Applied Economics, 2013, **45**, 1185–1200

Routledge
Taylor & Francis Group

Whose fault is it? Assigning blame for grade inflation in higher education

R. Todd Jewell, Michael A. McPherson* and Margie A. Tieslau

Department of Economics, University of North Texas, P.O. Box 311457, Denton, TX 76203-1457, USA

This study attempts to isolate the potential sources of grade inflation and to measure their relative importance. We incorporate existing models of grade inflation into a model of grade inflation at the department level. Our data comprise 1683 separate courses taught in 28 different academic departments by 3176 distinct instructors at a large public university over two decades. Our results suggest that incentives to inflate grades vary according to characteristics of academic departments. However, the vast majority (over 90%) of grade inflation observed in our data is estimated to be a result of either university-level factors or instructor-specific characteristics.

Keywords: grade inflation; university grade trends; fixed effect model; higher education

JEL Classification: A22; A20

I. Introduction

By now it is essentially universally accepted that grades at American colleges and universities have shown a general upward trend over the past several decades. Many studies have documented this trend, and Stuart Rojstaczer has compiled the most comprehensive 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 four-point scale per decade since the late 1960s, with grade inflation at private schools proceeding at a more rapid pace than at public institutions (Rojstaczer, 2008). In addition, Farley (1995), Cluskey *et al*. (1997), Grove and Wasserman (2004) and Bello and Valientes (2006) find evidence of rising grades in a variety of colleges and universities.

This phenomenon is known as grade inflation, and researchers from various disciplines have speculated about its causes. Conceivably, students are simply better and average grades have risen as a result of this. There is some evidence to the contrary, usually based on the observation that the Scholastic Aptitude Test (SAT) and American College Test (ACT) scores of entering students have not noticeably increased and may in some periods have declined (Birnbaum, 1977; Kolevzon, 1981; Cluskey *et al*., 1997). Others have argued that faculty members are inflating grades in response to the now widespread use of Student Evaluations of Teaching (SETs) in promotion, tenure and merit evaluations. 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

*Corresponding author. E-mail: mcpherson@unt.edu

http://www.tandfonline.com
http://dx.doi.org/10.1080/00036846.2011.621884

before (Prather *et al.*, 1979) have also been suggested, along with a host of others. This literature is summarized by Mulvenon and Ferritor (2005).

This study is an attempt to isolate the potential sources of grade inflation and to measure their relative importance. We add to the literature in two ways. First, we incorporate existing models of grade inflation into a model of grade inflation at the department level. Our model establishes testable hypotheses regarding the incentives for grade inflation at the department level. Second, we employ a rich data set to test hypotheses concerning grade inflation for different aspects of the undergraduate classroom experience. Our data allow for an analysis of grade inflation that may result from the behaviour of students, instructors, departments or the university itself. Although the existing literature is vast, statistical work to this point has been somewhat limited by available data. Our research involves data from 1683 separate courses taught in 28 different academic departments by 3176 distinct instructors at a large public university over two decades (1984 to 2005), a data set which covers a substantially longer time period and considerably more courses and departments than any other study. The unique nature of the data permits a much more comprehensive analysis of the determinants of grade inflation than has been possible previously. Our results suggest that department characteristics, especially size and the importance of undergraduate education, have an impact on grade inflation, but the vast majority of the grade inflation observed in our data is a result of either university-level factors or instructor-specific characteristics.

II. Previous Literature

Many studies have addressed the theoretical underpinnings of grade inflation. Kelley (1972), McKenzie (1975), Lichty *et al.* (1978) and Dickson (1984) are examples of economic models in which student or instructor efforts to maximize utility might lead to grade inflation. To summarize, instructors may have an incentive to inflate grades in order to improve their scores in the students' evaluations of their teaching. There are as well numerous empirical studies that have examined whether grade inflation in fact affects SET scores. Early studies include Voeks and French (1960), Nicholas and Soper (1972), Mirus (1973), Nelson and Lynch (1984), Aigner and Thum (1986) and Zangenehzadeh (1988). More recently, the interested reader can examine Germain and Scandura (2005), Hamermesh and Parker (2005), Isely and Singh (2005), McPherson (2006) and McPherson *et al.* (2009). While there are exceptions, the general conclusion of this part of the literature is that to some extent instructors can 'buy' better evaluation scores by inflating expected grades. A number of papers have made suggestions for addressing the problem of grade inflation. Zangenehzadeh (1988), McPherson (2006), McPherson and Jewell (2007) and McPherson *et al.* (2009) argue that SETs should be adjusted so that the advantage to instructors of inflating grades is mitigated. Nagle (1998) and Felton and Koper (2005) suggest the more direct approach of dividing students' grades by the average class grade so that grades are comparable over instructors and courses.

There is also a small literature that attempts to parse out the determinants of grade inflation, and an even smaller one that suggests palliative measures for grade inflation. Studies often discover evidence of differential grade inflation by subject. Most commonly, disciplines that are traditionally more quantitative such as economics, mathematics, psychology, chemistry and computer science exhibit less evidence of grade inflation, while courses such as art, English, music, speech and political science typically have higher rates of grade inflation (Prather *et al.*, 1979; Sabot and Wakeman-Linn, 1991; Cheong, 2000). Some contrary findings occasionally emerge. For example, Anglin and Meng (2000) find that some quantitative disciplines (biology, physics and chemistry) are among the most likely to inflate grades.

Other studies examine whether certain characteristics of instructors make them more or less likely to inflate grades. Kolevzon (1981) argues that low grade-inflation departments are more likely to have larger proportions of male faculty members, although the direction of causality is not examined. Sonner (2000) finds a positive correlation between average grades and the proportion of a department's classes taught by adjunct faculty. Dickson (1984) presents evidence that instructors in departments with fewer students per instructor are more likely to inflate grades, perhaps as a result of faculty concerns about job security. Dickson (1984) also finds no evidence that a faculty member's tenure status affects his or her likelihood of inflating grades. Cheong (2000) makes an interesting observation regarding trends in grades at the University of Hawaii-Manoa: Grade inflation seems to exist, and there is some evidence of a cyclical pattern, with spring semester grades higher on average than those in the fall.

Unfortunately, much of the existing literature on the determinants of grade inflation is rather limited in terms of hypothesis testing. Kolevzon (1981), Sabot and Wakeman-Linn (1991), Anglin and Meng (2000), Sonner (2000) and Grove and Wasserman (2004) all use simple t-tests or other comparisons of descriptive statistics to examine the issues. Studies that use regression methodologies suffer from other shortcomings. Dickson (1984) examines grades in over 600 undergraduate courses, but over only a single academic year. Cluskey *et al*. (1997) consider a 15-year period, but their data involve just four selected accounting classes. Cheong (2000) estimates separate regressions for student gender, by undergraduate versus graduate status, and by upper-division versus lower division courses with mean grade in all courses at the university as the dependent variable as a function of a time trend. Prather *et al*. (1979) consider individual students' grades over a 6-year period, but these data are not treated as a panel since the authors estimate a separate regression for 144 separate courses.

III. An Academic Department's Optimal GPA

Following past literature, this article views the issue of grade inflation as the outcome of economic processes. An economic analysis of grade inflation must consider the behaviour of economic agents at several distinct levels. First, there is the level of the individual student, as he or she makes decisions regarding how much time to spend on alternative activities. One option is for a student to spend time on coursework, a decision influenced by the existence of (and degree of) grade inflation. Next, individual instructors are faced with costs and benefits of inflating grades in their classes, in response to the behaviour of students and their own preferences and constraints. At the next level, departments must make decisions on how to allocate scarce resources between teaching, research and other departmental functions; these decisions will also be affected by the cost and benefit of inflating grades within and among different departments of the same university. Finally, decisions are made at the university level based on the cost and benefit of institution-wide grade inflation, to the extent that grade inflation impacts the number of applicants and a school's reputation.

McKenzie (1975) and Lichty *et al*. (1978) theoretically model the behaviour of students in the face of grade inflation, while Dickson (1984) models the behaviour of instructors and Chan *et al*. (2007) model the behaviour of universities under similar inflationary grading. However, we are unaware of any existing theory that attempts to shed light on the decisions of departments within a given university. This article does not attempt to create a general mathematical model of the interrelated decisions of students, instructors, departments and universities; instead, we integrate the implications of existing theoretical models with those of a model of department behaviour in the presence of grade inflation.

In general, we assume that a department can indirectly control the number of its students by assigning higher or lower average grades. Specifically, higher average department grades are assumed to attract students to the courses taught in that department. This in turn affects the department's output of research, teaching and service in predictable ways. We assume that individual instructors will act in the best interests of departments; we do not attempt to model the mechanism whereby individual faculty members might be led to inflate or deflate grades for the department's sake. Beginning with a general model of department behaviour, assume that total department output (D) equals the aggregated amount of research output (R), teaching output (T) and service output (S). Specifically, a department's total output function is defined as the following, where $d(\cdot)$ represents a function that aggregates research, teaching and service production into a single output measure:

$$D = d(R, T, S) \tag{1}$$

Assume that each department maximizes total output each time period subject to exogenous monetary resources in that time period (M). The subscript on time is suppressed in the following model. However, a department's decision-making process is assumed to be period-by-period, where values from previous periods are by definition exogenous. Expression (2) defines a department's optimization problem, where P_r, P_t and P_s are the exogenous marginal (monetary) costs of one unit of research production, teaching production, and service production respectively, and P_s is normalized to 1.

$$\text{Max } D = d(R, T, S), \quad \text{subject to } M = P_r \times R + P_t \times T + S \tag{2}$$

To facilitate an analysis of department behaviour with respect to the inflation of average grades, we assume the research, teaching and service output in each period are functions of the number of students taking courses in the department (N) and exogenous characteristics of the department, faculty and students (the vector X, components defined later). Our model is based on the assumption that the number of faculty available in any time period is exogenous, as this number is established prior to each academic year. Therefore, the faculty resource is fixed when a department makes its choice of research, teaching and service. The fixed faculty resource is a component of the X vector of exogenous characteristics. In addition, a fixed faculty resource implies a fixed time resource to be used in the production of department output, a fact reflected in the specified productive relationships below.

The definition of research output varies by discipline and department, but it is invariably an aggregation of the total scholarly output of all faculty members. R is assumed to be a decreasing function of N since any increase in students will imply that less of the faculty resource will be available to produce research output. This follows from the assumption of a fixed faculty resource and the fact that additional students require an additional time commitment on the part of the department.

$$R = r(N, X), \quad \text{with } r'_n < 0 \tag{3}$$

Teaching output is the number of student credit hours, scaled by some measure of teaching quality. While increasing N may reduce teaching quality as the fixed faculty resource is spread over more students, we assume that the direct effects of increasing students on credit hours will offset the indirect effects on teaching quality. Therefore, T will be an increasing function of N since any increase in students will lead to increased student credit hours.

$$T = t(N, X), \quad \text{with } t'_n > 0 \tag{4}$$

Service output is the aggregation of time spent by faculty members in department and university service positions including student advising. As is the case with teaching, additional students require additional service activity; thus, service will be an increasing function of N, given the fixed faculty size in any period.

$$S = s(N, X), \quad \text{with } s'_n > 0 \tag{5}$$

This study concentrates on the influence GPA may have on departmental behaviour. As discussed above, a department can increase the number of students it serves by inflating average GPA (G)

$$N = n(G), \quad \text{with } n'_g > 0 \tag{6}$$

Substitution of expressions (3)–(6) into the optimization problem of expression (2) implies that a department will maximize departmental output by choosing average GPA given the production function and resource constraint:

$$\begin{aligned} &\text{Max } D = d(r[n(G), X], t[n(G), X], s[n(G), X]) \\ &\text{subject to } M = P_r \times r[n(G), X] + P_t \times t[n(G), X] \\ &\quad + s[n(G), X] \end{aligned} \tag{7}$$

The optimization problem given in expression (7) can be solved by substitution. For purposes of tractability, assume that research, teaching and service each show constant (positive) marginal productivity in producing department output. Specifically, let $d'_r = MP^d_r = a$, $d'_t = MP^d_t = b$, and $d'_s = MP^d_s = c$. Implicitly, we also assume that department output is separable in research, teaching and service. A detailed derivation of the mathematical expressions in this section is given in the Appendix. The First-Order Condition (FOC) for maximization is listed below.

$$r'_n n'_g[a - cP_r s'_n n'_g] + t'_n n'_g[b - cP_t s'_n n'_g] = 0 \tag{8}$$

The FOC (8) relates optimal average GPA to the exogenous parameters M, P_r, P_t and X. Because the function for optimal GPA results from production maximization and GPA is modelled as a productive input, we refer to this function as a 'demand function' for GPA.

The expression implies that GPA will be increased to the point that the marginal benefit of increasing average GPA equals the marginal cost of increasing average GPA. It is clear that optimum G chosen by each department will be influenced by the nature of each department, as well as whether this department receives a higher payoff to research, teaching or service. For instance, PhD-granting departments generally place a greater emphasis on research output and may consider teaching and service output to be of lesser importance. In the context of FOC (8), this suggests that the marginal benefit of adding students (by inflating grades) may be smaller for research-focused departments. That is, the marginal productivity of research for such departments (a) is likely to be high, and the reduction in research output that results from an increase in the number of students (r'_n) may be larger for PhD than for non-PhD departments. The marginal cost of adding students may also differ according to a department's focus. Ultimately, whether PhD-granting departments are more or less likely to inflate grades is an empirical question.

By using comparative statics analysis, we can use the FOC to determine the expected sign of a change in the exogenous variables. For purposes of this study, we are interested in the change in average GPA when department, faculty or student characteristics change. Substitution of the optimum level of G (denoted G^*) into (8) and partial differentiation with respect to a component of X yields the following expression, where D''_{gg} refers to the second derivative of the maximized equation with respect to G:

$$\begin{aligned} \partial G^*/\partial X = &-n'_g(r''_{nx}[a - cP_r s'_n n'_g] + t''_{nx}[b - cP_t s'_n n'_g] \\ &- cn'_g s''_{nx}[P_r r'_n + Pt t'_n])/D''_{gg} \end{aligned} \tag{9}$$

Given the second-order condition, $D''_{gg} < 0$, and given $n'_g > 0$ by assumption, the sign of a change in any component of the X vector of department and faculty characteristics will be determined by the sign of the following relationship:

$$r''_{nx}[a - cP_r s'_n n'_g] + t''_{nx}[b - cP_t s'_n n'_g] - cn'_g s''_{nx}[P_r r'_n + P_t t'_n] \tag{10}$$

By assumption, n'_g, t'_n, s'_n, prices and marginal products all are positive. However, r'_n is negative, and r''_{nx}, t''_{nx} and s''_{nx} will vary according to the exogenous change. Relationship (10) represents a weighted sum of the change in marginal productivity of N in producing department output with respect to changing a component of the X vector. If the sum is positive, then $\partial G^*/\partial X > 0$, and an increase in the X component will lead to an increase in average GPA; if the sum is negative, then $\partial G^*/\partial X < 0$. Although many of the terms in (10) can be signed *a priori*, the total sign cannot be determined without knowledge of the relative sizes of prices and productivity, which will vary by the nature of the department's goals for research, teaching and service. Thus, the sign of $\partial G^*/\partial X$ is ambiguous and must be determined empirically, given that it represents a complex interaction between marginal productivity and prices.

We return to relationship (10) in the subsequent discussion of the empirical model and expected coefficient signs. However, a quick example may illuminate the intuition behind (10) and the difficulties involved in signing the expression. Consider a change in X that does not impact the marginal productivity of N in terms of service so that s''_{nx} equals zero. In this case, signing $\partial G^*/\partial X$ means signing the following:

$$r''_{nx}[a - cP_r s'_n n'_g] + t''_{nx}[b - cP_t s'_n n'_g] \tag{11}$$

The sign of (11) will be positive if the net marginal return to X is positive, and negative otherwise. Further assume that r''_{nx} and t''_{nx} are both positive; even under such simplifying conditions, the sign of (11) is unambiguously positive if and only if $(a - cP_r s'_n n'_g) > 0$ and $(b - cP_t s'_n n'_g) > 0$. Note that an increase in either $cP_r s'_n n'_g$ or $cP_t s'_n n'_g$ would increase the likelihood that relationship (11) is negative. Thus, high prices for research and/or teaching relative to the normalized price of service (P_r and P_t) could result in $\partial G^*/\partial X < 0$. Likewise, a department with a large commitment to service (i.e. large c relative to a and/or b) will be more likely to see $\partial G^*/\partial X < 0$.

IV. Data and Empirical Methods

The data set is composed of course-level observations for 21 academic years (1984–85 to 2004–05) 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 PhD in many of those programs. These data have the advantage of covering all UNT courses over the study period. As is common in studies of grade inflation, only undergraduate courses are considered. In addition, certain courses are excluded from the analysis because they are organized differently than traditional university courses and their grading systems may be nonstandard. Individualized classes, such as private music lessons, independent studies, honors research and theses, practica, driver's education, and internships and cooperative education, are not included. Similarly, student teaching, institutes and study tours, and field studies are not considered. Certain other courses are excluded since they may also have distinctive grading systems; these include activity-based physical education courses and lab sections in which a separate grade is entered from the classroom portion of the course.

Since we are especially interested in grade inflation at the department level, observations in departments that did not offer appropriate undergraduate courses over the entire sample period are excluded. Furthermore, courses taught in the summer or other terms outside of the regular semester are excluded because of concerns over comparability. To exclude outliers, very large (above 100 students) and very small (fewer than 10 students) classes are also excluded, which equates to deleting the top and bottom 5% of classes in terms of student size. Finally, instructors with less than four courses taught are excluded to facilitate estimation of instructor-specific effects. After making these exclusions, the useable data include 50 318 observed course sections of 1683 courses taught in 28 academic departments by 3176 instructors. These data represent 56% of the approximately 90 000 courses taught over the time period under study. Included departments and department sample sizes are listed in Table 2.

An empirical model of grade inflation

The above theoretical model suggests that average GPA in a department will be related to the characteristics of the economic actors involved. Specifically, the model predicts that average GPA will vary by the characteristics of the department itself, of the department's faculty and of the students whom the department serves. Table 1 lists summary statistics for the independent variables. To measure the characteristics of the department, we include the number of faculty members in each department each year (*Faculty size*), average SAT scores each year for students who take department courses (*SAT_dep*) and the percentage of students in a department each year who are either freshmen or sophomores (*Underclassmen_dep*). Available SAT data include average SAT scores for entering freshmen at each university in each year. We construct a department-specific SAT score (*SAT_dep*) as the average SAT score weighted by the proportion of students in that department that should have entered college in each year. *SAT_dep* is the average SAT score for a given department relative to the national average, and as such measures improvement of UNT students relative to all college students over time. The mean of *SAT_dep* for non-PhD departments (2.7) implies that students in our non-PhD sample had on average 2.7% higher scores on the SAT than the national average. The variable *SAT_class* (discussed later) is constructed in the same manner at the course level. In addition, we include time-invariant fixed effects for each department. Furthermore, we separate the sample into PhD departments and non-PhD departments since we expect the goals of each for research, teaching and service to be largely different for such departments.

Recalling the discussion of expression (10), the expected effect of changing an exogenous variable is difficult to sign *a priori*. Consider the effect of an increase in *Faculty size* on average GPA in a department. In this case, r''_{nx}, t''_{nx} and s''_{nx} will each likely be positive, since adding a faculty member will probably increase research, teaching and service productivity. The overall sign of (10) will be positive as long as the benefit of adding faculty members outweighs the cost, where the cost of adding faculty members is a function of prices, the importance of service in department output and service productivity. Of course, the sign of relationship (10) may differ between PhD and non-PhD departments in as much as the cost and benefit parameters vary over these departments. It will be similarly difficult to sign the effects of the other department-level, exogenous variables based on the model, but it may be reasonable to speculate that departments populated by students scoring well on the SAT will tend to assign higher grades on average. Similarly, if experience in the college environment and accumulation of knowledge helps students perform better, a department that has a greater proportion of freshmen and sophomores might be expected to assign lower average grades.

Table 1. Summary statistics by course ($n = 50\,318$)

	Mean	SD	Minimum	Maximum
Non-PhD departments ($n = 13\,826$)				
Dependent variable				
GPA	2.858	0.550	0.708	4.000
Departmental-level variables				
Faculty size	21.99	8.366	4	41
SAT_dep	2.693	4.402	−8.839	8.758
Underclassmen_dep	32.31	14.42	7.331	69.24
Course-level variables				
Class size	30.81	16.24	10	100
SAT_class	2.681	4.468	−9.871	8.974
Underclassmen_class	31.24	27.17	0	100
PhD departments ($n = 36\,492$)				
Dependent variable				
GPA	2.688	0.600	0.111	4.000
Departmental-level variables				
Faculty size	54.79	34.66	10	123
SAT_dep	2.013	4.585	−10.01	8.708
Underclassmen_dep	41.12	23.96	1.231	73.49
Course-level variables				
Class size	35.22	16.99	10	100
SAT_class	1.989	4.656	−11.59	8.974
Underclassmen_class	39.93	35.55	0	100

Given the wide range of instructor characteristics that may influence GPA, we include such characteristics in the form of an instructor-specific fixed effect for each of the 3176 distinct instructors in our sample. We include the following variables to measure student characteristics at the course level: *SAT_class*, *Underclassmen_class* and *Class size*. Although the effects of these variables are not modelled in this article, we can establish hypotheses based on past research (see, e.g. Carney *et al.*, 1978). We hypothesize that GPA should be directly related to the relative quality of students, as measured by *SAT_class*. Furthermore, as at the department level student grades should improve with experience and maturity, so classes with higher proportions of underclassmen should have lower GPAs. The relationship between the number of students (*Class size*) and the average grade in a particular course may be affected by differences in pedagogy at different class sizes. For example, instructors of relatively smaller classes may be able to provide more time to each student, both during class and outside of class. However, as class sizes increase, the time an instructor can spend per student necessarily declines. Other sorts of pedagogical changes may also occur as class sizes increase. In general, one might expect teaching methods to shift towards assessments that may encourage learning by rote. It is unclear *a priori*

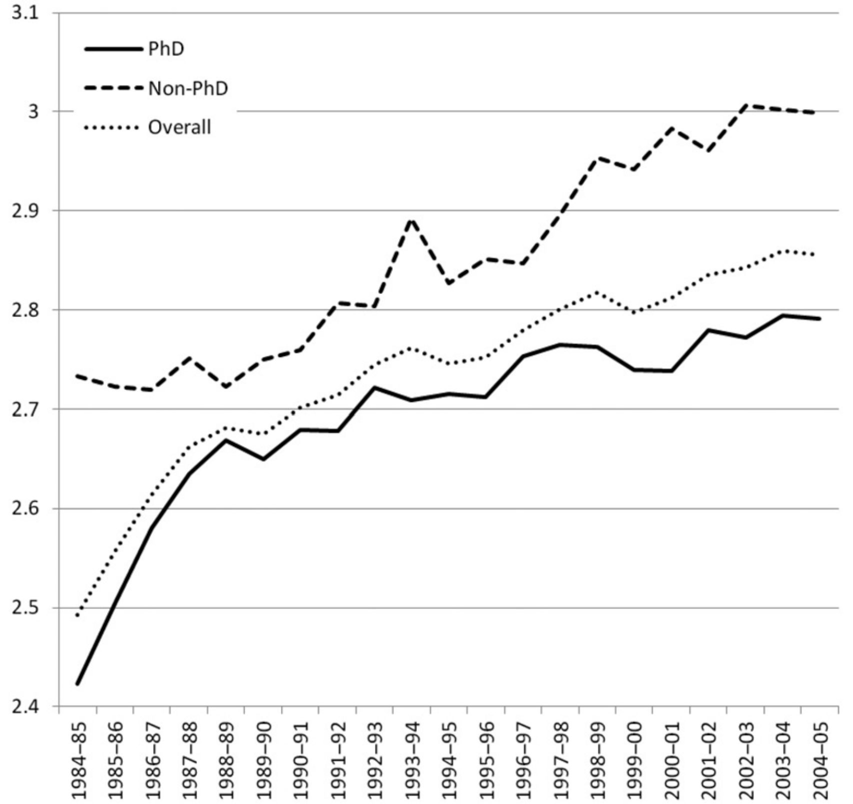

Fig. 1. Average GPA by academic year

how this might affect average grades in a given class. Note that the theoretical model treats the number of students in a department as endogenous to the department's decision to inflate grades. The number of students in each class is not endogenous to the department's decision, although it may be an endogenous variable for each instructor. We treat *Class size* as an exogenous variable, since we concentrate on the decision of the department.

Our dependent variable, *GPA*, is the average GPA in each course on a standard four-point system. Average GPA in all courses is 2.734, with PhD departments and non-PhD departments having averages of 2.688 and 2.858, respectively. Overall UNT average GPA increased from 2.493 in 1984 to 2.856 in 2005, an increase of 0.363 grade points and a combined increase of 14.6%. For PhD departments, the increase from 1984 to 2005 was 0.368 grade points, while for non-PhD departments it was only 0.266. Note that the overall increase and the PhD department increase is greater than the 0.307 average increase over roughly the same period reported by Rojstaczer (2008) for universities across the US, which may indicate that UNT has experienced more grade inflation than other schools and that much of this extra increase was driven by grades in UNT PhD departments. The time pattern of mean GPA at UNT for PhD and non-PhD departments is shown in Fig. 1.

An important purpose of this study is to analyse differences in grade inflation across departments. As an initial step, we report summary statistics by department in Table 2, where departments are ranked from lowest to highest average GPA. As expected, we see that quantitative disciplines tend to have lower average course grades; the five departments with the lowest average GPAs are mathematics, accounting, economics, chemistry and finance. Also, less quantitative disciplines tend to have higher GPAs; the five departments with the highest GPAs are teacher education, speech, dance, recreation, radio, television and film. This corresponds closely with findings reported in the existing literature (e.g. Prather *et al.*, 1979; Sabot and Wakeman-Linn, 1991; Cheong, 2000). Furthermore, there appears to be a correlation between a department having a PhD program and average grades. For instance, of the departments with the lowest seven GPAs only economics does not have a PhD program, and among the departments with the highest eight GPAs only teacher education and engineering have a doctoral program.

Table 2. Summary statistics by department (listed from lowest to highest average GPA)

Department	Variable	Mean	Department	Variable	Mean
Mathematics*	*GPA*	2.181	Management*	*GPA*	2.751
$n = 4767$	*Faculty size*	65.69	$n = 2010$	*Faculty size*	37.47
Accounting*	*GPA*	2.215	Computer Science	*GPA*	2.790
$n = 1699$	*Faculty size*	28.68	$n = 1604$	*Faculty size*	25.07
Economics	*GPA*	2.415	Philosophy*	*GPA*	2.826
$n = 1443$	*Faculty size*	20.84	$n = 924$	*Faculty size*	12.30
Chemistry*	*GPA*	2.521	Marketing*	*GPA*	2.841
$n = 569$	*Faculty size*	14.49	$n = 1616$	*Faculty size*	24.23
Finance and Real Estate*	*GPA*	2.541	Anthropology	*GPA*	2.849
$n = 2274$	*Faculty size*	31.02	$n = 598$	*Faculty size*	10.72
History*	*GPA*	2.549	Sociology*	*GPA*	2.901
$n = 2605$	*Faculty size*	38.83	$n = 973$	*Faculty size*	19.76
Political Science*	*GPA*	2.599	Journalism	*GPA*	2.903
$n = 1427$	*Faculty size*	29.17	$n = 1137$	*Faculty size*	16.22
Languages	*GPA*	2.602	Kinesiology	*GPA*	2.926
$n = 2827$	*Faculty size*	32.23	$n = 1777$	*Faculty size*	28.89
Psychology*	*GPA*	2.630	Engineering*	*GPA*	2.936
$n = 1229$	*Faculty size*	34.81	$n = 1139$	*Faculty size*	15.46
Physics*	*GPA*	2.632	Radio, Television and Film	*GPA*	3.123
$n = 924$	*Faculty size*	19.13	$n = 1194$	*Faculty size*	19.95
English*	*GPA*	2.718	Speech and Hearing Sciences	*GPA*	3.256
$n = 8282$	*Faculty size*	111.3	$n = 458$	*Faculty size*	9.120
Biology*	*GPA*	2.722	Dance and Theater	*GPA*	3.268
$n = 1166$	*Faculty size*	28.80	$n = 1405$	*Faculty size*	14.76
Business Computer	*GPA*	2.744	Recreation and Health	*GPA*	3.283
Information Systems*			$n = 721$	*Faculty size*	14.91
$n = 1210$	*Faculty size*	22.42	Teacher Education*	*GPA*	3.458
Geography	*GPA*	2.746	$n = 3788$	*Faculty size*	55.76
$n = 662$	*Faculty size*	10.50			

Note: *Indicates PhD department.

Long-run properties of the data

Before estimating our regression model, it is necessary to determine whether or not the variables are stationary, since models that include nonstationary variables cannot be estimated by traditional econometric methods. When testing for stationarity in panel data, one may choose from among a number of Dickey–Fuller (DF)-based unit root tests, such as the test of Im, Pesaran and Shin (IPS; 2003) or the Lagrange Multiplier (LM)-based unit root test of Im, Lee and Tieslau (ILT; 2005). While many researchers might choose the popular IPS panel unit root test, we employ the ILT test because it offers three important advantages over the IPS test, which are described below.

First, it is well known that DF-based unit root tests such as the IPS test are not well suited to testing data that contain a deterministic trend, as one might expect to find in our analysis (Schmidt and Phillips, 1992). This is because of the way that a series' level and trend are specified in the testing equations for DF-based unit root tests. Second, the ILT test has superior size and power properties over the IPS test. This is important when one considers that the low power of conventional unit root tests often leads researchers to conclude that variables are nonstationary when, in fact, they are stationary. Third, the ILT test allows for breaks in the level of the series under both the null and alternative, without dependence on nuisance parameters. As Perron (1989) showed, it is important to account for breaks, if they exist, since there can be a significant loss of power when testing for a unit root if one ignores existing structural breaks. In addition, the ILT test corrects for serially correlated errors and cross-correlation across panels.

The procedure for testing for a unit root using the ILT test is summarized as follows. We begin by testing for a unit root while allowing for two breaks in level, the locations of which are determined endogenously from the data. If we find that the data is not characterized by two breaks, we repeat the testing procedure allowing for one break in level. If no level breaks are found to exist, we apply the no-break LM unit root test of Schmidt and Phillips (1992). In this way, we jointly determine the number and location of breaks, if they exist, as well as the

Table 3. Panel unit root tests

Variable name	Test statistic
GPA	−75.21***
Faculty size	−81.49***
Faculty size^2	−86.08***
SAT_dep	−99.56***
SAT_dep^2	−127.57***
Underclassman_dep	−53.93***
Underclassman_dep^2	−84.31***
Class_size	−74.95***
Class_size^2	−65.14***
SAT_class	−56.14***
SAT_class^2	−84.30***
Underclassmen_class	−116.36***
Underclassmen_class^2	−64.32***

Note: ***Denotes significance at the 1% level.

appropriate number of lags to employ to control for any correlation in the errors, while also testing for the presence of a unit root in the series. We note that the test statistic follows a standard normal distribution and a significant test statistic indicates that the series is stationary. Table 3 presents the results of the panel unit root tests applied to the variables of our analysis. In all cases we are able to reject the null of a unit root at the 99% level of confidence or better, implying that the series are stationary in levels. Thus, we are able to proceed with estimation of our model using all of the variables expressed in levels.

Estimation methods

After ruling out the potential of data that are not stationary, we employ standard panel data techniques to estimate the determinants of average GPA using an unbalanced panel of UNT courses from 1984–85 to 2004–05, including the above-mentioned independent variables that measure characteristics of departments, faculty and students. In order to capture any nonlinearities in the effects of the independent variables on *GPA*, we also include squared terms for all independent variables. Furthermore, we include a linear time-trend (*Trend* and its square, *Trend*^{2}) to capture university-level grade inflation that is not measured in the time pattern of other independent variables. Panel data techniques allow for estimating a fixed effect for each instructor. The demand function for optimal GPA implied by expression (8) can be restated as the following data-specific equation:

$$GPA_{it} = (\alpha + u_j) + D_{kt}\delta + S_{it}\sigma + Trend_t + Trend_t^2 + \varepsilon_{it} \quad (12)$$

where i indexes course, k indexes department and t indexes year. The instructor fixed effect, u_j, represents the extent to which instructor j (who teaches course i at year t) has GPAs higher or lower than the overall average net of the influence of other independent variables. The vectors D_{kt} and S_{it}, represent characteristics of department k in year t ($Faculty\ size_{kt}$, SAT_dep_{kt}, $Underclassmen\_dep_{kt}$, and squared terms) and student characteristics in course i in year t ($Class\ size_{it}$, SAT_class_{it}, $Underclassmen\_class_{it}$, and squared terms), respectively. δ, σ and τ represent vectors of parameters to be estimated, and ε_{it} represents the random error term.

The panel nature of the data also allows for estimating department-level fixed effects. This option gives the researcher the ability to include department-specific constant terms, similar to the instructor-specific constants in Equation 12. Including department-specific effects will improve the explanatory power of the empirical model if there are some unobserved department characteristics that explain differences in average GPA across departments. For instance, departments may have distinct (time-invariant) preferences towards 'student-centeredness' that will impact the cost and benefit of adding students. In terms of Equation 12, including department fixed-effects would imply including those effects in the vector D_{kt}.

V. Results

Table 4 gives results from two regressions based on Equation 12, including instructor-specific and department-specific effects both of which are time invariant. *Regression 4.1* includes only courses from those departments that do not award the PhD, while *Regression 4.2* includes only courses in PhD-granting departments. Such departments are separated due to expected differences in the costs and benefits of adding undergraduate students. Unsurprisingly, a Chow-type test (results available from the authors) indicates that the PhD and non-PhD samples should not be combined. For brevity, the coefficients of the instructor-specific and department-specific effects are not reported, but they are available from the authors. The SEs reported in Table 4 are corrected for clustering on department, instructor and course. Some studies (see, e.g. Carney *et al.*, 1978; Anglin and Meng, 2000) employ percentages of As and Bs awarded as a dependent variable instead of average GPA. The models reported in Table 4 were also estimated using grade percentages. Because these estimations do not show results that are substantively

Table 4. Regression results: instructor-specific and department-specific effects

	Regression 4.1: Non-PhD departments $n = 13\,826$		*Regression 4.2*: PhD departments $n = 36\,492$	
	Coefficient	SE	Coefficient	SE
Trend	−0.00875[a]	0.0058	0.03694***	0.0036
Trend^2	0.00057***	0.0002	−0.00125***	0.0001
Faculty size	0.00331[a]	0.0038	−0.00414***	0.0009
Faculty size^2	0.00006	0.0001	0.00003***	0.0001
SAT_dep	0.02191***	0.0065	0.01406***	0.0037
SAT_dep^2	−0.00217***	0.0006	−0.00207***	0.0003
Underclassmen_dep	−0.00372*	0.0020	0.00637***	0.0016
Underclassmen_dep^2	0.00005*	0.0000	−0.00007***	0.0000
Class size	−0.00996***	0.0008	−0.00877***	0.0005
Class size^2	0.00005***	0.0000	0.00005***	0.0000
SAT_class	−0.01420***	0.0054	−0.02150***	0.0031
SAT_class^2	0.00191***	0.0006	0.00191***	0.0003
Underclassmen_class^2	−0.00631***	0.0005	−0.00790***	0.0003
Underclassmen_class^2	0.00003***	0.0000	0.00005***	0.0000
	r^2 (within) = 0.1187		r^2 (within) = 0.0927	

Notes: [a]Linear and squared terms jointly significant at the 1% level.
*** and * denote coefficient significant at the 1 and 10% levels, respectively.

different than those reported here, they are excluded from the article. However, these results are available from the authors.

Time trend

The estimate of the trend suggests that there is grade inflation at the university level. For non-PhD departments, the inflation is increasing at a slightly increasing rate; for PhD departments, grades are inflating at a decreasing rate. Based solely on the time trend, the results indicate that average GPA at UNT increased by 0.188 points from 1984–85 to 2004–05, accounting for 52% of UNT's overall increase. The time trend is picking up the effect of time-varying factors not measured by other independent variables. More specifically, this result suggests that more than one-half of the observed grade inflation at UNT over the sample period would have occurred even if none of the explanatory variables had changed. Thus, the time trend either reflects changes in excluded variables or a general, institution-wide trend towards increases in average grades.

One variable that is excluded in the estimation is overall growth in UNT's student population. Specifically, the student population was approximately 14 000 in 1984–85, and it grew to over 25 000 in 2004–05. As is the case at many universities, growth in the student population has outpaced growth in the faculty resource at UNT. Perhaps the time trend is measuring the grade inflation associated with marketing the university to students or to a general relaxation of course requirements allowing existing resources to absorb the increased student population. In addition to an overall increase in student population, UNT also experienced a cultural change over the sample period, moving from a teaching-focused university (especially undergraduate teaching) to a more research-focused institution with lower teaching loads and higher research expectations. Perhaps the university-level change in focus also resulted in an institution-wide change in the costs and benefits of increasing average grades. It is also interesting to note that UNT is continuing its movement towards being a more research-centred university, as well as increasing the size of the student population. One might expect that university-level grade inflation will continue to increase at UNT and may continue to be above the national average.

Department-level variables

For non-PhD departments, *Faculty size* has a positive effect on *GPA*, which increases as the number of faculty grows. *Faculty size* and *Faculty size*^{2} are jointly significant at the 1% level. The marginal effect on increasing the number of faculty in a department equals 0.006 grade points at the non-PhD-sample mean of *Faculty size*. At first, as the average quality of students in a department's courses improves relative to the national average, GPA rises. Above SAT scores five percentage points above the national average (the 53rd percentile of the sample), further improvements in student quality lead to lower average grades. It may be the case that the presence of substantially better students allows more challenging material to be presented or that better students are attracted to more difficult courses and majors.

Interestingly, the relationship between the proportion of freshmen and sophomores in a department and grades is U-shaped. As departments' proportions of underclassmen rise, average grades fall until about the 63rd percentile, after which further additions of underclassmen lead to higher grades. Presumably, while a lack of experience and maturity among freshmen and sophomores does lead to lower grades, at some point departments' cost-benefit calculus changes. For instance, service departments (i.e. those that teach substantial numbers of non-majoring students) may find it less costly to assign higher average grades in some courses. Similarly, a department with courses in the university core curriculum may perceive pressure to give out relatively higher grades.

For PhD departments, *Faculty size* has a U-shaped effect on *GPA*, with the marginal effect on increasing *Faculty size* is negative up to the 71st percentile (68 faculty members) and positive for only the largest 30% of departments-course observations. However, only three departments (English, Math and Teacher Education) have yearly *Faculty size* greater than 68, so the effect of faculty growth is negative for 14 of the 17 PhD departments represented. For the mean PhD department, the marginal effect of increasing *Faculty size* is extremely small, equal to −0.001 grade points. It is curious that the effect of adding faculty on GPA should be negative for most PhD programs, since one might expect that adding faculty would encourage the department to increase average grades so as to grow the overall size of the department, as is the case for non-PhD departments. On the other hand, the effect of adding faculty on G^* can be negative for departments that have a low payoff to teaching production. Noting that $r'_n < 0$, relationship (10) can be rewritten as the following summation:

$$(ar''_{nx} - cr''_{nx}\, P_r s'_n n'_g) + (bt''_{nx} - ct''_{nx}\, P_t s'_n n'_g) + |cn'_g s''_{nx}\, P_r r'_n| - cn'_g s''_{nx} P_t t'_n \qquad (13)$$

For PhD programs that are highly focused on research, $a > cP_r s'_n n'_g$, so the first summand is positive. However, since undergraduate teaching is given low value in PhD departments, it is likely that $b < cP_t s'_n n'_g$ and that the second summand is negative. The third and the fourth summand are positive and negative respectively. Thus, expression (13) is the sum of two positive and two negative summands, and the sign is more likely to be negative the smaller is b.

The relationship between GPA and the average quality of students taking departmental courses (as measured by *SAT_dep*) has a similar inverted-U shape to that of non-PhD programs. However, the effect of increasing the proportion of underclassmen on average grades is rather different for PhD programs. Instead of the U-shaped relationship observed for the non-PhD programs, average grades first rise with higher proportions of freshmen and sophomores, thereafter falling. This result suggests some sort of aspect of PhD-granting departments for which we have not otherwise accounted. For example, it might be the case that increasing the proportion of underclassmen allows such departments to fund additional graduate students. Furthermore, the additional underclassmen may largely be taught by graduate teaching fellows. If so, a department may discover that the benefits of inflating grades outweigh the costs. At higher proportions of underclassmen, however, research faculty may increasingly be needed to teach introductory-level classes. At this point the cost of grade inflation may outweigh the benefits, and the upward trend of average grades may reverse itself.

The results indicate that department-level measures are related to average grades in theoretically predictable ways. Over the sample period, the average department at UNT has seen a slight increase in faculty members, a substantial increase in the average quality of students and a modest decrease in the percentage of lower division students. Despite the statistical significance of the department-specific effects in our estimation, changes in these factors over time appear to account for less than 5% of UNT's overall grade inflation during the sample period. Thus, we are forced to conclude that department-level choices do not appear to be driving grade inflation at UNT. Instead, department-specific effects largely explain differences in average grades among departments.

Instructor-specific effects

Although we do not report the estimated instructor-specific fixed-effects, the estimation allows us to analyse the impact of instructor characteristics on grade inflation at UNT. Specifically, the results from Table 4 (Regression #2) allow us to estimate u_j for each instructor, and we can evaluate the pattern of u_j over time since the composition of instructors change each year. In 1984–85, the mean of u_j was −0.0828, and by 2004–05, the mean of u_j had increased to 0.0631. Thus, average GPA in our sample of UNT courses is estimated to have increased by 0.1459 grade points due solely to unobservable instructor characteristics. Although we cannot establish the specific

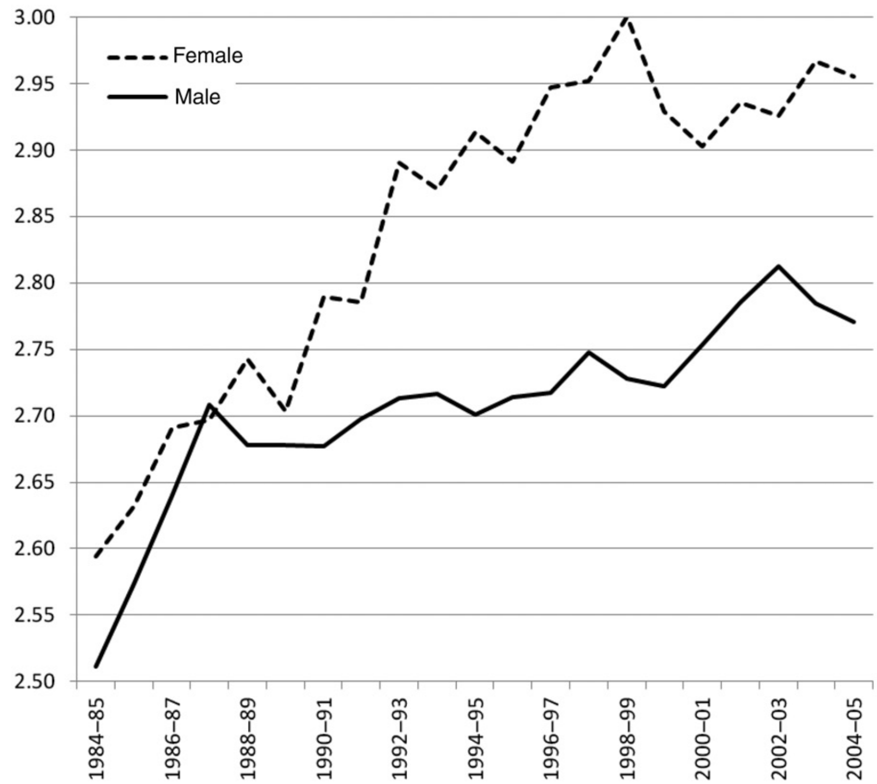

Fig. 2. Average GPA by instructor gender

characteristics that led to this instructor-level grade inflation, it appears that the instructor effects were responsible for approximately 40% of UNTs grade inflation over the sample period.

The literature has examined the growing importance of SET scores in promotion, tenure and merit raise evaluations, and how this might induce some faculty to inflate grades in an effort to 'buy' higher SET scores (e.g. Isely and Singh, 2005; McPherson and Jewell, 2007). Researchers have presented evidence that female instructors (e.g. Kolevzon, 1981) are more likely to inflate grades. In many institutions of higher learning, there have been increases in the number of female instructors over the past several decades. In the case of UNT, the female-male ratio for instructors has increased from 0.27 in 1984 to 0.94 in 2005. As shown in Fig. 2, average grades given by female instructors have increased more rapidly than those of male instructors over this same period. Although it is impossible to discern the nature of these instructor-specific effects, one could speculate that they are at least partially picking up the change in gender composition of instructors and the relationship between an instructor's gender and his or her incentive to inflate grades. Certain time-varying characteristics of instructors also may affect average grades. For example, after the first semester, new instructors typically adjust their grading to fit their institution's norms. This might mean that departments that are growing rapidly by frequently adding new faculty might have greater rates of grade inflation, *ceteris paribus*. Accounting for such effects directly is beyond the scope of this article, but we did duplicate our analysis excluding each instructor's first semester. These results (available from the authors on request) are not markedly different from those presented below, suggesting that any bias from this is likely to be very small.

Course-level variables

Results with respect to the course-level variables are remarkably consistent across the two regressions. In both non-PhD and PhD departments, courses that have more students tend to have lower GPAs, although the marginal effect (at the sample mean) is only −0.007 for PhD and −0.005 for non-PhD departments. The estimated marginal effect of a change in *Class size* is positive for the largest 2% of courses in the sample. As suggested earlier, this may reflect that the individual attention that the average student receives decreases as the number of students rises. That is, instruction may become less effective at larger class sizes. With respect to grade inflation, this result suggests that the increase in mean *Class size* in our sample from 29.8 in 1984–85 to 34.5 in 2004–05 actually caused average grades to decrease; thus, blame for UNT grade inflation cannot be laid at the feet of forces that have increased the number of students in the average course.

In both non-PhD and PhD departments, courses with better students have lower GPAs, while the effect of increased student SAT scores is positive in non-PhD departments for values of *SAT_class* above the 52nd percentile and in PhD departments above the 73rd percentile. At sample means, the marginal effect implies that a 1% increase in SAT decreases GPA by 0.005 for non-PhD departments and 0.014 for PhD departments. While this effect is small in magnitude, the direction of the effect is counterintuitive. In any event, this suggests that were students at UNT not increasing in quality, grade inflation would be an even more significant phenomenon than it is. This may also suggest that average SAT score is not an ideal measure of student quality at the course level.

While it is true that the estimated coefficients on $Underclassmen\_class^2$ are positive and statistically significant for both samples, practically all departments are on the downward-sloping portion of the function. Thus, increasing the percentage of underclassmen in any particular class has the effect of lowering average grades for both for PhD departments and non-PhD departments. As discussed earlier, this may be the result of a process by which students become more knowledgeable and experienced about what it takes to succeed in a college course, and perhaps generally more mature and attentive to their studies. The trend in *Underclassmen_class* is slightly negative over time, suggesting that changes in this factor have increased average GPA at UNT. However, the magnitude is small and is more than outweighed by the negative effect of changes in *Class size* and *SAT_class*.

Predicted grade inflation by department

As discussed above, department-level measures are significantly related to average GPA at any point in time. This result suggests that the time pattern of these factors influences the time pattern of average GPA, which further suggests that the rate at which average grades change over time will be influenced by department-level variations in these factors. Table 5 presents predicted GPA for each department at five points in the time period: 1984–85, 1989–90, 1994–95, 1999–00 and 2004–05. The estimated predicted values include the estimated effect of department and course characteristics, the overall UNT time trend, the department-specific fixed effects and the instructor-specific fixed effects for instructors teaching courses in a given department during each specific time period. The table also reports the overall rate of grade inflation by department over the entire time period.

Table 5 illustrates that predicted grade inflation differs rather markedly between departments. As a general rule, at any point in time departments offering doctorates assign lower grades than those that do not. Regarding grade inflation, for the most part the results for UNT departments presented in Table 5 are consistent with those reported in previous research. Specifically, as Prather *et al.* (1979), Sabot and Wakeman-Linn (1991) and Cheong (2000) found, disciplines with a more quantitative focus (such as computer science, physics, mathematics and engineering) tend to inflate grades at a lower rate. On the other hand, departments with little or no quantitative focus, such as English, history and journalism, have inflated grades to a much greater extent.

Table 5 also illustrates differences between PhD and non-PhD departments at UNT in terms of how grade inflation has varied over time. Interestingly, most of the predicted grade inflation for PhD departments over the entire sample period occurred in the years 1984–85 to 1989–90 (8.8% for the average PhD department). Predicted grade inflation in the same 5-year period for the average non-PhD department was only 1.3%; in the remaining 5-year intervals, non-PhD departments showed a fairly stable average grade inflation rate of about 3%. Average predicted grade inflation since 1990 for PhD programs dropped to 2.3%, 0.9% and 1.2% in the remaining 5-year periods. Thus, although PhD programs show higher grade inflation over the sample period, much of the difference is due to grade inflation in the mid-1980s, and non-PhD departments have actually inflated grades at a higher rate than PhD departments since 1990. While the causes of this cannot be conclusively determined, it is interesting to note that the 1980s saw UNT's first effort to move away from its roots as a primarily teaching institution towards a more research-oriented focus. The university even changed its name from North Texas State University to the University of North Texas in 1988. It is possible that PhD-granting departments entered a phase during which there was a push to add students, and grade inflation may have resulted from this pressure.

VI. Conclusion

Our model leads us to expect differential grade inflation by academic department, and our findings

Table 5. Predicted grade inflation (listed from highest to lowest predicted inflation)

	1984–85	1989–90	1994–95	1999–00	2004–05	Inflation
All non-PhD departments	*2.721*	*2.757*	*2.845*	*2.920*	*3.005*	*10.45%*
Geography	2.350	2.826	2.701	2.707	2.746	16.87%
Journalism	2.716	2.817	2.943	2.968	3.167	16.61%
Languages	2.462	2.371	2.526	2.707	2.842	15.42%
Recreation and Health	3.034	3.315	3.147	3.340	3.409	12.35%
Dance and Theater	3.123	3.182	3.246	3.334	3.425	9.67%
Economics	2.344	2.303	2.464	2.453	2.555	8.99%
Radio, TV and Film	3.084	3.109	3.094	3.044	3.317	7.55%
Anthropology	2.850	2.799	2.834	2.895	2.924	2.62%
Kinesiology	2.840	2.850	2.996	2.901	2.870	1.04%
Speech/Hearing Sciences	3.223	3.194	3.368	3.350	3.224	0.03%
Computer Science	2.820	2.702	2.706	2.880	2.798	−0.78%
All PhD departments	*2.449*	*2.665*	*2.728*	*2.752*	*2.786*	*13.78%*
English	2.205	2.675	2.745	2.896	2.984	35.33%
Business Computer Information Systems	2.581	2.636	2.795	2.805	2.947	14.15%
History	2.320	2.484	2.594	2.626	2.644	13.99%
Biology	2.555	2.759	2.643	2.828	2.893	13.25%
Marketing	2.704	2.723	2.890	2.911	3.018	11.59%
Teacher Education	3.169	3.371	3.602	3.619	3.514	10.86%
Accounting	2.151	2.129	2.279	2.297	2.379	10.61%
Chemistry	2.319	2.435	2.587	2.672	2.560	10.40%
Political Science	2.416	2.522	2.749	2.644	2.633	8.98%
Management	2.625	2.719	2.798	2.804	2.842	8.24%
Sociology	2.711	2.813	2.887	3.084	2.906	7.18%
Mathematics	2.074	2.196	2.225	2.136	2.193	5.72%
Finance and Real Estate	2.403	2.548	2.452	2.568	2.541	5.71%
Physics	2.482	2.722	2.765	2.669	2.522	1.62%
Philosophy	2.660	2.827	2.853	2.889	2.677	0.66%
Engineering	2.836	3.060	2.957	2.801	2.805	−1.09%
Psychology	2.574	2.845	2.824	2.553	2.388	−7.23%

support this expectation. Further, we find substantial differences in grade inflation trends by department and for PhD and non-PhD departments. The average PhD department assigns lower grades than the average non-PhD department, perhaps because doctoral programs realize lower marginal returns from adding students than other departments. However, the average rate of grade inflation for a PhD department is greater than for non-PhD departments, but this implication is driven by a high rate of grade inflation for PhD programs early in the sample period. In addition, we find a positive relationship between the size of a department in terms of the faculty and average grades for non-PhD programs, whereas for the doctoral-granting departments growth in the faculty size lowers grades for all but three departments. This may indicate that departments that have more of a teaching focus may find that the marginal benefit of adding students by inflating grades outweighs the marginal cost. When more research-oriented departments grow, they are more likely to add faculty members who would rather have fewer students in order to spend more time doing research.

Unsurprisingly, increases in the average quality of a department's students tend to increase average grades. However, at high levels of quality there is some evidence that further increases in quality lower average grades – this may suggest that departments eventually adjust to better students by teaching more challenging material. The results regarding increasing the proportion of underclassmen are also interesting. For non-PhD programs, increasing the proportion initially lowers average grades; this is unsurprising if student performance tends to increase with experience and maturity. Eventually, departments seem to find that raising grades is in their best interests, perhaps because at some point the department has become a service department. The pattern is reversed for the PhD departments – for these, increasing the proportion of freshmen and sophomores initially increases average grades, but later the relationship turns negative. This finding may have something to do with the ability of PhD departments to hire graduate students to teach these underclassmen.

While departmental characteristics are significant determinants of grade inflation, these factors are relatively small in magnitude. Of the variation in grades that our regressions explain, less than 5% results from departmental differences. Our estimates indicate that the main determinants of grade inflation at UNT are the time trend (explaining 52%) and differences specific to individual instructors (40%). The time trend is picking up factors that lead to general grade inflation at the university level. This sort of inflation may result from national or regional trends in competition for students and public funding formulas or other policies that may encourage universities to add students. As such, a university such as UNT may find it counterproductive to actively discourage grade inflation. Our results also suggest that individual instructors find it rational to inflate grades for reasons specific to themselves. This may in part reflect the now nearly universal use of student evaluation scores as inputs into tenure, promotion and merit raise decisions. A university wishing to reduce grade inflation may need to base evaluations of teaching on a broader array of metrics.

References

Aigner, D. J. and Thum, F. D. (1986) On student evaluation of teaching ability, *Journal of Economic Education*, **17**, 243–65.

Anglin, P. M. and Meng, R. (2000) Evidence on grades and grade inflation at Ontario's universities, *Canadian Public Policy*, **26**, 361–8.

Bello, A. L. and Valientes, D. M. (2006) Grade inflation: fact or myth?, Draft Paper, Department of Economics, University of the Philippines, Los Baños.

Birnbaum, R. (1977) Factors related to university grade inflation, *Journal of Higher Education*, **48**, 519–39.

Carney, P., Isakson, R. L. and Ellsworth, R. (1978) An exploration of grade inflation and some related factors in higher education, *College and University*, **54**, 217–30.

Chan, W., Hao, L. and Suen, W. (2007) A signaling theory of grade inflation, *International Economic Review*, **48**, 1065–90.

Cheong, K. S. (2000) Grade inflation at the University of Hawaii-Manoa, Department of Economics Working Paper No. 00-2.

Cluskey, G. R. Jr, Griffin, N. and Ehlen, C. (1997) Accounting grade inflation, *Journal of Education for Business*, **72**, 273–7.

Dickson, V. A. (1984) An economic model of faculty grading practices, *Journal of Economic Education*, **15**, 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*, The Center for Faculty Development, Mid-Career Fellowship Program, History Department.

Felton, J. and Koper, P. T. (2005) Nominal GPA and real GPA: a simple adjustment that compensates for grade inflation, *Assessment and Evaluation in Higher Education*, **30**, 561–9.

Germain, M. and Scandura, T. A. (2005) Grade inflation and student individual differences as systematic bias in faculty evaluations, *Journal of Instructional Psychology*, **32**, 58–67.

Grove, W. A. and Wasserman, T. (2004) The life-cycle pattern of collegiate GPA: longitudinal cohort analysis and grade inflation, *Journal of Economic Education*, **35**, 162–74.

Hamermesh, D. A. and Parker, A. M. (2005) Beauty in the classroom: professors' pulchritude and putative pedagogical productivity, *Economics of Education Review*, **24**, 369–76.

Im, K.-S., Lee, J. and Tieslau, M. (ILT) (2005) Panel LM unit-root tests with level shifts, *Oxford Bulletin of Economics and Statistics*, **67**, 393–419.

Im, K.-S., Pesaran, H. and Shin, Y. (IPS) (2003) Testing for unit roots in heterogeneous panels, *Journal of Econometrics*, **115**, 53–74.

Isely, P. and Singh, H. (2005) Do higher grades lead to favorable student evaluations? Another look at an old controversy, *Journal of Economic Education*, **36**, 29–42.

Kelley, A. C. (1972) Uses and abuses of course evaluations as measures of educational output, *Journal of Economic Education*, **3**, 13–18.

Kolevzon, M. S. (1981) Grade inflation in higher education: a comparative study, *Research in Higher Education*, **15**, 195–212.

Lichty, R. W., Vose, D. A. and Peterson, J. M. (1978) The economic effects of grade inflation on instructor evaluation: an alternative interpretation, *Journal of Economic Education*, **10**, 3–11.

McKenzie, R. B. (1975) The economic effects of grade inflation on instructor evaluations: a theoretical approach, *Journal of Economic Education*, **6**, 99–106.

McPherson, M. A. (2006) Determinants of how students evaluate teachers, *Journal of Economic Education*, **37**, 3–20.

McPherson, M. A. and Jewell, R. T. (2007) Leveling the playing field: should student evaluation scores be adjusted?, *Social Science Quarterly*, **88**, 868–81.

McPherson, M. A., Jewell, R. T. and Kim, M. (2009) What determines student evaluation scores? A random effects analysis of undergraduate economics classes, *Eastern Economic Journal*, **35**, 37–51.

Mirus, R. (1973) Some implications of student evaluations of teachers, *Journal of Economic Education*, **5**, 35–7.

Mulvenon, S. and Ferritor, D. (2005) Grade inflation in higher education: isolated or systematic?, *International Journal of Learning*, **12**, 55–62.

Nagle, B. (1998) A proposal for dealing with grade inflation: the relative performance index, *Journal of Education for Business*, **74**, 40–3.

Nelson, J. P. and Lynch, K. A. (1984) Grade inflation, real income, simultaneity, and teaching evaluations, *Journal of Economic Education*, **15**, 21–37.

Nichols, A. and Soper, J. C. (1972) Economic man in the classroom, *Journal of Political Economy*, **80**, 1069–73.

Perron, P. (1989) The great crash, the oil price shock, and the unit root hypothesis, *Econometrica*, **57**, 1361–401.

Prather, J. E., Smith, G. and Kodras, J. E. (1979) A longitudinal study of grades in 144 undergraduate courses, *Research in Higher Education*, **10**, 11–24.

Rojstaczer, S. (2008) Grade inflation at American colleges and universities, June 24. Available at www.gradeinflation.com (accessed 19 September 2011).

Sabot, R. and Wakeman-Linn, J. (1991) Grade inflation and course choice, *Journal of Economic Perspectives*, **5**, 159–70.

Schmidt, P. and Phillips, P. (1992) LM tests for a unit root in the presence of deterministic trends, *Oxford Bulletin of Economics and Statistics*, **54**, 257–87.

Sonner, B. (2000) A is for 'Adjunct': examining grade inflation in higher education, *Journal of Education for Business*, **76**, 5–8.

Voeks, V. W. and French, G. M. (1960) Are student-ratings of teachers affected by grades? The report of three studies at the University of Washington, *Journal of Higher Education*, **31**, 330–4.

Zangenehzadeh, H. (1988) Grade inflation: a way out, *Journal of Economic Education*, **19**, 217–26.

Appendix: Model Derivation

a. Start with expression (7).

$$\text{Max } D = d(r[n(G),X], t[n(G),X], s[n(G),X]),$$
$$\text{subject to } M = P_r \times r[n(G),X] + P_t \times t[n(G),X] + s[n(G),X] \tag{7}$$

b. Substitute in for $s[n(G),X]$ from constraint assuming implicit function theorem holds.

$$\text{Max } D^0 = d(r[n(G),X], t[n(G),X], s[n(M - P_r \times r[n(G),X] - P_t \times t[n(G),X]]), X]) \tag{7.1}$$

c. FOC for maximization.

$$\mathrm{d}D^0/\mathrm{d}G = d'_r r'_n n'_g + d'_t t'_n n'_g - d'_s s'_n P_r r'_n (n'_g)^2 - d'_s s'_n P_t t'_n (n'_g)^2 = 0 \tag{7.2}$$

d. Plug in constants for marginal products.

$$\mathrm{d}D^0/\mathrm{d}G = ar'_n n'_g + bt'_n n'_g - cs'_n P_r r'_n (n'_g)^2 - cs'_n P_t t'_n (n'_g)^2 = 0 \tag{7.3}$$

e. Rearrange terms, gives expression (8) in text.

$$\mathrm{d}D^0/\mathrm{d}G = r'_n n'_g [a - cP_r s'_n n'_g] + t'_n n'_g [b - cP_t s'_n n'_g] = 0 \tag{8}$$

f. SOC for maximization.

$$\begin{aligned}\mathrm{d}^2 D^0/\mathrm{d}G^2 = {} & ar''_{nn}(n'_g)^2 + ar'_n n''_{gg} + bt''_{nn}(n'_g)^2 + bt'_n n''_{gg} \\ & - cP_r s''_{nn} r'_n (n'_g)^3 - cP_r s'_n r''_{nn}(n'_g)^3 \\ & - 2cP_r r'_n n'_g n''_{gg} - cP_t s''_{nn} t'_n (n'_g)^3 \\ & - cP_t s'_n t''_{nn}(n'_g)^3 - 2cP_t t'_n n'_g n''_{gg} \\ = {} & D''_{gg} < 0\end{aligned} \tag{8.1}$$

g. Take partial derivative of FOC (8) evaluated at optimum G (G^*).

$$\begin{aligned}\partial \text{FOC}/\partial X = {} & D''_{gg}(\partial G^*/\partial X) + an'_g(\partial r'_n/\partial X) \\ & + bn'_g(\partial t'_n/\partial X) - cP_r r'_n (n'_g)^2 (\partial s'_n/\partial X) \\ & - cP_r s'_n (n'_g)^2 (\partial r'_n/\partial X) - cP_t t'_n (n'_g)^2 (\partial s'_n/\partial X) \\ & - cP_t s'_n (n'_g)^2 (\partial t'_n/\partial X) = 0\end{aligned} \tag{8.2}$$

h. Rearrange terms, gives expression (9) in text.

$$\partial G^*/\partial X = -n'_g (r''_{nx}[a - cP_r s'_n n'_g] + t''_{nx}[b - cP_t s'_n n'_g] - cn'_g s''_{nx}[P_r r'_n + P_t t'_n])/D''_{gg} \tag{9}$$