Grading Policy

It can be a challenge to assign letter grades to students in an appropriate and principled way. The following analysis tries to provide an answer without guarantee to be optimal. Nevertheless, it is the guiding principle implemented in my classes. Comments are welcome.

Why letter grades are needed at all?

Transcripts need to be simple and easy to understand. Letter grade is the crucial index that demonstrates and differentiates the "performance" of students in a course.

What does the “performance” mean?

There are two types, the “absolute” performance and the “relative” performance. For “absolute” performance, all lectures in a course needs to be standardized and coordinated. This includes set up the same required tasks, the knowledge points and the proficiency level for lectures of all instructors. For “relative” performance, students are compared with other students.

Which type of performance is used?

The “relative” performance is used unless announced otherwise. There are indeed a few courses at UCSD that are coordinated. But more resource is needed to guarantee consistency. Indeed, even if everything is standardized, it is still subject to instructors’ understanding of how deep topics need to be discussed, which inevitably cause unfairness across lectures. After all, it may not be a good idea to standardize everything in the university, where the diversity is acclaimed. On the other hand, there is no consistency issue if “relative” performance is adopted.

How to evaluate the “relative” performance?

To measure the “relative” performance, a ranking for students is created and letter grades are distributed according to historical records. For example, the historical distribution of Math20C UCSD is 25%A, 34%B, 31%C, 5%D, 5%F.

How is the ranking created?

A common assumption is that the performance can be measured by a weighted score of the grades of homeworks (HW), midterms (MID) and final (FI). In Math20C UCSD, the weights are 0.2*HW+0.4*MID+0.4*FI. However, different weights may result in different ranking and thus different letter grades. We can not really tell who is better when student...
X has higher grades in midterms but lower grades in final than student Y. The true interpretation of letter grades is that student X is better than student Y under the current weights.

**What do the grades really mean?**

Grades of HW, MID and Fi are used to create the ranking. Let us set back and ask what the grade, say the final grade, really means? The grade depends on two factors, both your proficiency (Prof) and the exam difficulty level (Level). Thus, it is a function

\[ \text{Grade} = f(\text{Prof}, \text{Level}). \]

Within each exam, the factor Level is fixed for all students and thus we can compare the performance of students in a relative sense based on the Grade. But we cannot judge a student in the absolute sense. For example, a Grade of 90% does not mean very good since it may be simply because the exam is easy. Similarly, a Grade of 50% does not mean bad.

**What is the best difficulty level?**

The exam should be designed so that the grades contain most information to distinguish students. Here, I use real grades from Math20C Fall15 at UCSD to demonstrate the relation between grades and difficulty level. Difficulty level perceived by instructors and students can be different.

Figure 1 is the histogram for midterm1. It shows that there are more than 50 students obtain score 0.9-1 and almost 40 students obtain 0.8-0.9, etc. The grades for midterm1 has power to detect who gets A, B, etc. and to determine who fails. But it is relatively hard to determine who get A+ instead of A because there are many scores too close to the perfect score 1.

Figure 2 is the histogram for midterm2. It is very easy such that nearly 200 students get scores between 0.9-1. It has little power to detect who excel. But it does provide valuable information as to who performs not good enough or is close to fail.

Figure 3 is the histogram for final exam. Without any adjustment or curve, it is a nice bell-shaped distribution such that there are some students excel, some fail and most in the middle. The final exam (extremely hard perceived by students) is most powerful in detecting students’ performance. I can easily find students with A+ and I can also easily find students with F. The mean of final grades is 0.4 (not
Figure 1: Difficulty level of Midterm1: Median (by instructor); Hard (by students).

Figure 2: Difficulty level of Midterm2: Easy (by instructor); Easy (by students).
0.5) indicates that the exam is a little bit hard. An easier exam with mean 0.5 will be preferred for the grading scale 0-1.

Figure 3: Difficulty level of Final: Hard (by instructor); Overwhelmingly Hard (by students).

**What difficulty level to expect in each exam?**

The statistics of each exam already taken will be analyzed. The difficulty level will be adjusted dynamically to maximize the power of detecting students’ performance. The best strategy is to do as good as you can for each task (e.g., HW, MID, FI)

**What if there is an accident?**

The only excuses acceptable are the death of immediate family members or you have to go to the emergency room with doctor’s notice presented. Other excuses are not acceptable for the sake of fair game. Here is an explanation. Accidents can affect Grade. Indeed, Grade can be written as a function of Prof, Level and Accident.

\[
\text{Grade} = f(\text{Prof}, \text{Level}, \text{Accident}).
\]

However, Accident is a random variable that can happen to anyone in different ways. If Grades are changed for those who are good at making excuses, it is unfair for other students. The whole grading policy will be broken if Grades can be changed in a casual or random fashion.