

The Effects of Grading and Teaching Practices on Students' Perceptions of Grading Fairness

Michael E. Gordon
Rutgers Business School

Charles H. Fay
Rutgers University

To examine the antecedents of perceptions of grading fairness, approximately 600 college students were surveyed about the prevalence and desirability of 1) teaching practices that assisted students to prepare for examinations, and 2) common test scoring manipulations used to transform poor scores into acceptable ones (e.g., curving low scores upward). Students also described the fairness of the grading they had experienced. Regression analysis revealed that grading fairness was predicted best by exposure to the teaching practices rather than the scoring practices. Results are discussed in terms of the possible effects of these teaching and grading practices on grade inflation.

Keywords: grading practices, grade inflation, grading fairness

Grading has long been recognized as one of “the least pleasant tasks facing a college instructor” (Frisbie, Diamond & Ory 1979, p. 2). Matters have not changed: “Grading is one of the least liked, least understood and least considered aspects of teaching” (Green & Emerson 2007, p. 495). Texts dealing with the evaluation of students speak to grading from the teacher’s perspective and recognize that grading fairness is a significant component of pedagogy. By contrast, little has been written that addresses students’ perspectives about grading fairness. The purpose of this exploratory study is to identify specific teaching and grading practices that are associated with students’ perceptions that they have been evaluated fairly by their teachers.

Fairness

University students’ comments about assessment practices more often deal with fairness than other issues such as validity (Sambell, McDowell & Brown 1997). Students at all performance levels were less satisfied with assessment outcomes

when they were concerned about grading fairness (Nesbit & Burton 2006). Perceived grading fairness is a critical determinant of students’ judgments about an instructor’s ethics (Kurher, 2003), of their aggressiveness toward the instructor, and of student motivation and learning (Chory-Assad 2002).

Although a great deal is known about the consequences of grading fairness, little is known about the specific teaching and grading practices that engender these perceptions. Rodabaugh (1996) identified college teachers’ practices that were the bases for students’ perceptions about the fairness of their educational experiences, many of which pertained to the assessment of students, including using multiple assessments and applying the same performance standards to all students. Little research other than Rodabaugh’s (1996) investigation has been devoted to college students’ perceptions of grading fairness. In this exploratory study the prevalence of specific college teaching and grading practices is examined. Students reported the frequency of their exposure to a variety of teaching and grading practices. Opinions about the fairness with which they were evaluated were also solicited. Students’ sense of fair treatment was found to be more closely related to their instructors’ teaching practices that assist them to prepare for tests than to their use of scoring manipulations that have the effect of raising grades.

Correspondence should be sent to Michael E. Gordon, Management and Global Business, Rutgers Business School, 94 Rockefeller Road, Janice H. Levin Building, Piscataway, NJ 08854-8054, USA. E-mail: gordon@business.rutgers.edu

Teaching Practices

Students' examination performance, and hence their grades, are influenced by the type and amount of their preparation. Various teaching practices assist students to prepare for examinations. Study guides offer a structured means for students to determine how well they know the material that the exam will cover and, in general, facilitate performance on assessments based on content understanding (Wood 1989). Also, practice tests facilitate progress toward the attainment of learning goals (Davies 1986). By offering students the opportunity to answer questions about a sample of subject matter (although not the items that will appear on the actual test), practice tests appear to improve academic performance (Perlman 2003). Finally, review sessions help to reassure students by offering a summary of the content areas to be examined and to reduce anxiety that interferes with test performance (Sahadeo & Davis 1988).

Grading Practices

Teachers rely on different distribution rules when awarding grades. Meritocratic rules distribute grades based on academic achievement, whereas particularistic rules evaluate students on the basis of individual characteristics or personal circumstances (e.g., the need to pass a course in order to graduate may effectuate the "gentleman's C"). Although assessment experts agree that grades should not be based on nonacademic matters (e.g., Frisbie, Diamond & Ory 1979), Guskey & Bailey (2001) report that "nearly every teacher wants to add individual exceptions" (p. 141) to standard grading procedures.

Empirical research that identifies the rules used to award grades at the college/university level is negligible. A nationwide survey of 4,200 U.S. undergraduates revealed that approximately 79% had witnessed particularistic grading on the part of faculty (Braxton & Mann, 2004). Brookhart's (1994) literature review found that high-ability students preferred meritocratic grading practices, whereas less able students preferred particularistic grading practices. Several particularistic grading practices that have the effect of raising grades are discussed below.

Curving grades. In a formal sense, curving grades involves assigning grades with reference to a predetermined distribution, often the normal distribution. In actuality, curving grades appears to be a particularistic practice invoked when a substantial number of students perform poorly on an examination, thus impelling some teachers to convert unacceptable scores into more acceptable scores. The exact procedure used to carry out this transformation is not clear—it may, for example, simply entail lowering the cutoff scores for certain letter grades or inflating students' test scores with a fudge factor that produces a grade distribution more to the teacher's liking.

Retaking examinations. Students may be asked to re-take a failed test, redo a poorly executed assignment, or submit additional work to earn more satisfactory grades. These tactics may be rationalized as a way of producing mastery of the content and experience of a course. Teachers are unlikely to offer these opportunities to all students, especially, we suspect, to those seeking to raise a relatively high grade (e.g., B+) to an even higher grade (e.g., A-).

Discarding the lowest grade. Some teachers allow for one misstep by dropping the lowest test score from each student's record prior to calculating the final grade. This oft-criticized practice incorrectly assumes that measurement error always is negative. Dropping the lowest grade permits (indeed, encourages) students to "tank" one test.

Grading on the basis of improvement. Some educators preach that "teachers should try to understand students' conception of fairness" (Walvoord & Anderson 1998, 109) and recommend that teachers should simply announce that they "reserve the right to raise a grade when a student's work shows great improvement over the course of the semester" (99). Justification for this grading practice stems from the assumption that scores collected at the beginning of a marking period are less representative of student learning than test results gathered at the end of the course (Guskey & Bailey 2001). Instructors are advised to "hold back a 'fudge factor' of 10 percent or so that can be awarded to students whose work shows major improvement over the semester" (Walvoord & Anderson 1998, p. 99).

METHOD

Participants

Two undergraduate samples participated in this study. The first consisted of all 193 students enrolled in three sections of Principles of Management, a core course offered by a business school located at a large state university. These students were juniors majoring in a business discipline who received course credit for completing a survey, 187 of which were usable.

The second sample consisted of 473 students of the 1,283 enrolled in the multiple sections of an organic chemistry course offered at the same university. This course was required for students majoring in a variety of academic disciplines, including pharmacy, chemistry, animal sciences, and biological sciences. These students volunteered to complete the survey. A total of 426 usable surveys were obtained from this group.

Measures

An anonymous paper-and-pencil questionnaire was administered during a regularly scheduled class meeting in both of the courses.

Grading fairness. In the absence of established measures of grading fairness, 14 statements were written to assess three aspects of grading fairness: distributive, procedural, and interactive justice (Rodabaugh 1996). To measure distributive justice students were asked whether their experience suggested that teachers awarded grades that were commensurate with what they had learned (e.g., “Grades reflect what students learn in a class,” “Students who learn the most receive higher grades than students who do not learn as much,” and “Students who learn very little in a class may receive the same grade as students who learn a lot”). Statements about procedural justice were based on the precepts of correctness and consistency (e.g., “Teachers will correct obvious grading errors,” “Students can check to see whether their exam grades are accurate,” and “Teachers use the same procedures to determine the grades for all students in a class”). The social sensitivity of teachers when dealing with grading issues—i.e., interactional justice—was assessed with statements written about their willingness to share information (e.g., “Teachers explain the grades they award,” and “Teachers don’t seem interested in discussing grades with students”) and their interpersonal style (e.g., “Teachers are unpleasant when discussing grades with students” and “Teachers are respectful of students who have questions about grades”).

A single item was used to determine students’ overall sense of the grading fairness that they had experienced in college: “Overall, teachers are fair when it comes to grading students.” This and the justice statements were presented in Likert format using a 5-point response continuum (Strongly Agree = 5, Neither Agree Nor Disagree = 3, Strongly Disagree = 1). Negatively worded items were reverse scored so that high scores reflected a strong sense of justice.

Teaching and grading practices. Students estimated the percent of their college courses in which each of a number of teaching and grading practices was used: “Estimate the percent of courses in which a particular practice is used by writing a number from 100 (all of your courses) to 0 (none of your courses).” For teaching practices, students were asked, “In *approximately* what percentage of your courses do teachers” provide students with study guides and practice examinations and conduct review sessions prior to examinations. For each respondent, the mean of the estimated percentages was used as an index (TP) of exposure to the three teaching practices.

Participants estimated the percent of courses in which teachers used each of six common particularistic practices that modify examination scores before awarding grades: dropping the lowest test score before calculating the final grade in a course, giving students a chance to take another test if they performed poorly on the original exam, allowing students to do extra work in order to raise their final grade, curving grades upwards when scores are low on examinations, counting the last test in a course more heavily than the first test if there is steady improvement in examination

scores throughout the term, and modifying grades awarded to students whose circumstances may have interfered with their performance on tests. For each respondent, the mean of the six percentages (GP) is an index of exposure to the six grading practices.

Background items. To increase participants’ sense of anonymity, only three categorical background items were included. Because students’ evaluations of grading fairness are related to the grades they have received (Dalbert, Schneidewind & Saalbach 2007), participants reported their grade point average by selecting one of the following response categories: 2.49 or less; 2.50–2.99; 3.00–3.49; or, 3.50–4.00. Two dummy variables were created, GPA1 and GPA2, with the combined lowest two categories of grade point averages (less than 2.99) assigned the value of 0 on both dummies.

Students educated in the United States or in a foreign country may have become accustomed to different teaching and grading practices. Because such expectations created in foreign high schools could influence student perceptions of grading fairness in an American college, students indicated whether they attended the majority of their high school years in the United States (HS = 1) or outside of the United States (HS = 0).

A student who spends a lot of time studying presumably expects to perform better on examinations than those who devote little time to study. Therefore, students who prepare extensively may feel differently about grading practices that attenuate resulting distinctions in grades (e.g., curving grades upward). Participants reported how much time they devoted each day to their studying for all of their classes: less than 2 hours; 2–3 hours; 4 or more hours. Two dummy variables, ST1 and ST2, were created, with the lowest time category assigned the value of 0 on both dummies. Finally, because student ratings of their instructors typically differ across academic majors (e.g., Barriga et al. 2008), and this might influence perceptions of fair and appropriate grading, a dummy variable (Sample) identified business (0) and chemistry (1) students.

RESULTS

Table 1 presents descriptive statistics for the teaching and grading practices as well as the background items for the combined samples. On average, students reported more exposure to teaching practices intended to prepare them for tests than grading practices intended to raise poor grades, although curving low grades upward was the most frequently experienced practice.

The 14 justice items were subjected to exploratory principal axis factor analysis with oblique rotation ($\delta = 0.2$). Because of concerns that the resulting structure might differ in the two samples, the Sample dummy was included as a 15th variable in the analysis. Four factors with eigenvalues

TABLE 1
Descriptive Statistics for Background Variables and
Teaching and Grading Practices

Background Variables	
Grade Point Average	
2.99 or less	14.5%
3.00–3.49	46.7
3.50–4.00	38.8
High School Education	
In the United States	89.9%
Outside of the United States	10.1
Study Time	
Less than 2 hours	38.6%
2–3 hours	37.9
4 or more hours	23.5
Teaching Practices ^a	
Provide study guide	48.51% (28.81)
Provide review session	59.60 (29.56)
Provide practice test	49.79 (28.44)
Grading Practices	
Drop the lowest score	22.82% (23.51)
Substitute another test for a poor score	6.93 (14.76)
Allow extra work	21.90 (23.68)
Curve low grades upward	63.08 (26.18)
Discount first tests if there is upward trend in scores	32.21 (30.15)
Modify grades depending on special circumstances	12.52 (22.11)

Note. Entries for the background variables are percentages of the total sample. Entries for teaching and grading practices are means; numbers in parentheses are standard deviations.

^aPrior to combining the data for the two samples, the mean of each of the teaching and grading items was calculated separately in each of the samples. The Pearson r between the two sets of nine means was 0.78 ($p < .02$, $df = 7$, 2-tail), indicative of a similar pattern of reported exposure in both samples.

greater than 1.0 were extracted, the first three of which were interpretable as interactional justice, distributive justice, and procedural justice, respectively. The Sample variable did not load strongly on any of the factors, suggesting that the factor structure was similar in both student groups.¹

A scale representing each of the three factors was constructed. Coefficient alpha was 0.73 for both the interactional and distributive justice scales, and it was 0.63 for the procedural justice scale. Such modest reliabilities reflect the fact that the responses to these scales portray students' varied experiential histories in dissimilar classes taught by different instructors. To test whether these scales reflected the fairness of teacher behaviors, each scale score was correlated with the single item assessing overall grading fairness. All correlations (interactional justice = .40, distributive justice = .41, and procedural justice = .37) were statistically significant ($p < .01$, 2-tail).

Each of the justice scales was regressed on the set of predictor variables (see Table 2). Although all regression equations were significant, the R^2 s ranged from .050 to .094.

TABLE 2
Regression Results (N = 583)

Predictor Variable	Criterion Variable		
	Interactional Justice	Distributive Justice	Procedural Justice
Teaching practices	.161***	.176***	.253***
Grading practices	.131***	-.017	.043
HS	-.058	-.066	-.084*
GPA ₁	-.001	.096	.024
GPA ₂	-.021	-.002	-.021
STD ₁	-.031	.025	.107*
STD ₂	.086	.084	.044
R ₂	.072	.050	.094

Note. Entries are standardized beta's.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

The modest reliability of the justice measures may account for the relatively poor explanatory power.

TP was statistically significant in all of the analyses. When each of the justice measures was regressed on the set of three teaching practices comprising TP, only providing study guides and review sessions prior to exams were statistically significant. Further, GP was a significant predictor of interactional justice only. Dropping the lowest test score, allowing students to do extra work to raise their grades, discounting first tests if there was upward trend in scores, and curving grades upward were all statistically significant when the GP items were regressed on the interactional justice scale.

The HS dummy had a significant negative regression coefficient for procedural justice. Students who went to high school abroad perceived greater procedural justice than American students. Lastly, STD₁ was a significant predictor of procedural justice. Students who studied between 2 and 3 hours per day perceived greater procedural justice than students who studied less than two hours per day.

DISCUSSION

Understanding the bases for students' perceptions of classroom fairness is necessary for effective teaching. Our findings suggest that college students' perceptions of grading fairness arise from instructors' efforts to assist them to perform well on examinations that determine course grades rather than from particularistic practices that convert unacceptable test scores into more acceptable ones. All measures of justice were significantly and positively related to teaching practices that afforded students the opportunity to prepare themselves better for examinations, especially providing review sessions and study guides. When students have greater opportunity to prepare for tests, it is more likely that they will take credit for the grades they receive. Such internal attributions about their performance may engender perceptions of grading fairness.

¹Interested readers may request the pattern matrix resulting from this analysis by writing to the first author.

Only particularistic grading practices were related to interactional justice. Students apparently view a teachers' social sensitivity in terms of their willingness to adjust poor test scores based on the special circumstances of individuals. For example, allowing students to perform additional work to shore up poor grades may rectify the effects of a "bad day at the office," during which a major test was administered. Such consideration likely influences the perceived quality of the interpersonal treatment they receive and, hence, interactional justice. However, it is also true that particularistic grading practices involve granting special consideration to individual students that potentially violate the consistency rule underlying procedural justice. This may account for the fact that the procedural justice scale was unrelated to the grading practices index. Finally, because the distribution of grades is more likely to be just if based on fair procedures, it is not unexpected that the distributive justice scale also was unrelated to particularistic grading practices.

The generalizability of the present findings should be tested with students in other types of courses, e.g. liberal arts or humanities. Because grade distributions differ predictably across academic disciplines (Barriga et al. 2008), students' expectations about grading fairness may differ as well.

Little is known about the influence of instructor status on the teaching and grading practices they employ (DeBoer, Anderson & Elfessi 2007). However, because student evaluations of teaching (SETs) are correlated with the generosity of expected grades, and because administrative decisions about faculty teaching often are based on SETs, untenured and adjunct faculty tend to award higher grades than their tenured colleagues (Kezim, Pariseau & Quinn 2005). Research is needed to test the reasonable expectation that tenured faculty are less reliant on particularistic practices than those without similar job security.

Implications

The association between teaching practices and justice perceptions has real implications. To help students achieve their best, teachers should consider the availability of study guides when selecting a textbook. Review sessions also might be offered. The first author conducts these outside of the regular class period. These sessions are based solely on substantive questions posed by students about course material (not, "Will there be anything on the test about . . .?"), and they are particularly useful for students who are reluctant to speak up in class (e.g., foreign students and students enrolled in large sections).

Given their comparatively minor relationship to student perceptions of grading fairness, teachers should reconsider reliance on the grading practices incorporated in the present study that may distort achievement or artificially lessen observed differences in merit. These grading practices undermine the major purpose of grading, viz., to communicate the achievement status of students. Various audiences have an

interest in unvarnished assessments of student achievements, the most significant of which is the students themselves who must be able to gauge their real strengths and weaknesses. Lastly, grades act as an incentive to learn for many students. Reliance on particularistic grading practices (e.g., discarding the lowest test score) has the effect of reducing these incentives (e.g., blowing off one exam).

Finally, public discussion of grading fairness is on the rise in the wake of widespread concern about grade inflation—i.e., an observed increase in high grades awarded to college students without a commensurate increase in genuine academic achievement. Each of the teaching and grading practices has the effect of raising grades. While the teaching practices offer students the opportunity to earn higher grades by virtue of more systematic and meaningful preparation for examinations, the particularistic grading practices *could* contribute to grade inflation.

It is unlikely that many college and university professors deliberately behave in a manner that they expect students to interpret as unfair. Because fairness is in the eye of the beholder, however, instructors require student feedback to identify teaching and grading practices that actually engender perceptions of fairness or, in the worst case, that avoid inadvertent reliance on practices that produce perceptions of unfairness. The typical SET does not provide specific information about grading techniques, thereby exacerbating the problems of classroom assessment that "may be a weak link in the drive toward improving American education" (Cizek, Fitzgerald & Rachor 1995/1996, 162). However, students appear to recognize the fairness of additional preparation as a way of improving grades but don't generally relate particularistic grade modification to fairness. From a justice perspective, weaknesses that result in grade inflation do not appear to be on the part of students.

REFERENCES

- Barriga, A. Q., E. K. Cooper, M. A. Gawelek, K. Butela, & E. Johnson. 2008. Dialogue and exchange of information about grade inflation can counteract its effects. *College Teaching* 56(4): 201–209.
- Braxton, J. M., & M. R. Mann. 2004. Incidence of student response to faculty teaching norm violations. *New Directions for Teaching and Learning* 99: 35–40.
- Brookhart, S. M. 1994. Teachers' grading: Practice and theory. *Applied Measurement in Education* 7: 279–301.
- Chory-Assad, R. M. 2002. Classroom justice: Perceptions of fairness as a predictor of student motivation, learning, and aggression. *Communication Quarterly* 50: 58–77.
- Cizek, G. C., S. M. Fitzgerald, & R. E. Rachor. 1995/1996. Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment* 3: 159–179.
- Dalbert, C., U. Schneidewind, & A. Saalbach. 2007. Justice judgments concerning grading in school. *Contemporary Educational Psychology* 32: 420–433.
- Davies, D. 1986. *Maximizing examination performance*. London: Kogan Page.

- DeBoer, B. V., D. M. Anderson, & A. M. Elfessi. 2007. Grading styles and instructor attitudes. *College Teaching* 55: 57–63.
- Frisbie, D., N. A. Diamond, & J. C. Ory. 1979. *Assigning course grades*. Urbana, IL: University of Illinois Office of Instructional Resources. (ERIC Document Reproduction Service No. ED285496)
- Green, K. H., & A. Emerson. 2007. A new framework for grading. *Assessment and Evaluation in Higher Education* 32: 496–511.
- Guskey, T. R., & J. M. Bailey. 2001. *Developing grading and reporting systems for student learning*. Thousand Oaks, CA: Corwin Press.
- Kezim, B., S. E. Pariseau, & F. Quinn. 2005. Is grade inflation related to faculty status? *Journal of Education for Business* 80: 358–363.
- Kurher, T. L. 2003. College student views: A profile of the ethical professor. *College Teaching* 52: 153–160.
- Nesbit, P. L., & S. Burton. 2006. Student justice perceptions following assignment feedback. *Assessment & Evaluation in Higher Education* 31(6): 655–670.
- Perlman, C. L. 2003. *Practice tests and study guides: Do they help? Are they ethical? What is ethical test preparation practice?* (ERIC Document Reproduction Service No. ED480062) http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1b/57/df.pdf (accessed May 14, 2010).
- Rodabaugh, R. C. (1996). Institutional commitment to fairness in college teaching. *New Directions for Teaching and Learning* 66: 37–45.
- Sahadeo, D., & W. E. Davis. 1988. Review—don't repeat. *College Teaching* 36: 111–113.
- Sambell, K., L. McDowell, & S. Brown, S. 1997. "But is it fair?" An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation* 23: 349–371.
- Walvoord, B. E. F., & V. J. Anderson. 1998. *Effective grading: A tool for learning and assessment*. San Francisco, CA: Josey-Bass.
- Wood, K. D. 1989. *The study guide: A strategy review*. Paper presented at the Annual Meeting of the College Reading Association, Philadelphia, PA. (ERIC Documentation Reproduction Service No. ED322472)

Copyright of College Teaching is the property of Taylor & Francis Ltd. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.