools do not bother to adjust an applicant's undergraduate GPA with these two pieces of important contextual information.[7]

The second way I see the myth played out among some of my colleagues is in discussions of seminar paper grading. We have a grading system that has a mandatory median, even for seminars. Thus, our system is premised, at least to a large extent, on the relativity of grades. Nevertheless, I will hear professors say that the papers in their seminar really deserved higher grades but the mandatory median prevented them from giving all of the papers the grades they deserved. This adherence to the non-relativity myth would be harmless except for the fact that the professor then apologizes to certain seminar students by sharing these sentiments, and the affected students end up feeling cheated about not getting the grades they now believe they truly deserved.

MYTH NO. 7: FEWER GRADATIONS IN A GRADING SYSTEM ARE BETTER THAN MORE GRADATIONS

I think of this as mostly a faculty-held myth, although some students buy into it as well. My law school has a grading system with lots of gradations, which I like. We can give scores of anywhere from 65 to 100, including every whole number in between. I have heard many colleagues over the years suggest that we ought to have a "simpler" system with fewer gradations, perhaps A, B, C, and D. The most extreme version of this is that we should go to a pure pass/fail system.

Speaking as one exam grader, I see plenty of costs to a system of fewer gradations, but no benefits. The attraction of more gradations is that nobody is forced to use them all, but the additional gradations are always there for graders who believe that their raw points put them in a position to make finer distinctions. In effect, a system with more gradations allows graders to

---

7. *See generally* David Kaye, *An "A" Is an "A" Is an "A": An Exploratory Analysis of a New Method for Adjusting Undergraduate Grades for Law School Admissions Purposes*, 31 J. LEGAL EDUC. 233 (1981).

design their own scale with fewer gradations if they choose. For example, if the grader believed that she could do no better than to separate exams into five different categories of quality, then under our system the grader could simply use just 5 of the 36 possible points on the system, such as 70, 77, 84, 91, and 98. In other words, even in a system with lots of gradations, nobody ever forces graders to make finer distinctions than they believe they can make.

While there are no disadvantages to having more gradations, as explained above, there are two significant drawbacks to having a grading system with fewer gradations. One is a student incentive problem[8] and the other is a grader line-drawing problem. The incentive problem has arisen in certain of our lawyering skills courses that use a modified pass/fail system that includes a "high pass," "pass," "low pass," and "fail." Once a student decides that he cannot achieve a "high pass," the student has a tendency to do the least amount of work possible in order to still get by with a pass. Within our normal 36-point system, the "pass" grade covers an enormous landscape of 75 to 89. Thus, the modified pass/fail system creates the unfairness that a student who would have received an 89 in the straight-number system will be treated the same as a student who would have received a 75. Two of our three Pretrial professors were so fed up with the incentive problem of the modified pass/fail system that a year ago they switched the grading of their Pretrial courses to the usual 36-point scale that we use in non-skills courses.

A common complaint of a grading system that includes lots of gradations is as follows: "Can any professor really say that there is a difference between the exam that received an 89 and the one that received a 90?" There are two responses to that complaint: First, the professor cannot say that there is really a difference, but the slight difference in final grades is nevertheless based on the professor's best available information, namely the professor's raw score points. Second, the beauty of the multi-gradation system is that the difference between an 89 and a 90 only ends up mattering a little bit: just a single point in a system that has 36 of them.

Contrast that result with the line-drawing problem that arises in a system with only a couple of gradations: when the grader only has a couple of gradations, the stakes are raised at each gradation. Thus, for example,

---

8. The incentive problem is significantly reduced, but not eliminated, in courses where the grade is based on a single final exam. As an example of how the problem is not completely eliminated, I think of my approach to the state bar exam, which had about an 85% pass rate at the time I took it. This was a grading system with two gradations: pass and fail. Knowing that all I needed to do was avoid being in the bottom 15% of test-takers certainly had a detrimental effect on my overall incentive to devote additional hours to studying for the test.

contrast the fates of the same students with an 89 and a 90 in our school's modified pass/fail system: one gets a "pass" and the other a "high pass." Here a minor difference in raw points suddenly ends up mattering a lot, even though by all rights it should not. One response that I have heard to this problem is that the grader will look for "clusters" and "natural breaks" and will avoid the line-drawing problem that way. However, a given raw-point distribution may or may not conveniently break down this way.

MYTH NO. 8: IT IS MATHEMATICALLY AND EQUITABLY JUSTIFIABLE TO DEVIATE FROM A PURELY LINEAR CONVERSION OF RAW TO ACTUAL POINTS

This is purely a faculty-held myth, and it is probably a myth of which students are generally unaware. It is also a myth that is somewhat complex to explain, since it involves delving into some math concepts. To understand this myth, you must first understand a little bit about the mechanics of grading. Generally, when a professor grades an exam, the professor gives raw point scores to each question on the exam, totals up the raw points for each exam, and then somehow "converts" the raw point figure to a final grade on the relevant grading scale, 65-100 in the school where I teach.

I cannot tell you the number of times that I have been asked by either adjunct professors or new full-time professors how to convert raw points to our system's actual grades, taking into account also the constraints of our mandatory median. The conversion is actually quite simple: 1. Line up the raw scores from high to low (or low to high); 2. Pick the middle raw score and assign it an actual score within our permissible upperclass median range of 82 to 84 (most teachers choose "84"); and 3. Choose a conversion factor from raw points to actual grades and then convert all of the raw points to actual grades by applying that conversion factor.[9]

The confusion usually sets in with Step No. 3, but even that step becomes fairly simple once the professor sees that the conversion factor is doing nothing more than determining the relative width of the grade curve.[10] The

---

9. Thus, imagine a class with 11 students whose raw scores ended up as: 34, 37, 37, 41, 44, 49, 51, 52, 56, 56, and 59. The middle score is a 49, which becomes an 84. If the conversion factor chosen is 1 = 1 (that is, one raw point above or below the raw median equals one actual point above or below the actual median), then the scores would come out as follows: 69, 72, 72, 76, 79, 84, 86, 87, 91, 91 and 94. If the grader thinks this curve is too wide, then using a 1 = .5 raw-to-actual conversion, the grades would end up: 76, 78, 78, 80, 81, 84, 85, 86, 88, 88, and 89.

10. Our system gives little guidance on what the appropriate width of the curve should be. The one piece of guidance we are given on curve width is that approximately 10% of the grades should be 90 or above. There is also somewhat vaguer guidance with respect to the bottom of the curve.

rank-ordering and relative distance between the exams has already been determined by the assignment of raw scores. Granted, the assignment of raw scores to each of the exams may have been far from a perfect process, but what professors need to appreciate is that however imperfect that process may have been, it is still the best information they have about the relative standing of the various exams.

What I have found fascinating in my conversations with various senior colleagues about how they convert their raw points to actual grades is that many, if not most, of my senior colleagues do not maintain a consistent conversion factor throughout the range of the curve. In other words, these colleagues make raw point determinations about the relative distance between and among exams, and then deviate from these raw-point determinations for no apparent reason other than an uneasy sense that their allocation of raw points to the various exams was itself a less than precise process.

There are three common sources of deviation from a pure linear conversion. One is to give a more favorable conversion factor to raw scores at the very top (to reward the "best" exams)[11] or at the bottom (to avoid especially low grades).[12] The second deviation from a linear conversion is to use a higher conversion factor for scores above the median than for those below it, thus creating a kind of "single-tailed" curve.[13] Both of these kinds of deviation are often justified by some reference to the inherent randomness of the initial assignment of the raw points.

---

11. A colleague once explained to me that he sometimes gives an extra point or two to the exam with the highest raw score in order to recognize the fact that it was the highest score. Other than a general philosophy that the rich should get richer, I never could understand the rationale for that bonus.

12. Many colleagues tell me that they deviate from a pure linear conversion whenever that conversion would cause an exam to end up with less than a final grade of 70, which is the cutoff for passing. The argument is that as to these exams, the grader should make an "independent assessment" of whether the exam should pass. After almost 10 years in the business, I still don't know how a colleague can make such an "independent assessment" of competence when there is no external and objective benchmark (beyond the raw-point distribution of the exams, which is the standard that is being overruled) against which to measure whether an exam "passes" or "fails" in the abstract. If there were such an objective benchmark, then why is it that today so few law students fail courses whereas 40 years ago perhaps one-third of each class flunked out? Do we believe that the quality of our student bodies has changed that dramatically, or is it that the "objective" benchmark against which we are supposed to measure passing or failing really isn't so objective after all? Perhaps a better argument for deviating from a linear conversion at the line between passing and failing is that the stakes are dramatically different than at other points on the grade scale. In other words, since our grading system arbitrarily creates a "cliff" at the point on the scale between a 70 and 69, we should give students the benefit of the doubt if they end up fairly close to a 70 under a purely linear conversion of raw to actual points.

13. To use my note 9 example, the grader might decide to use a 1 = .5 raw-to-actual conversion for all exams *below* the median, but a 1 = 1 conversion for all exams *above* the median. Thus, the final distribution of the exams in that example using this conversion approach would be: 76, 78, 78, 80, 81, 84, 86, 87, 91, 91, and 94.

What these graders fail to appreciate, however, is that even though we may know there is some randomness in the initial assignment of the raw points, we cannot say which exams benefited and which suffered as a result of the randomness. Without knowing that (and if the graders did know it, they could simply correct that randomness by fixing the raw points themselves), the deviation from a linear conversion of raw-to-actual grades is hardly a "solution." Instead, these deviations simply add yet a second form of randomness—this one completely avoidable—to the grading process.

The third common deviation occurs in seminars, where after deciding what a paper's "true" score is (Myth No. 6), the professor gives that "true" score to every paper above the median, but then gives the median or below to papers that are in fact very close in raw score to the papers that ended up receiving significantly higher grades.

Thus, for example, imagine a seminar with five papers. On the first read, the professor assigns the five papers the following scores: 99, 97, 93, 92 and 91 (the scores these papers "deserve"). Instead of complying with the mandatory median of 84 by moving the entire scale down by 9 points (which would, using a one-to-one conversion factor, yield a 90, 88, 84, 83, and 82),[14] the professor instead gives a 99 and 97 to the first two papers, and then 84's to the other three. Thus, the original 93 paper, which was just 4 raw points away from the original 97 paper, ends up with a score (84) that is 13 grade points away from where the 97 paper ended up (97). As noted above, some professors will tell the students what their original "true score" was, and then apologize for the final score that the student received by blaming the mandatory median. The real unfairness here is not in the mandatory median, but rather in the professor's exaggerating the relative distance among the papers.

This third form of deviation from a linear conversion is fueled primarily by the graders' adherence to three other, related myths about grading: first, that grade numbers have absolute rather than merely relative meaning (Myth No. 6); second, that a forced median is an undesirable limit on the graders' freedom to give students all the grade wealth they deserve (Myth No. 9); and third, that in any event, student grades should of course be higher in a small-class or seminar setting (Myth No. 10).

---

14. The grader, of course, could use conversion factors other than 1 = 1, see note 9, to affect the width of the overall curve.

MYTH NO. 9: MANDATORY MEDIANS AND OTHER SUCH RESTRICTIONS
ON A GRADER'S AUTONOMY ARE UNDESIRABLE AND UNFAIR

More than two years ago, I was able to appreciate the pervasiveness of this myth, held by both students and faculty, when a colleague and I attempted to further standardize the grading practices at our school by switching from a mandatory median to a mandatory mean.[15] As best my colleague and I could gather from reading the legislative history of our mandatory median and from talking to faculty members who were here at the time it was instituted, that system was meant to respond to what was perceived to be a collective action problem among faculty that related to grading.

Prior to the mandatory median, there were no constraints on our faculty's grading discretion other than the "common law." In those laissez-faire days, certain faculty members were doling out either consistently less or consistently greater "wealth" to students, if grades can be thought of as a form of wealth. This pattern had at least three detrimental effects: 1.) it tended to generate resentment among faculty members against those colleagues who were perceived to be currying favor with students by giving them more than a "normal" share of grade-wealth; 2.) for first-year students, it introduced a random element into a student's grade average based on the happenstance of which professor the student was assigned to for a particular course; and 3.) for upperclass students, it created an artificial incentive to choose courses that were taught by consistently higher-grading professors or to avoid those taught by consistently lower-grading professors.

What prompted my colleague and me to propose a mandatory mean instead of a mandatory median is that even a mandatory median gave professors significant discretion to dole out varying amounts of grade-wealth in their classes while still complying with the median. The reason that these variations were troubling is that relative class rank, not absolute GPA, is what ends up mattering the most to students. Therefore, when one teacher gives more total grade-wealth than other teachers, the effect is the same as if that generous teacher were deducting points from other teachers' students.

By conducting an empirical study of grading patterns across professors, we learned that these variations in the distribution of grade-wealth were not just theoretical, but in fact regularly occurred in practice. However, when we attempted to further normalize grading practices across classes with the

---

15. The median is the middle score within a group; the mean is the arithmetic average of the scores within a group. Thus, in a class of five students with scores of 80, 81, 84, 92, and 98, the median would be 84 and the mean would be 87.

institution of a mandatory mean, we were quickly deluged with several arguments supporting the myth against an increased standardization of grading practices.

The most common argument against the imposition of a mandatory mean was that perhaps the current differences in grade means among professors were in fact justified by the differing performances of students in different classes. The simple response to this is maybe so, but how in the world would anyone ever be in a position to know that? There are really only two things that a professor grading a set of exams can determine about those exams with any degree of certainty: the relative rank ordering of those exams and the approximate raw-point distance apart of those exams from one another.

As to any other assessment of those exams, we simply lack sufficient information. Are we in a position to assert that our exams are better overall than those being graded by another professor who teaches a different course? Are we in a position to assert that our exams are better overall than those being graded by another professor who teaches the same course? If we are not in a position to assert that our students' overall performance was better or worse than the overall performance of some other class, then why should we be able to distribute more (or less) grade-wealth than some other professor distributes to his or her class?

We might believe that we are at least in a position to assess how the performance of this class compares with that of classes from previous years in which we have taught this course. I am personally dubious about the likely accuracy of even this sort of assessment, given memory lapses, differences in the particular performance-measuring device we use from year to year, and even differences in how well we taught the same course.

. Suppose, however, that we graders were in a position to make a relative assessment of a particular class's performance as against others we have taught in the same subject area. To take an extreme example, suppose we used the very same multiple-choice exam from year to year (most of us, of course, do not have such a reliable way to compare class performance from year-to-year). Even in the case of the identical multiple-choice exam, however, our ability to compare our classes' scores from year to year would not provide us with sufficient information to determine an appropriate grade mean in any given year.

The reason is that we would still lack at least two key pieces of information to make an informed judgment about how much total grade-wealth to distribute to our class relative to those taught by other faculty. First, we do not know how well students are performing in other classes. It may be that the entire student body is stronger or weaker, in which case the rise and fall that we are seeing in our classes' performances would similarly be

reflected in other professors' classes. Second, even if we know something about the performance of students in classes we are not teaching, we do not know how other faculty will choose to assess that performance. To put this point another way, we lack information not only about the relative performance of our particular class compared to others but also about the way in which that performance is being assessed.

I can certainly see how faculty members might believe that an ability to distinguish between the performance of their own classes from year to year puts them in a position to distribute more or less total grade-wealth from year to year. The problem with this logic, however, is that it ignores the reality that when we give our own class more or less total grade-wealth than that given by a colleague, we are making an implicit (if unwitting) statement about the relative effort or performance of our group compared to theirs. Yet, as noted above, we lack the appropriate information about the other groups and the other graders to make that judgment.

The only objective and discernible information that we have about the relative differences in the quality of various groups of students is the past performance of those particular students relative to the rest of the group. Consistent with this reality, under our failed mandatory-mean proposal a particular group's past performance would have determined what the permissible mean range would be for that class.[16]

A second common argument that we heard against our proposal for further standardization of grading across classes was the timeless faculty favorite: What about academic freedom? The response to this argument is that academic freedom does not include making relative determinations, in the absence of necessary information, about how much total grade-wealth that we can distribute to our students compared to that given out by a colleague. Under our mandatory mean proposal, we graders would have retained our freedom to determine the only two facts about which we have reliable information concerning our students' performance: their rank-order

---

16. For example, if a class of 10 students had overall grade averages of 77, 77, 80, 81, 86, 88, 88, 90, 91 and 92, then the mandatory mean for that class would be an 85. The existing mandatory median system at the school where I teach does take into account the past performance of each class, but in a strange way. For upperclass courses, the usual permissible median range is 82-84. If the median grade average of students in a particular class is higher than 84 or lower than 82, then the permissible median range is extended up or down (but not moved wholesale) to reflect the abnormal class. For example, if the median student grade average in a class were 79, then the permissible median range for that class would not be 82-84, but 79-84. What is strange in this system is that the professor for such a class still has the option to give an 84 median instead of being restricted with such an abnormally low-grade class to a median range of, say, 79-81. It is similarly odd that in a class with a median grade average of 87, the median range in this system is 82-87, thus giving the professor the discretion to give a median as low as 82 in a class of historically high performers.

within the group being assessed and their approximate distance from one another.

The third and final common faculty argument we heard against our proposal for a tighter standardization of grading was as follows: Doesn't this all suggest a precision about grading which just isn't realistic? In order to answer this question, we must distinguish between precision of grading within a course and precision of grading across courses. As to precision of grading within a course, our proposal had no effect one way or the other. We graders would all still be free to make whatever relative rank-order and distance determinations that we feel we could make about the students within our class.

As to precision of grading across courses, our proposal was in fact premised on an assumption of *im*precision about grading, not precision. That is, our proposal assumed that any disparate grade-wealth distributions across courses would necessarily be imprecise and random in a world where the graders lacked relevant information from which to make an informed judgment about the relative level of total grade-wealth that we ought to distribute to our class. The appropriate response to this assumption of imprecision, we felt, was to normalize as much as possible the relative grade-wealth distributions across classes.

Students' complaints about mandatory medians or other grade-standardization devices are typically that these grading restrictions prevent professors from giving them the high grades that they truly deserve, and that they foster competition among students. However, students fail to remember that unregulated grading in their course must also mean unregulated grading across the board. What guarantee would any student have in an unregulated grading system that he would end up with the most generous graders? If the student did not, then the lack of a forced curve would suddenly become a burden rather than a benefit for that student. In a system of completely unregulated grading, there would be winners and losers among students, but the differences in outcomes would likely be based more on randomness and strategic course-selection than on verifiable differences in group performance across classes.

There are two forms of unfairness that can occur in an unregulated grading environment. First, teachers could disadvantage their own students by giving out a disproportionate number of low grades. Second, and conversely, teachers could disadvantage *other* teachers' students by giving their own students a disproportionate number of high grades, thereby hurting the class ranks of students that are not in their classes.

MYTH NO. 10: STUDENTS IN SMALL CLASSES AND SEMINARS SHOULD
GENERALLY GET HIGHER GRADES

This myth, held by both students and professors, is a common one. We professors would probably be averse to admitting a couple of possible reasons, be they conscious or subconscious, why we like to give higher grades in a smaller class. First, we generally get to know the students better in smaller classes and are more aware of how hard they are working, thus making it only natural for us to want to give the students as much grade-wealth as possible for their efforts. Just because we know more about the effort put out by each of our students in a small class, however, we still know nothing about how their effort or performance stacks up against those students in classes that we are not teaching, large or small.

Second, it is only natural for us to believe that, compared to most of our colleagues, we can inspire a greater effort from our students than those same students would put out in other classes. We might think that this is especially true in a small-class setting. Because the students in my small class work harder, the subconscious logic might go, of course they deserve more total grade-wealth than that which is being given by my less-inspirational colleagues. The problem with this logic, aside from the human tendency to self-delusion, is that we cannot all be right. Even if we all thought we were right and acted on that belief, the effect of the higher means for students in small classes would be canceled out except to the extent that certain students took more small classes than others. Yet it would seem odd to foster a system in which students' overall class standing should be in part a function of how many small classes they take.

The small-class argument reminds me of an analogous issue that existed in the bank law department where I once worked. That law department consisted of about eighty lawyers, divided into eight substantive sections of five to fourteen lawyers. Every year around raise time the General Counsel would create an average raise per lawyer, the same for each section, that would determine how much total raise money was given to each section head. The section heads were free to distribute that money within their section in any way they saw fit, but they were all limited to the same average raise per lawyer.

Invariably, one or more section heads would grumble to the General Counsel about how hard *their* section had worked during the past year and how their section should therefore receive a greater raise per lawyer than other sections. Just as invariably, the General Counsel would ask these disgruntled section heads whether they could provide him with any reliable information which demonstrated that their lawyers worked better or harder

than those of other sections. The raise amounts were never changed.

Some might try to fault the bank analogy to a law school grading system by pointing out that whereas the bank's system was a zero-sum game, the grading system need not be. The fact is, however, any grading system *is* a zero-sum game to the extent that higher grades in one class necessarily deflate the currency on which overall class ranks will be based. In other words, when the absolute grade numbers of a student's classmates go up, that student's own relative class rank will go down.

Another argument I have heard to justify higher grades in small classes is that the small sample size makes it statistically inappropriate to apply a forced grading curve of any kind. While a small sample size increases the problem of randomness in an application of a forced median or mean, applying the forced standardization (particularly when adjusted to account for the past performance of those in each group) is still superior to a system of unregulated grading for small classes.

Everyone knows that the absence of forced grade standardization in small classes would invariably lead to lots of grade-wealth-per-student being distributed in these small classes, along with all of the attendant collective action problems that occur in an unregulated grading environment. I cannot see why there is anything distinct about a small class (other than the natural tendency of the professor to give higher grades, discussed above) that should cause students who are in it to receive on average higher grades than their peers in large classes. Some faculty will insist that students work harder in smaller classes, but I fail to see how an individual professor is in a reliable position to assess this.

I have heard some faculty and students argue that even if all small classes are not exempt from a mandatory grade standardization, at least writing seminars should be. The unique argument about seminars goes something like this: "But some students are better at writing papers than they are at writing exams." This is probably true, and if so the converse must also be true: "Some students are worse at writing papers than they are at writing exams." Mandatory grading regulations allow a professor to reward students' paper-writing performance relative to one another within the group. Such mandatory systems regulate only the total grade-wealth to be distributed, but each faculty member retains the freedom to distribute that wealth based on the faculty member's relative rank-ordering and spacing of the students within the group.

Perhaps it bears mentioning that the faculty committee which created my school's current mandatory median system specifically considered the exclusion of seminars from the mandatory median rule but ultimately decided against it. That Committee report noted in part: "The Committee

unanimously believes that seminars should be treated the same as regular courses for grading purposes. The tremendous discrepancy between grades in seminars and regular courses and among seminars is perceived as a grossly inequitable part of our current system, and we agree."

CONCLUSION

Grading myths and folklore will always be with us, at least as long as grades still matter to students, and law professors do not come from statistics backgrounds. As noted at the outset of this essay, I don't expect that the reader will necessarily agree with all, or even most, of the propositions put forward here about what I consider to be grading myths. However, if this essay causes its readers to re-think even one of their existing assumptions about the subject of law school grading, then it will not have been written in vain.