

Grading on the Curve: An Oxymoron?

James K. Brewer
Florida State University

ABSTRACT. This paper makes the case that grading on the curve is not grading, but quota setting, and is tantamount to having no standards, which are crucial to grading. Particular attention is given to the nonacademic reasons instructors give for curve grading while criticizing the use of theoretical distributions in assigning grades.

Evaluating students in any course is probably one of the most problematic, vexing, and time consuming of an instructor's responsibilities. The fact that it is required by most institutions and must be based on a set of standards only adds to the strain of this highly subjective activity. There can be no one "right" way for a human to pass judgment on the competence of another human and only for the power-mad could this ever be an enjoyable activity.

In almost all courses each student is given a single grade at the end of the course (according to Websters' New Collegiate Dictionary a grade is, "A mark indicating a degree of accomplishment in school"). It is clear that if grades must be assigned according to standards, then standards are the single most crucial element in any grading practice. There are basically two forms of standards: (1) Those which address individual accomplishments or performance and (2) those which are norm-referenced and result in grades based, in part, on the performance of other than the individual being graded. The performance in both cases is assessed by some form of assessment instrument, like a series of tests, etc., and may be qualitative or quantitative. The latter form is commonly referred to as "grading on the curve" and it is this writer's contention that this form of standard-setting or grading can result in nonexistent standards and is not grading, but "quota setting," which has no place in an academic setting.

Grading on the Curve

The most common form of norm-referenced grading involves standards which specify, directly or through some abstract curve like the normal distribution, that a certain percentage of each type of grade will be given regardless of how poorly or how well the students perform individually. This establishes, a priori, a certain "quota" for each grade.

Under such quota setting, regardless of the performance level of the students, a certain percentage of them will get A's and a certain percentage will get F's, etc. (if using a letter-based system common in the USA). Now suppose everyone performs at exactly the same level on the assessment instrument. This means they must be assigned different grades although performance is identical. This cannot be grading and, in fact, is tantamount to having no standards. To illustrate the absurdity of this with an analogy, suppose quota grading was used on oranges in Florida and it was a year in which all oranges were bad. Automatically, a certain percentage of these bad oranges would be labeled "Grade A" and the public would be subjected to Grade A, but bad oranges. As an additional illustration, suppose a non-jury trial judge had set as a quota an equal number of "guilty" verdicts as "not guilty" verdicts for each day on the bench. Suppose further that the judge heard half the cases before noon and they were all found "not guilty." This means that every case in the afternoon must be declared "guilty" regardless of the merits of the case. Everyone would decry such a quota system in the grading of oranges or for judging guilt, but this is precisely what is being done to students if grades are assigned on a curve.

Even if the case of every student performing at or approximately at the same level does not often occur, it does not make sense to assign a student a grade based on what other students do or do not do. Where would our system of jurisprudence be if such logic was followed? We would be led to conclusions like, "Since there are so many serious crimes, anyone not guilty of first degree murder should go free," or, "Because this society is so crime-free, we should hang someone who jaywalks."

Reasons Given for Grading on the Curve

When someone who grades on the curve is asked why they would engage in quota setting, the response is often one of the following.

- 1) It helps offset any poor items which may have been used to assess the performance.

This response is, in effect, saying that the instructor is choosing to punish or reward the students because the instructor is inept or incompetent at item writing. Grades cannot be given to compensate for instructor ineptitude—they are to reflect student accomplishment only. The instructor should seek help with item writing and analysis rather than assigning grades to offset his or her weakness.

- 2) It helps minimize cheating since the student knows his or her grade will be lower if others are higher.

Here the instructor is using grades as a club to solve a societal problem which reflects a gross misuse of grading. Also, if such curve grading had the desired effect it would minimize cooperation in study and discussion groups, in students tutoring others, and it would foster a self-centered, selfish attitude in which no student would aid another in learning. This is contrary to what instructors should want in any learning environment. Besides, cheating can be reduced by increased security, but such security measures will not restore a curve-graded class to one with a helpful, supportive attitude.

- 3) It helps compensate for assessment instruments which are too hard or too easy.

Although related to the first reason above, this reason is more general in nature. Like the first it justifies grade curving because of a weakness of the instructor, not the student. Clearly an instructor does not know an instrument is too easy until after the results are in; if the instructor had known in advance the instructor would have redesigned the instrument. The basis for the "too easy" judgment is only that the students did well. If the instructor knows enough to equate "doing well" with "too easy," the instructor surely knows enough to have identified this as an easy instrument before it was given. To now punish students who do well by curving the grades is to always equate "doing well" with "too easy." This same logic is true for an instrument judged as "too hard," except the instructor now rewards students for "doing poorly." An instructor who can identify "doing well" and "doing poorly" could have set standards for individual performance and avoided quota setting altogether. If an instructor is truly surprised either way when the students' performance is assessed, then this instructor needs help with assessment instrument construction or has little insight as to instruction effectiveness. In any event, curving will promote sloppy and thoughtless test construction since the instructors believe erroneously that curving will compensate for their ineptitude.

- 4) I am unsure of what level of performance reflects what grade, so I am unable to set standards for individuals performance.

This reason is most often given by beginning instructors or graduate assistants who are assigned their first class and left to their own devices concerning grading and standards. Although understandable, it is lamentable since no one should be allowed to grade until they know what their standards are. As has been demonstrated, curve grading is tantamount to having no standards, so resorting to curving grades is only a "cop-out" and represents an abdication of an instructor's grading responsibility. So where do standards come from? Instructors have two sources for standards: a) they can borrow someone else's standards or b) make up their own. The new or unsure instructor should choose the former, temporarily, and the experienced or confident instructor should select the latter. Since standards are based on the subjective judgment of the instructor, they may always be modified to make the user more comfortable, but they must be fixed prior to any grading situation and cannot be altered by the performance of students involved in that grading situation.

The Bell-Shaped Curve

There are those who take grading on the curve one step further and specify that grades be assigned according to the normal (bell-shaped) curve regardless of the performance of the students. This is adding insult to injury in that these instructors are not only saying that the grade a student gets depends on the performance of other students, but they are also saying that grades must be given according to a theoretical distribution of performance scores. Even if scores on the performance instrument did "pile up" in a symmetrical, bell-shaped distribution (as one might expect according to the Central Limit Theorem—depending on the assessment instrument type), there is no justification for assigning grades according to it. Likewise, there would be no more or less justification for assigning grades if the scores "piled up" flat as a pancake. To believe that the shape of a set of scores endows the instructor with the knowledge of who should get what grade borders on the mystical and is closely related to numerology. For example, what if a "nice" bell-shaped distribution of scores resulted and all the scores were between 94 and 100 on a 100 point assessment scale? With a standard deviation of 1.0 and a mean of 97, for instance, would any rational person call 94 an "F" grade simply because it is 3 standard deviations below the mean? Expecting to see a bell-shaped curve is one thing, but assigning grades based upon it is something else.

Instructors who cannot or will not produce standards for individual performance often are drawn to the abstract, bell-shaped curve because it relieves them of the responsibility for assigning grades according to standards; however its use simply confirms that the instructor has no standards. Confusion and misconceptions concerning the normal curve and the Central Limit Theorem

are rampant in the activity of grading. The greatest disappointment in this regard is that some statisticians who should know better grade on this curve and concern themselves not with standards, but with questions like, "How many students should take the assessment instrument before I can justify using the normal curve?" Being concerned about the number of students necessary to justify curve grading is akin to the early clerics being concerned about how many angels could stand on the head of a pin. Even if the number of students is infinite, there is no justification for this or any mathematical curve in the assigning of grades.

Conclusion

The value of setting standards for individual performance cannot be overrated and the value of "grading on the curve" cannot be underrated. Standards, although subjectively established, must be the result of the instructor's willingness to declare what kind of performance constitutes what grade - even if other instructors do not share this view, and even if the standards were set because they simply "felt right" to the instructor. As has been illustrated above, the setting of standards and the assigning of grades on anything other than individual performance, e.g., curving the grades, is tantamount to having no standards and thus no grading system. The justifications for such curve grading are not academic justifications, but rather reflect societal or instructor deficiencies which must not be addressed by assigning grades. Likewise, practices like making grades dependent on attendance, attitudes, etc., are equally deplorable since this amounts to using grades as a weapon against students for nonacademic reasons. If an instructor has to use the threat of a low grade on students to make them attend the instructor's class, the instructor is the problem, not the student.

It may well be that the poor performance of students, particularly in the USA, on mathematics as well as other topics is due in part to the philosophy of grading reflected in "grading on the curve." With such a philosophy, the concern is with relative performance rather than absolute performance and it should be no surprise that a student may perform relatively well, but be a poor performer in an absolute sense (i.e., when using standards). If we in the USA persist in the widespread usage of curve grading we should expect to see diminished absolute performance since we are tacitly saying we have no standards and without standards there can be no improvement.

James K. Brewer, Florida State University, 307 Stone Building, Tallahassee, FL
32306-3030