A COMPARATIVE ANALYSIS OF GRADING PRACTICES BY DISCIPLINE WITHIN A COLLEGE OF BUSINESS

Michael M. Barth, The Citadel
Jun Liu, Georgia Southern University
William H. Wells, Georgia Southern University

ABSTRACT

This research tests for differences in average grades awarded across six business disciplines in the business college at a southern regional university. Although complaints about college grade inflation have existed for over one hundred years, there has been increasing interest in recent years in grade inflation within the U.S. higher education system. Most published research on grade inflation has focused on systematic inflation at the institutional level, but relatively little has been published relative to cross-disciplinary grade inflation and deflation within a particular institution.

Differences in grading standards between disciplines may reflect better teaching, but they may also reflect lax standards. Lenient grading practices can have unintended negative consequences and are a cause for concern. On the other hand, relatively higher grades for a particular discipline that reflect better teaching should be encouraged and emulated. It is important to identify and evaluate differences in grading practices because unwarranted inconsistencies in grading practices have potential adverse effects upon faculty, students and the institution as a whole.

Using regression analysis and controlling for a number of potential causal factors such as student GPA, withdrawal rates and instructor experience, we compare average grades for nearly 400 classes in six business subdisciplines. We find that the average grades given out in three of the six subdisciplines within this college of business were systematically higher than in the other three, even after controlling for these other explanatory factors. While not offering any conclusions as to why these particular differences exist, there is a general discussion of factors that can explain higher (or lower) average grades from one discipline to the next. Although this research does not attempt to reach conclusions as to the reason for inconsistencies between disciplines at this university, it does provide a framework that other institutions can use to begin to at least identify systematic differences between grading practices within their own academic programs.

INTRODUCTION

The purpose of this study is to determine if systematic grade inflation or compression exists across academic disciplines at this university's college of business administration. Grade inflation is typically defined as a systematic increase in grades and grade point averages (GPA) without a concomitant increase in performance. Grade compression, of course, would be the opposite effect of reduced grades for the same level of academic performance. Expressions of concern about collegiate grade inflation have been around for over 100 years (Gordon, 2006), but there has been a recent surge of interest in this subject. Much of this recent discussion has centered on long-term upward trends in average GPAs across university systems, but there has been relatively little discussion about cross-discipline inflation within the same university.

Cross-disciplinary grade inflation is an issue that affects all areas of a university and requires a collective response if it is to addressed effectively (e.g., Briggs, 2007; Dresner, 2004; Gordon, 2006; Smith & Coombe, 2006). Allocation of scholarship dollars is based to some degree on student performance, and unwarranted inflation in one discipline misallocates resources from one group of students to another. Deans' list and similar academic honors rolls may be distorted by disparate grading practices as well. Grading practices may affect student recruitment or retention, as students may forego a more rigorously graded discipline in favor of a less rigorously graded discipline without realizing the potential long-term costs. For example, if potential employers are not satisfied with the academic results of the less rigorous discipline or program, that can have long-term implications for students in that discipline or program. Students rely on faculty to not only know that those types of issues can arise, but to also act in the best interest of the students by maintaining quality standards that maximize the value of education. However, faculty members often perceive that there are disincentives towards rigorous academic programs. For example, there often is a perception that rigorous academic standards lower student evaluations of teaching (SET), which in turn has a direct impact on promotion, tenure and merit pay. Although the research linking grading leniency with SET scores is mixed (Gump, 2007), it nonetheless continues to be an article of faith with some faculty members that higher grades lead to higher SET scores. Whether or not this link exists, the perception of a link between grade inflation and SET scores provides an incentive towards grade inflation.

The process of examining grading practices by academic unit is a journey, not a destination. Our research uses regression analysis to test for differences in average grades across disciplines while controlling for a variety of faculty effects, student effects, and discipline-specific effects that might explain systematic increases or decreases in the average class GPA. While this research attempts to explain some of the observed differences in average GPAs, it cannot explain all observed differences. It can identify certain predictors of student achievement, but it cannot identify all predictors of student achievement, nor is it meant to identify or establish grading norms. It was intended to reduce some of the murkiness surrounding observed differences in grading practices as well as to spur intracollegial debate about standards and practices among faculty across the disciplines.

The scope of this research is limited to testing for discipline-specific inflation or compression, although eventually it may be possible to generalize the model into a more generic model that could provide benchmarks for normative grading practices. It is important to stress that this model cannot be used to measure teaching effectiveness. There is no measure of student learning, only student grading, included in the model that we present. We cannot measure the level of learning by comparing grades, only the level of instructor assessment of learning. However, the presence of systematic differences between disciplines leads logically to that next stage, determining why those differences exist. And while it may be easier to say that higher grades in one discipline relative to another are simply the effect of grade inflation, it is equally plausible that the differences represent more effective teaching.

Our results are unique to this institution. However, the contribution of this research is that the process that we present can be used by other institutions that are concerned about the potential for systematic grade inflation or compression in their academic units. Following a brief review of the extant literature, a methods section describes the process we undertook, followed by a results and a conclusions section.

REVIEW OF LITERATURE

Many researchers suggest that higher nominal grades are the result of factors other than simple grade inflation. Lanning and Perkins (1995) argue that the "Mastery of Learning" model in Colleges of Education results in each student receiving an A for any assignment. This model allows students to repeat course work until a concept or assignment is mastered, and the desired or required grade (an A) is achieved. Under this model, final grades are differentiated through non-academic measures, such as attendance and participation. Under that model, then, inflated grades can actually be a sign of higher standards, as students are not allowed to simply get by with lower quality work. Therefore, the presence of higher grades does not necessarily imply lower standards. All that it indicates is that there is, indeed, a difference.

Kohn (2002) and Boretz (2004) suggest institutional factors that may contribute to higher grades. Kohn (2002) notes that instructors may be less stingy with grades than in the past, leading to systematic inflation across the educational system. Also, in some institutional models, students may be allowed to take relatively more courses in their major areas and thus avoid courses in other areas that might be more difficult or which the student has less interest and motivation to study. Boretz (2004) concludes that the trend in allowing more retests and revisions of assignments leads to higher grades, and this is again a policy that differs from institution to institution. Further, Kohn (2002) and Boretz (2004) suggest that extended course withdrawal dates make it easier for students to drop classes in which they are struggling, which would generate a relatively higher number of W grades which would not appear in the average GPA for a class.

Changes in student demographics may produce higher performing students. Non-traditional students make up an ever increasing proportion of today's classroom. Kwon and Kendig (1997) indicate that an older, more mature student population may actually perform better in the classroom than younger, traditional students. Another important demographic is socio-economic. As more students rely on scholarships to fund their educations, they must work harder to maintain eligibility. Potter and Nyman (2001) recognize that state aid packages push students to earn higher grades. Under this type of incentive system, higher grades are more meaningful to students because they have a direct economic impact today.

Empirical evidence supporting a trend in higher grades is offered by Hanson, Quinn and Wells (2002) and Kezim, Pariseau and Quinn (2005). Additionally, both studies find higher average grades for courses taught by adjunct faculty than for courses taught by traditional tenure and tenure track faculty. This inflated grading by adjunct and temporary instructors is often interpreted as an attempt to obtain higher student evaluations, but there are other reasons that may contribute. Adjunct faculty may hold at most a masters degree in the discipline and may or may not have attended a doctoral-granting institution. These faculty members may not have been held to the higher standards that doctorally-qualified faculty have been held to at those institutions that grant doctoral degrees. Adjunct faculty may also have less of an identity with institutional standards than tenured or tenure track faculty who have created longer term ties between themselves and their institutions.

Interestingly, Marsh and Roach (2000) provide evidence that higher grades, in and of themselves, do not guarantee higher teacher evaluations. Germain and Scandura (2005) also explore the relationship between grades and faculty evaluation. A common conclusion is that evaluations are positively related to grades, but there may be quality factors involved. If better teachers teach better, then the students of better teachers should exhibit higher average grades (assuming equal standards) than the students of marginal teachers. Germain and Scandura (2005) extend the maturity factor discussed in Kwon and Kendig (1997) to include previous life experiences, personal development and previous relationships with instructors with success in the classroom. The same individual characteristics that lead to better classroom performance may also create a more favorable view of the instructor. Therefore, their reasoning is that higher evaluations are not created by higher grades, but share a common source.

Felton and Koper (2004) and Nagle (1998) provide alternatives to traditional grades that adjust for the level of difficulty in individual courses. Their solutions compute ratios of individual course grades to the class GPA, thereby providing an index of relative grades rather than an absolute measure. Their index models provide a means by which course difficulty and instructor grading scales are factored into individual measures of success. Simply put, earning a B in a course where the average grade is a C means more than earning a grade of B where the average grade is a B.

These studies illustrate some of the pitfalls inherent in determining the causal roots of grade inflation and compression. There are often solid reasons behind differences in average grade levels between instructors, between disciplines within the same college, between colleges within the same university, and between different universities. While grades for one group may be relatively higher because of higher quality instructors or higher caliber students, they also may be higher because of looser standards or grade inflation.

The objective of our study is to look at differences within our own college to identify where there are systematic differences in grading practices between disciplines, operating under the assumption that the cultural factors, the socioeconomic factors, and the institutional-specific factors affect each of our six disciplines in roughly the same manner. We do acknowledge that there are significant differences in the experience level, training, and other demographic factors between the faculty members in the different disciplines, and we attempt to control for those as best we can.

The goal of our study is to ascertain whether we are all using the same basic standards, and if not, to identify the sources of differences. In the following section, we describe the steps we took to isolate those intracollege differences. If the difference is attributable to better teaching practices, that difference should be emulated and celebrated. If the difference is unexplainable, then that difference can at least be identified. Grading practices continue to be the prerogative of the individual instructor, but the success of students is first and foremost a team effort, and an effective team must be working together toward a common goal.

METHOD

We use ordinary least squares (OLS) regression to test for faculty effects, student effects, and discipline-specific effects that might explain systematic differences, either increases or decreases, in average class GPAs for a sample of nearly 400 upper division classes within the six major disciplines of the College of Business Administration at this institution.

Our hypothesis is that the average GPA for a given course can be explained by instructor demographics such as tenure/tenure track status and experience; class-specific information such as the caliber of the students, as measured by past student performance, by class size and capacity; and by trends over time that reflect changing student demographics. Although we use a basic OLS model, we also include several spline functions to address a number of non-linear relationships that we observe between the class average GPA and our explanatory variables.

Our dependent variable is the class average GPA (CLASSGPA), using the standard scale with A=4, B=3, etc. Class GPAs were calculated for 397 courses over the span of eight semesters. The minimum class size to be included in the study was 25 students. Table 1 provides some basic descriptive statistics on the dependent variable as well as the explanatory variables described in the following subsections.

Table 1: Summary Statistics By Discipline								
		COBA	ACCT	ECON	FINC	LOGT	MGNT	MKTG
VARIABLE	N	397	71	12	77	31	80	126
	Mean	2.72	2.52	2.39	2.36	2.61	2.86	3.02
CLASSGPA	Minimum	1.63	1.68	1.79	1.63	1.93	2.30	1.94
	Maximum	3.76	3.30	3.07	3.48	3.36	3.64	3.76
	Mean	2	2.3	2.2	1.8	2.6	2	1.9
EXPER	Minimum	1	1	2	1	2	1	1
	Maximum	5	4	3	3	5	4	4
	Mean	9.1	6.2	2.5	9.0	6.2	11.5	10.7
TIMES	Minimum	1	1	1	1	1	1	1
	Maximum	31	24	5	26	13	31	30
	Mean	22%	17%	0%	4%	6%	63%	17%
NONTENURE	Minimum	0	0	0	0	0	0	0
	Maximum	1	1	0	1	1	1	1
	Mean	2.69	2.84	2.70	2.68	2.58	2.60	2.68
CUMGPA	Minimum	2.30	2.37	2.42	2.30	2.38	2.31	2.36
	Maximum	3.25	3.25	2.81	2.89	2.77	2.82	2.91
	Mean	113	107	107	115	112	114	114
TOTALHRS	Minimum	81	81	95	88	100	100	100
	Maximum	130	130	121	128	127	127	130
	Mean	41	38	36	38	40	43	46
CLASSIZE	Minimum	25	25	35	35	30	28	27
	Maximum	80	52	42	50	60	60	80
	Mean	94%	93%	93%	98%	94%	94%	94%
FULL%	Minimum	47%	54%	77%	65%	65%	62%	47%
	Maximum	134%	108%	114%	134%	123%	113%	117%
	Mean	6%	9%	8%	8%	4%	3%	5%
W%	Minimum	0%	0%	0%	0%	0%	0%	0%
	Maximum	35%	35%	16%	29%	13%	15%	19%
	Mean	17%	18%	15%	17%	18%	18%	16%
LATEREG	Minimum	0%	0%	4%	2%	3%	3%	3%
	Maximum	100%	100%	24%	48%	68%	50%	44%

Explanatory Variables for Instructor Characteristics

We include three variables to pick up instructor-specific differences: EXPER to measure the instructor's time at this institution; TIMES to measure the amount of experience in teaching a particular course; and NONTENURE to identify employment status.

Instructor experience (EXPER) is the number of semesters that the class instructor has been at this institution, starting with the Fall 1998 semester when the institution converted from the quarter system to the semester system. While this does not exactly measure the instructor's true length of time at the institution, it does approximate the amount of experience under the semester system. Some research has identified a tendency for average grades to decline over time as an instructor gains experience (e.g., Kezim et. al., 2005). We include a second variable, TIMES, to identify the number of times that each instructor has taught a particular course in the past. While EXPER picks up the overall length of time that the instructor has been at the institution since semester conversion, TIMES is actually measuring the instructors experience with a specific course. Whether the instructor is a newly minted PhD or a grizzled veteran, there may be less certainty about the appropriate rigor or testing standards when engaged in a new preparation. Our final instructor-specific variable is NONTENURE is an indicator variable that picks up differences between tenured or tenure track faculty (our base) and temporary faculty and/or adjuncts, which are coded as 1. As cited earlier, a number of past studies have shown that grades are typically higher for non-permanent faculty.

Explanatory Variables for Class Characteristics

CUMGPA is the average cumulative grade point average for all of the students enrolled in the course as of the beginning of that semester. We hypothesize that past performance is the best predictor of future performance. Since the classes we include in our study are junior- and senior-level, the individual students have had time to amass a statistically valid cumulative GPA prior to the class in this study. If a particular class is relatively full of high-achieving students, then one would expect that the average grade given in that class would be relatively higher as well. That is, "A" students tend to be "A" students in both their preparatory classes as well as their major classes.

We also include a variable to try to capture those classes that are more likely to have graduating seniors in the student ranks. The variable TOTALHRS measures the average cumulative hours for all the students in a class. We hypothesize a possible link between the average hours and the average course grade. Graduating seniors should have a relatively higher level of motivation to pass the last set of courses and matriculate to post-academic life. There may also be a tendency for the instructor to "go easy" on graduating seniors. Finally, there is some element of survivor bias. Graduating seniors have been through the system and weak performers have been winnowed out along the way.

We expect to find a link between the size of a class (CLASSIZE) and the average grade, but the direction of the relationship between class size and CLASSGPA is uncertain. Smaller classes, with more individualized instruction, may generate higher average grades, creating a negative relationship. However, it is also possible that students tend to flock to the "easy" courses, which would suggest a positive relationship between class size and average grade. A somewhat related variable is meant to measure the demand for a class. The variable FULL% is the number of students enrolled in the class divided by the number of seats available for that class.

The percentage of students that voluntarily withdraw from the class (W%) is expected to affect CLASSGPA, but again the direction of the effect is uncertain. On the one hand, if there is a high withdrawal rate, that presumably would eliminate many of the D and F grades and inflate CLASSGPA. On the other hand, a high withdrawal rate may be indicative of a rigorous course or perhaps a poor instructor, which would suggest a negative correlation with CLASSGPA. Both effects may be present, in which case we might expect a nonlinear relationship between grades and withdrawal rates.

Student attitudes and efforts are hard to measure from the available data. One variable that may be an indicator of student attitude is the time at which they registered for the class. The variable LATEREGIST is the percent of students that late-registered for the course after the initial registration period. The university database includes a variable that shows the date that a student registered for a particular class, although it only shows the last date that a change occurred. If a student signed up early for a particular Fall semester class in February during open registration and then switched to a different section of the same course during open enrollment the following August, that would show up as a late registration. A student that failed a pre-requisite course or that decided at the last minute to take a particular class would show the same late enrollment date. While the majority of students sign up for classes during the normal registration period and do not change their registration to a different time slot, there does exist some amount of section swapping, and there is more swapping in some classes and courses than in others. We expect that a high number of late registrants is an indicator of relatively lower commitment by students, on average, and therefore a relatively lower CLASSGPA.

We include a simple time variable (TREND) to isolate long-term trends in average course grades. An important feature of the student body at this university is the systematic increase in standards of admission over the period of study. Admission standards have increased, and SAT scores have risen over the past five years. Although we recognize that studies have shown that the SAT score is a relatively inaccurate reflection of the ability of students, the change in the average SAT score does suggest that the demographics of the student body have changed, presumably for the better, but changed nonetheless. If that is so, then average class grades may be increasing (or decreasing) over the period of this study.

Explanatory Variables for Each Discipline

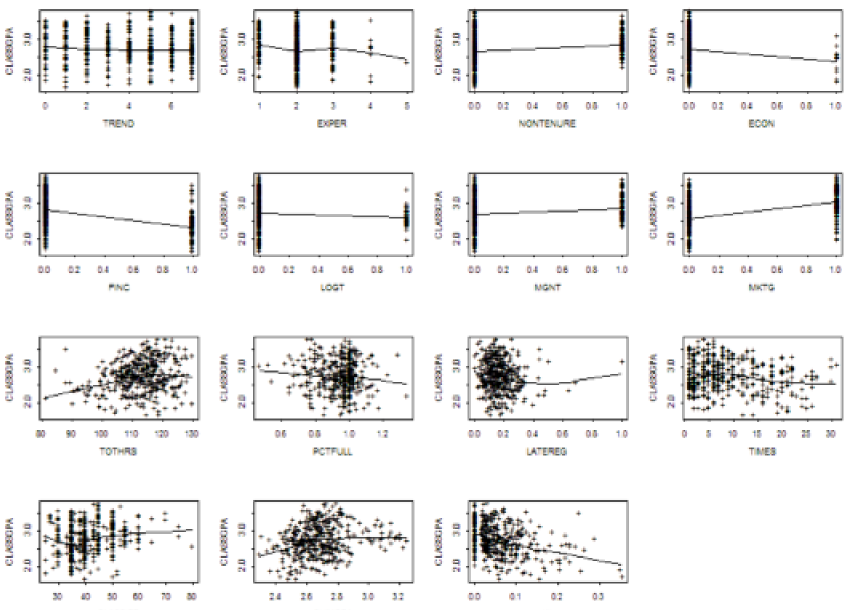
Our final set of explanatory variables to identify the specific academic disciplines. The six academic disciplines included in this study are Accounting (ACCT), Economics (ECON), Finance (FINC), Logistics (LOGT), Management (MGNT) and Marketing (MKTG), the last five of which are coded using dummy variables. As shown in Table 1, the number of classes differed significantly from one discipline to the next. The Accounting discipline was chosen as the base discipline partly because of its order in the alphabet, but primarily because of the nature of the coursework. Several of the other five disciplines have one or more sub-disciplines (e.g., Fashion Merchandising within the Marketing discipline), and two of the disciplines have relatively few observations. The Accounting program at this university has the advantage of being relatively large and relatively homogenous, which makes it a more desirable base.

Regression Model

Based on the means, minima and maxima for the explanatory variables included in Table 1, certain relationships immediately pop out. There are obviously a greater number of course offerings in Marketing. The

CLASSGPA values for the Logistics, Management and Marketing courses are noticeably higher than for the course offerings in the Accounting, Economics and Finance areas. In addition to differences in the average grade between the disciplines, there are differences in withdrawal rates, class sizes, use of non-tenure track faculty, and the average number of times that faculty members have taught a course. Table 1 also shows that the cumulative GPAs of the students enrolled in the Accounting program are on average higher than in the other programs.

Figure 1: Scatter Plot of Explanatory Variables



The relationships between the explanatory variables and CLASSGPA are plotted in a series of XY graphs in Figure 1, with trend lines superimposed. After reviewing the graphical relationships, we noted that some of the relationships appear to be nonlinear. Further complicating the analysis were multicollinearity issues, as many of the independent variables were linearly dependent to some degree. This analysis required us to go beyond simple OLS to measure those nonlinear relationships that appear in the data. Our final model uses linear regression spline functions to model the potentially nonlinear effects in CUMGPA, TOTALHRS, W%, TIMES and CLASSIZE suggested in Figure 1. Eubank (1988) provides a detailed explanation and discussion of regression spline functions, but the basic premise is to split the nonlinear variable into pieces and to then estimate the regression line for that particular variable in segments.

The final regression model is as follows:

$$\begin{aligned} CLASSGPA = {} & b_0 + b_1 EXPER + b_2 TIMES + b_{2.1}(TIMES - \kappa_2) + b_3 NONTENURE + \\ & b_4 CUMGPA + b_{4.1}(CUMGPA - \kappa_{4.1}) + b_{4.2}(CUMGPA - \kappa 4.2) + \\ & b_5 TOTALHRS + b_{5.1}(TOTALHRS - \kappa_5) + b_6 CLASSIZE + \\ & b_{6.1}(CLASSIZE - \kappa_6) + b_7 FULL\% + B_8 W\% + \\ & B_{8.1}(W\% - \kappa_8) + b_9 LATEREGIST + b_{10} TREND + b_{11} ECON + \\ & b_{12} FINC + b_{13} LOGT + b_{14} MGNT + b_{15} MKTG + \epsilon \end{aligned}$$

where $(x)_{++} = x$ **if** $x > 0$ **and** $(x) = 0$ **if** $x \; 0 \leq$. For example, the value for κ_2 is 26.89, so (TIMES – κ_2) would be zero if the value of the TIMES variable was 25 and would be 3.11 if the TIMES variable was 30 (30 – 26.89 = 3.11). The result is that a new slope is created for values above the κ values. These κ's are known as the *knots*. The linear spline function we used in this model is a special case of regression splines of degree *p* ($p \geq 1$), which is defined as a piecewise polynomial of degree *p* that is smoothly connected at its knots. In this study we used linear spline functions ($p = 1$) mainly because of the relatively simple functional forms suggested in Figure 1 and the interpretability of linear spline functions. In this study the knots are treated as parameters and estimated jointly with the regression coefficients. The estimated knots for these five variables are given in Table 2 and the estimated nonlinear effects are plotted in Figure 2.

Table 2: The Estimated Knots

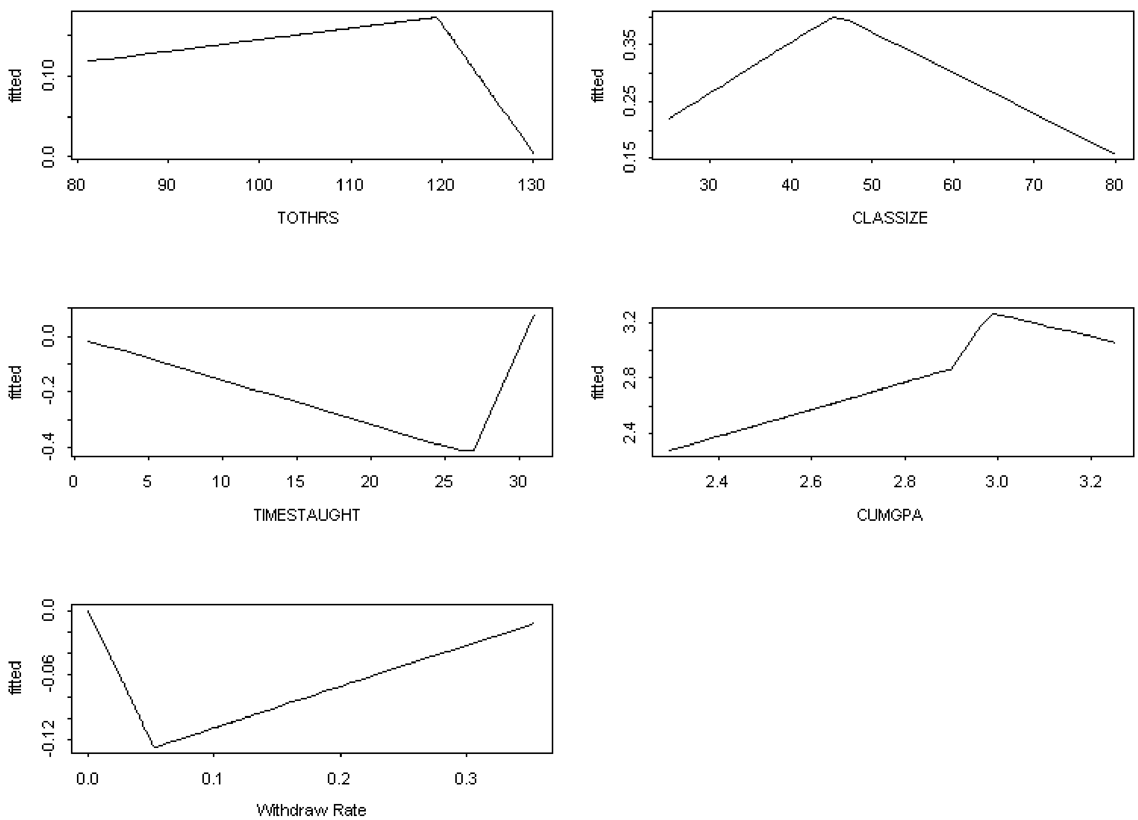
Knot	Value
κ_{11} (TIMES)	26.89
κ_{12} (TOTHRS)	119.44
κ_{131} (GPA)	2.90
κ_{132} (GPA)	2.98
κ_{14} (SIZE)	45.56
κ_{15} (W)	0.052

RESULTS

The regression results are provided in Table 3. The resulting residual standard error is 0.2926 and the multiple coefficient of determination is 0.5844. From Table 3 we can see that while many of the variables are statistically significant, others are not. It is well-observed that in practice, the explanatory variables in a multiple regression are usually correlated and the orthogonality assumption rarely holds exactly.

When violations of the basic regression assumptions are severe, the variability of the estimated coefficients will be inflated and the regression results may be misleading and inconsistent from one sample to the next. As a test of multicollinearity, we calculated the condition index and the variance-decomposition matrix. Our multicollinearity testing indicates that there are some potential multicollinearity problems with model. For more details on the diagnostic procedure see Belsley, Kuh and Welsch (1980). The results of that analysis suggest that there are three possible linear dependencies, the first involves FULL%, CLASSIZE, and (CLASSIZE – κ_6), the second involves TOTALHRS, (TOTALHRS – κ_5) and CUMGPA, the third involves (CUMGPA – $\kappa_{4.2}$) and (CUMGPA – $\kappa_{4.2}$).

Figure 2: The Estimated Nonlinear Effects



We noted that some of the variables with interdependence such as FULL%, TOTALHRS and (TOTALHRS – κ_5), are statistically insignificant in the regression model. The practical effect of multicollinearity among these independent variables is that it may produce poor estimates of the individual model parameters. However, even if the individual parameter estimates are unstable, this does not mean that the full model is a poor predictor. The goal of this research is not so much to say that these variables are more or less predictive than those variables in explaining grade levels, but rather to identify which of the subdisciplines may have higher than average or lower than average predicted values, given all of the inputs. In that context, the multicollinearity issue here is relatively minor. The preferred cure for multicollinearity problems is to increase the number of observations, and over time we may be able to do so as more data becomes available.

The results show that the cumulative student GPA (CUMGPA) is a good baseline measure of student performance. Overall, the cumulative GPA is positively associated with class average GPA, but the effect is nonlinear. When CUMGPA is less than or equal to 2.90, the coefficient is very close to 1 (0.9890) as expected; when it is higher than 2.90, the effect increases from 0.99 to 4.86; however when CUMGPA is very high (higher than 2.98), the effect becomes slightly negative (-.79). This makes intuitive sense, because when the cumulative GPA is very high, generally it becomes difficult to maintain the class average GPA at the same high level.

There is a positive difference between non-tenure track faculty compared to tenured/tenure-track faculty, even after controlling for the other variables. The results show that the average class GPA is 0.155 points higher for the non-tenure track faculty than that of the courses taught by tenured/tenure-track faculty. The relatively heavier use of non-tenure track faculty in the Management area could explain some of the differences in average class grades observed in Table 1.

The effect of the withdrawal rate W% on CLASSGPA is nonlinear. When the withdrawal rate is 5.22% or lower, CLASSGPA decreases with the increase of the withdrawal rate (the coefficient of W% is -2.44). However, after the withdrawal rate reaches 5.22%, CLASSGPA starts to grow with withdrawal rate, as evidenced by the positive coefficient of (W%– κ_8) which is 2.83, meaning that the combined effect of W% and (W%– κ_8) is 0.39 at withdrawal rates over 5.22%. At first, it appeared counter-intuitive that the parameter estimate for W% was negative, but when combined with the nonlinear effect, it makes more sense. Our interpretation of this interesting observation is that the "normal" withdrawal rate, which is under five percent, generally reflects either the exit of the better students who decide to eliminate the course from their schedule or the retention of a higher percentage of lower-performing students. When the withdrawal rate becomes higher, though, the class GPA tends to increase, which is evidence of a weeding out process as more and more marginal students withdraw. However, we also note that Table 1 shows that the average withdrawal rates are materially different from discipline to discipline, and it might be that the true relationship is still partially obscured by multicollinearity problems.

The effect of the number of times the instructor taught the same course (TIMES) on CLASSGPA is nonlinear. When the instructor is relatively new to the course (taught 27 times or less), the models shows that CLASSGPA drops about 0.016 points for each additional time the instructor teaches the course. This observation supports prior research that shows a tightening of grading practices as an instructor gains experience. However, it is also interesting to note that the TIMES effect becomes positive when the instructor is very experienced at teaching the same course. Specifically, our data show that when the instructor has taught the course for more than 27 times, CLASSGPA increases about 0.12 points for each additional time the instructor teaches the class. One possible explanation for this is that, the more experience an instructor has, the more likely that s/he will do a good job in teaching the material and students are more likely to learn the subject better. However, another hypothesis that has been floated in the literature is that grade inflation may be associated with instructor burnout. We do not posit an explanation for the change, but simply note its presence in our data.

Table 3: Regression Results					
Source	DF	SS	MS	F-Value	P-value
Model	21	45.16	2.15	25.11	.000
Error	375	32.11	0.09		
Total	396	77.26			
Variable	Coefficient Estimate	Standard Error	t Value	P-value	
(Intercept)	-0.551	0.555	-0.99	0.3213	
TREND	0.002	0.007	0.22	0.8259	
EXPER	-0.027	0.025	-1.09	0.2785	
NONTENURE	0.155	0.044	3.55	0.0004	
ECON	0.038	0.098	0.39	0.6983	
FINC	0.095	0.066	1.45	0.1476	
LOGT	0.397	0.079	5.02	0.0000	
MGNT	0.562	0.072	7.81	0.0000	
MKTG	0.758	0.064	11.81	0.0000	
PCTFULL	-0.064	0.143	-0.45	0.6563	
LATEREG	0.039	0.166	0.23	0.8155	
TOTHRS	0.001	0.003	0.51	0.6070	
TOTHRSk	-0.017	0.010	-1.72	0.0868	
TIMES	-0.016	0.003	-6.36	0.0000	
TIMESk	0.140	0.047	2.96	0.0033	
SIZE	0.009	0.003	2.64	0.0087	
SIZEk	-0.016	0.006	-2.61	0.0095	
GPA	0.989	0.157	6.28	0.0000	
GPAk1	3.873	1.518	2.55	0.0111	
GPAk2	-5.654	2.086	-2.71	0.0070	
W1	-2.442	0.880	-2.78	0.0058	
Wk	2.827	1.075	2.63	0.0089	

The relationship between class size and class GPA is positive up to about 46 students, and then becomes negative at sizes greater than that. This effect is counter-intuitive, for we thought that as the class size increases, students receive less attention from the instructor so the class GPA is expected to be lower. The multicollinearity issues mentioned before and the interaction with the FULL% variable, which was statistically insignificant, may be to blame. Although we did note an increase in class GPA as TOTALHRS increased, the results were not statistically significant. Therefore, while the theory may sound great, the statistical evidence does not support the idea that instructors go easier on graduating seniors.

The absence of a grade inflation trend is also interesting. The average SAT for students at this university has been increasing in the past a few years, but our analysis shows no discernible trend in grades. One plausible explanation is that instructors tend to subconsciously "curve" their grades to meet historical norms. Another explanation is that the marginal effects are too small to be reflected in the limited data that we have to work with. Yet another interpretation is that, even though the demographics appear to be changing, the standards do not.

Finally, we note that the average grades in Logistics, Management and Marketing are statistically significantly higher than the average grades in Accounting, even after making corrections for student abilities and for instructor experience. Economics and Finance are not statistically significantly different. Interestingly, while the simple mean difference between Accounting and Marketing average GPAs shown in Table 1 are 0.50 points different, the regression results show that after making these corrects, the difference is even greater (0.76 points). The differences observed in the regression model between Accounting and Logistics and Accounting and Management are also higher than the differences in means shown in Table 1. While it can be argued that Accounting is a tougher discipline, we also note that the average GPA of the students in Accounting is higher to begin with. Arguably, the weeding out process has already produced higher caliber students in Accounting, yet the average grades do not reflect that, on a relative basis to some of the other disciplines. We do not conclude that they are too high and we do not conclude that they are too low, but we do conclude that they are different, and the reason for the difference is an area for further exploration.

CONCLUSIONS

The results that we report here are of course specific to this institution. However, the research method that we used can be readily applied to any university or college. This type of research is important because it gives more insight into the grading practices among disciplines in business school. That insight is important because the students are competing for honors, rewards, scholarships, and financial success in their careers after college and simple grade inflation in one discipline can unfairly disadvantages students in the other disciplines. By the same token, if there is simple grade deflation, that can also unfairly disadvantage students in one discipline relative to another.

Our study found systematic differences in the average grades given in different disciplines, even after controlling for student variables and instructor variables. We found that the average class grades are higher in the Management, Marketing and Logistics disciplines than in the Accounting, Economics and Finance disciplines, even after controlling for a number of specific factors. Perhaps these disciplines are staffed with better teachers, perhaps they simply give higher grades, or perhaps there is some other explanation. The "why" is an important question, but not one that can be answered here. The purpose of this research is to present a method for identifying differences between disciplines, not to make value judgments on those differences.

If grade inflation in a particular discipline is warranted (i.e., the grades are higher because the teachers are better) then that should be recognized and rewarded. Nothing in this study really measures the degree to which learning takes place, only the relative level of instructor evaluation of that level of learning. The presence of systematic grading differences between disciplines is something that the faculty as a whole should be aware of, because it then leads to better evaluation of the reasons behind these differences. If the grade inflation is attributable to lax standards, then there should be some kind of correction made. If the grade inflation is attributable to higher standards and higher quality instruction, though, the reason for that success and the methodology employed to achieve it should be shared with the other disciplines.

Self-examination of grades and grading practices is a vital aspect of faculty governance. Grading practices are one component of the annual evaluation of faculty and departments. Unexplained grade inflation in one discipline is unfair to the students in the other disciplines and may create perverse incentives for students to choose majors that have inflated grading. Excessive compression may also result in students losing out on opportunities for scholarships, grants, and graduate school appointments. Students are not in a position to fully understand quality differences between majors and must rely on the university community to guide and advise them in these matters. While we do not advocate either inflation or compression, we do advocate that faculty practice self-examination on a continuous basis, and look at not only themselves but also at their colleagues. College teaching is, after all, a team sport.

REFERENCES

Belsley, D. A., Kuh, E., & Welsch, R.E. (1980). *Regression Diagnostics*. New York: John Wiley & Sons.

Briggs, C. L. (2007). Curriculum Collaboration: A Key to Continuous Program Renewal. *Journal of Higher Education*, 78(6), 676-711.

Boretz, E. (2004). Grade inflation and the myth of student consumerism. *College Teaching,* 52(2), 42-46.

Dresner, J. (2004). Grade inflation…Why it's a nightmare. Retrieved October 3, 2007, from http://hnn.us/articles/6591.html.

Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression.* New York: Marcel Dekker.

Felton, J. & Koper, P. (2004). Real GPA and nominal GPA: A simple adjustment that compensates for grade inflation. *Journal of the Academy of Business Education Proceedings*, 5, 1-7. Retrieved May 20, 2006, from http://www.abe.villanova.edu/proc2004/proceed2004.html.

Germain, M.L. & Scandura, T. (2005). Grade inflation and student individual differences as systematic bias in faculty evaluations. *Journal of Instructional Psychology,* 32 (1), 58-67.

Gordon, M. (2006). When B's are better. *Chronicle of Higher Education*, 52(49), 55-55.

Gump, S. (2007). Student evaluations of teaching effectiveness and the leniency hypothesis: A literature review. *Educational Review Quarterly*, 30(3), 55-68.

Hanson, D., Quin, F., and Wells, W. (2002). Grade inflation at the college level: An empirical examination of GPA. *Journal of Business, Industry and Economics*, 2, 37-43.

Kezim, B., Pariseau, S. E., & Quinn, F. (2005). Is Grade Inflation Related to Faculty Status? *Journal of Education for Business,* 80(6), 358-363.

Kwon, I.W. & Kendig, N. L. (1997). Grade inflation from a career counselor's point of view. *Journal of Employment Counseling,* 34(2), 50-55.

Lanning, W. & Perkins, P. (1995). Grade inflation: A consideration of additional causes. *Journal of Instructional Psychology,* 22 (2), 163-168.

Marsh, H. & Roche, L. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology,* 92(1), 202-227.

Nagle, B. (1998). A proposal for dealing with grade inflation: The relative performance index. *Journal of Education for Business,* 74(1), 40-43.

Potter, W. & Nyman, M. (2001). Be careful what you wish for: Analysis of grading trends at a small liberal arts college. *College and University: The Journal of the American Association of Collegiate Registrars,* 76(4), 9-14.

Smith, E. & K. Coombe (2006). Quality and qualms in the marking of university assignments by sessional staff: An exploratory study. *Higher Education*, 51, 45-69.