# Instructor-Specific Grade Inflation: Incentives, Gender, and Ethnicity* 

R. Todd Jewell, University of North Texas<br>Michael A. McPherson, University of North Texas

Objectives. This study attempts to isolate instructor-specific measures that may be sources of grade inflation and to measure their relative importance. Methods. We estimate a fixed-effects model, using by far the most extensive data set related to grade inflation ever assembled. Our data comprise 48,038 courses taught by 1,871 distinct instructors at a large public university over a two-decade period. Results. Our results suggest that female faculty members are the most likely to inflate grades, while ethnicity has a lesser effect. Conclusions. Characteristics of instructors, in particular gender, affect the degree of observed grade inflation, controlling for student- and department-specific effects.

Over the past several decades, grades at American colleges and universities have shown a general upward trend. Stuart Rojstaczer (formerly of Duke University) has compiled what are perhaps the most complete data in this area. Using information from 29 schools, he shows that grade point averages (GPAs) have increased approximately 0.15 points on the usual 4 -point scale per decade since the late 1960 s, with grade inflation at private schools proceeding at a more rapid pace than at public institutions. ${ }^{1}$ Numerous other studies, including Farley (1995), Cluskey, Griffin, and Ehlen (1997), Grove and Wasserman (2004), Bello and Valientes (2006), and Jewell and McPherson (2009) have found evidence of rising grades in a variety of colleges and universities.

The increase in grades without a concomitant increase in student quality has become known as grade inflation, and researchers from various disciplines have speculated about its causes. ${ }^{2}$ A long list of researchers, including Nelson and Lynch (1984), Zangenehzadeh (1988), Isely and Singh (2005), Germain and Scandura (2005), McPherson (2006), and McPherson, Jewell, and Kim (2009) have argued that faculty members are inflating grades in response to the

[^0]now widespread use of student evaluations of teaching (SETs) in promotion, tenure, and merit evaluations. ${ }^{3}$

Unfortunately, much of the existing literature on the determinants of grade inflation is rather limited. Many studies, including Kolevzon (1981), Sabot and Wakemann-Linn (1991), Anglin and Meng (2000), Sonner (2000), and Grove and Wasserman (2004), use simple descriptive statistics to examine the issues. While there are studies that use regression methodologies, in general these also suffer from important shortcomings that call into some question their conclusions. Most of the shortcomings in the previous literature stem from the paucity of data that cover a large enough variety of courses, instructors, and time periods. For example, Prather, Smith, and Kodras (1979) examine 144 individual courses over a six-year period. The data of Cluskey, Griffin, and Ehlen (1997) cover a 15 -year period but only consider senior-level accounting courses. Cheong (2000) examines data covering a 12 -year period and that are from 40 academic departments. However, data on individual courses and instructors are evidently largely absent-only very rudimentary regression analyses are employed.

Other studies have examined whether certain characteristics of instructors make them more or less likely to inflate grades. Kolevzon (1981) suggests that male faculty members may be less prone to inflate grades than their female colleagues, but finds no evidence that a faculty member's tenure status affects his or her likelihood of inflating grades. Although their principal focus was on department-level effects, Jewell and McPherson (2009) found evidence of significant instructor-specific effects.
The present study is an attempt to examine rigorously the influence of characteristics of individual instructors on grade inflation. We employ a rich data set to test hypotheses concerning grade inflation for different aspects of the undergraduate classroom experience. Our research involves data from 48,038 course sections taught in 38 different academic departments by 1,871 distinct instructors at a large public university over two decades (1984-2005), a data set that covers a substantially longer time period and considerably more courses and departments than any other study. ${ }^{4}$ The unique nature of the data permits a much more comprehensive data analysis of the determinants of grade inflation than has been possible previously. We find that while instructor ethnicity appears to have little effect on the propensity to inflate grades, instructor gender does. In particular, in our sample, female instructors give significantly higher grades than their male colleagues.

[^1]
## Data and Empirical Methods

Following past literature, the present article views the issue of grade inflation as the outcome of economic processes. We follow the model of Dickson (1984). In that model, instructors' utility depends positively on the level of job security afforded by teaching, negatively on the effort an instructor must expend in teaching, and positively on student effort and ability. In this framework, a faculty member may perceive that issuing higher grades will lead to greater job security through higher evaluations from students and peers. Utility maximization suggests that instructors have an incentive to adjust grades as long as the marginal utility that accrues due to increased job security exceeds the increased cost to the instructor because of increased instructor effort in teaching (and, presumably, decreased research output). There is also at some point, presumably, a cost in terms of faculty members' reputations to excessive inflating of grades.

The Dickson model suggests that the degree of grade inflation may vary by characteristics of the department (e.g., the faculty-student ratio), and of the course (such as the level and size of the class). Some empirical evidence from earlier studies supports these predictions. For instance, the more quantitative disciplines are sometimes found to exhibit less evidence of grade inflation (Prather, Smith, and Kodras, 1979; Sabot and Wakeman-Linn, 1991; Cheong, 2000). Dickson (1984) finds that departments with smaller class sizes are more inclined to issue higher grades. Jewell and McPherson (2009) find an inverse relationship between number of faculty in a department and average grades issued; this may indicate a reorientation toward research as departments add faculty. Jewell and McPherson (2009) also find that departments respond to growth in numbers of students by inflating grades.

Our focus is on the effects that characteristics of instructors may have on grades; these effects are generally predicted by the Dickson model. For example, level of experience and rank affect instructors' assessments of their job security, and instructors at different points in their career paths may have different costs of teaching. While difficult to observe, observed grades also depend on teaching effectiveness and each instructor's marking standard. Our article focuses in particular on instructor gender and ethnicity. These characteristics may affect observed grades in several ways. Student effort may differ according to the gender or ethnicity of instructors, or somehow gender or ethnicity of the instructor attracts students of different ability. It is also possible that teaching effectiveness differs by instructor gender or ethnicity. For example, if we observe that the grades in classes taught by female instructors are higher than comparable classes taught by men, it could suggest that women are more effective teachers. While these are possible, it may be more likely the case that instructors face different incentives according to their gender or ethnicity. If it is the case that women and minority faculty face or feel that they face differential pressures, they may have incentives to issue higher grades to their classes. For example, a female faculty member may feel that she must excel
as a teacher in order to be eligible for promotion, tenure, or merit increases; if so, she may be inclined to inflate grades to a greater extent than her male colleagues. Ginther and Khan (2004) find evidence that female faculty are less likely to receive tenure than men. As noted above, Kolevzon (1981) found that low-grade-inflation departments are more likely to have larger proportions of male faculty members. Nonwhite faculty members may face similar circumstances. For example, Nettles et al. (2000) find that black faculty members are less likely than whites to have received tenure and promotion. McPherson and Jewell (2007) present evidence that after controlling for other observable characteristics of the students, instructor, and course, minority faculty receive lower evaluation scores from students. If ethnicity-based discrimination exists in promotion, tenure, and merit decisions, or if nonwhite faculty members believe that it does, these faculty may gain more utility from grade inflation than their white colleagues.

Our data set comprises course-level observations for 21 academic years (1984-1985 to 2004-2005) at the University of North Texas (UNT). UNT is a large, comprehensive, state-funded university with more than 25,000 undergraduate students. UNT has academic programs in all traditional subjects and awards the Ph.D. in many of those programs. A particular advantage of these data is that they cover all UNT courses over the study period. Following the grade inflation literature, only undergraduate courses are considered in the analysis. Certain courses, such as private music lessons, independent studies, honors research and theses, practica, driver's education, and internships and cooperative education, are excluded from the analysis because they are organized differently than traditional university courses and their grading systems may be nonstandard. For similar reasons, student teaching, institutes and study tours, and field studies are not considered. Other courses may also have distinctive grading systems and are excluded from the analysis (e.g., activitybased physical education courses and lab sections in which a separate grade is entered from the classroom portion of the course).

Since we are interested in the effects of gender and ethnicity, instructors with missing information on any of these measures are excluded. Furthermore, courses taught in the summer or other terms outside of the regular semester are excluded because of concerns over comparability. Finally, instructors with fewer than four courses taught are excluded to facilitate estimation of instructor-specific effects. State law allows very small class sizes (fewer than 10 students) only in exceptional cases; we exclude these as outliers. This is equivalent to deleting the bottom 5 percent of classes in terms of student size. We also carried out the analysis without the largest 5 percent of classes; this does not affect our results in any important way.

After making these exclusions, the usable data include 48,038 observed course sections taught in 38 academic departments by 1,871 distinct instructors. These data represent more than 50 percent of the approximately 90,000 courses taught over the time period under study.

## An Empirical Model of Grade Inflation

As noted above, the Dickson (1984) model predicts that average GPA will vary by the characteristics of instructors, of students in the classes, and of the department itself. In terms of instructor characteristics, we are interested in the relationships between grade inflation and gender and ethnicity. With respect to ethnicity, we split the sample into whites and nonwhites. As discussed above, we hypothesize that instructors who are not white males may be more inclined toward the issuing of higher grades than their white male colleagues.

We include the following variables to measure student characteristics at the course level: SAT, Students, Pct_ $1^{s t} \_y r, P_{c t} 2^{n d}{ }_{-} y r$, and $P c t-3^{r d}-y r$. While an average SAT score for each of our 48,000 courses would be the optimal way to control for changes in student quality, such data are not available. However, we do have access to average SAT scores for entering freshmen at each university in each year. We construct a course-specific SAT score as the average SAT score weighted by the proportion of students in that course that should have entered college in each year. That is, for a given course $S A T$ is calculated as the percentage of freshmen in that course times the average SAT for individuals who would be freshmen in that year plus the percentage of sophomores in that course times the average SAT for individuals who would be sophomores, and so forth. SAT is the average SAT score for a given class relative to the national average, and as such measures improvement of UNT students relative to all college students over time. While an imperfect measure, $S A T$ is included mainly to control for any impact that student quality may have on grade inflation. The mean of $S A T$ for the estimation sample is 2.5 , which implies that UNT students taking sampled courses had on average 2.5 percent higher scores on the SAT than the national average. We hypothesize that, other things equal, GPA should be directly related to the relative quality of students, as measured by the $S A T$ variable.

Students measures course size in terms of the number of students. The relationship between the number of students and the average grade in a particular course may be affected by differences in pedagogy at different class sizes. For example, instructors of smaller classes may be able to spend more time with each student, both during class and outside of class. However, as class sizes increase, the time an instructor can dedicate to each student declines. Other pedagogical changes may also occur as class sizes increase. In general, one might expect teaching methods to shift toward assessments that may encourage memorization. It is unclear a priori how this might affect average grades in a given class. ${ }^{5}$ Pct_ $I^{s t}-y r, P_{c t-} 2^{n d}-y r$, and Pct_ $3^{r d}-y r$ represent the percentages of each class that are made up of first-, second-, and third-year

[^2]students. On the supposition that student grades improve with experience and maturity, classes with higher proportions of nonseniors should have lower GPAs.

Given the wide range of departmental characteristics that may influence GPA, we include departmental characteristics in the form of a departmentspecific dummy variable for each of the 38 distinct departments in our sample. The impact of time on GPA, or how grade inflation changes over time, is measured with the yearly time trend (Trend). An important purpose of this study is to analyze differences in grade inflation by instructor characteristics. The instructor-level measures are gender and ethnicity. Given data limitations, we employ a white/nonwhite distinction for ethnicity. ${ }^{6}$ As an initial step, we report summary statistics by gender and ethnicity in Table $1 .{ }^{7}$

Our dependent variable, GPA, is the average GPA in each course on a standard four-point system. Average GPA in all sampled courses is 2.812; however, GPA for the 1984-2005 period increased from 2.574 to 2.949 , an overall increase of 0.375 . Note that this increase is greater than the 0.307 average increase over roughly the same period reported by Rojstaczer (2008) for universities across the United States, which may indicate that UNT has experienced more grade inflation than other schools.
Table 1 provides additional descriptive information about our data. White males taught nearly 60 percent of the 48,038 classes in our sample. White females, nonwhite males, and nonwhite females taught, respectively, 28.6 percent, 8.0 percent, and 4.4 percent of all classes. Without addressing grade inflation directly, Table 1 provides information on the average grade assigned by each category of instructor. Classes taught by nonwhite males have the lowest average GPA: 2.65. White females assigned the highest grades of any group.

## Estimation Methods

We use panel data techniques to estimate the determinants of average GPA using an unbalanced panel of UNT courses from 1984-1985 to 20042005, including the above-mentioned independent variables that measure characteristics of departments, faculty, and students. The standard approach to estimating a model using panel data with instructor-specific effects would be to use a random-effects estimator, since variables such as gender and ethnicity do not change over time in our data set. ${ }^{8}$ Unfortunately, a Hausman test

[^3]TABLE 1
Summary Statistics

| Variable | White Males ( $n=28,311$ ); Instructors $=1,005$ |  | White Females $(n=13,733) ;$ <br> Instructors = 585 |  | Nonwhite Males ( $n=3,857$ ); Instructors $=174$ |  | Nonwhite Females ( $n=2,137$ ); <br> Instructors = 107 |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Mean | $S D$ | Mean | $S D$ | Mean | $S D$ | Mean | SD |
| GPA | 2.74 | 0.57 | 2.90*** | 0.58 | 2.65*** | 0.55 | 2.86 *** | 0.60 |
| Trend | 9.85 | 5.82 | 11.25*** | 5.80 | 11.48*** | 5.57 | 12.92*** | 5.45 |
| SAT_class | 2.44 | 4.38 | 2.71*** | 4.45 | 2.92*** | 4.13 | 3.35*** | 3.92 |
| Students | 44.73 | 45.01 | 39.63*** | 38.65 | 45.85* | 37.81 | 36.58*** | 30.10 |
| Pct_1 ${ }^{\text {st }}$ - yr | 15.49 | 23.57 | 17.29*** | 25.41 | 13.81*** | 21.53 | 14.38** | 23.02 |
| Pct_2 ${ }^{\text {nd }}$ _yr | 17.41 | 16.89 | 18.56*** | 17.34 | 15.92*** | 15.47 | 16.12*** | 15.52 |
| Pct_3 ${ }^{\text {rd }}$ - yr | 24.44 | 16.08 | 25.08*** | 16.48 | 24.42 | 14.97 | 24.95 | 16.02 |

[^4]indicates that the random-effects model is inappropriate. ${ }^{9}$ When the randomeffects estimator is inappropriate, an alternative is to estimate the model using a fixed-effects estimator. However, fixed-effects estimation can only be used with time-varying covariates, as the influence of time-invariant variables are subsumed into the estimated fixed effects.

Since the instructor-specific characteristics in which we are interested (gender and ethnicity) do not vary with time, we are forced to estimate their effects on grade inflation in a different way. As pointed out by Wooldridge (2002:267), even though the effect of time-invariant variables cannot be identified at any single point in time, differences in the marginal effect of time-invariant variables can be identified by interacting these variables with a time trend. ${ }^{10}$ In our case, it is possible to tease out information on differences in rates of grade inflation over time by interacting time-invariant instructor characteristics (gender and ethnicity) with the time trend and estimating a standard fixed-effects model. The resulting estimation will produce a fixed effect for each instructor and an estimated rate of change of GPA that varies by instructor characteristic. The estimation results can then be used to predict the time pattern of GPA and grade inflation for each instructor characteristic or group of characteristics.

In order to estimate the time pattern of GPA, we first separate instructors into one of four categories based on gender and ethnicity: white males, white females, nonwhite males, and nonwhite females. ${ }^{11}$ Given this gender-ethnicity categorization, we estimate the following equation:

$$
G P A_{i t}=\left(a+u_{j}\right)+D_{k} \mu+D_{k} \operatorname{Trend}_{t} \beta+C_{j} \operatorname{Trend}_{t} \delta+S_{i t} \sigma+\varepsilon_{i t}
$$

The instructor-specific fixed effect, $u_{j}$, represents the extent to which instructor $j$ has GPAs higher or lower than the overall average, net of the influence of other independent variables. $D_{k}$ is a vector of dummy variables indicating the department in which course $i$ is taught at time $t$. This vector is also interacted with the time trend to measure any grade inflation that may result from unobservable department characteristics. $C_{j}$ represents a vector of dummy

[^5]TABLE 2
Regression Results: Ethnicity and Gender Trends

| $N=48,038$ |  |  |
| :---: | :---: | :---: |
| Variable | Coefficient | SE |
| Trend (for white males) | $0.0119^{* * *}$ | 0.0025 |
| White female trend | $0.0044^{* * *}$ | 0.0011 |
| Nonwhite male trend | 0.0009 | 0.0017 |
| Nonwhite female trend | $0.0073^{* * *}$ | 0.0027 |
| Students | $-0.0017^{* * *}$ | 0.0001 |
| SAT_class | -0.0013 | 0.0009 |
| Pct_ $1^{\text {st }}$ - yr | $-0.0037^{* *}$ | 0.0001 |
| Pct_2 ${ }^{\text {nd }}$ - $y$ r | $-0.0040^{* *}$ | 0.0001 |
| Pct_3 ${ }^{\text {rd }}$ - yr | $-0.0031^{* * *}$ | 0.0001 |
| $R^{2}=0.1213$ |  |  |

***Significant at 1 percent level.
variables indicating the gender-ethnicity category into which instructor $j$ fits, which is interacted with the time trend to capture the category-specific rate of grade inflation over the sample period. The vector $S_{i t}$ represents characteristics of course $i$ in year $t . \mu, \delta, \beta$, and $\sigma$, represent vectors of parameters to be estimated, and $\varepsilon_{i t}$ represents the random error term.

## Results and Discussion

The results in Table 2 provide several new insights into the causes of grade inflation. ${ }^{12}$ Interesting patterns emerge, particularly with respect to instructor gender and ethnicity. The variable Trend represents the time trend in GPA for white male instructors, and the remaining trend interactions are interpreted as differences from the white male GPA trend. The most interesting result is that female instructors appear to inflate grades at a statistically significantly greater rate than their male counterparts. Given the estimated gender-ethnicity category rates of grade inflation, we can estimate the effect for all females, all males, all whites, and all nonwhites. Unsurprisingly, the inflation rate of all females is significantly different from that of all males. The estimated rate of change in GPA for male instructors is 0.013 and 0.024 for female instructors. Regarding instructor ethnicity, the inflation rates of white and nonwhite

[^6]instructors are not statistically different; the estimated rates of change in GPA are 0.016 and 0.020 for white and nonwhite instructors, respectively. ${ }^{13}$

The nature of the estimation method precludes the inclusion of genderspecific or ethnicity-specific fixed effects (as these do not vary over time), but by calculating predicted values for each group for each period one can better understand changes in average grades over the sample period. These predicted values are presented in Figure 1. Over the sample period, female instructors have higher predicted grades on average than males. Overall predicted grades issued by female instructors increased by 12.1 percent, while male instructors' grades rose by 9.5 percent. Regarding instructor ethnicity, nonwhites have lower predicted grades on average in all periods except one. This result is somewhat surprising, since the expectation may be that nonwhite instructors give higher grades to combat discrimination on the part of students or the administration. While the average predicted grade for nonwhite instructors rose at a faster rate over the period ( 15.1 percent vs. 12.0 percent), the previously discussed estimated trends in Table 2 are not statistically different by ethnicity.

Taken together, these results may provide some support for the contention that female faculty members may face greater pressure to issue higher grades, perhaps to counter perceived or actual discrimination or as a result of perceived job insecurity. However, it should be noted that the percentage of women of the total faculty has increased dramatically over the sample period-from 21.7 percent in the 1984-1985 academic year to 43.5 percent in 2004-2005. Given that women on average tend to issue higher grades than males in all periods, some part of the finding that female instructors' grades have increased at a more rapid pace than males may be because the proportion of women in the faculty has risen. In any case, female instructors do tend to give higher grades than their male colleagues.

It is interesting that nonwhites evidently do not seem to perceive similar pressure. However, the nonwhite category is rather heterogeneous, grouping together nonwhite American instructors and nonwhite instructors who are citizens of other countries. These subgroups may have very different motivations and incentives with respect to the issuance of grades, but our data do not permit an examination of this possibility. As was the case with female faculty members, the proportion of nonwhite faculty members in the total faculty has steadily increased over the sample period (from 6.7 percent to 16.6 percent). While this issue bears further consideration in future research, we find little evidence in our data that an instructor's ethnicity has any significant effect on the grades he or she gives.

Grading behavior also varies according to characteristics of each class; as the number of students in each class increases, average grades assigned fall.

[^7]FIGURE 1
Predicted Values Gender and Ethnicity


TABLE 3
Predicted Values: Departmental Groupings

| Departmental Grouping | Trend=0 | Trend=20 | Inflation <br> Rate (\%) |
| :--- | :---: | :---: | :---: |
| Biology, Chemistry, Math, Physics, Computer <br> Science, Engineering, Speech and <br> Hearing Sciences | 2.479 | 2.638 | 6.33 |
| Accounting, BCIS, Finance, General <br> Business, Management, Marketing | 2.571 | 2.700 | 4.90 |
| Teacher Education, Educational Psychology, <br> Special Education, Kinesiology | 3.112 | 3.299 | 6.01 |
| Art, Art History, Dance and Theater, English, <br> Language, History, Philosophy, RTVF, | 2.585 | 3.073 | 18.88 |
| Journalism | 2.634 | 2.826 | 7.29 |
| Applied Technology, Economics, Behavioral <br> Science, Communication, Criminal <br> Justice, Geography, Political Science, <br> Psychology, Anthropology, Social Work, <br> Sociology | 2.612 | 2.924 | 11.94 |
| All departments (37) |  |  |  |

Unsurprisingly, classes with larger proportions of first-, second-, and thirdyear students receive lower grades, holding other factors constant. There is also some evidence that average grades assigned decline slightly as the average quality of students in each class (as measured by SAT scores) rises, echoing the results reported in Jewell and McPherson (2009). These authors interpret the result to mean that if student quality had not been improving grade inflation would have been an even larger problem. It may also imply that average SAT scores is a less-than-ideal measure of student quality.

Finally, it is useful to consider how grades may have changed by department over the sample period. We control for departmental effects in the regression analysis above using dummy variables and trends for each department. It is also instructive to examine predicted values of grades by department. Because a number of departments have a rather small sample size, we place each into a group with other similar departments. The predicted values are presented in Table 3. Arts and Humanities departments in particular and Social Sciences departments to a lesser extent have experienced the greatest rates of increase in average grades, while Business, Education, and Math, Science, and Engineering departments have seen the lowest rates of increase. This is broadly consistent with the earlier literature (see, e.g., Prather, Smith, and Kodras, 1979; Sabot and Wakeman-Linn, 1991; Cheong, 2000). ${ }^{14}$

[^8]
## Conclusion

That grades at American colleges and universities have risen over the past several decades is by now beyond dispute. What factors may have caused this phenomenon has been the subject of much academic research. This article adds to this literature by using a substantially larger data set than has ever been previously used to examine the determinants of grade inflation, particularly the characteristics of instructors.
A number of factors contribute to the phenomenon of grade inflation, including the number and type of students in each class. In addition, certain disciplines seem to inflate grades differentially. We find, consistent with the earlier literature, that the arts and humanities (and to a lesser extent the social sciences) tend to have inflated grades at a higher rate than the sciences and other disciplines.

Our focus, however, is on the effects of characteristics of instructors. Our results indicate that certain individuals seem to be especially inclined toward assigning higher and higher grades over time. While we cannot know what motivates particular individuals, it is interesting that women instructors issue higher grades on average in every period. On average female faculty members also seem to inflate grades, although some part of this may be the result of the dramatic increase in the proportion of the faculty that is female. It is possible that despite our efforts to control for student quality, female instructors are on average teaching better students. Similarly, students in classes taught by women may, for some reason, put forth more effort. Female instructors could also be more effective teachers than their male colleagues. Still, these explanations strike us as unlikely. Instead, because of real or perceived discrimination on the part of students or administrators, female instructors may feel it necessary to excel as teachers in order to be eligible for promotion, tenure, or merit increases.

As noted earlier, the earlier literature has found that women are less likely to receive tenure and be promoted. These results suggest that some significant part of inflation can be attributed to this characteristic of instructors. We find very little evidence of differential grade inflation by ethnicity. In all periods except one, nonwhite instructors give lower average grades than do whites; furthermore, there is no significant difference by ethnicity in the rate at which grades have increased. Nevertheless, greater disaggregation by ethnicity than is permitted by our data would be a useful avenue for future research and might help illuminate whether or not faculty members from particular ethnic groups feel any pressure to grade differently.

## REFERENCES

Anglin, P. M., and R. Meng. 2000. "Evidence on Grades and Grade Inflation at Ontario's Universities." Canadian Public Policy 26(3):361-68.

Bello, A. L., and D. M. Valientes. 2006. Grade Inflation: Fact or Myth? Draft Paper. Los Baños, CA: Department of Economics, University of the Philippines.
Birnbaum, R. 1977. "Factors Related to University Grade Inflation." Journal of Higher Education 48(5):519-39.
Cheong, K. S. 2000. Grade Inflation at the University of Hawaii-Manoa. Department of Economics Working Paper No. 00-2.

Cluskey Jr., G. R., N. Griffin, and C. Ehlen. 1997. "Accounting Grade Inflation." Journal of Education for Business 72(5):273-77.

Dickson, V. A. 1984. "An Economic Model of Faculty Grading Practices." Journal of Economic Education 15(3):197-203.
Farley, B. L. 1995. "A' Is for Average: The Grading Crisis in Today's Colleges." Issues of Education at Community Colleges: Essays by Fellows in the Mid-Career Fellowship Program at Princeton University. Princeton, NJ: Center for Faculty Development, Mid-Career Fellowship Program, History Department.
Germain, M. L., and T. A. Scandura. 2005. "Grade Inflation and Student Individual Differences as Systematic Bias in Faculty Evaluations." Journal of Instructional Psychology 32(1):58-67.

Ginther, D. K., and S. Khan. 2004. "Women in Economics: Moving Up or Falling Off the Academic Career Ladder?" Journal of Economic Perspectives 18(3):193-214.
Grove, W. A., and T. Wasserman. 2004. "The Life-Cycle Pattern of Collegiate GPA: Longitudinal Cohort Analysis and Grade Inflation." Journal of Economic Education 35(2):162-74.
Isely, P., and H. Singh. 2005. "Do Higher Grades Lead to Favorable Student Evaluations? Another Look at an Old Controversy." Journal of Economic Education 36(1):29-42.
Jewell, R. T., and M. A. McPherson. 2009. Whose Fault Is It? Assigning Blame for Grade Inflation in Higher Education. Draft Paper. Denton, TX: University of North Texas, September.
Kolevzon, M. S. 1981. "Grade Inflation in Higher Education: A Comparative Study." Research in Higher Education 15(3):195-212.
McPherson, M. A. 2006. "Determinants of How Students Evaluate Teachers." Journal of Economic Education 37(1):3-20.

McPherson, M. A., and R. T. Jewell. 2007. "Leveling the Playing Field: Should Student Evaluation Scores be Adjusted?" Social Science Quarterly 88(3):868-81.
McPherson, M. A., R. T. Jewell, and M. Kim. 2009. "What Determines Student Evaluation Scores? A Random Effects Analysis of Undergraduate Economics Classes." Eastern Economic Journal 35:37-51.
Mulvenon, S., and D. Ferritor. 2005. "Grade Inflation in Higher Education: Isolated or Systematic?" International Journal of Learning 12(6):55-62.
Nelson, J. P., and K. A. Lynch. 1984. "Grade Inflation, Real Income, Simultaneity, and Teaching Evaluations." Journal of Economic Education 15(Winter):21-37.
Nettles, M. T., L. W. Perna, E. M. Bradburn, and L. Zimbler. 2000. Salary, Promotion, and Tenure Status of Minority and Women Faculty in U.S. Colleges and Universities. NCES-2000-173. U.S. Department of Education National Center for Education Statistics, March.

Prather, J. E., G. Smith, and J. E. Kodras. 1979. "A Longitudinal Study of Grades in 144 Undergraduate Courses." Research in Higher Education 10(1):11-24.
Rojstaczer, S. 2008. "Grade Inflation at American Colleges and Universities." Available at [http://www.gradeinflation.com](http://www.gradeinflation.com).

Sabot, R., and J. Wakeman-Linn. 1991. "Grade Inflation and Course Choice." Journal of Economic Perspectives 5(1):159-70.
Sonner, B. 2000. "A is for 'Adjunct' Examining Grade Inflation in Higher Education." Journal of Education for Business 76(1):5-8.
Wooldridge, J. M. 2002. Econometric Analysis of Cross Sectional and Panel Data. Cambridge, MA: MIT Press.

Zangenehzadeh, H. 1988. "Grade Inflation: A Way Out." Journal of Economic Education 19(3):217-26.


[^0]:    *Direct correspondence to Michael A. McPherson, Department of Economics, P.O. Box 311457, Denton, TX 76203-1457 [mcpherson@unt.edu](mailto:mcpherson@unt.edu). Author McPherson will share all data and coding for replication purposes.
    ${ }^{1}$ See Rojstaczer (2008).
    ${ }^{2}$ It is certainly possible that students are simply better and average grades have risen as a natural result of this. However, there is some evidence to the contrary, usually based on the observation that the SAT and ACT scores of entering students have not noticeably increased and may in some periods have declined. See, for example, the work of Kolevzon (1981), Birnbaum (1977), and Cluskey et al. (1997).

[^1]:    ${ }^{3}$ Institutional factors, such as allowing students to drop courses at later dates than before and degree programs that permit students to take nontraditional (and perhaps easier) courses than before, have also been suggested (Prather, Smith, and Kodras, 1979), along with a host of others. This literature is neatly summarized by Mulvenon and Ferritor (2005).
    ${ }^{4}$ As noted previously, grade inflation seems to be at least somewhat more serious at private institutions. It is nevertheless a significant phenomenon at public institution such as ours. While each institution is unique, our results may be representative of the broader phenomenon. However, only an interinstitution study would allow us to speak to this.

[^2]:    ${ }^{5}$ It is possible that Students is endogenous. This would be the case if instructors who give higher grades attract larger classes. However, there are systemic constraints to dramatically increasing a particular instructor's class size. Typically, when an instructor's course fills, additional students are shunted to sections taught by other instructors rather than by finding a larger room or by increasing an instructor's teaching load.

[^3]:    ${ }^{6}$ We estimated the model removing foreign-born instructors from the analysis. This did not change our results dramatically; as a result we present the all-inclusive results here. The results from this more limited sample are available on request from the authors.
    ${ }^{7}$ Another instructor-level measure is tenure status. Our data only include rank as of 2005, so no allowance can be made for changes in tenure status.
    ${ }^{8}$ Although a tenure-track faculty member can change his or her rank over time, our data do not include this information.

[^4]:    Statistically different from comparable figure of white males at the 10 percent level
    **Statistically different from comparable figure of white males at the 5 percent level.
    *** Statistically different from comparable figure of white males at the 1 percent level.

[^5]:    ${ }^{9} \chi^{2}{ }_{(75)}=245.99$, significant at any conventional level of significance.
    ${ }^{10}$ Wooldridge's example uses time dummies, but the logic holds for a time trend as well. He uses a specific example of estimating the gap in wages for men and women over time using a fixed-effect estimator. He states (p. 267): "we can estimate how the gender gap has changed over time, even though we cannot estimate the gap in any particular period of time." In other words, one cannot identify the marginal effect of being male on wages, but one can estimate the difference in how wages change over time between men and women. Analogously, we cannot estimate the marginal effect of "white male," but we can estimate (and test) the difference in the rate of change in GPAs over time for white males versus other gender-ethnicity categories. We also estimate the model using Wooldridge-type time-dummy interactions with the gender/ethnicity categories. The resulting inflation rates show little difference to those from the time trend model; we choose the time trend model mainly for ease of discussion. A full set of predictions from both models are available from the authors.
    ${ }^{11}$ Our sample includes 1,005 individual white male instructors, 585 white females, 174 nonwhite males, and 107 nonwhite females. More specific subdivisions result in very small sample sizes.

[^6]:    ${ }^{12}$ The estimation also includes dummy variables for each department as well as departmentspecific trends. For brevity, the coefficients of the department-specific effects are not reported, but they are available from the authors.

[^7]:    ${ }^{13}$ Test statistic for male versus female trend is $F(1,46086)=10.41$, significant at 1 percent level. Test statistic for white versus nonwhite trend is $F(1,46086)=1.34$, insignificant as any conventional level of significance.

[^8]:    ${ }^{14}$ It is possible that instructors in departments with high student-faculty ratios may feel freer to give lower grades than instructors in departments that have fewer students. Our data do not permit us to control for this possibility.

