Relative Rank: A Remedy for Subjective Absolute Grades

Nicholas L. Georgakopoulos
Relative Rank: A Remedy for Subjective Absolute Grades

Nicholas L. Georgakopoulos

Two-thirds of law schools admit to standardizing some of their faculties' grades by using various "curves." That is a sad testimony to the problem of grading in this age of subjectivity. The need for "curving" shows that schools treat identical grades from different professors differently. But curves' scarcity shows how restrictive faculties take curving to be. This short essay describes a slight change that would allow schools to increase the comparability of their grades without the limitations of curves.

Grades appear to show absolute skill, quality, or learning. Excellence, mediocrity, and failure are not relative terms. Graders, however, are subjective. The skill to which a grade may correspond—let alone the dimensions of legal skill—is not defined identically by every professor. In a world where each grade of each student received individual attention, this might not be a problem. However, subjective grading is a problem if each student's grades are processed en masse, such as by averaging. The problem is aggravated if students are competing—for rank, honors, prizes, or the career prospects to which they lead—based on the average grades they receive. These averages may not represent a student's average evaluation if that student's set of graders did not evenly cover the spectrum of grading tendencies.

---

* Associate Professor, University of Connecticut School of Law. I wish to thank Steven Utz, Leonard Orland, Richard Pomp, James Stark and Robert Whitman for their valuable comments throughout the course of the proceedings of the Grading Committee and Dean Hugh Macgill for prodding my thoughts on the topic by appointing me to the committee. All findings and opinions in this article are my own, however, and should not be attributed to the Grading Committee or the University of Connecticut School of Law.

As a further complication, evaluation subjectivities may correlate in
groups of courses; for example, criminal law grades might tend to
differ in the same way from business law grades. If so, students who
prefer one or the other subject will have sets of grades that are not
comparable. Fields of law may draw on different talents, such as quali-
tative versus quantitative reasoning, rule versus relationship orienta-
tions, or formal versus intuitive reasoning. If the holders of different talents
are represented with different strengths in the student body—perhaps
because they are given different incentives to enter law school—this
problem of grading discrepancies may, despite its innocence, be wide-
spread and permanent. For these reasons, students’ performance should
not be compared on the basis of average grade.

Schools seem to consider this a minor problem, albeit one that
often justifies a centralized response, such as curves. A school’s suc-
cess, however, does not depend directly on the accuracy with which it
signals its students’ legal skills. Judges, law firms, and other employers
are, thus, left on their own to decode transcripts that can be obscure,
innocently deceptive, or even manipulative. Some students have faith
that employers are aware of grading idiosyncrasies. Some are demoral-
ized by the possibility of grade-shopping by other students.\textsuperscript{2} The greatest unfairness that I have observed at my law school is the reduced
relevance of upper-class grades.\textsuperscript{3} It is practically impossible for slow
learners to overcome a poor showing in their first year. This problem is
bound to exist in every school that standardizes a smaller proportion of
upper-class than of first year grades.

The problem is not in the subjectivity of grades but in the aggrega-
tion of incomparable ones. A simple fix is offered here. Convert abso-
lute grades into relative ones, insofar as they are used for relative con-
clusions. Grades become relative when each student’s grade in each
class is adjusted in relation to the average grade in that class (also
taking into consideration grade dispersion). An A is converted from a
mark of excellence to a measure of a student’s superiority over the
average student in a given course. The resulting grades can be used as

\begin{itemize}
\item[2.] Id. at 423.
\item[3.] First year courses are graded on an informal but broad curve (from A to D or D+).
Upper-class grades tend to be much more concentrated, primarily because of a reduction in the
proportion of low grades. If the upper-class grades are half as dispersed, they can induce half
as much grading difference. Effectively, each upper-class year is given a 50% weight compared
to the first year. A poor performance in the first year cannot be overcome (even with grade-
shopping). This is only marginally better for slow learners than awarding all upper-class courses
A’s, which would effectively preserve first year rank.
\end{itemize}
relative measures for rank, honors, or prizes.

This proposal combats subjectivity with relativity. It is a minimal imposition on faculty and administration, since it requires no change in the conduct of any grader. The burden is borne exclusively by the registrar’s computer. Letter-based grades, GPA, and GPA-based ranking can all stay. After all, average GPA should change to reflect changes in the average student. By measuring rank based on students’ success relative to others who take the same courses, relative ranking counters the biases that grading idiosyncrasies create.

Part I of this essay explains the mechanics of relative rank. Part II discusses its principal advantages and Part III its disadvantages. Part IV offers some examples from actual conversions of absolute to relative grades. The conclusion recommends this simple proposal—which would not have been feasible before widespread computer use—as a way to allocate legal talent efficiently and to maximize the use of that talent.

I. THE MECHANICS AND IMPLEMENTATION OF RELATIVE RANKING

If relative ranking were implemented, grades—as far as they are used for relative ranking—would lose all their abstract and absolute meaning. The ranking of students would be based on how each grade compares with every other grade in the same course. In order to achieve this, after the professor has graded the class, the computer can calculate the average grade. For example, in a 20-student class with 10 B’s (GPA value of 3) and 10 A’s (GPA value of 4), the average grade would be 3.5.

Grades must be compared to the average grade to take relative meaning. But how must the distance of each letter grade from the average be measured? Is it the same to get a B+ in a class with a B average where no A’s were awarded, as it is to get a B+ in a B-average class where several A’s were given? Relative grades must also be ad-

4. Letter grades are only used as an example. This proposal can apply to any quantifiable grading system, whether it consists of actual numbers (as a 100-point scale) or qualitative evaluations which are converted to quantities (as letter grades usually are, or as other attempts to depart from quantifiability, such as an Honors Pass/Pass/Fail system, may readily be). Indeed, with the caveat of programming complexity, the relative ranking system can allow each professor in a school to use a custom grading system, as long as it can be translated to numbers. Some may use 100-point scales, some may use letter-based 4-point scales and some European 10- or 20-point scales. Relative ranks can be computed across the different grading systems without trouble.
justed for dispersion.

The usual measure of dispersion is standard deviation. Inasmuch as it uses the square root function, it may be considered a complex formula. However, it has been used so widely that practically everyone has heard of it. Most everybody has some vague understanding that it measures dispersion. Moreover, most computer programs include it as a function, which makes it easy to program.

The dispersion of letter grades should be measured in units of standard deviation (σ). For example, a B+ (3.5) in one class with a B-average but where A's were awarded may be a +.8σ, while in another where B+ was the top grade it may be a +1.2σ. In essence, the letter grade is converted to a relative grade.

After relative grades are computed for each course, they would be averaged the same way letter grades are averaged for GPA purposes. Thus a relative-GPA is computed for each student. Neither relative grades nor the relative-GPA should appear in transcripts if the goal is to clarify, not to confuse, transcripts.

Once a relative-GPA is calculated for each student, the students are ranked accordingly. Thus, each student receives a relative rank. Only relative rank should be added to the student's transcript.

A computer program can easily do this calculation and ranking. I implemented a test of relative ranking using a spreadsheet program.

5. The standard deviation of a class' grades requires one to have first computed the average (mean) grade of the class. The differences of each grade from the mean are squared and summed. The sum is divided by the number of students in the class. The square root of that number is the standard deviation. Because several students would receive the same grade, instead of adding the squared differences of all, say, B's, one would multiply the squared difference of a B from the average by the number of B's in the class. Thus, the shorter function would be to multiply each grade's squared difference from the mean by the number of students receiving that grade, and sum over all grades before dividing by the number of students in the class and taking the square root to get the standard deviation. One could argue that the actual standard deviation should not be used, but rather the estimate of the standard deviation of the theoretical distribution from which the classes' grades were drawn. This method could be called the best estimate of the dispersion of the notional curve the professor used in this class. This would require dividing by one less than the number of students in the class instead of by the actual number and would result in the standard deviation estimate instead of the actual standard deviation. The difference would be felt in the smallest classes only, where relative grades would become less dispersed. This is the appropriate reaction to the risk that a small class is more likely to present a higher and less diversified risk of classmates who are "undesirable" for relative rank purposes. See infra note 11.

6. I used the spreadsheet program Microsoft EXCEL. A spreadsheet is a data holding, processing, and presentation program with emphasis on the processing. Any grader who is interested the spreadsheet (in EXCEL version 5) I mention in the text may request one, by e-mail if possible. (NGeorgak@Story.Law.UConn.edu).
This spreadsheet used the actual grade distributions from several Fall 1993 courses from the University of Connecticut School of Law. Using corresponding distributions,7 it produces, or “rolls” random grades for eight students and allows the user to input the grades for two hypothetical students. A new set of random grades is drawn and the students are ranked.

Using this spreadsheet, one can get examples of how this system would work. In one hypothetical outcome, I had two students achieve the same GPA by getting the same grade not only in three courses they both took, but also a fourth same grade in different courses. Mr. Hyp O. Thetical and Ms. Para D. Gmatic received an A each in Legal History (where an A is converted to a relative grade of +2.2σ) and in Civil & Political Rights (where an A is +1.06σ) and a B each in Appellate Advocacy (where a B is a -1.44σ). They each got one more B, from different professors. Mr. Hyp O. Thetical got his in Bankruptcy where a B is worth +0.35σ, while Ms. Para D. Gmatic got hers in Business Organizations where it is worth -0.08σ (students in Business Organizations received higher grades than in Bankruptcy). The result is that both have the same nominal GPA. This is misleading because a B in Bankruptcy can be argued to be a better grade than a B in Business Organizations. Relative ranking makes the apparent equality disappear. Although in traditional GPA ranking, they tie for fourth (thus ranking in the top 40% out of the 10 students in the example), Mr. Hyp O. Thetical’s relative GPA ranks him third (top 30%) compared to Ms. Para D. Gmatic’s seventh (top 70%). Thus Mr. Hyp O. Thetical’s transcript might read:

*Hyp O. Thetical:*

- American Legal History: A
- Appellate Advocacy: B
- Bankruptcy: B
- Civil & Political Rights: A
- GPA: 3.5

  Rank: relative: top 30%
  raw: top 40%

The difference between the nominal and the relative rank shows

---

7. Those follow each course’s grade distribution, so that if no C’s were given in Legal History, for example, no C can be randomly rolled, and if 10% received C’s, the odds of rolling a C are 10%.
that he received a grade which was nominally worse than its relative strength. In other words, relative ranking compensates for having taken the “hard” grading Bankruptcy course. Ms. Para D. Gmatic’s transcript would be:

Para D. Gmatic:

<table>
<thead>
<tr>
<th>Course</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Legal History</td>
<td>A</td>
</tr>
<tr>
<td>Appellate Advocacy</td>
<td>B</td>
</tr>
<tr>
<td>Business Organizations</td>
<td>B</td>
</tr>
<tr>
<td>Civil &amp; Political Rights</td>
<td>A</td>
</tr>
<tr>
<td>GPA: 3.5</td>
<td></td>
</tr>
<tr>
<td>Rank: relative: top 70%</td>
<td></td>
</tr>
<tr>
<td>raw: top 40%</td>
<td></td>
</tr>
</tbody>
</table>

The exact form with which a school may decide to report or use the relative ranking may, of course, differ. The point is that relative ranking takes into account the different value that grades have when awarded in different courses.

II. PRIMARY ADVANTAGES OF RELATIVE RANKING

Relative rank alleviates grade “disparity,” the difference that professors’ grading attitudes may make in students’ ranking. This reduces the students’ incentive to shop for “easy” graders and the related incentive of the professor to be nice by giving high grades.

Example: A student is considering taking either Bankruptcy or Business Organizations. He notices that the Bankruptcy course traditionally offers lower grades than the Business Organizations course. Although he student may prefer to take Bankruptcy, the usual ranking scheme provides an incentive to sign up for Business Organizations. If the student does not, he may be relatively more successful in the Bankruptcy course (because, for example, he enjoys it more) but may still receive a C+ (notice that a C+ is only slightly below average in Bankruptcy according to the actual examples below). He risks even getting a C or D. By taking Business Organizations he can get a C even if he is just about the worst student in the class and may even get a B without
having to rise above average.

By making relative ranking depend on the standard deviation instead of the absolute difference from the average grade in the class, this system also avoids creating any incentives regarding choice of dispersion.

Example: Professor Concentrated and Professor Dispersed teach Business Organizations and both give grades with the same average. Professor Dispersed, however, uses a more dispersed scale, giving more A's and C's, and fewer B's by comparison. The different dispersion allows a student who does not want to expend much effort in Business Organizations to take it with Professor Concentrated and be assured of avoiding a C. Adopting relative ranking without adjusting by standard deviation would not cure this effect. Suppose Professor Concentrated gives 50% B's and 50% B+'s, while Professor Dispersed gives 25% A's, 25% B's, 25% B+'s, 25% C's and 25% C+'s. If we did not adjust for the different dispersion in the two classes, a B would be equivalent as a relative grade in the two (and Concentrated's B would be better than Dispersed's C+). By adjusting for the different dispersion in the two classes, the grades take their actual relative meaning. A B from Concentrated is about halfway between a C+ and B from Dispersed. The same undesirable incentives are produced by the raw-GPA ranking, where a student who fears a C will prefer Concentrated and a student shooting for an A will prefer Dispersed.

In this way, relative ranking accounts for both grade “inflation” (the increase in the average grade) as well as grade “compression” (the crowding of student’s GPAs in a very narrow range). The effect of grade inflation is eliminated because relative grades are adjusted for the average. The adjustment takes into account where letter grades are on

8. A C+ is 2.5 raw GPA points and a B+ is 3.5 in this and all subsequent examples. There are no minus grades.

9. Grade inflation, however, does survive as a problem—albeit mitigated—under relative rank because it causes the disproportionate size of the two tails of the grading distribution. Suppose the average grade in a class is B, the highest possible grade is an A, but the lowest (used) grade is a D. The D is twice as far from the average as the A. If the students who received A’s did not all perform equally well, the small latitude for differentiation that exists above the average grade implies that some superior students can claim their grades do not fully reflect their superiority. They are treated equally with less able students who still received A’s. But relative ranking ameliorates rather than aggravates this problem of grade inflation, since the
average in each course. Grade compression is eliminated because the
measurement of distance from the average in units of standard deviation
means that giving half the class B+'s and half B's results in equally
dispersed relative grades as would having a wider range of letter
grades.

Moreover, relative ranking does not give the shopping incentives
that can exist under a system where different types of courses are treated
differently for GPA or ranking purposes. If a school were to separate seminars from regular courses for ranking computations, for example, students would still have an incentive to shop for the most easy-grading seminars and the most easy-grading, large classes. This incentive to shop would exist under all types of rules, such as “maximum curves” (for example, a requirement that less than 50% of students receive A's). The students would still prefer the course which gives 49% A's to the course which gives 10% A's.

Finally, grades under the relative system are more predictable as far as the student is concerned, because relative grades depend on the constant quality of classmates, instead of the variable of each professor's idiosyncrasies. The alternative proposal to differentiate ranking among courses and seminars does not eliminate this unpredictability. Under all systems, however, the predictability of grades is low since the students do not know what grade they will receive.

Perhaps ranking overstates one of the many lawyerly skills and is poorly correlated with successful legal careers. This is not an argument against ranking, but in favor of more complex grading methods. If grades are correlated with some skill, rank will be too. If rank is correlated with some skill, a judge, law firm, or client trusting a student can rely on the school’s rank as an assessment of the probability that this

---

A's superior students receive in courses that do not display grade inflation will be worth more under relative ranking than in absolute ranking.

10. Classmates are many, and hence the risk of receiving “undesirable” classmates for relative rank purposes is diversified. Each professor, by contrast, represents a much larger undiversified risk of “undesirable grading tendencies” under absolute ranking. A student who will take 20 professors in law school has diversified grading risk over 20 possibly correlated professorial tendencies under absolute ranking, but over perhaps 200 to 400 uncorrelated students under relative ranking. Note that third year and second year students can take courses together, increasing their totals. The proper calculation of standard deviation mitigates the risk of an undiversified small class of undesirable (top) students. See supra note 5.
trust is not misplaced (relative to another student). Improving ranking will help rank better reflect reality, without precluding further improvements in grading.

The strongest argument for having law schools provide the clearest information possible about students should be familiar to all. Like producers of goods and issuers of securities, schools have easier access to information that those interested in their students (the notorious "market") want. If schools were to refuse to provide it, the alternative would be ludicrously wasteful. Each legal employer would have to become trained in decoding each school's idiosyncratic transcript. The existence of regional schools implies that a different solution has emerged. Local law firms specialize in local schools' transcripts. A national market in legal talent requires better allocation of that talent. The only way to get there is to signal student quality simply and clearly.

III. PRIMARY DRAWBACK

This solution to the arbitrary nature of letter grading is not perfect precisely because grades become absolutely relative. A class of good students will be hurt because the comparison will not reveal the absolute quality (and the opposite is true for a class of bad students). This drawback—the risk of which is much smaller for the student than that of meeting a professor with undesirable grading tendencies—is overcome by looking at the transcript and, if it is not abolished, at the absolute ranking.

My few test runs, however, indicate that the rankings probably would change most in the middle of the student body, where the greatest concerns for grade-strategizing appear to exist. If that is true, little would be gained by maintaining absolute ranking since it would only resolve the—probably rare—problem created by the class of top or bottom students. Moreover, the difference between the two rankings, which may be the result of a host of factors, may end up being misused. A relative rank higher than the raw absolute rank, for example, may be used to infer willingness to take difficult courses. Even more dangerous is the opposite, of inferring from the low relative rank (compared to the raw rank) that the student was grade-shopping.
There is no perfect solution. We should be ready to accept the remote danger that a class of top students might be hurt in order to eliminate the more likely injustice of some courses systematically giving higher grades than others. The advantages of relative rank far outweigh the risk of the rare undervaluation.

The relativity of the proposed ranking system, which is its major advantage as well as its major drawback, is perfectly suited to a faculty which prefers to allow different professors to use different grading policies, or more importantly, to have different philosophies about law teaching and grading. Not only does it eliminate the need for the stifling curves many schools are forced to use, but it should also reduce concerns (whether of students or of professors) that certain professors or courses use different grading standards. I stress that even a slight difference in grading standards will affect a raw ranking. The more a school experiences crowding of students' GPA's in a very narrow range ("grade compression"), the more is this problem aggravated. Given grade compression, all students' nominal GPA's are very close. A small change in GPA makes a large change in ranking. Therefore, even a slight difference in grading standards will influence student rank dramatically, penalizing the students of the "harsher" professors, and creating incentives for the students to grade-shop and for professors to raise their grades.

The relativity of relative rank may be a drawback, but it should not be over-emphasized. Trying to remedy the effects of a class of top students (or one of bottom students) leads immediately to the idea of adjusting the relative grade of each class by the incoming students' quality. For example, a class of top students, as determined by first-year performance, would all have their relative grades adjusted upward. Not only would such a proposal greatly complicate relative ranking without sufficient justification, but such a measure would also hurt the prime beneficiaries of relative ranking-students who do not perform well in their first year. Slow learners will receive little relief if they are made to carry their tarnished past into every grade they receive.

IV. RELATIVE GRADE EXAMPLES FROM ACTUAL COURSES

Some actual examples from the Fall 1993 grades at the University
of Connecticut School of Law illustrate the importance of converting nominal grades to relative grades for the purpose of comparing students’ performance. The following table gives the grade distributions and the conversion to relative grades for several courses, with course titles having been changed.

Roman Law:

1 A and 5 B+’s given. An A=+2.2σ, a B+=−0.44σ. The rare A is a very good relative grade. If the standard deviation estimate, discussed supra note 5, were used, an A would be +2.0σ and a B+ would be -0.40σ. The adjustment reduces the dispersion of the small class’ grades to mitigate the risk of disadvantageously good classmates.

Bankruptcy:

2 A’s, 10 B+’s, 11 B’s, 11 C+’s, 7 C’s, 1 D+, 1 D, 1 F. A=+1.6σ, B+=+1σ, B=+0.3σ, C+=-0.3σ, C=-0.9σ, D+=-1.6σ, D=-2.2σ, F=-3.5σ. Notice that an A is not as great a relative grade here as in Roman Law. But the relative ranking makes the professor’s harshness disappear since even a B is a plus. It is not the case that an F should not be such a bad a relative grade (-3.5σ). It is further from the bulk of the students than the A’s are. Using the standard deviation estimate in such a large class does not produce a noticeable difference.

Criminal Procedure:

7 A’s, 5 B+’s, 5 B’s. An A=+1, B+=−0.1σ, B=−1.3σ. When B+ is the average grade, a B is not a relatively good grade (it is worse than a C in Bankruptcy). It is not as bad a grade as a B would be if there were even more A’s (compare Legal Postmodernism B’s which are even worse relative grades).

Seminar:
9 A’s. An A=0σ. If there is no differentiation in grading, it is useless for relative ranking. The class is effectively converted to a Pass/Fail course.

Corporate Finance:

1 A, no B+’s, 5 B’s, 2 C+’s, 2 C’s. An A=+2.2σ, B=+0.4σ, C+=0.5σ, C=-1.4σ. The fact that one student excelled by being far above the rest of the class gives this A great weight for relative rank purposes. A C+, however, is not a very bad grade (it is better than a B in Criminal Procedure).

Legal Postmodernism:

10 A’s, 7 B+’s, 3 B’s. An A=+0.9σ, B+=0.5σ, B=-1.9σ. Giving many A’s decreases their worth. Getting a B in such a course is devastating (worse than a Bankruptcy D+). Relative ranking should encourage the Professor to separate the cream of the crop from the simply solid students. If the standard deviation estimate (described supra, note 5) were used, the harshness of this B would be somewhat alleviated and it would become a -1.8σ.

Compare the same letter grade in the different courses. Also compare which letter grades get the same relative grade across courses. A C in Corporate Finance is worth the same as a B in Criminal Procedure. A B+ in Bankruptcy helps the student quite a bit (it is a +0.3 relative grade), while a B+ is not helpful in Roman Law, where there are no lower grades. I believe this relative ranking is less unfair than nominal ranking, which considers the worst grade (B+) in Roman Law better than a relatively respectable grade (B) in Bankruptcy.

V. CONCLUSION

Relative ranking is a simple measure with the potential to clarify the signals that transcripts give about students’ ability. Perhaps the only
reason that students have not traditionally been ranked by relative grade is that, until recently, computing relative grades and rank was not practically possible. Technological progress has trivialized the awesome computational requirements of relative ranking.

The immediate beneficiaries of relative rank will be the schools which adopt it first, however like Cassandra's such a prophesy may be given that my law school has no plans to adopt it. If the faith in incentives that lawyer-economists have is correct,\footnote{For such articles by this author see, e.g., Insider Trading as a Transactional Cost: A Market Microstructure Justification and Optimization of Insider Trading Regulation, 26 CONN. L. REV. 1 (1993); Frauds, Markets, and Fraud-on-the-Market: The Tortured Transition of Justifiable Reliance from Deceit to Securities Fraud, 49 MIAMI L. REV. 671 (1995); Elements of a General Theory of Securities Regulation, ARGYRADES MEMORIAL VOLUME 318 (Athens, 1995); Why Should Disclosure Rules Subsidize Informed Traders?, 16 INT'L REV. L. & ECON. 417 (1996).} those schools should end up noticing a marked preference for their students by employers and, perhaps, a better fit between students' preferences and the courses they take.

The remote beneficiary of relative rank may be the legal system—and not simply because the facilitation of a national legal market will allocate legal talent better. Student populations may be skewed with respect to the strength of those who have different skills—perhaps because students with different talents receive different incentives to enter law school (compare formal versus intuitive reasoning, rule versus relationship orientations, qualitative versus quantitative thinking, and alternative career paths given each). If those skills were required in different proportions by courses in different fields, such as business versus criminal law courses, we should expect that even the same professors see more excellence in one type of course and less in the other. Raw ranking biases the legal system against using the disadvantaged skill. If students' rank influences opinions, those who have preferred the structurally harder courses will appear less promising. They will not receive their share of clerkships, good jobs, and other shortcuts to prominence, and hence they will under-contribute to society. The system will fail to deploy them with the greatest effectiveness, and will waste their very valuable contributions.