Grade Inflation: A Systematic Approach to Fair Achievement Indexing In Higher Education

ABSTRACT

Grade inflation is a problem that effects many higher learning institutions. We show that it is possible to modify the current ranking system in order to more accurately rate students throughout a university and curb the problem of grade inflation. Using ideas of standard deviation, mean, and median our model can justly rank students into appropriate deciles. Our model handles universities that use the plus/minus system as well as those universities that do not. Since a student is rated by analyzing the "average" grade in that student's class, our model can deal with the idea of class difficulty as well as class size. The changing of one student's grade does not drastically change the members of the individual deciles due to our use of a mean/median analysis of the class. This mean/median procedure takes into consideration the skewness of a class's grades. Suppose A Better Class University (ABCU) has a particular class in which the majority of the students received an A. Our model ranks these students as being average in that class and thus lowers the overall rating for that student. We have written a relatively simple algorithm that can be easily implemented in many different university settings.

BACKGROUND INFORMATION

A Better Class University (ABCU) has had the benefit of a few wealthy alumni for the past few years. Thanks to their generous contributions, the dean has compiled a scholarship program which will award a hearty grant to the top students in the school. Also, there are a number of additional grants possible for students in other deciles. However, the 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 at Duke university was 3.3 in 1997, up from the 1969 mean of 2.7. It also shows that Duke is not alone in this trend.

The problem of grade inflation is a concern of universities across the country. The idea is that the 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 A's has increased dramatically over the last few decades. In 1995 The Yale Daily News reported that A's and B's now constitute 80% of all grades received at Yale. According to the New York Times (4 June 1994) nearly 90% of grades at Stanford are now A's or B's, and an estimated 43% of grades at Harvard are A's while Princeton shows 40% of all grades as A's. This situation has led many universities to seek new methods for ranking student performance.

With so many students receiving high grades, the meaning of "above average" (the usual meaning of a B) and "excellent" (A) performance has been clouded. For example, Student 1 receives an A in course X where nearly everyone received an A. In this case the average grade for course X was an A or A-, so student 1's performance should be considered only "average" for this class. Student 2, on the other hand, receives an A in course Y where the average grade was a C. This grade implies student 2 clearly performed "above average." Thus, in a qualitative analysis of the students' "ranking," student 2 should be ranked higher than student 1 despite their identical grades. Of course there are many theories explaining the forces driving grade inflation. Some say 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. But whatever the cause, a major problem arises when scholarship foundations or graduate schools try to distinguish exactly who deserves to be in the top 10%, etc. For this reason an alteration of the current quantitative ranking system is necessary.

One possible approach to such a ranking problem is to use a system of quality points which are based on comparative performance within each class. In this system, a student's quality points for a given course can be calculated based on their performance relative to overall class performance. To obtain this objective, overall class performance may be measured by finding the mean or median of all the grades in a course. Then, the student's individual performance can be quantitatively measured by calculating his grade in terms of standard deviations from the mean/median. It is apparent that 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 class size and median class grade next to the student's grade in each class. Then, at the bottom of the transcript, a summary telling in how many classes the student surpassed the median, met the median, or performed lower than the median (Boston Globe, 4 June 1994). Although it was not implemented this year, Duke University's faculty was considering using an "achievement index" (AI) to rank students. The factors considered in this index 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. By implementing this system, Duke hoped to rank students with their average AI in comparison to their classmates' instead of assigning standard grade point averages, thus stabilizing (or perhaps counteracting) the GPA problem described above; however, this proved to be too big a step for faculty to accept at this time.

Unfortunately, most arguments currently opposing this type of indexed system appear qualitative rather than quantitative. For example, in an article in The Chronicle at Duke, Professor Robert Erickson explains the AI system and possible alterations on the index, but concludes that, "the reason this plan won't work is that the faculty will not agree to have their grading system tinkered with." Many faculty fear that implementing such a system may put their students at a disadvantage until such a system became more widely used, and thus students may seek to attend other schools. No one wants to be the guinea pig.

There is one quantitative question that arises when determining the best index. When determining the "average" performance in a class for comparative grading, should the index use the class's mean or median grade? Statisticians have not found a decisive answer to this question yet. 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 normal. The following model attempts to find a solution to this problem of ranking in a system of inflated grades.

ASSUMPTIONS

- The sample data of 70 points effectively represents the entire population, consisting of approximately 2000.
- Past performance has no effect on performance in any given class.
- Results from one semester can be extrapolated to the cumulative ranking system.
- Class difficulty level and class size 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.

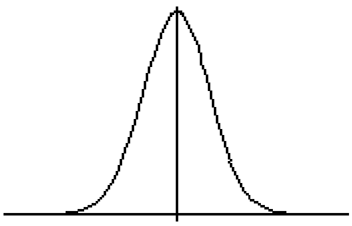
MOTIVATION FOR MODEL

The main goal of the question is to develop the concept of an index ranking system to find a way to order, or compare, students with only slightly varying grades. Clearly some other qualitative description must be used to describe a student's performance. Since the desired outcome is a ranking of students, an index rating students' performance relative to other students is germane. The main concern in this inflated grading system is that letter grades have lost meaning. Thus, to better rate a given student it is important to redefine what each of his/her grades really means. The above example of student X and Y both receiving A's illustrate this point. To arrive at this end, we must first determine how a particular class performed overall. Here we must decide whether to use the median or the mean of a class's grades as a measure for comparison for a given student. This problem will be dealt with later. Assuming we have determined which estimator to use, a point system must be designated to represent a student's relative performance. Since standard deviation is a measure that has been used by statisticians everywhere and stood the test of time, we choose to denote a student's standing by measuring how many standard deviations from the given estimator the student's grade lies. Thus, if he/she receives a grade that is 2 standard deviations above the estimator, two quality points will be rewarded. If the student's grade equals the mean/median, he/she will receive no points. On the other hand, if the student falls below the estimator, say by 1.2 standard deviations, negative points are issued (-1.2).

OUR MODEL

Since we have decided that an indexing system of comparison is a valid method of determining student ranking, we must 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, the best method is to determine which is more effective for a particular class. 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, measuring the skewness of the distribution of grades in a given class may help to determine which estimator to use for grade comparison in that particular class.

Figure 1



Through trial and error we determined a possible skewness coefficient of 0.2. If a course's distribution has a skewness greater than our determined coefficient, than the median is a better estimator. On the other hand, if the skewness remains small enough, than the distribution is close to normal, and the mean is a better estimator. The standard system used for rating students simply takes each individual grade for a given student, weighted by credit hour, and computes the students grade point average by finding the student's average grade per credit hour. Clearly, in an environment of grade inflation this system would show little discrimination between students. In the indexing system, after determining the most appropriate estimator for comparison, we may define a student's relative performance in terms of standard deviations, as described above. The quality points awarded for a given class may then be weighted by multiplying by the courses' credit hours to assign appropriate distinctions such as between one and five credit hour courses. Finally, a student's overall index may be computed by summing total points and dividing by total credit hour. This index may then be used to rank students based on more comparative performance than the standard system currently allows.

The current ranking system in most universities maps letter grades to a 4.0 or 5.0 scale. The process can be viewed over one semester and extrapolated to a cumulative approach of the life of student at a university. The procedure for calculating the index for a student is as follows:

Note: We are viewing this on a 100 point scale, but the relationship is exactly analogous.

Let G_i = Student Grade for class i, for some G_i, Î [0,100] and i=1,2,...,n where n is the number of classes. sLet C_i= Number of Credit Hours for Class i

$$Index = \frac{\sum_{i=1}^{n} G_i(C_i)}{\sum_{i=1}^{n} C_i}$$

Under our ranking system, the procedure is as follows:

Let G_i = Student Grade for class i, for some G_i, Î [0,100] and i=1,2,...,n where n is the number of classes. Let A_i={a_0, a_1, a_2, ..., a_n,} be a set of n grades for class i. Let S=Skewness Coefficient for A, m =mean of A_i, c = median of A_i, and s = standard deviation of A_i. S is defined as the third moment about the m divided by s ^{3}.

Case 1: If |S| < 0.2, then choose m as the estimator.

$$Index = \frac{\sum_{i=1}^{n} \frac{G_i - \mu}{\sigma}(C_i)}{\sum_{i=1}^{n} C_i}$$

Case 2: If |S| >= 0.2, then choose chi as the estimator.

ANALYSIS

First, some justification of our assumptions about class size and difficulty level are necessary. When dealing with the issue of class size, recalling the desired outcome is important. Since the model is seeking a comparative ranking system, the number of students in a given class is not the significant issue, rather each student's performance relative to each other student in a course is the important factor. For this reason, class size is not directly involved in the model. The difficulty level of a given course, although not explicitly dealt with in determining a student's index rating, is indirectly accounted for in our model. The difficulty level of a course is reflected in the overall class performance. For example, one would suspect that a relatively easy course tends to have a large percentage of high grades, while a more challenging class may 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.

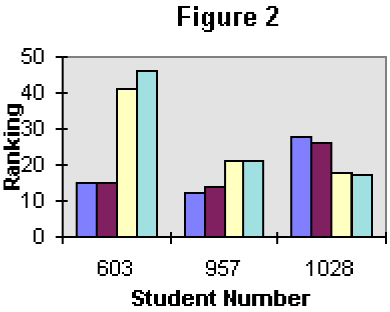
In the analysis of our data we have seen a number of cases where the system used had a great effect on the ranking of particular students. Looking at the change in deciles, we chose three students on whom to focus our attention.

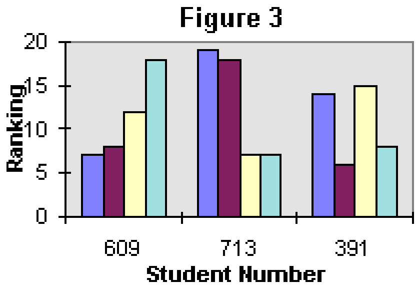
Student 603 is ranked 15^{th} under the current plus/minus system (Figure 2). The student's grades are as follows: B, B, A, A. Obviously, under the current system, the student's rank does not change for a system using straight letter grades. However, using the index ranking system Student 603's rank drops dramatically. In our plus/minus system, the student's rank falls three deciles from the 30^{th} percentile to the 60^{th} percentile. That is, the student's rank drops from 15^{th} to 43^{rd}. Using our straight letter system, the drop is even more drastic. The student falls to 5^{th} and plummets to the 70^{th} percentile.

In order to better understand how such an event could occur, we took a closer look at the courses which Student 603 took. The first two courses he/she received a B for a final grade. However, the value of that B is what is under consideration in our model. The grades in the first class are sixteen A's, thirteen B's, and one C. Due to the large skewness coefficient, the estimator in use is the median. Hence, since the median is equal to an A (96.3), it can be seen that the student's grade is actually "below average." The

second class has an average of an A- (91.3). Therefore, once again, receiving a B is a relatively low grade. In both of the other two classes, the mean/median grade is an A. 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 (Table 1 - Summary.)

It should be noted that the notion of confidence intervals is directly related to this index system. From Table 1, we see that the highest rated student scored an average of 2.118 deviations from the estimator. In studies of confidence intervals, the standard deviation is used to measure the degree of certainty for an estimate. For example, when a distribution is viewed two standard deviations from the mean, it is proven that the majority (95%) of probability is found within that area. Conversely, the lowest rated student deviates from the estimator by 3.063 deviations.





Note:

- The first box represents the current plus/minus system.

- The second box represents the current straight letter system.

- The third box represents our index system under plus/minus.

- The fourth box represents our index system under straight letter.

Student 957 is ranked 12th under the current plus/minus system. The student's grades are as follows: A, A, A-, B+, A, A, A, B+, A, B, A. In the first six classes, the lowest overall grade out of all of the students is a C. The mean/median for these classes 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 in ranking, it is not as drastic as the drop we observed for Student 603. Looking more carefully at the other courses, we focus on the tenth class. Here, the mean/median is calculated to be C+ (79.4), while Student 957 received a B. On average, under the plus/minus system, he/she was greater than the estimator by 0.2507.

Similarly, Student 1028 is ranked 28th under the current rating system. However, when his grades are compared with those in his respective classes, his rating approval drops to 22nd. Once again, the strong level of grades throughout his classes lowers the "value" of the grades that he/she received.

Another important set of data points to examine is those students that fall out of the top 10% due to the new rating system. Since the importance of this system is to better determine those students worthy of scholarship and advancement, this distinction is key.

Student 609 is currently ranked 7^{th} under the plus/minus system and 8^{th} under the straight letter system (Figure 3). The student's grades are: A, A, B+, A. However, due to the relatively high average grades in these classes, the student suffers a loss in ranking under our system. According to our rating, this student is 12^{th} under the plus/minus system. This drop causes Student 609 to fall out of the first decile into the 20^{th} percentile. The importance of an accurate system can be seen in this example. Under the current system, this student has a high probability of receiving a portion of the scholarship proceeds. However, under our ranking system, his/her chances drop dramatically.

A similar event occurs for both Students 713 and 391. Referring to both Figure 3 and Table 1 (sorry… this table is currently unavailable), we can see that Student 713 benefits from the implementation of our system. His/her grades are as follows: B, A-, A, B, B. Carefully examining his classes, it can be seen that the mean/median for each of his classes are below his/her respective scores. Due to this, the student's ranking rises from 19^{th} to 8^{th}. As in the case for Student 609, the new system of ranking has a large effect on the awarding of scholarship.

Focusing on the sensitivity of our model, we need to ensure that our model is not too susceptible to a single grade change in a class. Upon changing a single grade, we ask: How does this effect the overall rankings of this person and the overall rankings of other students? Since students' grades are dependent on other students' grades, it would make sense for a single change to possibly effect other students' grades across the board.

The most extreme scenario where the changing of a grade has an effect on a class is the case where the class size is small and the grades are skewed. For example, let's take the case of a class with two students. Student X and student Y have A's in this class. Thus, the mean/median would be an A. If student X's grade changes to an F, then the mean/median becomes a C. This increases student Y's index since student Y is no longer "average" in this class, but above average. This could potentially alter the rank of a few students. This scenario is the most extreme case, and is also extremely rare.

With a much larger class, a similar grade change would have minimal effect. Let class A have 20 students where the distribution is largely 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}. Since this class is highly skewed, we would use the median (A-) instead of the mean. Suppose a student who received an A+ should have received a F. Once the correction is made, the grades would still be skewed. We would still use the median which would still be approximately an A-. Thus the only person who would be affected would be the one student who thought he/she is receiving an A+ while instead is receiving an F.

In certain circumstances, a grade change could potentially change the median of a skewed class. Since the class is skewed, the median will not change by much and thus not constituting a major problem. A few students will possibly move ranks, and in some cases, will move deciles.

If the distribution is more normal, then the mean is used, and as long as the class size is relatively large, then the effect of a grade change is minimal.

STRENGTHS

Perhaps one of the foremost issues in this problem is feasibility of actual implementation. Since the algorithm used to compute a students' index rating uses information already available to the registrar's office, implementing this system essentially involves the introduction of a computer program that would take the data already being entered in the current system.

Another important aspect of our index system involves faculty approval. As discussed in the introduction, one major problem faced by universities trying to implement new ranking systems to counteract grade inflation has been faculty resistance. 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 perhaps certain 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 class. For example, consider classes X and Y in which all students earn a letter grade of an A. Does this mean that a student receiving an A in either of these courses should be rated equally, since the student performed equally well relative to overall class performance? 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 class X also received A's in their other courses while students in class Y had a wide range of grades in their other courses, performing "average" in class

X is theoretically more difficult than in class Y, and thus awarding equal points for both classes does not effectively differentiate the rank the model desires.

Another uncertainty may arise when dealing with courses that have many sections offered at one time. The case may arise 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 in our index model lies in the value of our skewness coefficient, since we used a trial and error method to decide on this value. Also, since our data is limited to fairly small class sizes, different coefficients may be found to work better for larger class sizes. Rare cases where 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 be equal to the estimator, resulting in the issuance of a zero quality point. Thus, a student can never be rewarded for doing well or hurt by doing poorly in such a situation.

FUTURE MODELS

Clearly 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 class 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 class is full of straight A students and a particular student performs "above average" in that course according to our index system, he/she should be awarded higher points than a student performing "above average" in a class 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 best estimator for a given class, or perhaps 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.