

Notes on Grading

Steven J. Zeil

December 28, 2012

Contents

1	Normalized Scores	2
2	Ranking	2
3	Assigning Letter Grades	3
4	Interpreting Your Scores	3

The technique that I use to compute grades is often unfamiliar to students but is intended to be as impartial and objective as possible, given the fact that the A,B,C,D,F scale is, by definition, a subjective rating by the instructor.

1 Normalized Scores

I do not employ the 100 point-anything-less-than-70-is-a-failure scale familiar to most students from their days in grade school. From a testing and grading standpoint, that scale really has nothing to recommend it except familiarity, and most people have different traditions about what grade constitutes an A, a B etc. Working to this scale requires an instructor to make many more subjective judgements - every assignment, quiz, or other graded item must be designed beforehand to try to yield the desired level of numeric performance. That's an extremely difficult task, which is why so many instructors wind up applying arcane and often arbitrary "curves" afterwards.

Instead I normalize all scores so that, no matter how easy/hard the assignment or how picky/lax the grading, the class's scores get mapped into a comparable range. This is essentially the same technique that is used on the SATs, ACTs, GREs and other national standardized tests.

There are different techniques for normalizing scores, and the topic of how to do so properly belongs in a class on statistics. The best-known normalization formula, and the one I use for exams and other situations where the number of scores above and below the average are likely to be equal, is the "z-score":

$$\frac{(x - \text{avg})}{\text{sd}}$$

where x is the student's score, avg the class average, and sd is the class standard deviation (a measure of how widely spread the class scores have been). More information on this score and why it is useful can be found in any statistics book.

For programming assignments, where experience has shown that the scores tend to be skewed, I have found that the formula

$$1 - \frac{(\text{max} - x)}{\text{sd}}$$

where max is the largest score achieved by the class, provides a more appropriate normalization.

2 Ranking

The normalized scores for the various assignments are averaged together to provide a composite number that can be used to rank students in terms of overall performance. Similar rankings are produced from the normalized scores on the midterm and final exams.

It is the ranking that is the point of this whole process. Over the course of a semester, I will become very familiar with the work of some students, both strong and weak, and will have a pretty good idea of

what grade-level they are performing at. By computing a reliable and statistically valid ranking, I can then determine where the students whose work I am less certain of should fall in relation to those others.

3 Assigning Letter Grades

There are some teachers who employ this sort of normalization and then give "A"s to a certain percentage of students, "B"s to a certain percentage, and so on. I deplore the use of such quotas - the simple fact is that I have seen classes in which over half of the students did outstanding work and deserved "A"s, and I have seen classes in which very few met the expected level of performance (B).

Therefore, I use the normalized scores as described above simply to place students into a statistically valid ranking. I examine the total work turned in by students at various points within the ranking to determine whether that student has performed overall at an level of meeting my expectations for anyone successfully completing the course (B), exceeding those expectations (A), failing to meet those expectations but demonstrating enough competence to move on to subsequent courses (C), etc. This establishes the bounds for each letter grade within the overall class rank, and the remaining students can be assigned letter grades accordingly.

Is the final process of assigning letter grades a subjective one? Of course it is, but **all** letter grading is subjective, and to claim otherwise would be intellectually dishonest. Any instructor is *expected* to have enough professional expertise to judge what constitutes acceptable, poor, and good work. The only real question is *when* they exercise that judgement.

Instructors who employ the 100/70 rule are being subjective, first in choosing the 70% threshold and then again in designing their assignments and tests to meet that rule. Instructors who employ quota systems are being subjective in establishing those quotas. I prefer to withhold my subjective judgement until the assignment or exam results are in, when I have the most information available on which to base my decision.

4 Interpreting Your Scores

After the midterm exam, I will publish a preliminary letter-grade estimate for your exam score and assignment scores to that point.

In the meantime, keep in mind the basic rules:

- Positive normalized scores generally mean you are doing above average work.
- Negative normalized scores generally mean you are doing below average work.
- How far away you are from the average is measured in standard deviations. being a full standard deviation away from the average (+1.0 or -1.0) is significant. As a general rule, 70% of the class will fall within the range -1..+1. 95% of the class will fall in the range -2..+2.