Grade Inflation: A Systematic Approach to Fair Achievement Indexing

Amanda M. Richardson
Jeff P. Fay
Matthew Galati
Stetson University
421 N. Woodland Blvd.
Deland, FL 32720
vgalati@bellatlantic.net, jfay@steton.edu

Advisor: Erich Friedman

Background

Constantly rising grade-point averages at universities across the nation have made it increasingly difficult to distinguish between "excellent" and "average" students. For example, *The Chronicle of Higher Education* found that the mean grade-point average (GPA) at Duke University was 3.3 in 1997, up from the 1969 mean of 2.7. It also found that Duke is not alone in this trend.

Average grades have consistently increased while the system of measurement has remained unchanged. Receiving an A in a course does not necessarily denote exceptional performance, since the percentage of students receiving As has increased dramatically over the last few decades. In 1995, the *Yale Daily News* reported that As and Bs constituted 80% of grades at Yale. According to the *New York Times* (4 June 1994), nearly 90% of grades at Stanford were As or Bs, and an estimated 43% of grades at Harvard and 40% at Princeton were As. This situation has led many universities to seek new methods for ranking student performance.

Some say that expectations and grading difficulty have dropped from the faculty point of view, while others argue that the quality of student has been on the rise. Whatever the cause, a major problem arises when scholarship foundations or graduate schools try to distinguish exactly who deserves to be

The UMAP Journal 19 (3) (1998) 315–322.

in the top 10%, etc. For this reason, an alteration of the current quantitative ranking system is necessary.

One approach is a system of *quality points* based on comparative performance within each course. In this system, a student's quality points for a given course can be calculated based on performance relative to students' overall performance in the course. To obtain this objective, overall course performance may be measured by the mean or by the median of all the grades in the course. Then the student's individual performance can be measured in terms of standard deviations from the mean/median.

Many schools are looking for a feasible system to reach this goal. For example, in 1994 the faculty at Dartmouth voted to include on a student's transcript the course size and median course grade next to each grade, plus a summary telling in how many courses the student surpassed the median, met the median, or performed lower than the median (*Boston Globe*, 4 June 1994). This year, Duke University's faculty considered (but did not implement) using an "achievement index" (AI) to rank students. The factors considered in this index would include the course's difficulty level, the grades received by all the other students taking the course, and the grades those students received in other courses.

Most arguments against this type of indexed system appear qualitative rather than quantitative. For example, in an article in *The Chronicle* at Duke, Prof. Robert Erickson explains that "the reason [the AI system] won't work is that the faculty will not agree to have their grading system tinkered with." Many faculty fear that such a system may put their students at a disadvantage until such a system became more widely used, and hence students may seek to attend other schools. No one wants to be the guinea pig.

A quantitative question that arises when determining the best index is, when determining "average" performance in a course, should the index use the course's mean grade or its median grade? Some argue for the median, since it is more robust, while others argue for the mean, since it is a better estimator when the distribution is close to a normal distribution.

Our model attempts to find a solution to the problem of ranking.

Assumptions

- The sample data that we made up, with grades for 68 students, effectively represents the entire population of 2,000 students at ABC College.
- Past performance has no effect on performance in any given course.
- Results from one semester can be extrapolated to a ranking system cumulative over semesters.
- Course size and difficulty level do not change the model's effectiveness.

- The system is implemented at the administrative level, after standard grades are reported by professors.
- In comparing a plus-minus grading system (including A+, A, A−, B+, etc.) to one with only straight letter grades (A, B, C, etc.), the latter's A would encompass the former's A+, A, and A−.
- The model may compare students at large universities with those at small colleges without loss of generality.

Motivation for the Model

The goal is to find a way to order, or compare, students with only slightly varying grades; so an index rating students' performance relative to other students is germane. To arrive at this end, we must first determine how students in a particular course performed overall. Assuming that we have determined which estimator (mean or median) to use, we measure a student's standing by how many standard deviations from the given estimator the student's grade lies. Thus, if the student's grade is 2 standard deviations above the estimator, the student receives 2 quality points; if the student's grade equals the mean/median, the student receives 0 points; and if the student's grade falls below the estimator, say by 1.2 standard deviations, negative points are issued (−1.2).

The Model

We now determine the best index to use in implementing such a system. The question of using mean or median as our standard of comparison is of utmost importance. Perhaps instead of choosing one or the other for all courses, we should determine which is more effective for a particular course. The mean is the preferable estimator for data resembling a normal distribution (**Figure 1**) but is more sensitive to outlying data than the median when the distribution is strongly skewed. Thus, the skewness of the distribution of grades in a given course can help to determine which estimator to use for grade comparison in that particular course.

Through trial and error, we decided that if a course's distribution has skewness greater than 0.2, we would use the median as the estimator. If the skewness is smaller than that, the distribution is sufficiently close to normal that we can use the mean as the estimator.

In our indexing system, after determining the most appropriate estimator for comparison, we define a student's relative performance in terms of standard deviations. The quality points awarded for a given course are then weighted by the courses' credit hours. Finally, a student's overall index is computed by summing total points and dividing by total credit hours.

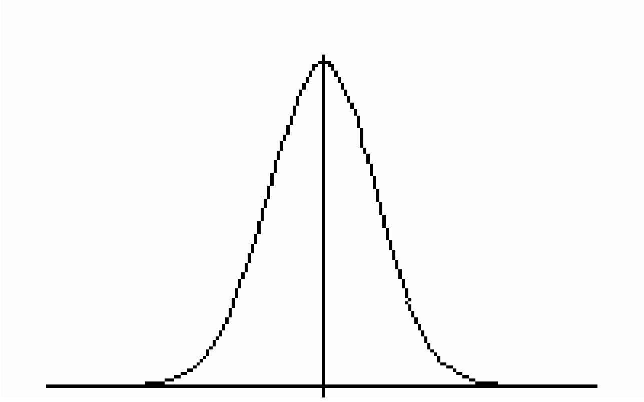

Figure 1. A normal distribution.

Let G_i be the student grade for course i, $G_i \in [0, \text{MaxGrade}]$, where course i has C_i credit hours. Then

$$\text{GPA} = \frac{\sum G_i C_i}{\sum C_i},$$

where the sums are taken from $i = 1$ to the number n of courses taken by the student.

Our procedure is as follows. Let $A_i = a_0, a_1, a_2, \ldots, a_{n_i}$ be all of the n grades for course i. Let μ be the mean of A_i, χ its median, and σ its standard deviation. The skewness S is defined as the third moment about the mean divided by σ^3:

$$S = \frac{\sum (a_i - \mu)^3}{\sigma^3},$$

where the sum is over all n_i students in the course.

If $|S| < 0.2$, we use μ as the estimator; if $|S| \geq 0.2$, we use χ as the estimator. We define

$$\text{Index} = \begin{cases} \dfrac{\sum \dfrac{G_i - \mu}{\sigma} \cdot C_i}{\sum C_i}, & \text{if } |S| < 0.2; \\[2em] \dfrac{\sum \dfrac{G_i - \chi}{\sigma} \cdot C_i}{\sum C_i}, & \text{if } |S| \geq 0.2. \end{cases}$$

Analysis

First, we offer some justification of our assumptions about course size and difficulty level.

Since the model seeks a comparative ranking system, the number of students in a course is not the issue; rather, each student's performance relative to each other student in the course is the important factor. Hence, course size is not directly involved in the model.

The difficulty level of a course, although not explicitly dealt with in determining a student's index rating, is accounted for indirectly in our model, in the overall course performance. For example, one would suspect that a relatively easy course would tend to have a large percentage of high grades, while grades in a more challenging course might be more evenly distributed or even tend toward lower grades. While the standard grading system would rate students purely on the grade they received, and thus not take difficulty level into account, the comparative nature of our indexing system causes inherent dependence on this factor.

Sensitivity

In the analysis of our data, we have seen a number of cases where the system used had a great effect on the ranking of particular students. We offer three students as examples of what can occur in our minicollege of 68 students. [EDITOR'S NOTE: We do not reproduce here the authors' complete table of grades, indexes, and ranks under the various systems.]

Student 603 is ranked 15th by GPA with plus/minus grades (**Figure 2**), with grades of B, B, A, A. This student's rank would be the same for GPA with straight letter grades. However, using our index ranking system, Student 603's rank drops dramatically. With plus/minus grades, the student's rank falls three deciles from the 30th percentile to the 60th percentile; that is, the student's rank drops from 15th to 43rd. With straight letter grades, the drop is even more drastic: The student falls to 5th and plummets to the 70th percentile.

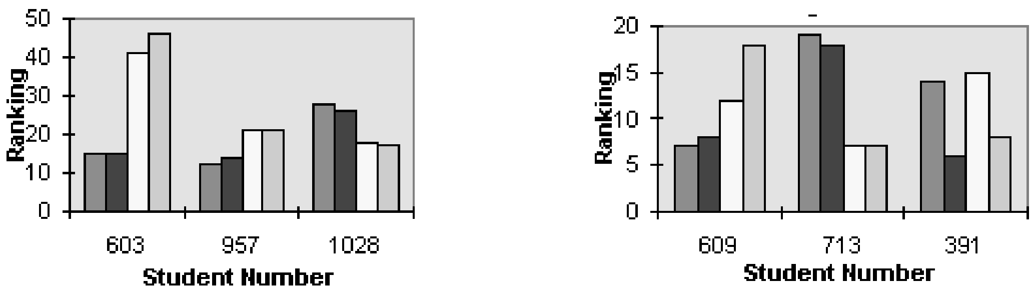

Figure 2. Rankings of some students under various systems. The leftmost bar is for GPA with plus/minus grades; the bar second from the left is for GPA with straight letter grades; the bar second from the right is for our index system with plus/minus grades; and the rightmost bar is for our index system with straight letter grades.

To understand better how such an event could occur, we take a closer look at the courses that Student 603 took. In the first two courses, the student received a B. However, the value of that B is what is under consideration in our model. The grades in the first course are 16 As, 13 Bs, and 1 C. Due to the large skewness coefficient, our formula uses the median, which is an A. So the student's grade is actually "below average." The second course has an average of an A−; here too, a B is a relatively low grade. In both of the other two courses, the mean/median grade is an A, and Student 603's As in these courses represent average performance in them. Our ranking system takes the average of the deviations from the estimator to find an index for the student. Specifically, the student has deviated from the estimator by a factor of 0.6387.

Student 957 is ranked 12th by GPA with plus/minus grades, with grades A, A, A−, B+, A, A, A, B+, A, B, A. In the first six courses, the lowest overall grade out of all of the students is a C, and the mean/median for these courses is A, A, A, B+, A, and A. It is obvious from these statistics that the student's ranking has been overestimated. Under our system, Student 957 is ranked 23rd and drops from the 20th percentile to the 40th percentile. Although this is a dramatic drop, it is not as drastic as for Student 603. Looking more carefully at the other courses, we focus on the tenth course. Here, the mean/median is C+, while Student 957 received a B.

Similarly, Student 1028 is ranked 28th by GPA. However, when his grades are compared with those in his respective courses, his ranking drops to 22nd. Once again, the strong level of grades throughout his courses lowers the "value" of the grades that he received.

What about students who fall out of the top 10% under the new rating system? The new system is supposed to determine better which students are worthy of scholarship and advancement, so this distinction is key.

Student 609 is ranked 7th in GPA with plus/minus grades and 8th with straight letter grades, with grades A, A, B+, A. However, due to the relatively high average grades in these courses, this student suffers a loss in ranking under our system, to 12th with plus/minus grades. This drop causes Student 609 to fall out of the first decile.

Student 713 benefits from our system. This student's grades are B, A−, A, B, and B, but the mean/median for each course is below the student's respective grades. Due to this, the student's ranking rises from 19th to 8th. As in the case for Student 609, the new system of ranking has a large effect on the awarding of scholarship.

We need to ensure that our model is not too susceptible to a single grade change in a single course. How does such a change affect the overall ranking of this person and the overall rankings of other students? Since students' rankings are dependent on other students' grades, a single change could affect other students' rankings across the board.

The most extreme scenario is when the course size is small and the grades are skewed. For example, let's take the case of a course with two students. Suppose that student X and student Y both receive As, so the mean/median is an A. If student X's grade changes to an F, the mean/median becomes a C. This change increases student Y's index, since student Y is no longer "average" in this course but above average. This change in student Y's index could potentially alter the rank of a few students. This scenario is the most extreme case but would also be extremely rare.

With a much larger course, a similar grade change would have minimal effect. Suppose that a course has 20 students with a distribution highly skewed towards the high end. The grades could be as follows:

A+, A+, A, A, A, A, A, A−, A−, A−, A−, A−, B+, B, B, B, B, C+, C, F.

We use the median (A−) instead of the mean. Suppose that a student who received an A+ should have received an F. Once the correction is made, the grades would still be skewed and we would still use the median, which would still be approximately A−. Thus, the only person who would be affected would be the one student whose grade was corrected.

In certain circumstances, a grade change could potentially change the median of a skewed course. Since the grades in the course are skewed, the median will not change by much and thus the change will not constitute a major problem. A few students may move ranks, and in some cases, deciles.

If the distribution is more normal, then we use the mean; and as long as the course size is relatively large, the effect of a grade change is minimal.

Strengths

Implementing our system would involve just the introduction of a computer program that would use data already in the current system in the registrar's office.

Since our index system is implemented at the administrative level, professors would not have to alter their methods of grading at all. The index system is merely a new method of interpreting the grades currently issued by professors.

Another strength of the index model relates to the issue of grade inflation itself. One problem with grade inflation is that it may not be universal. In other words, certain departments or colleges may be more or less affected by grade inflation. Thus, for employers or graduate institutions seeking the best candidates from a wide variety of undergraduate institutions, our model using the index system takes away the problem of comparing universities with varying levels of grade inflation.

Weaknesses

One possible flaw with our index model is the lack of consideration of the quality of student in a given course. For example, consider courses X and Y in which all students earn a letter grade of an A. Our model gives each student the same quality points. Perhaps consideration should be given to the quality of students taking a given course. If all the students in course X also received As in their other courses, while students in class Y had a wide range of grades in their other courses, then performing at the "average" level in course X is theoretically more difficult than in course Y, so awarding equal points does not effectively differentiate as we would desire.

Another uncertainty may arise with courses with multiple sections. It may happen that higher-level students all take a certain section of the course, thus making the comparisons of course performance invalid. For this reason, larger universities especially may need to group all sections of a given course before computing the mean/median and calculating the comparative index to rate each student's performance in that course.

Another weakness is our trial-and-error choice of 0.2 as the value of skewness that determines which formula to use for the index. Our data are limited to fairly small course sizes, and different values may be work better for larger course sizes. In rare cases when only one student is enrolled in a course (say, for a senior project or independent study), the student's grade would determine the mean/median and thus would always equal the estimator, resulting in the issuance of zero quality points. Thus, a student can never be rewarded for doing well or hurt by doing poorly in such a situation.

Future Models

More research into the effects of different skewness factors is necessary before implementation of such an index system. Also, including a method of evaluating the quality of students in a given course would further help the comparative ranking idea essential to this model. Looking at each student's grades outside a given course may help determine the quality of student. Thus, if a course is full of straight-A students and a particular student performs "above average" in that course according to our index system, that student should be awarded higher points than a student performing "above average" in a course full of students who have lower grades outside that course.

Also, further investigation into the choice of mean and median may yield more effective determination of which is the better estimator for a given course or perhaps for a particular university as a whole. Consideration may also be given to the different effects an index system has in large universities compared to smaller colleges and private universities.