Moving Past Grading on the Curve – The Right Way

Michael B. Shattan

This paper was completed and submitted in partial fulfillment of the Master Teacher Program, a 2-year faculty professional development program conducted by the Center for Teaching Excellence, United States Military Academy, West Point, NY, 10996

Key Issues

Since the 1970s the community of college teaching scholars and practitioners has been calling for an end to the common practice of normative grading, otherwise known as "grading on the curve" (Bresee 1976), (Michaels 1976). In normative grading, a distribution of course grades is typically set prior to the beginning of the semester. For example in a class of 10 students a teacher may determine that he or she will award 2 As, 3 Bs, 4 Cs, and 1 D or F (or some other distribution) prior to meeting the students assigned to their class. Student evaluation then becomes a manner of comparing student achievements relative to one another to determine the proper rank ordering. While this practice does have its advantages and is still somewhat common on college campuses today, it certainly has significant drawbacks that have spurred a desire among many college professionals to move to a criterion-based grading system instead.

The guiding document at the United States Military Academy that dictates how student assessment will be carried out is the Dean's Policy and Operating Memorandum 02-1 which states that "the goal (of student assessment) is not to rank order cadets against each other based on any preconceived concept of an appropriate grade distribution. Instead, instructors challenge cadets to meet announced standards of performance and assign grades based on their success in doing so." (DPOM 2-1). Thus, for our institution as well as many others norm-based grading is not an option and finding justification for the academy's stance is not hard to find. Normative grading can have several negative effects on the learning environment including discouraging cooperative learning between students and robbing lower achieving students from the motivation to work harder to improve their grades. Indeed, with normative grading students will naturally view each other as the competition instead of teammates working through the rigors of the academic program together. For the unfortunate average student who finds himself in a class where he or she struggles, real effort in the class would be futile as they would never be able to surpass their classmates to achieve a decent grade (Birney 1964). Additionally, normative evaluations prevent the instructor from gaining any real feedback as to whether or not their teaching is being successful. Regardless of whether the entire class achieves nearly perfect scores on an exam or performs miserably, the final grades will come out the same.

History of Practice

The alternative to normative grading championed in the literature is absolute or "criterion" based grading. This grading pedagogy has slowly gained steam in the academic

community and gained widespread use in the 1990s. The literature of the 1990s indicates a growing interest in criterion based grading but still views normative grading as very appropriate for certain types of courses and situations (Hammons 1992). Additionally, Cross' 1993 survey of educators at Virginia Polytechnic Institute found a nearly even split between the practices of normative-based and criterion-based grading practices and anecdotal evidence seems to indicate that college campuses across America still see a mix of practices today (Cross, Frary and Weber 1993). However, more recent literature seems to overwhelming support criterion-based grading.

Practice Variations

Sadler identifies two primary ideals that any criterion-based system should attempt to achieve:

- "Students deserve to be graded on the basis of the quality of their work alone, uncontaminated by reference to how other students in the course perform on the same or equivalent tasks, and without regard to each student's previous level of performance
- At the point of beginning a course of study, students deserve to know the criteria by which judgments will be made about the quality of their work." (Sadler 2005)

To achieve these ideals Sadler identifies 4 types of grading models that universities have adopted. Each grouping represents a basic category of grading methodology with several variations falling within it. These models include: *achievement of course objectives* in which course grades are determined by tallying the number of course objectives met by each student; *Overall achievement as measured by score totals* the most common approach in which a final percentage of marks achieved by the student is translated into a letter grade by pre-determined cutoff ranges; *Grades reflecting patterns of achievement* in which a certain number of A's or B's must be earned throughout the course to earn a particular final grade; and *specified qualitative criteria or attributes* in which student work is assessed based on a predetermined set of qualitative, somewhat generic criteria which is then aggregated in some way to a final grade. For example, criteria for this last category for a written paper might include communication, style, substance, and relevance. While each model is unique all of them attempt to achieve the ideals of criterion-based grading.

Beneficial Attributes

The benefits of criterion-based grading are clear. By achieving the ideals stated above we can truly use assessment to foster cooperative learning among cadets, to motivate each cadet to work hard knowing that their individual efforts can truly pay off. Additionally, we can use the resulting scores and grades as one form of assessment to determine the level of learning students achieve in the classroom. Criterion-based learning offers improved feedback to the students as well. Historically students have not believed that their grades are reliable indicators of their achievement in a course and are more about their "degree of luck in guessing material to be tested, the section leader, and individual test taking abilities." (Birney 1964). A transparent and

criterion-based grading system can be very effective at dispelling this misconception and actually increase the level of learning the classroom as well (Rust, Price & O'Donovan 2003), (Nelson 1994).

Controversial Aspects

Despite the potential benefits criterion-based grading, many educators either choose not to pursue this option or attempt one of the models outlined by Sadler but do not achieve the desired end-state. Those who choose not to pursue criterion-based grading usually point the extensive amount of time and effort that can be required to develop and defend criteria for the achievement of levels A, B, C and D and additional time required to properly score student work using this grading system (Hammons 1992). Those who do adopt a criterion-based system and do take the time to develop proper criteria for their grade level may still frustrate students if they do not communicate to their students what is expected of them to achieve success in each criterion identified. For example, a teacher who identifies the criteria listed above (communication, style, substance, and relevance) for a research paper may still confound their students if they do not explain what achieving excellence in "substance" consists of. Thus, to achieve the ideals of this grading system the instructor must communicate standards of achievement for each criterion as well as list what each criterion will be.

Additional pitfalls in establishing a truly criterion-based grading system can be hidden in the details of how a teacher scores individual submissions and tests throughout the semester. Unless each assessment of student work is criterion-based with clearly defined standards of achievement for each criterion it is impossible to aggregate these scores to a final grade and legitimately claim that your grades are truly based on absolute student performance (Sadler 2005). During Academic Year 2003, The Department of Physics took on the challenge of achieving a fully criterion-based grading system by taking the time to clearly define Department standards for achieving grades A, B, C, and D on each graded event throughout the semester and rewriting all rubrics and grading sheets for these events to ensure they matched the published standards. (Nelson 2003) The results were encouraging. Instructors found grading events such as lab reports and the final exam faster and more consistent between teachers. Additionally, the resulting student grades better resembled a normal distribution than had previously been achieved while the course average remained nearly unchanged. With a population size of 880 students a normal distribution should be expected and the instructors were pleased to see these results.

Conclusion

The call for fully embracing criterion-based grading rings loud across the literature pertaining to college evaluation and assessment. The United States Military Academy has made criterion-based grading mandatory for all departments and we are implementing it today. However, any criterion based grading system can quickly fall short of achieving its desired end-

state if care is not taken to ensure that students understand not only the criteria upon which they will be graded, but also the standards of performance that represent the distinct levels of criterion achievement that we associate with letter grades. Implementing an effective criterion-based grading system is not easy. It takes careful forethought on the part of the teacher to ensure that criteria are clearly defined for each learning outcome we desire our students to achieve, that those criteria have clearly defined standards of achievement that is communicated to the students, and that the link between learning objectives, criteria, standards, and assignments is evident It can be a lengthy process to develop but one that can truly result in increased student motivation, higher performance, and better feedback for teachers and students alike.

References:

Birney Robert C. (1964) *The Effects of Grades on Students.* The Journal of Higher Education, (35)2, 96-98.

Bresee, Clyde W. (1976) *On "Grading on the Curve"*. The Clearing House, 50(3), 108-110.

Dean's Policy and Operating Memorandum 02-1. (16 August 2004), available at http://usma-portal/dean/Pages/DPOMs.aspx.

Cross, Lawrence, Robert Frary and Weber, Larry. (1993) *College Grading: Achievement, Attitudes, and Effort.* College Teaching, 41(6), 143-148.

Hammons, J. O., & Barnsley, J. R. (1992). *Everything you need to know about developing a grading plan for your course (well, almost*). Journal on Excellence in College Teaching, *3*, 51-68.

Michaels, James W. (1976) *A Simple View of the Grading Issue.* Teaching Sociology, 3(2), 198-203.

Nelson, Craig E. (1994). *Critical Thinking and Collaborative Learning*. New Directions for Teaching and Learning, (59), 45-58.

Nelson, Raymond. (2003) *Managing Expectations in Introductory Physics—Grading Without a Curve*, Presentation at Center for Teaching Excellence Brown Bag Luncheon, West Point, NY.

Rust, Chris, Margaret Price & Berry O'Donovan. (2003). *Improving Students' Learning by Developing their Understanding of Assessment Criteria and Processes.* Assessment & Evaluation in Higher Education, (28)2, 147-164

Sadler, Royce D. (2005). *Interpretations of criteria-based assessment and grading in higher education.* Assessment & Evaluation in Higher Education, (30)2, 175-194.

Annotated Readings:

Bracey, Gerald W. (1994) *Grade Inflation?* Phi Delta Kappan, 76(4), 328-331.

In this article Bracey depicts the trials he experienced while being forced to grade using a normal distribution at Stanford University in the mid 1960s. Among other problems a key issue he illustrates is commonly being forced to include test questions based on trivial portions of the course in order to force differentiation. Asking only questions from the core subject material would result in almost all correct answers and extreme difficulty in forcing a normal distribution. Interestingly he also reports that in the early nineties the number of high school students achieving grades of A+, A or A- inched up while the corresponding SAT scores for these students inched down.

Brookhart, Susan M. (1994) *Teacher's Grading: Practice and Theory*. Applied Measurement in Education, 7(4), 279-301.
This article presents a review of research that has been conducted regarding teachers' grading practices, and then discusses the results in light of evaluation and motivation theory. The research included in the Brookhart review focused primarily on teaching practices in secondary education and classified each research project in terms of analytical framework used, subjects involved, method of research and a summary of findings. Key results of the Brookhart review included the confounding of effort and achievement in grading practices. Brookhart's research suggests that teachers tend to want to offer good grades as a way to motivate students to work hard (i.e. if you work hard you'll get a good grade) whether or not the student's level of achievement matches their work input. Parents, however, tend to view the grades that they see on a report card as achievement-based only. This leads to communication issues between parents and teachers and calls into question how reliably teachers can judge effort in the first place. The tendency to include both effort as part of a student's final evaluation tended to taper as students got older.

Green, Kris H. and Emerson, Allen. (2007) *A new framework for grading.* Assessment & Evaluation in Higher Education, 32(4), 495-511.
The article outlines the frustrations college teachers typically find with current grading practices and in particular focuses on the difficulties teachers face in determining whether or not they should include qualitative data in student evaluations as well as quantitative data. Green and Emerson do a good job of outlining the characteristics that make up a good grading system and then present their newly developed grading framework called COGS: Categorical Objective Grading System.

Guskey, Thomas R. (2000) *Grading Policies that Work Against Standards … and How to Fix Them.* NASSP Bulletin, December, 1-12.
This article was published by the North Central Association: Commission on Accreditation and School Improvement, a body that accredits secondary schools and universities and gives guidance on how schools can successfully implement a standards-based reform initiative. The article discusses four grading policies that prevent this reform: "grading on the curve", selecting valedictorians, using grades as a form of punishment, and using zeros in grading. Of most interst is Guskey's comments on "grading on the curve." These comments highlight the negative impacts of this practice including discouraging cooperative learning among students and the loss of meaning in the grades in terms of what the students actually know.

Haigh, Martin. (2007) *Sustaining learning through assessment: an evaluation of the value of a weekly class quiz.* Assessment & Evaluation in Higher Education, 32(4), 457-474.
Haigh argues that commencing each class with a quiz emphasizing the previous class' work and supports it with immediate feedback encourages students to revise their notes ahead of each class session and to undertake more reading and keep pace with course progression. This practice reduces the necessity for any spoken review of the previous week's work, provides guidance on the status of current student learning, and creates a knowledge platform upon which deeper learning may be constructed. It also tends to keep students motivated with continuous feedback and rewards them for their efforts as well as reinforcing good studying practices. Haigh's

conclusions are drawn from the grades and end-of-course surveys from 372 students in a number of courses at Oxford Brookes University, UK.

Kirby, Bernard C. (1962) *Three Error Sources in College Grading.* The Journal of Experimental Education, 31(2), 212-218.
In this article Kirby outlines three main sources of random error in college grading and conducts an analysis to quantify them. The three sources Kirby identifies are hard vs. easy instructors, the "cutting point" error (grade cutoffs between letter grades) and lucky vs. unlucky guessing. Of these three sources the most pronounced is hard vs. easy instructors. To quantify this random error Kirby takes a statistical sample of the average grade given by 206 instructors of freshman and sophomore classes at San Diego State College. Their average grades varied over two full letter grades from a low of 1.82 to a high of 3.88. His results show the importance of a common understanding among instructors of what level of achievement is expected for the attainment of each letter grade.

Trotter, Eileen. (2006) *Student perceptions of continous summative assessment.* Assessment & Evaluation in Higher Education, 31(5), 505-521.
This paper describe the effect assessment has on student motivation, their approach to learning and the change to their learning environment. The case study Trotter uses to explore these effects is a tutorial system for a business taxation course at the University of Salford, UK. The tutorial system is very analogous to the graded homework practices currently carried out by many academic departments and it is not surprising that the conclusion reached is that while continuous summative assessment may be time-consuming to the administrator, the rewards of an enhanced learning environment for students outweigh the additional burden on staff.