

SECTION ON GRADING FROM APRIL 2005 ANNUAL REPORT OF EPC

Full Report at <http://www.unc.edu/faculty/faccoun/reports/2004-05/R05EPC1.htm>

EPC is charged with reporting on grading each year. Last year's report presented an update to the extensive EPC report of February 2000 (the Turchi report), which had shown that there is a consistent upward progression in average grades, a pattern that was shown to have continued. Among other things, last year's report recommended consideration of alternatives to traditional grade-point average (GPA) that take into account discrepant grading practices across courses.

Discrepancies in Grading Practices. Average grades vary widely across departments as well as across instructors and courses within departments. Some of this variation is systematic: (1) Average grades are lowest in the natural sciences, highest in the humanities, with the social sciences in between. Evidence for this pattern at Carolina is presented in the Turchi Report. Analyses of grading at other institutions give similar results⁷. (2) Average grades are higher in upper-level classes with a small number of students than they are in lower-level classes with a large number of students. Evidence of this pattern is seen in ongoing analyses of grading at Carolina and also from those at other institutions⁸.

Uses for Grades. We believe that it valuable to consider three different types of uses for grades. Our analysis draws on past studies of grading at UNC, on published research on grading, as well as our discussions as a committee. The types of uses are:

1. **Comparing performance.** One common use of grades is to compare performance by different students and by the same student in different courses. This use has been endorsed by Faculty Council: "High grades should be used for the one purpose of signaling outstanding academic achievement (Faculty Council, 1976)." Further examples of the comparative use of grades can be seen in the use of GPA in awarding distinction upon graduation, in the use of GPA in admissions, in use of GPA as a screening device by business recruiters visiting campus, and in the routine request by prospective employers that GPA be accompanied by a class rank based on GPA. The comparative use of grading is also important to students trying to understand the significance of their own level of performance.
2. **Mastery.** A second use of grades is to indicate mastery of content at defined levels.
3. **Motivation.** A final use of grades is a motivational tool for specific student behavior (e.g., class attendance or class participation) or to reward students for improvement so that they continue to apply themselves to their studies.

EPC believes that grades are employed to varying degrees for all three of the uses described above; it also believes that different uses of grades may be appropriate in different courses. As a committee, EPC has not attempted to evaluate which of the above uses is most appropriate or under what circumstances different approaches to grading should be taken. This categorization is useful because it provides a framework for

⁷ Johnson, V.E. **Grade inflation: A crisis in college education.** New York, NY: Springer.

⁸ University of Washington, Office of Educational Assessment:
<http://www.washington.edu/oea/uwrepts.htm>

assessing the consequences of discrepancies in grading. We address the importance of grading discrepancies on each of the above described uses starting with the least consequential case.

Discrepancies in grading practices across courses do not seem particularly consequential for the use of grades as a motivational tool. When grades are used for this purpose, their meaning seems inherently bound to the communicative context between the instructor and the student. While instructors may be more or less deft in using grades as a motivational tool, grading differences across courses are unlikely to undermine the motivational use of grades as long as the instructor assigning the grade makes the meaning clear to the student receiving the grade.

How discrepancies in grading practices affect the use of grades to convey levels of mastery is less clear. When this use is dominant, discrepancies simply indicate that across courses different proportions of students achieved different predefined levels of mastery. Grading discrepancies potentially create problems concerning how information about the setting of predefined levels of mastery is communicated. Some curricula have very well defined goals associated with particular courses that may allow a standard set of expectations about the goals associated with courses. This situation seems most likely to exist in professional programs and in introductory courses in curricula that have a cumulative progression of courses where mastery of material in one course is essential for understanding material in the next. For more varied, non-progressive curricula the predefined levels of mastery are difficult to assess for those outside of the course. In that case, grading discrepancies across courses are problematic because they may reflect differences in the ambitiousness of goals across courses rather than differences in levels of student achievement.

Discrepancies in grading practices are manifestly problematic for the comparative use of grades. In the worst case, a student's GPA can be seen as conveying more information about what courses the student took than about how much he or she learned. Most faculty find little joy in the comparative use of grading but it is without doubt viewed as an important part of grading by students and by those outside the University who use grades for evaluative purposes. Furthermore, the University facilitates and encourages this use by providing information on class rank and by awarding distinction based on GPA.

The notion that GPA should be adjusted based on the courses that a student has taken is common in high schools where regular courses are graded on a four point scale, honors courses are graded on a five point scale, and advanced placement courses are graded on a six point scale. While this scaling procedure reflects a serious effort to grapple with problems created by averaging non-comparable grades, EPC believes that it is not appropriate at the college level because it requires a prior valuation of the merit of courses. Such a valuation may be possible for highly standardized curricula, such as those in many high schools, but would not apply easily to the diverse plans of study and unique course offerings that are the distinctive intellectual reasons for seeking a university education.

An alternative for dealing with grading discrepancies that does not require a prior valuation of courses is to compute an adjusted GPA that takes into account all the grades

assigned in a course. The idea is to find an aggregate statistic that treats a high grade as more significant in a class in which few high grades are given than in a class where many high grades are given. This idea is appealing because it preserves the policy of leaving grading practices up to the instructor in a course, but then takes those practices into account before aggregating grades from different courses taught by different instructors. The most straightforward way of doing this is to convert each individual grade into a deviation from the average in the class in which the grade was assigned before computing a student's GPA. This approach incorporates a linear model that is common to a large number of applied statistical procedures. While this model has merit, it also has deficiencies. It assumes that differences between grades have a constant meaning both within and between classes. Thus, the difference between A- (3.7) and B (3.0) is assumed to be the same as the difference between B (3.0) and C+ (2.3) for a given instructor and across instructors. This assumption is probably false and is better replaced by the ordinal assumption that for any instructor A- is higher than B and B is higher than C, etc. A more serious deficiency is that a student's adjusted GPA can be reduced by taking a course in which high grades are given even if the student earned an A in the class. Thus, this grade adjustment mechanism creates an incentive for high-achieving students to avoid taking classes in which high grades are given because doing so can lower adjusted GPA regardless of how well the student does in the class. EPC believes that selection of courses should be governed by how the content of courses relates to students' goals and interests. Grade-based incentives for course selection – whether the current incentive to avoid classes with relatively low grades or a newly created incentive to avoid courses with relatively high grades – should be minimized.

An alternative to the linear adjustment method for GPAs, that preserves its positive features but avoids its negative features, has been developed and evaluated by a statistician, Valen Johnson, while he was on the faculty at Duke University. The method aggregates grade information into an adjusted GPA while treating grades as providing ordinal (relative) information and while taking into account the grade distribution in the course in which the grade was assigned as well as how the students in that class did in their other classes. Unlike the linear model discussed above, this method does not penalize a student who gets an A in a class where a great many As are assigned (though the student also derives little benefit from such a grade).

EPC has undertaken preliminary research on the use of Johnson's method for computing adjusted GPAs, particularly as it applies to awarding distinction and highest distinction at Carolina. That research has provided provocative evidence that adjusted GPA provides a more valid aggregation of information about the relative performance of students than does traditional GPA. Based on these findings, EPC has the following recommendations:

1. Comprehensive analyses of the use of different methods for computing adjusted GPAs should be completed and disseminated to the University community. To the extent practicable, students should be allowed to access information about how their own performance is assessed by adjusted GPA as well as by traditional GPA.
2. Consideration should be given to determining criteria for the awarding of distinction and highest distinction within each of the larger units that award Bachelor's degrees (e.g., the College of Arts and Sciences, School of Journalism and Mass Communication,

School of Nursing, etc). Patterns of grading between those units are sufficiently different that the use of a common set of GPA cutoffs for the awarding of distinction may not be appropriate.