Grade inflation

by Steve Falkenberg
Department of Psychology
Eastern Kentucky University
Richmond, KY 40475

What is the issue?

Grade inflation is an issue because university administrators and other individuals concerned with outcome assessment in higher education have conducted a variety of statistical analyses and identified some "alarming" trends. The average grade in some courses has gone up significantly. In fact enough courses are showing this trend that average GPA of college students in the US has gone up. Along with this trend, students have come to expect higher grades. At present, in certain courses and departments, the average student expects to get a B or A in the course and in fact most students get A's or B's in the courses. Students receiving grades of C or less feel the instructor has evaluated their performance as less than satisfactory. It is argued by individuals concerned about "Grade Inflation" that at one time, a grade of C was an average grade and grades of B and A were reserved for recognition of above average or truly exceptional performance.

The fear is that our courses have been "dumbed down" or that Universities are lowering standards and reducing requirements and that this accounts for the higher grades. The concern is that if the requirements for getting a college degree are being lowered, the meaning of a college degree may change, and a college degree may have less value. Further concern about this issue comes from the perception, (or mis-perception), on the part of some faculty that students are less well prepared or less qualified than they used to
This trend in grades is accompanied by a corresponding degree inflation in some disciplines. In psychology for example, the entry level degree has for many years, been the Ph.D. In recent years, however, as degree production has outstripped growth in the field, Ph.D. candidates are having more and more difficulty in obtaining positions right out of grad school. Many of the new Ph.D.'s find it necessary to take a post doc for a year. The post doc makes them more marketable and they can compete more effectively for jobs than the new Ph.D.'s. As a result, departments advertising a position frequently have applicants with a Ph.D. and at least one post doc to choose from. By default the entry level for the field has become a Ph.D. and at least one post doc.

**Statistical concepts.**

The grade inflation issue arose from observations based on statistics. The primary concern is that the distribution of grades has changed. Technically, the concern is that the distribution is more negatively skewed than it should be, that is that there are too many A's and B's and not enough D's and F's (see diagram). The observation that the distribution has changed, however, does not tell us the cause of the change. Demonstration that more A's and B's are awarded is one thing. Figuring out why is quite another.

Part of the confusion concerning grade inflation arises from the notion that grades should be normally distributed. In point of fact, many psychological characteristics of individuals appear to be normally distributed in the population and many of the variables thought to influence psychological processes appear to follow the normal distribution. As a result psychologists frequently assume that the variables in their studies are normally distributed. This assumption then becomes the basis of the statistical treatment of the data in psychological research. When psychologists develop
tests to measure the psychological characteristics of individuals in the population, they frequently base the design, validation, and scoring on similar statistical assumptions. It is natural that someone would try to apply similar logic to the development and scoring of academic achievement tests. In point of fact the function of the classroom achievement test is completely different from the function of selection and achievement tests (SAT, ACT, GRE, GMAT, etc.) designed by psychologists. In technical terms, classroom tests are "ideographic" while the selection and achievement tests designed by psychologists are "normative". The logic and statistical assumptions used in normative test design, development and validation do not apply.

In the 1950's and 1960's it was popular for university faculty to "curve" the grades in their courses. The assumption was that grades (like all the other psychological variables that were being studied by scientific psychology at the time) should be normally distributed. Early on, some faculty forced the grades into a normal distribution by applying a "true curve." Under this plan, regardless of the grades earned, the top 10% received A the next 20% were B's, the next 40% were C's, the next 20% were D's and the bottom 10% were F's, regardless of the student's actual level of performance. The popularity of curving grades was relatively short lived. Psychologists and psychometricians quickly realized that the process was based on faulty assumptions and bad logic and most had abandoned the process by the early to mid 1960's. However, one still sometimes hears students and faculty say things like, "There should be no more than 10% A's in a college course."

One of the biggest problems with curving was that under a true curve, 10% of the students had to fail regardless of how much they had learned. Even if all the students in the course had mastered the material to a satisfactory level, the lowest 10% would still be failed. As a result, curving was only used to adjust grades upward.
Another problem with curving is that it is based on the assumption that grades are a random variable. The normal distribution is the distribution of a random variable. Without getting bogged down in the technical definition of a random variable, if it isn't random it isn't a random variable. The only way that the underlying distribution for students scores on the final exam in a college course can be normally distributed is if the students are a random sample from the population. If you only administer the test to students who have enrolled in the course or students who have completed the course, the sample is clearly not random and a very different (and unknown) distribution may apply. As soon as a faculty member teaches the students something, their performance on the test is no longer random but rather a reflection of the outcome of the teaching-learning process.

One of the reasons grades cannot be normally distributed is that a variety of selective pressures operate to take less qualified students out of the pool. Grades may be close to normally distributed in a university with unrestricted admission on the pretest for the first course in which the students enroll after entering the university. If students are tested after having had 3 or 4 months of the course, their scores will be higher. After the first course or first semester, a bunch of the poorer students will fail and leave the university. As a result of these factors, scores of second semester freshmen should be higher and the grade distribution more negatively skewed. This selective process continues through graduate school and as a result, the higher the level course the greater percentage of A's and B's that can be expected.

In the hey day of statistical grade adjustment, a number of attempts were made to identify "appropriate" distributions for college courses at each level. It was hoped that by statistically adjusting the skew, the curve could be made to fit the increasingly select groups of students in higher level courses. There was no objective criterion for developing these distributions and as a result, this
process met with limited success.

While the statistical assumptions underlying curving were flawed, an even greater flaw in the logic of statistical grade adjustment had been overlooked. Student's grades in a college course are a result of a number of factors including:

1. the student's mastery of the material; how much the student has learned, (the more the students learn, the higher the grades will be),
2. the student's ability (which as indicated above is only likely to be normally distributed if the students are a random sample of the population and they are not),
3. the student's motivation,
4. the student's background (how much of the material and the prerequisites had the student mastered prior to enrolling in the course),
5. the student's effort,
6. the effectiveness of the instructional methods,
7. the quality of the instructional aids and resources (texts, computer simulations, videos, demonstrations, practice tests, lab exercises, homework problems and assignments, etc.),
8. the ability of the faculty member to captivate the student's interest, make the material relevant to their lives and culture, and keep the students engaged in the learning process.

Changes in any of these variables can change the shape of the grade distribution in important and relevant ways. It is next to impossible to separate these factors and each factor calls for a different response. Two examples follow:

- Improvements in instructional technology may result in better mastery of course content. In that case statistical adjustment of grades to reduce the number of high grades would be completely inappropriate.
- If many students are making very low grades on exams because
the instruction is of poor quality, statistical adjustments to the grades (such as curving to the highest grade) can be used so that a given group of students will not be penalized, but it is imperative that university administrators take steps to assure that students are provided with instructional opportunities which will enable them to master the material. In other words, when a faculty member has to curve his/her grades, it should be taken by the administration as a red flag to indicate possible instructional deficiencies.

What Grades Mean

Grades reflect a student's ability. The first law of psychology is that the best predictor of future performance is past performance. When a straight A student enrolls in your course, it is a real good bet that she/he will get an A in your course as well. When a struggling student enrolls in your course, it is likely she/he will struggle in your course as well. Grades are so highly correlated with intelligence that you can virtually give an IQ test to your students and assign the grades on that basis. There is clearly a problem here. While more able students can be expected to learn more, more quickly, grades should reflect more than ability.

Grades reflect a student's standing in the class. This is the interpretation on which curving and statistical adjustment are based. The notion is that tests, being imperfect measures of a student's knowledge, can be used to rank students but cannot be used to accurately quantify a student's knowledge. Class standing is then taken as an indicator of the amount the student has learned and grades are based on class standing. This amounts to an admission that grades have been assigned out of ignorance. Instructors should decide what they want the students to know and what they want them to be able to do and then develop tests to determine if they have developed the required knowledge and
skills. Then all students who meet the specified standards would pass, regardless of class standing.

**Grades sometimes reflect a student's previous knowledge or background.** Students frequently come into a course with very different backgrounds. Grades may then reflect the student's background rather than what they learned in the course.

**Grades should reflect a student's mastery of the material.** Grades can be taken as a certification of level of mastery of the course material. Under this interpretation, the grade becomes an evaluation of competence. The student may use a variety of strategies to develop the competence. For example, she/he may attend your course, do the assignments, and take your tests to develop mastery of the material. However, if the student can pass the competence exam without regular attendance or participation in the activities, it should be irrelevant to the grading process.

The problem with this approach is that it requires the establishment of standards against which the competence of the students can be evaluated. Many college courses have no defined objectives, no measurable goals, and are not designed to enhance the students competence in any domain. These tend to be the courses that the students object so strongly to being required to take and which no one would take if they were not required.

The advantage of the mastery approach is that faculty who dedicate themselves to developing specific competence in the students can concentrate on the development of effective instructional techniques which will maximize the number of students achieving mastery. It should be noted that by default, this will result in higher grades.

**Grades reflect the success or failure of the instructor.** B. F. Skinner said that it is not the student who fails a class. When the student
does poorly it is because the instructor failed to motivate the student, failed to captivate her/his interest, and failed to provide appropriate learning activities and instruction which would enable the student to master the content. Of course, no teacher can reach all the students in her/his classes. But it should be the objective of every faculty member to reach a higher percentage of them each semester. This necessitates getting to know the students, finding out where they are coming from, what their background is, and where they are starting from. It then necessitates designing learning activities which will lead them from where they are to mastery of the material. While it is clear that in the context of the modern higher education institution, highly individualized instruction is prohibitive, careful use of technology and experience with the needs of students over the years can enable a faculty member to develop a course which will reach a large percentage of the students. This will of course lead to higher grades. It is to be expected that faculty who have been teaching for some time will get better at it and will reach more students. Older faculty should be expected to turn out students with a higher level of mastery and hence higher grades. If a faculty member has been teaching a course for twenty-something years and the high grade on her/his midterm is still 35%, something is seriously wrong.

**Possible causes for the change in the grade distribution.**

There are several possible reasons why the distribution of grades might have become more negatively skewed.

- Grade distributions could become more negatively skewed because the course has been "dumbed down" and the courses are easier and therefore more students are getting A's and B's. A number of factors can lead to this. One of the most common is that faculty who are ineffective at teaching, are poor
communicators, are unable to explain the material or who have difficulty writing good test items find that large numbers of students are failing their exams. They then must "curve".

- Grade distributions could become more negatively skewed because the faculty have developed or learned more effective teaching methods and are more effectively motivating students to learn the material. In this case the students will learn more and hence grades will be higher. This is the good kind of grade inflation. This is the kind that universities should encourage. Higher education has a long and noble history. Surely, we have developed more effective methods of teaching than were available half a century ago. Even a cursory comparison of a psychology text from the turn of the century with a psychology text of today reveals dramatic improvements in the pedagogy that are reflective of advances in our understanding of human learning and cognitive processes.

- Grade distributions could become more negatively skewed if there have been changes in measurement techniques and methods. Something as simple as switching from what psychologists call "recall" items to "recognition" items will result in higher scores on exams designed to test the same material.

- Grade distributions could become more negatively skewed if the university allows the development of courses and curricula which largely duplicate the content of other courses which the students have already taken. In this case the students in a course will already have had the material in one or more other courses prior to enrolling in the course. If the students already know all the material when they enroll in the course, grades will be high. This occurred in psychology during the hey day of behaviorism when half to two thirds of every course in the department was training in basic principles of classical and operant conditioning. At that time it was a common complaint of students that the courses covered the same content over and over again. A quick check for this would be to give a form of the final exam on the
first day of the class and if a significant number of the students can pass (or ace) the final before taking the course, revise the course to include new content or eliminate the course from the curriculum.

- Grade distributions could have changed because of a shift in the nature of higher education and its role in society. In the 1930's and 1940's only the top 5 - 10 percent of any highschool graduating class went on to college or university. When a college professor walked into a college classroom, he/she faced the intellectual elite of the nation. These students were highly intelligent, well read, had mastered the content of their previous course work. They were motivated by intellectual pursuits. Today, approximately the top half of every highschool graduating class go on to college. When a college professor walks into a college class she/he does not address the intellectual elite but rather average to above average students wishing to pursue a career or delay entry into the job market. Large numbers of the students at regional universities have no designs on traditional professions (doctor, lawyer, etc.), and have little interest in esoteric academic debates. Today the top 5 to 10 percent of high school graduating classes go to ivy league schools and professors at regional universities rarely see students of the caliber which was common 75 years ago. The function of a regional university (or even a land grant university) is different in many ways from that of the old traditional university of a century ago. When a professor at MIT walks into a first year chemistry class, she/he is looking at the major movers and shakers of the next generation of chemists. When a professor at a regional university walks into a first year chemistry class, most of the students are fulfilling a prerequisite for a career oriented program. The standards of mastery for these two groups must be very different. Failure to understand this and to make the necessary adjustments in the teaching-learning process is a prime cause of bitterness and burnout among faculty.
What should be done about grade inflation?

1. If a faculty member gives all A's give her/him a big raise. Any professor who can get all her/his students to master college algebra for example, should be rewarded for meritorious performance. Of course this assumes than the grades reflect mastery of course content and that appropriate standards have been established and appropriate assessment methods have been implemented. That brings us to some other things that can be done about grade inflation.

2. Take steps to assure that grades reflect mastery of course content. Establish appropriate performance and mastery standards for courses and assure that tests assess mastery. If this cannot be done, the benefit of the course to students and to society is questionable.

3. Eliminate courses, majors, curricula, options, and programs which teach the same thing in every course, just under a different title and with different applications.

4. If a faculty member has been teaching 20 years and they still give tests in which the high grade is a 58 out of 100, that faculty member needs to be sent to remediation. It is clear that either the teaching methods they are using are ineffective or the tests are inappropriate.

5. Identify faculty who grade on a curve and send them for remediation. It is understandable that a first or second year faculty member would misjudge the level and ability of the students and write some tests that the students can't pass. But faculty who do not show substantial improvement in this area, will probably need outside intervention and retraining to overcome the deficit in test preparation skills.

6. Encourage the development of nationally standardized mastery exams for each unit of study in your discipline. Encourage
nationally recognized panels of experts in the field to develop the standards and objectives for each unit and prepare the examinations and testing procedures. A typical college course should consist of 8 to 16 mastery units with 8 to 16 mastery exams. Students should be permitted to repeat units until they have mastered them and should be required to master prerequisite units before being permitted to sit for exams over later units.

7. Set specific objectives. Decide what you want the students to know and be able to do at the end of the course. Teach them to do it. Test them over it. Yes this means you should teach the test, but that won't be a problem if you are sure the test covers what they need to know. This also means abandoning unrealistic expectations and fantasies about how much the students can learn in a semester. It is unrealistic to expect that students will know things that they have not been taught or that they will be able to do things they were never taught to do. When students appear to have learned something that wasn't taught in your course, you can be sure that they learned the material or developed the skill somewhere other than in your course. If you are going to test over what they learned in some other course, or some other context, make that information or skill a specific prerequisite of your course. Much of the cultural bias in college courses could be eliminated this way.

8. Abandon the fantasy of faculty psychology. You cannot develop mental faculties like muscles and you can't teach students to think independent of the content they are thinking about. What you can do is improve the students ability to think and reason about a specific class of problems. Unfortunately, this improvement will not generalize to other classes of problems. Nearly 100 years of psychological research has failed to demonstrate that learning generalizes from one learning situation to another. The goal of education cannot be to improve the individuals thinking and reasoning skills. It must be to broaden the students experience by teaching them how to think
and reason about a greater variety of problem types and to correctly analyze problems extracting the isomorphisms and determining which type of problem they are dealing with. When students appear to be able to apply what they learned in one situation to another, they have merely demonstrated that they learned how to solve this type of problem in some other course, or some other context.

**Summary**

We should design our courses and curricula in such a way that we have specific goals and objectives for each course. We should next design our testing and grading procedures to enable us to determine if the students have mastered the knowledge and skills taught in the course. Then we should focus our attention on developing instructional technology, instructional methods, and learning experiences and activities which will captivate the students interest, motivate them, and enable them to master the material. In short, it should be our goal that every student get an A in every course.

*Copyright © 1996 Steve Falkenberg*

This article may still be found at http://people.eku.edu/falkenbergs/grdinfla.htm