AN ASSESSMENT OF GRADE INFLATION IN HIGHER EDUCATION

DAVID M. COMPTON AND BRENDA METHENY

Georgia College & State University

Summary.—Considerable research has been conducted on grade inflation and its pervasiveness. Given the significance of grade inflation on the quality of the educational experience and the reputations of colleges and universities, efforts to assess its presence and underlying causes should be supported and solutions developed. Because periodic changes in average grades in the short term may be anomalous, mean grade point averages (GPAs) for 262 undergraduate courses at a Liberal Arts college were examined for trends across a 10-yr. period. Analysis showed higher grades appeared within two of the colleges' four academic units, although the reasons remained unclear. Tentative explanations are explored.

Currently, administrators, faculty, and staff at many colleges and universities have concern about grade inflation, which may be defined as a significant increase in average or mean course grades over a number of years, e.g., 10 years, without a concomitant increase in students' performance or achievement. Short-term changes in average course grades normally are considered anomalies (Goldman, 1985).

Although actual empirical data on the issue are limited (Bearden, Wolfe, & Grosch, 1992), there are reports of grade inflation (e.g., Goldman & Hewitt, 1975; Streinta & Elliot, 1987; Young, 1990; Cluskey, Griffin, & Ehlen, 1997). Anecdotal evidence from public colleges and universities is widespread and has generated considerable political debate. For example, the mean grade point average (GPA) at the University of California at Berkeley rose from 2.95 to 3.10 in a period of 10 years (Gose, 1997). Finally, between 1964 and 1996, the mean GPA at the University of Washington rose from 2.31 to 3.12.

Even at more competitive institutions as defined by U.S. News and World Reports (1999), grade inflation has been a cause for concern. As Gose (1997) noted, in 1969 the mean GPA at Duke University stood at 2.70. By 1996, the mean GPA had risen to 3.30. In 1994, 42% of the grades awarded at Georgetown University were A's. Further, institutional attempts made to curb perceived grade inflation often have the opposite effect. For example, after Dartmouth College began listing the median grade for each class next to the student's own grade, the mean GPA rose from 3.23 to 3.28
within four years. The problem apparently begins in high school or earlier (Department of Education, 1994; Ziomeck & Svec, 1997).

While grade inflation has been a concern for at least three decades (Cluskey, et al., 1997), the problem appears acute in education of teachers and related areas, the humanities, and other traditionally defined “Liberal Arts” areas. Conversely, areas with a knowledge-base amenable to more objective testing methods, such as disciplines in the sciences, mathematics, and business, appear to be largely immune to the phenomenon (Becker, Geer, & Hughes, 1968; Oldenquist, 1983; Addy & Herring, 1996).

The present investigation was conducted to measure grade possible inflation at a particular Liberal Arts college and to identify the key fields. Finally, given that grade inflation was recorded, an additional goal became framing the issue in a form worthy of departmental and campus-wide discussions and proposing a tentative solution.

**Method**

**Data Collection**

Student enrollment during the assessment period rose from 4,900 to approximately 5,500 students. In 1997, the composition of the student body was approximately 37% male and 63% female. The racial composition of the student body was 73% European-American, 23% African American, 1% Hispanic-American, 1% Asian-American, and 1% Native American.

The initial dataset consisted of the grades from 363 freshman through senior classes offered at a middle-sized liberal arts university located in the southeastern United States between 1987 and 1998. In all, 101 courses were not considered because they were deleted from the curriculum or the time between successive offerings of the course exceeded two calendar years. As a result, all analyses were based on a dataset of 262 courses’ grades from the College of Arts and Sciences ($n = 1541$), School of Education ($n = 39$), School of Business ($n = 39$), and the School of Health Sciences ($n = 30$). Data for courses in Military Science ($n = 12$) were included but considered separately as part of the analyses of grade inflation by school or college.

**Data Analysis**

Data were analyzed using analysis of variance or analysis of covariance where appropriate (see Results section). When Academic Units were compared, there was concern that the number of courses considered as part of each of the four academic units was uneven, i.e., unequal sample sizes. Within the context of a factorial analysis of variance this, of course, means that the number of courses in different cells was unequal and not proportional across rows or columns. As a tentative solution, the data were analyzed using a least squares analysis of variance. This adjustment does not correct for
statistical independence. Unfortunately, there is no consensus on how to proceed with hypothesis testing when the research design involves nonorthogonal analyses (Wilkinson, 1990). The within-group variances, however, were approximately equal. Where appropriate least squares means ($M_{LS}$) are reported. Last, in all analyses, an alpha level of .05 was considered statistically significant.

**RESULTS**

Visual inspection (see Fig. 1) and preliminary analysis of the dataset suggested a difference in GPA between lower- and upper-level courses. This was confirmed by an analysis of the main effect of course level ($F_{1,260} = 15.24$; lower level $M_{LS} = 2.71, SD = 0.47$; upper-level $M_{LS} = 2.94, SD = 0.47$). This result was not surprising when mediating factors such as student attrition, transfer of students, intrinsic interest in course material and class size ($M = 25.8$ lower-level vs $M = 17.9$ upper-level) are considered. Thus, where appropriate, course level was included as a covariate in all subsequent analyses.

The data by academic unit are presented in Fig. 2. Examination of these data with campus unit and years as independent variables and course level as the covariate gave a main effect for campus unit ($F_{4,256} = 33.84$). Post hoc analysis with Tukey tests ($\alpha = .05$) indicated that pooled across the assessment period, the mean GPA in College of Arts and Sciences courses ($M_{LS} = 2.67, SD = 0.44$) was significantly lower than that of the Schools of
Health Sciences, Education, and of the Military Science courses ($M_{LS} = 3.25, 3.35,$ and $3.48$, $SDs = .31, .22,$ and $.30$, respectively). A similar pattern was observed between the School of Business ($M_{LS} = 2.73$, $SD = .32$) and the Schools of Health Sciences, Education, or in the Military Science courses.

In addition to the main effect of campus unit, a main effect of years ($F_{4,1024} = 5.74$) was also detected. As can be seen in Fig. 1, collectively the mean GPAs rose from 2.81 in 1988-1989 and appeared to level off during the 1994-1995 academic period at 2.89. Examination of the polynomial contrasts (Cohen & Cohen, 1983) indicated that this trend was linear ($F_{1,256} = 16.49$).

Consideration of the main effects, however, must be tempered in light of a significant interaction for campus unit $x$ years ($F_{16,1024} = 8.34$), the trend of which was linear ($F_{4,256} = 15.59$). Because the primary purpose was the assessment of higher grades, decomposition of the interaction involved separate examination of each campus unit. With the exception of the analysis of the College of Arts and Sciences data, the covariate, class size, was significant and met the assumption of homogeneity of slopes (Keppel & Zedeck, 1989; Wilkinson, 1990). As can be seen in Fig. 2, grades in the College of Arts and Sciences remained fairly constant across the 10-yr. period ($F_{4,612} = 0.92$, ns). When course level was included as a covariate, similar trends were
observed in the three remaining schools (largest $F_{4,112} = 1.16$, ns, School of Business).

Visual examination of Fig. 3 suggests that, within certain academic units, higher grades appear. As seen in Fig. 3, two of the four academic units experienced an over-all rise in mean GPA. Based on this observation, a look at the characteristics of the various units for grades was carried out.

![Net Change by Academic Unit](image)

**Fig. 3.** Mean grade point average differences by academic unit for the academic periods 1988-1989 to 1996-1997

Review of the majority of the departments or programs suggested that grading practices within these units have remained fairly stable over the 10-yr. assessment period. Departments and programs where stability in grades was observed included the Departments of Accounting, Art, Chemistry and Physics, Mathematics and Computer Sciences, English, Speech and Journalism, Modern Foreign Languages, History and Geography, Music and Theater, Government and Sociology, Psychology, Economics and Finance, Information Systems and Communications, Management, Marketing, Music Therapy, and Foundations and Secondary Education, Early Childhood Education, and Business Administration programs. That is, all $F$ values for change over 10 years were nonsignificant.

Departments and programs for which significant changes over the 10-yr. assessment period were detected included the Department of Biology and Environmental Sciences ($F_{4,88} = 4.85$), Nursing ($F_{4,32} = 7.21$), Special Education and Administration ($F_{4,36} = 8.06$), Middle Grades Education program ($F_{4,24} =$
4.22), Health, Physical Education, and Recreation Health \((F_{4,20} = 3.62)\), and Physical Education courses \((F_{4,32} = 10.99)\), and Military Science courses \((F_{4,44} = 6.49)\). Grades within the Department of Biology and Environmental Sciences, however, seemed to decrease across time.

**Students Withdrawing from a Course**

Because the date to drop a course used at the university was relatively late, the relationship between the percentage of students who dropped a given course and the average GPAs for each of the two year blocks was estimated using Pearson product-moment coefficients. These statistically significant values (.21 to .28) are presented in Table 1. As can be seen in the table, relevant associations between the percentage of students who dropped and the average GPAs were observed, but only in the 1988-1989 through 1994-1995 periods.

**TABLE 1**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.21*</td>
<td>.28*</td>
<td>.28*</td>
<td>.22*</td>
<td>.13</td>
</tr>
</tbody>
</table>

*p < .01, two-tailed test.

**DISCUSSION**

A number of factors could contribute to a rise in average GPA. Among those factors under the control of the professor or instructor are easier grading practices. Teaching evaluations play a substantial role in raises, tenure, and promotion. As Gose (1997) noted, younger and nontenured faculty may grade more leniently and thereby curry more favorable course evaluations. Nelson and Lynch (1984) found support for evaluation-induced grade inflation and a positive correlation between easier grading and the evaluation of teaching effectiveness. When examining issues related to tenure and promotion, Zangenehzadeh (1988) reported that students’ expectations about grades, attitudes toward the course, and the instructor can influence teaching evaluations. In one examination of this issue on the local level, the correlation between teachers’ over-all effectiveness rating and the expected grade in the course was \(r_{e_i} = .42\) (Compton, unpublished observations). Thus, the expected grade in a course accounted for more than 17.5% of the variance in rated teaching effectiveness. Tenured senior faculty may also choose the easier path associated with lenient grading as a way of insulating against students’ and parents’ complaints. Further, in the face of falling real incomes, a factor related to merit raises, faculty have adopted more lenient grading
practices to get better evaluations (Nelson & Lynch, 1984). Last, grades in upper-level courses were significantly higher than those in lower-level courses. If course level has a marked influence on students’ evaluations of faculty performance as has been suggested (Goldberg & Callahan, 1991; Dreger, 1997), then it is possible that this too could lead to more lenient grading practices.

Institutional factors may also contribute to grade inflation. For example, at the departmental and institutional level, perceived and actual increases in faculty demands may stimulate grade inflation (Kolevzon, 1981). Additional factors may include later dates for dropping courses, use of the excused incomplete, the option of repeating a course for a higher grade, and, of course, the recruitment of better students. In fact, late dates to drop courses and the use of incomplete courses to protect the GPA have been suggested before as possible contributing factors to grade inflation, as indexed by analysis of trends (Zangenehzadeh, 1988). In the present study, late dates for dropping a course show some effect.

Other investigators have suggested avoidance of failing members of targeted minority groups (e.g., Zangenehzadeh, 1988). The net effect would be lowered course standards for some or all of the students. Kolevzon (1981) found a disproportionate increase in the number of male students in departments with grade inflation. Last, a possibility that could be explored by departments is whether the instructor is teaching in a primary area of training. One particularly relevant argument is that the recruitment of higher quality students has led to an increase in class grades. Similar arguments can be made for improvements in teaching faculty or improvements in pedagogy. However, grade inflation may be the result of selected faculty who shirk their responsibilities to provide accurate assessments of their students and discriminate average performances from other categories (Cluskey, et al., 1997).

One dramatic change over the period was the campus-wide increase in the quality of our students as defined by the Scholastic Aptitude (Achievement) Test or SAT. Data on the average SAT scores for two periods, 1989 and 1997, were considered. During this period of time, the mean SAT score for the campus increased from 808.58 (Verbal = 386.50, Mathematics = 422.08) to 956.38 (Verbal = 480.29, Mathematics = 476.09), an increase of 147.80 points! Examination of the areas in which significant increases in average GPA were observed indicated that, within these areas, the average SAT score rose between approximately 85 and 147 points. However, the Educational Testing Service “recentered” the score scale on the SAT in 1995, and first published recentered scores for the Class of 1996. Briefly, recentering places national average SAT scores in both Verbal and Mathematics scales near 500, the center of the SAT’s 200-800 scale. Recentering did not
affect the relative standing of test takers to each other, expressed as the percentile rank.

Where the increase has been observed, the mean GPAs may at first blush seem disconcerting. On closer analysis, many of these courses are field-centered or experientially-oriented courses, in which class performance is defined in terms of achieving a definable performance criterion. For example, if the student meets the minimum standards for the course, they receive a grade of B. Stellar performances are rewarded with a grade of A. Marginal students are unlikely to enroll in these courses because they fail to meet pre-admission standards. The standards of courses vary from area to area but often include a minimum GPA. Thus, the prospective students who perform poorly in the core curricula are denied access to these upper-level criterion-based courses.

Interestingly, Hadley and Vitale (1985) have also proposed the use of criterion-referenced grading. Hadley and Vitale have also advocated a 13-point system grading using pluses (+) and minuses (−) in addition to the grades A through F and the development of models designed to measure students’ outcomes. Presumably, information from one of our requirements for graduation, the Major Field Tests (Educational Testing Service, 1999) as an exit examination, is comparable to the latter recommendation. However, until there is a contingency between exit examination scores and effort, the validity of these data are suspect.

A Modest Proposal

When dealing with inflation, one informative strategy used by economists is to frame the issue by indexing the data (Grieves, 1982) to evaluate students’ performance in the context relative to peers’ performance. Further, according to Grieves, indexed grading strategies provide incentive for instructors to assign course grades in a manner more in line with the traditional meaning of the grades, e.g., A = superior performance, B = above average performance, etc. Table 2 provides an example of this strategy using one z-score formula, where \( z = \frac{x - \mu}{\sigma} \) (Hays, 1988). Calculations such as those in Table 2 could easily be included as part of the student’s grade report.

Limitations of the Present Investigation

While the present investigation provides some useful benchmark, the limitations are important. Control for faculty differences in experience and teaching ability was not available. The lack of SAT scores of the students whose grades were considered in our analyses is another limitation. When comparing academic units, there were large differences in sample size. Between-groups homogeneity of variance was observed. Even so, with unbalanced designs statistical independence is an issue currently without resolution (Wilkinson, 1990; see Hays, 1988, for a good discussion of this issue).
TABLE 2
AN EXAMPLE OF \( z \) SCORES USED IN CONJUNCTION WITH COURSE GRADES

<table>
<thead>
<tr>
<th>Course</th>
<th>Student A</th>
<th>Student B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grade/Mean</td>
<td>( z ) Score'</td>
</tr>
<tr>
<td>Psychology 101</td>
<td>B (3.0)/2.34 0.66</td>
<td>B (3.0)/2.88 0.12</td>
</tr>
<tr>
<td>Mathematics 101</td>
<td>c (2.0)/2.00 0.00</td>
<td>A (4.0)/3.25 0.75</td>
</tr>
<tr>
<td>Chemistry 102</td>
<td>B (3.0)/2.01 0.99</td>
<td>B (3.0)/3.15 -0.15</td>
</tr>
<tr>
<td>Biology 102</td>
<td>A (4.0)/2.55 1.45</td>
<td>B (3.0)/3.77 -0.77</td>
</tr>
<tr>
<td></td>
<td>3.00</td>
<td>0.78</td>
</tr>
</tbody>
</table>

\( z \) score calculations were based on a standard deviation (\( \sigma \)) of 1.00.

Finally, grade trends alone may not provide the most meaningful assessment of grade inflation but should be viewed in conjunction with independent measures of students’ performance. Unfortunately, at our institution, the only other summary assessment of students’ achievement was the Major Field Test.

REFERENCES


Accepted February 28, 2000.