Local grade inflation and local proportion of withdrawals

Jonathan Kuhn
Purdue University North Central

Aaron Warren
Purdue University North Central

Diane Maletta
University of Notre Dame

Alan Branford
The Flinders University of South Australia

ABSTRACT

Educational institutions must be accountable to communities in general and students in particular, and fair and consistent assessment is an important component of this. If assigned grades for one group of students are higher than grades for a similar group of students, then grade inflation is said to be localized to the first group relative to the second. Local grade inflation is essentially a form of favoritism; one group of students is favored over another group of students. Identifying the behavior of local grade inflation involves comparing local grade point averages (LGPAs) of the grade distributions of different groups of students. Local grade point averages are calculated from the grades of individual students. Grade distributions for 7,500 class sections from a small public Midwestern University, from fall 1998 to fall 2007, were collected and analyzed. Statistically significant (p-value < 0.01) categorical explanatory variables for LGPAs were compared and contrasted with statistically significant categorical explanatory variables for local proportions of withdrawals (LPWs). Statistical analysis found clear evidence (p-value < 0.01) that both LGPA and LPW are significantly different for different explanatory variables such as courses and instructors, as well as subjects, departments and academic course levels, but not for instructor academic qualifications, gender, and job category, nor for academic year, academic semester and class time period. Moreover, the R^2 measures of fit of model to data for one-variable, two-variable and multi-variable LGPA-dependent analysis of variance models were mostly larger than for equivalent LPW-dependent one-variable, two-variable and multi-variable models.

Keywords: local grade inflation, grade distribution, local grade point average (LGPA), local proportion of withdrawals (LPW)

INTRODUCTION

Although much has been said about longitudinal grade inflation, little has been said about non-uniform grade inflation across a campus---local grade inflation. Longitudinal grade inflation studies measures changes in student grade distributions over time, while accounting for other possible variables such as class size or academic course level. A local grade inflation study measures differences in grades between groups of students, where, in addition to class size and academic course level, time is also relegated to the status of another variable.

If grades for one group of students are higher than grades for a similar group of students, then grade inflation is said to be localized to the first group relative to second. More specifically, measuring local grade inflation involves comparing local grade point averages (LGPAs) of the grade distributions of two different groups of students. Local grade inflation essentially measures favoritism, whether one group of students is unfairly ranked over another group of students. Although some differences in grades between groups are inevitable, mainly due to variations in instructional styles or the students themselves it is difficult to reconcile significant local grade inflation between groups of students.

Longitudinal grade inflation and local grade inflation generally exist at the same time (except in extreme cases) but counteract one another. On the one hand, severe longitudinal grade inflation leads ultimately to all students getting an A and an inability to academically distinguish one student from another. Severe longitudinal grade inflation defeats the purpose of grading, of academically distinguishing one student from another and of ranking students. Since all students are ranked equally, all are granted an A, grades for one group of students cannot be higher than grades for another similar group of students and so, in this extreme case, there is no local grade inflation. On the other hand, severe local grade inflation implies there are huge differences in grades for one group of students relative to a similar group of students. Since severe local grade inflation implies grades are unequal, in particular, all cannot be A, there cannot also be at the same time severe longitudinal grade inflation. With little or no longitudinal grade inflation, it is possible to clearly rank and grade students. Even with little or no longitudinal grade inflation, though, students may still not be ranked fairly with respect to one another at one time period if local grade inflation exists.

Both LGPA and local proportion of withdrawals (LPW) depend on the mix of grades in any given group of students. The entire data set of students is divided into groups of students by a categorical explanatory variable in a model of the data: membership of students in a group is defined by this explanatory variable. For example, explanatory variable “course” divides the data set into different courses, where each group consists of students who took the same course. Eleven explanatory variables are considered in this study, namely course, subject, department, academic course level, instructor, instructor gender, academic qualifications, job category, academic year, academic semester and class time period. The notion of "local" applies to any group of students created by these eleven categorical explanatory variables. The numerical variable class size was used as a weighting variable in the model.

Grade distributions for 7,500 class sections from a small public Midwestern University over fall 1998 to fall 2007 were collected and analyzed. This paper describes a statistical investigation into local grade inflation and differences in local proportion of withdrawals in these data. The central questions considered were:

- Does this data set provide evidence for local grade inflation? In what way(s)?
- Is there evidence of differences in local proportion of withdrawals? In what way(s)?

- What are the similarities and differences between these two?

LITERATURE REVIEW

Introduction

Both longitudinal grade inflation (Kohn, 2002; McSpirit, et al., 2000a and 2000b; Germain and Scandura, 2005; Stone, 1995; Gose, 1997; Shoichet, 2002; Eiszler, 2002) and local grade inflation (Prather et al, 1979; Kolevzon et al., 1981; Nelson and Lynch, 1984; Sabot and Wakeman-Linn, 1991; Kuh and Hu, 1999; Johnson, 2003; Bar and Zussman, 2005; Bowen, 2006) have been extensively considered in the literature. When local grade inflation is considered, it is often considered in relationship to variables such as student evaluations (Remmers, 1930; Nelson and Lynch, 1984; Johnson, 2003) or course choice (Prather, Smith and Kodras, 1979; Sabot and Wakeman-Linn, 1991; Bar and Zussman, 2005; Walton et al., 2008; Bar, Kadiyali and Zussman, 2009) or even instructor incomes (Nelson and Lynch, 1984). Furthermore, local grade inflation is typically considered with respect to instructor and/or department (Johnson, 2003), but not with respect to all four groupings of course, instructor, department and college, as it is in this paper.

Withdrawal--or, conversely, persistance--rates have also been widely analyzed. A number of similar theoretical models have been proposed which typically involve various socio-economic-academic variables, including, typically, student grade point average (GPA) to explain withdrawal rates (Spady, 1970; Tinto, 1975; Bean, 1980; Pascarella, 1980; Bean and Metzner, 1985; Pascarella, Duby, and Iverson, 1983; Simmons-Welburn and Welburn, 2001; Hausmann et al., 2007; Boddy, 2010). These models or variations of these models have been analyzed using data drawn from various types of academic institutions. (Pascarella, Duby and Iverson, 1983; Braxton, Brier and Hossier, 1988; Kember and Harper, 1987; Napoli and Wortman, 1998; Thomas, 2000; Perry et al., 2005; Stupnisky et al., 2007)

Some withdrawal papers suggest possible theoretical models describing withdrawal rates (Spady, 1970; Tinto, 1975; Bean, 1980). Other papers provide an assessment of withdrawal rates based on survey data, often between 300 to 1000 students (Bean and Metzner, 1985; Pascarella, Duby, and Iverson, 1983; Simmons-Welburn and Welburn, 2001; Hausmann et al., 2007), but very rarely based on a huge dataset of 307,672 student academic and administrative records of one entire college campus over an almost ten-year period, as is the case in this paper. Furthermore, data is often analyzed using regression and/or structural models (Pascarella, Duby and Iverson, 1983) but not models with nested variables as is done in this paper.

Most often, withdrawal rate is considered the dependent variable and local grade inflation (and other variables) independent variables. (Bean and Metzner, 1985) Only rarely is the reverse relationship considered in the literature and, if it is, the relationship is often hidden using bi-directional correlation. (Kuh and Hu, 1999; Walton et al., 2008) In this paper, local grade point average (LGPA) and local proportion of withdrawals (LPW) are considered separately, as two dependent variables for two separate models; each model has a different dependent variable but the same set of explanatory variables.

Longitudinal grade inflation versus local grade inflation

"(Longitudinal) Grade inflation is an increase in reported grades unwarranted by student achievement" (Stone, 1995, p 3) Grade inflation, particularly longitudinal grade inflation, has appeared in the literature for at least 80 years and is often associated with instructors attempting to improve their teaching evaluations. (Remmers, 1930) For example, Eiszler (2002) compared 983,491 expected grades and teaching evaluations, using linear and polynomial regression, at a mid-sized public university in 1980-1999 and found higher expected grades significantly predicted higher teaching evaluations, even after controlling demographic and other possible confounding variables. "The percentage of students expecting A/A– grades increased steadily by a total of more than 10 percentage points during the 1990s after remaining stable during the 1980s. Student ratings of teaching gradually, but steadily, increased by more than one-tenth of a point after remaining relatively stable during the first half of the 1980s. The predictive relationship between student ratings of teaching and expected grades was significant even after variables related to alternative explanations were statistically controlled." (Eiszler, 2002, p 483)

In contrast to this, Shoichet (2002) reports researchers at the Education Department's National Center for Education Statistics analyzed the huge 1999-2000 National Postsecondary Student Aid Study dataset of 16.5 million undergraduates and found less severe grade inflation than many would suppose; in particular, 14.5% A and more than one third C or below. Furthermore, like Kuh and Hu (1999), they found grade inflation occurred more severely in selective four-year institutions than at other institutions.

Economics also seems to play a significant role in longitudinal grade inflation. Stone (1995) gives anecdotal evidence which suggests since higher education funding depends on enrollment, this encourages enrolling as many students as possible, and so lowering academic standards and, effectively, causing grade inflation. Germain and Scandura (2005) suggest, based on a general theoretical model, "(longitudinal) Grade inflation may be due to consumerism by universities that now compete for students." (p 58) They suggest rather than focus on improving teaching evaluations to keep students happy, faculty should focus on accommodating student's individual differences.

Gose (1997) gives anecdotal information which suggests in spite of academic worries about grade inflation, grades continue to rise because of pressure from students, administrators and other faculty members. Kohn (2002) advocates rather than worry about grade inflation, do away with grades, and rank students using other more complete measures of competence, although these other measures are not made clear. Bowen (2006), a university president, reports with anguish his own failures to stop grade inflation at his college where some departments had an average of A-. The point here is Bowen was not concerned about an increase in grade inflation universally across all departments, but in grade inflation local to some departments, which is a central concern in this paper.

Local grade inflation often compared with variables other than withdrawal rates

A review of the literature suggests local grade inflation alters student behavior in a number of different ways. Kolevzon (1981) compared the grades of 20 departments from a 4-year university with 8,500 full-time undergraduate students between two academic years, 1969-70 and 1975-76—specific sample sizes were not given. The author found increased average grade inflation, but with higher grade inflation local to the sociology and psychology departments and lower grade inflation rates local to the music and art departments. Female instructors marked easier than male instructors. Male students were attracted to easier subjects more than female students. Lower grade inflation occurred in departments using objective exams rather than in departments using papers as grading tool. There is no mention of a comparison of grade inflation with withdrawal rates in Kolevzon's paper.

Nelson and Lynch (1984) find not only is there a positive association between local grade inflation and teaching evaluations but also between local grade inflation and instructors' incomes. "First, there is evidence to support the contention that the evaluation process will produce some grade inflation; that is, easier grading is positively correlated with teaching evaluations. Second, there is evidence to support the hypothesis that faculty will adopt somewhat easier grading policies if they are experiencing falling real incomes from teaching." (Nelson and Lynch, 1984, p 21) Nelson and Lynch (1984) analyzed 10,658 student evaluations from 13 undergraduate lower division economics courses between winter 1973 and fall 1979 using a simultaneous equations model.

An extensive study by Prather, Smith and Kodras (1979) covering 144 undergraduate courses was based on 125,000 final course grades at a large public university from 1966 to 1975. It shows that rather than a homogeneous systematic change in grading patterns over time, grade inflation is local to individual courses and programs caused by students choosing less challenging courses and programs. Prather, Smith and Kodras' paper focuses on students choosing lenient courses, which could be thought of as effectively withdrawing from challenging courses, but this is clearly different from students first attending but withdrawing from challenging courses, as is the case in our paper. A number of demographic and other variables were included in the multiple regression linear models used to describe the grade inflation data in Prather, Smith and Kodras' paper.

Bar and Zussman (2005) and, later, Bar, Kadiyali and Zussman (2009) analyzed a variety of data, including 800,000 student records over a period 1990 to 2004, 63,540 web site visits between May 2002 and December 2004 as well as a survey of around 500 economics students in spring 2006 at Cornell University. They found significant increases in student enrollment into leniently graded courses after the introduction of a web site giving instructor grade distributions. They called this compositional, rather than local, grade inflation and, interestingly, separated compositional grade inflation from traditional grade inflation, something we do not do. However, again, there was no discussion of local withdrawal rate.

Sabot and Wakeman-Linn (1991) studied, using probit functions, 376 student transcripts, application forms and a survey from Williams College in 1985-86. They found that students enrolled in courses with higher grades were more likely to take more courses from the same department. There was no direct discussion on withdrawals.

Theoretical models of withdrawal rates

Withdrawal rates are often associated with identifying why a student would withdraw from the education system as a whole, or, more specifically, a university, rather than from a particular instructor, course, department or college (within a university), as is the case in this paper. Spady (1970) appears to be one of the first to synthesize ideas from previous literature to create a theoretical socio-economic-academic flow-chart model of the various possible variables, including grade performance, which could lead to student academic dropout from a university. This paper also creates a flow chart, but rather than focus on social or economic factors which might influence a student to withdraw from college entirely, this study focuses strictly on the eleven factors listed above, all of which are academic and administrative factors internal to the campus.

Tinto (1975) follows up the work of Spady (1970) and others by introducing another possible theoretical socio-economic-academic flow-chart model to describe and explain the various factors that might lead to academic dropout. Tinto distinguishes between different types of dropouts, including voluntary withdrawal, which he supposes might be more socially influenced than forced dismissal, which might be more academically influenced. (Tinto, 1970, p 92) He also discusses two other related categories of transfers and persisters. Our paper deals with voluntary withdrawal only: it calculates withdrawal rate as a percentage of students who receive a W grade local to, for example instructor, course, subject, department and so on. More than this, failed (forced failure of) students who receive an F grade are included in the calculation of the local grade point average for local grade inflation.

Bean and Metzner (1985) describe another possible theoretical model which suggests withdrawal of nontraditional students is affected by mostly living off campus, being older, and mostly being part-time attendance. Our analysis is based on data which contains a large number of nontraditional students.

Our paper does not explore the emotional, social or economic reasons why a student might withdraw from college. Boddy (2010), in a short paper, suggests a Freudian projective technique might help identify the emotional reasons for why a student withdraws from a course. An example of this projective technique would be to ask a student who has withdrawn to give reasons, after looking at a cartoon of two students talking, why one student would withdraw from college. Simmons-Welburn and Welburn (2001) suggest more closely integrating library services, such as online information, the internet and printed holdings, with other student support services to encourage entering undergraduate students particularly from the lowest socioeconomic status levels to achieve higher academic levels. Again, these studies are viewed from a campus-wide perspective, unlike in the current study.

Analysis of withdrawal rates based on survey data

Analysis of the various socio-economic-academic flow-chart models of withdrawal often rely on relatively small-sized survey data of students. Braxton, Brier and Hossier (1988) elaborate on an earlier model by Tinto (1975) which suggests an entering student's background and commitment to academics not only increases their level of social and academic integration in a college but also outweighs their financial, personal or psychological problems and so reduces withdrawals. Results from an analysis of survey data from an initial 260 students taken over a three semester period in this paper finds a student's persistence depends not only on initial commitment, but also on a student's continued academic commitment. By way of comparison, our analysis is based on 307,672 student academic and administrative records of one entire college campus over nearly a ten-year period

Kember and Harper (1987) use discriminant analysis on survey data of 779 students to show the psychological surface (memorization) and globetrotting (jumping to conclusions) types of learning that are associated with student withdrawals. They analyzed a number of types of learning, as explained in an earlier study by Ramsden and Entwistle (1981). This ties in with the Braxton, Brier and Hossier (1988) paper's suggestion of focusing on a student's continued academic commitment. Rather than use discrimination analysis, this paper used one-variable, two-variable and multi-variable models with both nested and crossed variables to analyze the data.

Analyses of survey data of 1000 freshman in 1994-1995 by Napoli and Wortman (1998) confirmed the theoretical model by Tinto (1975) suggesting that the factors responsible for student persistence in college are primarily academic, and not social. Their study shows immediate term-to-term persistence is more important than the longer year-to-year persistence and students from larger, rather than smaller, campuses integrated better because of more available resources.

Perry et al. (2005) analyzed a voluntary survey of 524 students from a full-year psychology course and found a significant (p-value < 0.001) positive correlation ($r(359) = 0.26$) between GPA and voluntary withdrawal, but not departure (Perry et al., 2005, p 546). In fact, they find high academic control (those who believe they can effect and predict achievement outcomes) but also high failure preoccupation students are the least likely to withdraw from courses or quit college entirely.

Stupnisky et al. (2007) structural analysis of a voluntary survey of 802 first-year students shows, similar to Perry et al. (2005), high academic control significantly positively predicted GPA, controlling for high school grades, age and gender. More than this, they show GPA did not depend on self-esteem, but the reverse was true: self-esteem depended on GPA.

Pascarella, Duby and Iverson (1983) perform a hierarchical analysis on a survey of 579 non-residential student dependents (similar to Bean and Metzner (1985)'s nontraditional students) in 1979-1980, using Tinto (1975) theoretical model of college withdrawal, and finds these students different from residential students; in particular, family background, individual attributes and pre-college schooling ("person-fit environment" factors) play a less important role in dropout decisions of non-residential than residential students. They suggest an alternative model which fits these students better. Their paper has a flowchart of both Tinto's model and the alternative Pascarella et al.'s model. Pascarella, Terenzini and Wolfle (1986) employ causally linked structural regression equations on a survey of 763 student dependents in 1976-77, using Tinto (1975) theoretical model of college withdrawal, to find although a two-day student orientation had little direct effect on persistence, had a large indirect effect, through social integration and commitment to institution, to freshman year persistence. Pascarella (1980) also suggest strong student-faculty informal contact increases student persistence.

Thomas (2000) analyses 322 students at a four-year private liberal arts college in 1993 using single level path model and finds while number of student acquaintances is important, structural location of peers effects, among other things, persistence. GPA is included in this study and is found to be positively associated with student outdegree (number of friends a student says they have) as opposed to student indegree (number of students who say a student is their friend). GPA is also positively associated with withdrawal (from college, as opposed to a course).

Hausmann, Schofield and Woods (2007) used a multilevel model change technique to analyze three surveys completed by 365 students from a large public mid-Atlantic University and found a sense of belonging declined along with intentions to persist over an academic year. Persistence improved if students were made to feel more welcome by receiving letters and small gifts (for example, magnets and decals with the university's name and logo) from administration and faculty. Again, persistance is with regard to persistance at the college, rather than persistance in a particular course, or with a particular instructor.

Braxton, Jones, Hirschy and Hartley (2008) use four hierarchical linear regression analyses to study data from 408 first year students in eight colleges and universities and found some evidence to suggest faculty should engage students in active learning practices to improve student persistence. The models used in this paper used both hierarchical as well as crossed variables and were used on data from one college campus only.

Bean (1980) created a theoretical model based on work turnover (similar to suicide-based Tinto and Spady models), which was assessed using both multiple regression and path models applied to data collected from 1,195 students. Standardized beta weights and correlation were used; also, differences in student gender was also considered, unlike in our study.

Ishitani (2006) used event history modeling to analyze data from 4,427 diverse students found in a national survey called NELS: 1988-2000 Postsecondary Education Transcript and found first generation students had a higher dropout rate and took longer to complete their degree programs.

Local grade inflation associated with withdrawal rates

Walton et al. (2008), an educational policies committee of the academic senate for California Community Colleges, recognize the dangers associated with allowing different grade distributions for students taking the same course in the same department/program/college but with different instructors, or formats and lengths, in particular, related to academic rigor and of “course-shopping” by students. This report suggests and provides anecdotal evidence average GPA increases with increasing withdrawal rates because withdrawals are not included in the GPA. (Walton et al., 2008, p 19) This report suggests dialog within in an institution to compensate for different grading distributions.

Kuh and Hu (1999) compare College Student Experiences Questionnaire (CSEQ) survey data from two periods 1984-87 (52,256 dependents) and 1995-97 (22,792 dependents) across all American colleges and find longitudinal grade inflation effects academic institutions unevenly; in particular, mostly in research universities as well as selective liberal arts colleges and also for white, female and upper-division students. Like Walton et al. (2008), they suggest withdrawals, incompletes and repeat courses (replacing lower grade with higher one) will inflate grades in remedial and lower-level mathematics courses. (Kuh and Hu, 1999, p 298) This is in contrast to the current study which shows local grade point averages are inversely related to local proportion of withdrawals.

Johnson (2003) thoroughly analyzes both survey and transcript information to show local grade inflation; in particular, grading is most to least stringent for the natural sciences/mathematics, social sciences and then humanities, in that order. Johnson’s work shows there is a statistically significant positive correlation between grades and teaching evaluations, even controlling for variables such as gender, race, year, department, GPA and, most importantly, teaching effectiveness. Unlike Walton et al. (2008) and Kuh and Hu (1999), Johnson shows an inverse relationship between local (to departments) grade inflation and withdrawal rates.

Summary

There are many papers on either grade inflation or on withdrawal rates, but generally not on both where they are thoroughly compared and contrasted, as they are in this paper. Most papers base their analyses on datasets involving roughly 100 to 1000 students, substantially smaller than the huge data set of 307,672 student records gathered over an almost ten-year period given in this paper. A wide variety of statistical and probability models are used in other papers, but none like the one-variable, two-variable and multi-variable analysis of variance (ANOVA) models where variables are both nested and crossed, as appear in this paper.

METHOD

Data were collected on students (307,672 records, 17 variables), instructors (16,846 records, 12 variables) and course sections (91,878 records, 5 variables) at a small public Midwestern University from fall 1998 to fall 2007. This study was based on a single dataset of 7,500 grade distribution records (11 main variables) distilled from the student, instructor and course section datasets.

As shown in Figure 1 (Appendix), variables for the grade distribution records used in this study consisted of the eleven categorical explanatory variables course, subject, department, academic course level, instructor, instructor qualifications, instructor job, instructor gender, academic year, academic semester, class time period, and one quantitative explanatory variable class size. The two main dependent variables of interest were the restricted range continuous variables local grade point average, LGPA, and local proportion withdrawn, LPW. Class size was used as a weighting variable in the model. All variables were academic institution variables; socio-economic variables were not considered in this study.

A number of possible models derived from the variables in Figure 1 were fitted to the grade distribution records data. Nesting (hierarchies) restricted model choice for the data: although courses could be nested in either subject or academic course level, courses could not be nested inside both subject and academic course level at the same time in any one model, for example. Two hierarchical (or nested) groupings appear in Figure 1, including the course hierarchy of course, subject, department and academic course level grouping and the instructor hierarchy of instructor, instructor qualifications, instructor gender and instructor job grouping. This type of nesting structure is common in the education literature (Pascarella, Duby and Iverson, 1983). In addition to nested variables, model choice also depended on possible pairwise interactions between variables.

Roughly twenty-five percent of student, instructor and course section datasets were culled because of missing or corrupted information. Matching information from the student, instructor and class section databases to create the final grade distribution records dataset was difficult due to, for example, different variable naming conventions in different source datasets or variable names that changed over the almost 10 year time period of the study. Consequently, a number of variables, particularly related to the instructor part of the dataset have categories designated either "not available" or "other".

Course Hierarchy Explanatory Variables

The course hierarchy consisted of course, subject, department and academic course level. To clarify how these four explanatory variables appeared in a model of the data, consider an example dataset for eight class sections in Table 1 (Appendix). Here, a total of four courses (MA 223, STAT 213, STAT 301 and PSY 201) in three subjects (mathematics, statistics and psychology) were taught by five instructors (005, 031, 056, 093 and 111) in two departments (MSP and SSC) over two semesters, fall 2005 and spring 2006. One record corresponds to one class section. For example, the MA 223 course was taught four times (there were four class sections of MA 223), mathematics subjects were taught four times, instructor 111 taught three times, faculty from the MSP (Mathematics, Statistics and Physics) department taught seven times and there were six second-year courses (4 MA 223, 1 STAT 213 and 1 PSY 201) and two third-year courses (both STAT 301).

Course, subject, department and academic course level variables, are hierarchically related to one another, as shown in Figure 1. In particular, courses are nested in subjects since each course belongs to one exactly one subject. For example, all statistics courses such as STAT 213 and STAT 301 are designated as members of the statistics subject. In a similar way, subjects are nested in departments, which often house multiple subjects. Furthermore, courses are nested in course level: MA 223 and STAT 213 are both exclusively sophomore courses and STAT 301 is a senior course only. Academic course level has six levels in the study: pre-university, freshman, sophomore, junior, senior and graduate levels. Instructors, though, are not nested in courses or subjects or departments because, for example, an instructor may teach different mathematics courses, or may teach non-mathematics courses such as statistics courses (hence, different subjects) or even teach courses outside of his department such as economics courses.

A total of 673 courses in 69 subjects were taught by 338 instructors in ten departments from fall 1998 to fall 2007. Of 673 different courses given, 120 were taught only once, and one course, COM 113, was taught 266 times. Each course was taught an average of 28.69 times with standard error (sample standard deviation divided by square-root of sample size) of 0.478 times, although 49.8% of courses were taught 10 or fewer times over the almost ten-year period. Of 69 different subjects taught, all were taught at least twice and one subject, mathematics, was taught 626 times. Each subject was taught an average of 144.85 times with standard error of 1.556 times, although 49.8% were taught 101 or fewer times from fall 1998 to fall 2007. Of 341 different instructors who taught, 25 taught only one class whereas two instructors taught 102 times. Each instructor taught an average of 22.17 times with standard error of 0.22 times, although 49.6% taught 16 or fewer times from fall 1998 to fall 2007. Of 10 departments, faculty in business taught the most, 1435 class sections, whereas nursing taught the least, 326 class sections. Departments taught an average of 750.1 course sections with standard error 4.29 sections from fall 1998 to fall 2007. Freshman, sophomore, junior and senior undergraduate courses made up 40.5%, 30.3%, 18.8% and 7.3%, respectively, of 7,500 class sections given taught 1998 to fall 2007.

Instructor Hierarchy Explanatory Variables

The instructor hierarchy explanatory variables consisted of instructor, instructor academic qualifications (four categories: bachelors, masters, PhD and other), instructor gender (male or female) and instructor job (six categories: limited term lecturer, continuing lecturer, assistant professor, associate professor, professor and other). Again, nesting restricted the model choice for the data: instructor could be nested in one and only one of academic qualifications, gender or instructor job, in any one model.

Most class sections were taught by associate professors, 2,404 (32%), and limited term lecturers, 1,946 (25.9%), over fall 1998 to fall 2007 period of study. Most sections were taught by instructors with Masters, 3,760 (50.1%), and PhDs, 3,010 (40.1%), from fall 1998 to fall 2007. Most sections were taught by males, 4,712 (62.8%), and so 2,789 (37.2%) were taught by female instructors.

Other Explanatory Variables

Other possible explanatory variables used in models to fit the data, included academic year (10 years: 1998-99, 1999-2000, up to 2007-08), academic semester (two semesters: fall, spring), and class time period (three periods: morning, afternoon and evening), and class size.

The number of sections taught per year has more or less steadily increased from academic year 1998-1999, with 493 taught, to academic year 2006-2007, with 1,103 taught. Fewer class sections were taught in fall semesters, 3,443 (49.5%), than in spring semesters 3,516 (50.5%), from fall 1998 to spring 2007. More sections were taught in the afternoons, 3,010 (40.1%), than in either the mornings 2,450 (32.7%) or evenings 2,041 (27.2%) from fall 1998 to fall 2007 period.

Although student group size did not appear explicitly as a variable in the analysis of data, it did play the important role of appropriately weighting both local grade point average, LGPA, and local proportion withdrawn, LPW. Student group size (excluding withdrawals and incompletes) ranged from zero to 126 students, with an average of 17.39 students and standard error of 0.15 students per class section, from fall 1998 to fall 2007. Student group size (including withdrawals but excluding incompletes) ranged from zero to 145 students, with an average of 19.17 students and standard error of 0.16 students per class section over the fall 1998 to fall 2007 period.

Calculating Local Grade Point Average (LGPA) and Local Proportion Withdrawn (LPW)

Both LGPA and LPW were calculated for different (local) groups of students; specifically, not only for a single class section of students, but also for students in all class sections of a course, or all students of one instructor, or all students taking courses in one subject or all students taking courses in all subjects given by one department. In other words, the notion of "local" applied not only to a class section, but also to a course, instructor and the other explanatory variables. As a consequence of this, both LGPA and LPW were weighted according to the different sizes of local groups of students. To clarify how LGPA and LPW were calculated, it is helpful to revisit Table 1.

LGPA and LPW for a class section are weighted by class size. The first record in Table 1, section 001 of MA 223, is a mathematics course of 28 students taught by instructor 111, taught in fall of 2005-06. The frequency of letter grades A, B, C, D, and F for this section is 3, 11, 13, 1, and 0, respectively, for a total of 28 students. Assigning 4 points for each A, 3 points for each B, 2 points for each C, 1 point for each D and 0 points for each F, in this case,

$$\mathrm{LGPA} = \frac{4 \times 3 + 3 \times 11 + 2 \times 13 + 1 \times 1 + 0 \times 0}{28} \approx 2.57.$$

Also, since one student withdrew, the proportion of students who withdrew from this section,

$$\mathrm{LPW} = \frac{1}{29} \approx 0.03.$$

LGPA and LPW for an instructor are weighted by number of students taught by this instructor. The LGPA for an instructor must be weighted by aggregated class section sizes (excluding withdrawals and incompletes). For example, the LGPA for instructor 111 is calculated by a weighted combination of records 1, 4 and 5 in Table 1,

$$\mathrm{LGPA} = \frac{4 \times (3+4+5) + 3 \times (11+11+6) + 2 \times (13+12+9) + 1 \times (1+1+1) + 0 \times (0+1+1)}{28+29+22} \approx 2.57.$$

Also, the LPW for the instructor must be weighted by aggregated class section sizes (including withdrawals but excluding incompletes).

$$\mathrm{LPW} = \frac{1 + 3 + 0}{29 + 32 + 22} \approx 0.05.$$

LGPA and LPW for a subject are weighted by number of students who took courses in this subject. The LGPA for mathematics subject is calculated by combining records 1, 2, 3 and 5 in Table 1,

$$\mathrm{LGPA} = \frac{4 \times (3 + 3 + 1 + 5) + 3 \times (11 + 5 + 5 + 6) + 2 \times (13 + 5 + 4 + 9) + 1 \times (1 + 5 + 3 + 1) + 0 \times (0 + 0 + 1 + 1)}{28 + 18 + 14 + 22} \approx 2.45.$$

Also,

$$\mathrm{LPW} = \frac{1 + 6 + 3 + 0}{29 + 24 + 17 + 22} \approx 0.11.$$

The LGPA and LPW for other subjects, courses and other explanatory variables are calculated in a similar way.

Pairwise associations (dependencies) between categorical explanatory variables

Symmetric measure of association, Cramer's V, and directional measure of association, λ, were both used to check for pairwise associations between the eleven categorical explanatory variables, as shown in Table 2 (Appendix). Cramer's V is related to the chi-squared measure of association, χ^2, for a two-dimensional contingency table with R rows and C columns,

$$\mathrm{V} = \sqrt{\frac{\chi^2}{N\, min(R - 1, C - 1)}},$$

where, for $0 \le V \le 0.1$, $0.1 < V \le 0.3$, and $V > 0.3$, two variables are considered weakly associated, moderately associated and strongly associated, respectively. Lambda, λ, measures the proportional reduction in error for a two-dimensional contingency table with sample size N,

$$\lambda = \frac{\mathrm{E1} - \mathrm{E2}}{\mathrm{E2}},$$

where E1 = sample size (N) – largest of row totals and where E2 = sum of column differences, after subtracting the largest cell frequency from each column total. If $0 \le \lambda \le 0.4$, $0.4 < \lambda \le 0.6$, and $0.6 < \lambda \le 1.0$, two variables are considered weakly associated, moderately associated and strongly associated, respectively.

According to V and λ, instructor and course are strongly pairwise associated with most other explanatory variables. As shown in Table 2, instructor gender, instructor job and instructor qualifications all are perfectly associated with (V = 1) and, more than this, perfectly dependent (λ = 1) on instructor because of the hierarchal relationship between these variables. For example, knowing any instructor improves prediction of instructor's job by 100%. Also not too surprisingly, instructor is strongly associated ($V \geq 0.575$) with course, subject, department and course level. In fact, the variables subject and department are strongly dependent (λ = 0.833 and 0.986, respectively) on instructor: most instructors teach courses in one, possibly two subjects and usually in one department. There is also a strong association (V = 0.521) between instructor and course period. In a similar way, perfect association results (V = 1, λ = 1) occur for variables in the hierarchal relationship between course (C) and the other explanatory variables subject (S), department (D) and academic course level (L). Also, course is strongly associated ($V \geq 0.575$) with instructor, instructor gender, job and qualifications, and the latter variables are moderately

($0.537 \leq \lambda \leq 0.579$) dependent on course. Course is also strongly associated with academic semester and year (V = 0.478 and 0.562, respectively).

According to V, instructor gender, job and qualifications are strongly pairwise associated with subject and department. Subject is strongly pairwise associated with instructor gender, job and qualifications as well as class period (V = 0.574, 0.454, 0.447 and 0.331 respectively); also, department is highly significantly associated with instructor gender (V = 0.447).

RESULTS

Types of analyses used in study

Analyses of variance (ANOVA) were used on models of the data of two dependent variables, LGPA and LPW, and eleven categorical explanatory variables. The basis of this technique is the philosophy that the variability in the data is described either by the model (measured by the sum of squares of the model, SSM) or otherwise (measured by sum of squares of error, SSE). The fit of the model to the data is calculated by a ratio of an average of the former (the mean squares of the model, MSM = SSM / degrees of freedom of model) divided by an average of the latter (the mean squares of error, MSE = SSE / degree of freedom of error), called the F = MSM / MSE statistic. A model is considered a good fit to the data when a large proportion of the total variation is described by model rather than by error, that is, the probability of having such a large F (called the p-value) is small. As well as just assessing whether an entire (possibly multi-variable) model fits the data well, F statistics are also often used to assess whether the presence of individual variables in a multi-variable model improve the fit of the model, and whether the marginal F statistics for these individual variables are statistically significant. (Kendall et al., 1983, chapters 35 to 37)

Of four possible different types of ANOVA (Littell et al., 1991, chapter 4), only types III and IV analyses were used in this study. Type I and II analyses were not used because they both deal with orthogonal balanced (or more generally, proportional) data where number of observations for each cell is equal (or more generally, proportional with respect to relevant row or column totals—see Kendall et al., 1983, chapters 35 to 37; Pendleton, 1986. This was clearly not the case in this observational study because, for example, an unequal number of male and female instructors taught during the nine and half year period. Type I and II analyses are mostly used in experiments, where an equal (or proportional) number of observations can be assigned to each cell by the researcher. Type III and IV analyses (Driscoll and Borror, 2000; Langsrud, 2003; Freund, 1980; Searle, 1994; Yates, 1934) were used in this study because they both deal with unbalanced (non-proportional, non-systematic) data. More than this, both type III and IV analyses can deal with unbalanced data where, in particular, some cells have no observations; in fact, the type IV analysis was designed specifically to handle this messy data situation. There are many empty cells in the current study. For example, not all 341 instructors taught all 673 courses and so in a table of the instructors versus courses subset of data, there would be a great many empty cells each indicating where a specific instructor did not teach a specific course.

The problem an ANOVA has dealing with messy data is calculating p-values on data that simply is not there. An ANOVA essentially looks for statistically significant differences between the averages of different levels of variables and this becomes difficult if one or more of these levels have no data (empty cells), so no averages are associated with these levels to allow for a sensible comparison for differences. Understandably, both type III and type IV analyses work better with a few rather than many empty cells because then p-values are based on more rather than less data. Whereas there is only one type III analysis per model, there are many (which depend on the location of empty cells) type IV analyses per model. In spite of the many type IV analyses available, though, the default type IV analysis based on the original data cell locations was used in this study.

Both type III and IV analyses were conducted on the data and compared. When the empty cells do not alter marginal F statistics enough to change the significance/non-significance of variables according to both the type III and type IV analyses, the two different analyses tend to confirm one another. This strengthens the credibility of statistical significance/non-significance of the differences between averages related to, particularly, the main variables in the model. Type III and type IV analyses give identical results if there are no empty cells, if there is only one variable in the model or, in the case that there are two or more variables in the model, when all variables are hierarchical (nested) with respect to one another (which assures no empty cells; for example, in the case of the instructor nested inside instructor gender model, all instructors are either male or female). Also, the marginal F statistic for the interaction effect in a type III ANOVA is always identical to the F statistic for the interaction effect of a type IV ANOVA if the main effects which make up the interaction are present in the model.

Type III and type IV ANOVAs of one-variable, two-variable and multi-variable models are calculated. To compensate for the empty cells in the data, emphasis is placed on identifying those results which occur repeatedly in these different analyses as well as various descriptive statistics and graphs. The fits of the data to various model assumptions were also investigated and included checking residuals for normality and influential outliers. All analyses were conducted using SPSS 17.0 (SPSS Incorporated, Chicago, IL) and SAS 9.2 (SAS Institute Incorporated, Cary, NC).

Type III (and IV) ANOVAs of one-variable models

Coefficients of determination R^2 and model F statistics were calculated for type III (equivalently type IV since there is only one explanatory variable in each model) ANOVAs of all the 22 one-variable models that were possible consisting of one of the two dependent variables, LGPA and LPW, and one of the eleven categorical explanatory variables, as shown in Table 3 (Appendix). The R^2 statistic and F statistic are two different measures of how closely a model describes the data. If a model fits the data closely, then this model has the right mix of variables; more specifically, the explanatory variables (just one explanatory variable in this case) give a good explanation of the dependent variables in this model. If $0 \leq R^2 \leq 0.4$, $0.4 < R^2 \leq 0.8$, and $0.8 < R^2 \leq 1.0$, then a small, moderate and high proportion, respectively, of variability in data is described by a model. In addition, small p-values associated with the F statistic for a model, smaller than 0.01 (and starred* in Table 2), also indicates the explanatory variables give a good explanation of the dependent variable.

According to R^2, course and instructor are most associated with LGPA and LPW in one-variable models. According to the R^2 statistic, the model with explanatory variable "course" and dependent variable LGPA is the tightest fitting of all 22 models to the data because $R^2 = 0.615$ (sample size $N = 7500$) is the largest of all the R^2 values. In other words, without controlling for other variables but weighted by sample size, of the 11 possible explanatory variables given in this study, course alone explains 61.5% of the variability in LGPA. Explanatory variable "instructor" is also strongly associated with LGPA because $R^2 = 0.550$. Explanatory variable "course" is the most influential explanatory variable of LPW where $R^2 = 0.550$. Moderately well-fitting models, where $0.050 < R^2 < 0.500$, include the subject, department or academic course level explanatory variables, for either the LGPA or LPW dependent variables. The loosest-fitting models, where $R^2 \leq 0.050$, include either the year, semester or class time period variables or instructor gender, job or qualifications variables and this is true for either LGPA or LPW.

The model F statistics values are not informative in one-variable models. The LGPA is significantly (p-value < 0.01, $F(667, 6713) = 16.05$) different for different instructors. In fact, both LGPA and LPW dependent variables are significantly different for all 11 explanatory variables in Table 3, except one, the model with dependent LPW and explanatory instructor job variable. All (except one) of the 22 one-variable models describe the data more closely than a basic model where the dependent variable is equal to the average of all data. In other words, the model F statistic does not (except in one case) allow the one-variable models to be distinguished from one another, to identify which explanatory variable most affects either LGPA or LPW. The F statistics are not as informative as the R^2 values for this reason in this one-variable case.

Plots indicate non-normal error distribution for a few one-variable models. Normal probability plots and various scatter plots, not shown here, indicate some non-normality. In particular, these plots indicate there are a surprising number of class sections where the LGPA are a perfect 4.00. The ANOVA analyses used in this study, though, are generally robust to non-normality. (Miller, 1986)

Type III and IV ANOVAs of two-variable models

The R^2, model F and marginal F statistics are calculated for type III and type IV ANOVAs of all the 110 two-variable models that are possible consisting of one of the two dependent variables, LGPA or LPW, and two of the eleven categorical explanatory variables. Two varieties of two-variable models occur in this study: factorial (crossed) or hierarchical (nested). Two-variable factorial models have three explanatory variables, two main variables and an interaction variable. For example, variable instructor (I) is crossed with variable academic year (Y) to give interaction instructor x year and denoted IxY; more specifically, 338 (levels of) instructors are crossed with 10 (levels of) years to create 10 x 338 = 3380 possible instructor x year interactions. Two-variable hierarchical models have two explanatory variables, one nested inside the other. For example, variable course (C) is nested inside variable subject (S) to give C(S); more specifically, 673 (levels of) courses are divided up into 69 (levels of) subjects. Most, 96, of the 110 two-variable models are factorial, while the rest are hierarchical. Whereas the R^2 statistic and model F statistic both measure how closely a two-variable model fits the data model, a version of the F statistic, a marginal F statistic, is also able to measure how strongly any individual explanatory variable affects a dependent variable within a model. More specifically, in addition to calculating a model F statistic for all main, nested or interaction explanatory variables in a two-variable model based on the model sums of squares, marginal F statistics are calculated for each individual explanatory variable that are each based on sums of squares apportioned from the model sums of squares. More than this, sum of squares of an individual explanatory variable is assessed after accounting for sums of the squares of all other explanatory variables in a type III or IV analysis of the model. If the marginal F statistic resulting from the additional sums of squares due to the individual variable is larger than would be expected by chance alone, then the addition of this individual variable causes an improvement of the model's fit to the data, and so is a significant effect on the dependent variable. (Kendall et al., 1983, chapters 35 to 37; Pendleton, 1986)

Each element in Table 4 (Appendix) corresponds to one of the 110 two-variable models. The 55 elements above the diagonal in Table 4 are associated with LGPA-dependent models; the 55 elements below the diagonal are associated with LPW-dependent models. For example, the element at the intersection of the instructor row and instructor gender column, the I|G model, gives information on an ANOVA of these two nested variables with the LGPA dependent variable; in contrast, the element at the intersection of the instructor gender row and instructor column, the G|I model, gives information on an ANOVA of these two nested variables with LPW dependent variable. Each element contains three rows of information: R^2 in the top row and then two rows of a sequence of three letters each. Both rows of three letters correspond to row main variable, column main variable and interaction variable, in that order, but the first sequence of letters is associated with a type III analysis, whereas the second sequence of letters is associated with a type IV analysis. A variable is significant (p-value ≤ 0.01) if a letter appears, otherwise, if a letter does not appear, but a blank indicated by a dash appears instead, the variable is not significant (p-value > 0.01). For example, letter sequence I, C, IxC found in both second and third rows of the I|C model of Table 4 indicates main instructor (I) variable, main course (C) variable and interaction instructor x course (IxC) variable are all significant, using both type III and IV analyses, for this LGPA-dependent model. By contrast, the L, -, - letter sequence in the L|G model for this LPW-dependent model indicates, using both type III and IV analyses, that although academic course level (L) is significant, the instructor gender (G) and academic course level x gender (LxG) are not. Type III and type IV analyses for one model typically give different sums of squares and marginal F values, and so p-values for each of the variables in any particular model are different but not necessarily hugely different. For example, letter sequence I, C, IxC found in both the second and third rows of the I|C model of Table 4 does not indicate that the marginal F statistics associated with instructor (I) variable, course (C) variable and instructor x course (IxC) interaction variable are identical, but it does indicate these marginal F statistics are similar enough that the corresponding p-values, although different for type III and IV analyses, are both still smaller than the level of significance 0.01: type III and type IV analyses are said to be statistically similar to one another. When statistically similar, the location of empty cells alters marginal F statistics somewhat but not enough to change the significance/non-significance of variables, type III and type IV analyses confirm one another and so this strengthens the credibility of the statistical significance/non-significance of particularly the main variables in the model, in this case. If only one row of letters appears in an element in Table 4, type III and IV analyses give identical results: the marginal F statistics and p-values in the associated ANOVA of a model of the data are identical. For example, the I|G model in Table 4 is a hierarchal one, where instructor is nested in instructor gender and so type III and type IV analyses give identical results in this case.

The R^2 values are largest for courses and instructors, particularly for LGPA data in two-variable models. The closest fitting models, where $R^2 \geq 0.500$, are the ones which include either the course variable or instructor variable both with LGPA variable as the dependent variable and, to a lesser extent, LPW as the dependent variable. Moderate fitting models, where $0.100 < R^2 < 0.500$, include the subject, department or academic course level variables both with either LGPA as the dependent variable or, again to a lesser extent, LPW as the dependent variable. The loosest fitting models, where $R^2 \leq 0.100$, include either of the year, semester or class time period variables or instructor gender, job or qualifications variables (except when paired with the course or instructor variables or the subject, department or academic course level variables) with LGPA as dependent variable. The same is true, to a lesser extent, with LPW as the dependent variable. These two-variable model results agree well with the single-variable models given in Table 3, but provide better fits. For example, whereas in Table 3, $R^2 = 0.550$ for the LGPA-dependent model with the instructor explanatory variable, in Table 4, $R^2 \geq 0.550$ for the LGPA-dependent models which include the instructor explanatory variable paired with any of the other ten explanatory variables.

Instructors and interactions are highly significant variables in two-variable LGPA-dependent models. The closest fitting two-variable LGPA-dependent models in Table 4 are the ones where the variable instructor (I) is paired with one of the five variables: course (model I|C, $R^2 = 0.784$), academic course level (model I|L, $R^2 = 0.691$), year (model I|Y, $R^2 = 0.679$), class time period (model I|P, $R^2 = 0.633$) or subject (model I|S, $R^2 = 0.610$). In all of these cases, type III and type IV analyses are statistically similar, give slightly different but statistically significantly similar marginal F statistics, where, specifically, the marginal F statistics of the two main and one interaction variables are significant (p-value < 0.01). First, significant main variables imply LGPAs are significantly different for different instructors after controlling for one of the five other variables and the appropriate related interaction. In particular, the average of all instructor LPGAs is 2.81 with a large standard deviation 0.54 and a large interquartile range from first quartile 2.43 to third quartile 3.21. Similarly, the LPW weighted average is 0.078 with a large standard deviation 0.107 and large interquartile range from 0.045 to 0.114. Second, significant interactions indicate instructors change or adapt their LGPA according to different courses, course levels, years, class periods or subjects in a non-additive way. The looser-fitting two-variable models ($R^2 < 0.600$) where instructors are paired with instructor gender, job or qualifications are all nested (and so there is no interaction) or where instructors are paired with department or semester have non-significant interactions. Overall, instructors more closely describe LGPA when interacting with one other variable than when not interacting with another variable in two-variable models.

Instructors and interactions with instructors are significant variables in LPW-dependent two-variable models. The closest fitting LPW-dependent model with instructor is the one paired with course (model I|C, $R^2 = 0.629$). In general, all nine LPW-dependent two-variable models with instructor in Table 4 are looser fitting, have smaller R^2 values, than corresponding LGPA-dependent two-variable models with instructor. However, in most cases involving instructors, the marginal F statistics for the two main and one interaction variables are significant (p-value < 0.01) in statistically similar type III and type IV analyses of LPW-dependent models. So, although looser fitting than LGPA-dependent models, instructors again more closely describe LPW when interacting with one other variable than when not.

Courses, but not interactions with courses, are highly significant variables in two-variable LGPA-dependent and LPW-dependent models. More (13 of 19) two-variable models with course have larger R^2 values than the corresponding two-variable models with instructor. More (7 of 10) LGPA-dependent two-variable models with course have larger R^2 values than corresponding LPW-dependent two-variable models with course. Most (18 of 20) type III and type IV analyses are statistically similar, giving slightly different, but statistically significantly similar, marginal F statistics for two-variable models with course. Large differences in LGPAs for the 668 different courses has a mean of 3.11 with a large standard deviation of 0.567, and where LGPA has a large range from a minimum of 1.00 to a maximum of 4.00. More (6 of 14) two-variable non-nested models with course involve non-significant interactions than corresponding two-variable non-nested models with instructor (2 of 14): courses are less likely than instructor to interact with one other variable when describing LGPA or LPW.

ANOVAs are increasingly unreliable with a greater number of empty cells. Most (103 of 110) two-variable models have statistically similar type III and type IV ANOVA analyses which strengthens the credibility of the significance/non-significance of the two main variables in spite of empty cells in many of these models. However, seven models do give statistically different type III and type IV analyses: three LGPA-dependent models (I|L, C|Y and C|P) and four LPW-dependent models (L|I, L|S, S|J and C|J). All seven of these models consisted of one of either the instructor (I), course (C) or subject (S) variables, each with a large number of levels (338 instructors, 673 courses or 69 subjects) and so leading to a large number of empty cells. In general, ANOVAs of models describing data with more empty cells, specifically those with instructor, course or subject in particular, are more unreliable than models describing data with fewer empty cells. For example, the credibility of the significance/non-significance of either course or year or both in the C|Y model is somewhat suspect due to empty cells, which are in turn due to not all courses having been taught every year over the near-ten year study period. However, in all 14 hierarchal (nested) two-variable models, which involve only instructor, course and subject variables, there are no empty cells and so in these cases, type III and IV analyses gave the same sensible results.

Plots for some models indicate non-normal error distributions; in particular, normal probability plots and various scatter plots, not shown here, indicate some non-normality. Again, these plots indicate there a surprising number of classes where the LGPAs are a perfect 4.00. Again, the ANOVA analyses used in this study are generally robust to non-normality. (Miller, 1986)

Type III and IV ANOVAs of multi-variable models

The R^2, model F, and marginal F statistics are calculated for type III and type IV ANOVAs of all eighteen multi-variable models consisting of one of the two dependent variables, LGPA and LPW, and as many (in fact, it turns out to be exactly 26) of the main, nested and two-variable interactions of the eleven categorical explanatory variables as possible. Similar to before, the R^2 statistic and model F statistic both measure how closely a 26-variable model fits the data model. Also, marginal F statistics measure how strongly any individual explanatory variable of a model affects the dependent variable after accounting for the 25 other variables when using a type III or type IV analysis. As shown in Tables 5 and 6 (Appendix), significant (p-value ≤ 0.01) variables according to both type III and type IV analyses, are indicated by double asterisks (**), significant only according to a type III analysis are indicated by only one asterisk (*) and significant only according to a type IV analysis are indicated by a hash symbol (#).

There are only two R^2 and two model F values in multi-variable models. When LGPA is the dependent variable, $R^2 = 0.937$ (N = 7500), or 93.7% of the variability in the data is described by all nine 26-variable models. Also, $R^2 = 0.865$ (N = 7500) for all nine LPW-dependent 26-variable models. These extraordinarily high R^2 values are achieved because of the large number of variables (26) in each model. Also, $F(5336, 2044) = 5.73$ is significant (p-value < 0.01) for all nine models with LGPA as the dependent variable and $F(5406, 2086) = 2.48$ is significant for all nine models with LPW as the dependent variable. Both the R^2 and model F values are identical for models 1 to 9 for LGPA because the same five variables, course

(C), instructor (I), year (Y), semester (Sm) and class time period (P), and their interactions, appear in all nine models, but where course (C) is nested in either the subject (S), department (D) or course level (L) variables and instructor (I) is nested in either the instructor gender (G), instructor job (J) or instructor qualifications (Q) variables. This is also true for models 10 to 18 with LPW as the dependent variable. Larger R^2 values for LGPA-dependent models than for LPW-dependent models indicates models fit LGPA data closer than LPW data.

Instructors and some interactions with instructors are highly significant variables in multi-variable LGPA-dependent and LPW-dependent models. Confirming both the one-variable and two-variable cases, instructor is significant (p-value < 0.01) in all 18 models in Tables 5 and 6 using both type III and IV analyses. According to marginal F statistics, both LGPA and LPW are significantly (p-value < 0.01) different for different instructors, even after controlling for all other 25 variables in each model. In model 1, for example, even controlling for course, subject, instructor level of seniority (job), year, semester, class time period and various interactions, LGPA and LPW are significantly different for different instructors (nested in either job, gender or qualifications). Significant instructor interactions occur particularly with academic year (nested in either instructor job, gender or qualifications, IxY(.), and significant in all 18 models in Tables 5 and 6 using both type III and IV analyses), but also with class time period (IxP(.), for all nine LGPA-dependent models) and to a much lesser extent with course (3 of 9 LPW-dependent models). These results indicate not only that instructor LGPAs and LPWs changed or adapted to different academic years, but also instructors' LPWs adapted to different class time periods. Further confirmation that instructors' LGPAs change for different academic years is this: although instructor, year and instructor x year variables are all significant in the two-variable I|Y LGPA model, only the instructor and instructor x year (but not year) variables are significant in the current LGPA-dependent multi-variable models. In other words, although significant differences between LGPA for different years disappear (are explained) when a greater number of variables are introduced, the interaction between instructor and year remains significant, indicating this interaction is very strong indeed.

Courses, not interactions with courses, are significant variables in multi-variable LPGA-dependent models. However, courses are not significant variables in LPW-dependent models. The LGPA is significantly (p-value < 0.01) different for different courses when nested in subject, C(S), or department C(D), but not academic course levels C(L), after controlling for all other 25 variables in each model, using both type III and IV analyses, as shown in Table 5. Furthermore, courses do not statistically significantly interact with other variables in the LGPA models and only interact with instructors in a few cases for the LPW models. This indicates course, unlike instructor, affects only LGPA independently (additively) of the other variables in the model. Whereas the effect of instructor on both LGPA and LPW in multi-variable models does not diminish when compared to instructor's effect on LGPA and LPW in two-variable models, the effect of course diminishes somewhat on LGPA and very much so on LPW when comparing multi-variable models with two-variable models.

Both LGPA and LPW are significantly (p-value < 0.01) different for different course levels, subjects and departments in all 18 multi-variable models in Tables 5 and 6 using a type III (only) analysis. Most (42 of 44) two-variable models (except L|I and L|S) in Table 4 also contain significant academic course levels, subjects and departments. In Figure 2 (Appendix), LGPA increases for increasing academic course levels, from 1.65 pre-university courses, up to 3.39 for graduate courses; the largest jump occurs between pre-university 1.65 up to 2.52 for freshman courses. Also, the highest LGPA occurs for the education department with 3.64 and the lowest are the MSP department with 2.45 and SSC department with 2.42. Also, as shown in Figure 3, the corresponding LPW for the education department is 3%, and for the MSP and SSC departments, are 14.7% and 9.2% respectively. In Figure 3 (Appendix), LPW decreases from 43.4% in pre-university courses down to 3.9% for graduate courses; the largest decrease occurs between pre-university 43.4% down to 10.8% for freshman courses.

Course levels are more significant than subjects and departments in LGPA-dependent multi-variable models. Course LGPA is significant when nested in subjects or departments, but not when nested in academic course level, L, because academic course level describes a larger portion of the model variation than does either subjects or departments relative to the course variable; for example, in model 9, further analysis reveals (significant) marginal F = 12.96 for course level, C, and (non-significant) marginal F = 1.08 for C(L), whereas in corresponding model 6, smaller (but significant) marginal F = 3.67 for department, D, and (significant) larger F = 2.19 for C(D). Basically, academic course level has a bigger impact on LGPA than either subjects or departments: there is a greater difference in LGPA related to different academic course levels than related to different subjects or departments.

Instructor is a more significant variable than gender, job and qualifications in both LPGA-dependent and LPW-dependent multi-variable models. Although both LGPA and LPW are significantly (p-value < 0.01) different for different instructors, they are not significantly different for instructor gender, job or qualifications differences in multi-variable models. For example, in model 16 in Table 6, further analysis reveals (significant) marginal F = 1.57 for instructor nested in job, I(J), which is larger than (non-significant) marginal F = 0.12 for instructor job, J. In the one-variable models in Table 3, instructor gender, job or qualifications are all significant according to F statistics (except one, the model with dependent LGW and explanatory variable job), but all belong to very loose fitting models, where $R^2 \leq 0.013$. In the two-way models of Table 4, instructor gender, job and qualifications, when paired with instructor, are all significant in the LGPA-dependent models, but only instructor job, paired with instructor, is significant in the three LPW-dependent models, according to marginal F statistics. Furthermore, pairing instructor gender, job and qualifications with instructor (models I|G, I|J, I|Q, G|I, J|I, Q|I), dramatically improves model fit, where $R^2 = 0.326$ for LPW models and where $R^2 = 0.550$ for LGPA-dependent models. Instructor describes a larger portion of model variation than instructor gender, job or qualifications. Instructors act differently, are not constrained by gender and qualifications and, to a lesser extent, job, with regard to LGPA and, to a lesser extent, LPW.

Some interactions between years, semesters and class periods are significant variables in LGPA-dependent and LPW-dependent multi-variable models. Neither LGPA nor LPW are significantly different for different years, semesters or class time periods, individually, in all 18 multi-variable models in Tables 5 and 6. Although the academic year, semester and class time period variables do not significantly affect LGPA or LPW individually, various interactions with these variables, particularly with instructor (as discussed previously), are major effects on LGPA and LPW. More than this, the interaction between year and class period, YxP, significantly (p-value < 0.01) affects LGPA and the interaction between year and semester, YxSm, significantly affects LPW in multi-variable models. In the one-variable models in Table 3, the three variables years, semesters and class periods are all significant according to F statistics, but all these variables belong to very loose fitting models, where $R^2 \leq 0.023$. In two-variable models in Table 4, the variables year, semester and class time period are mostly significant and interactions between pairs of these three variables are not significant according to marginal F statistics.

However, model fit in all these cases (models Y|Sm, Y|P, Sm|Y, Sm|P, P|Y, P|Sm) is very loose ($R^2 \leq 0.042$). Further analysis not given here reveals LGPA is less affected by period than LPW is affected by semester. On the one hand, LGPAs are higher for evening classes than for either morning or afternoon classes over the years, indicating a lack of interaction between period and year. On the other hand, LPW changes quite dramatically over the ten years of the study. It changes from a low of 6% to a high of 12%, and is greater in the first half rather than the second half of the year for the first six years of the study, before reversing in the last four years, indicating an interaction between year and semester. Overall, variables year, semester and class time period and pairwise interactions between them do not affect either LPGA or LPW as much as either course or instructor.

Other significant interactions which occur in more than one model and affect LPW include department x semester, DxSm, and department x class time period, DxP. It is clear from a figure not given here more students withdraw during the fall rather than spring semester in the business, nursing, MSP and technology departments than in the other six departments. Also, more morning class students withdraw than either afternoon class students in all departments, except in the biology and education departments.

DISCUSSION

Summary of results

Statistical analysis found clear evidence that both LGPA and LPW are significantly different for different courses and instructors, as well as subjects, departments and academic course levels, but not for instructor academic qualifications, gender, and job category, nor for academic year, academic semester and class time period. More than this, the explanatory variables considered in this study were found to better describe LGPA than LPW.

Courses and instructors, far more than any other of the other categorical variables considered in this study, affected both LGPA and LPW. Any particular course and instructor a student takes, much more than any of the other variables, determines how well they will perform in a class and whether they will withdraw or not. Course is slightly more influential than instructor when considered alone, but becomes less influential when more variables are introduced into a model describing both LGPA and LPW. Courses are less likely than instructors to interact with other variables when describing LGPA or LPW.

LGPA and LPW are also significantly different for different subjects, departments and academic course levels. Academic course level has a larger effect on LGPA than either subjects or departments: there is a greater significant difference in LGPA related to different academic course levels than related to different subjects or departments. The LPGA increases for increasing academic course levels; LPW decreases for increasing academic course levels, implying an inverse relationship between LGPA and LPW. Highest LGPA (and lowest LPW) occurs for the education department and lowest LGPA (highest LPW) occur for MSP and SSC departments.

LGPA and LPW are weakly, if at all, affected by instructor gender, job and qualification or year, semester and class time period. Not only do instructor gender, job and qualifications not affect LGPA and LPW alone, they also do not affect LGPA or LPW in interactions with other variables. Instructor describes a larger portion of model variation than instructor gender, job or qualifications. Instructors act differently, are not constrained by gender and qualifications and, to a lesser extent, job, with regard to LGPA and, to a lesser extent, LPW. In contrast, courses become less significant, as they are constrained by subject, department and academic course levels when affecting both LGPA and LPW. It should be noted that the variables year, semester and class time period, do affect LGPA or LPW when interacting with other variables, particularly, instructor. Unlike courses, instructors change or adapt their LGPA and LPW according to the circumstances such as different years or class time periods.

Implication of results

Because there is such a large difference between LGPAs and LPWs for different courses, instructors, subjects, departments and academic course levels, it is very clear that a student should be (or is) very careful when choosing courses, instructors, subjects and departments if (when) concerned about GPA and withdrawal rates. It may therefore be argued that, to be fair to students and to assess them in a reasonable way, there should be a greater consistency between the LGPAs of the different courses, instructors, subjects, departments and academic course levels.

But in what way should the institution approach this? Possibly, it might be an idea to encourage faculty to try to at least target their LGPAs to within at most two standard deviations of the average LGPA of other instructors on campus. Or, if one department has a compelling reason to believe that grades in their courses is likely to be much higher or lower, then consistency could start at the department level, where instructors are encouraged to target their LGPAs to within at most two standard deviations of the average LGPA of other instructors within their department. Possible reasons may include whether a department offers primarily service courses or freshmen-level courses, whether a department houses programs that have specific admissions standards beyond university admissions requirements, and whether the subject matter treated by the department is generally deemed more difficult for students to master (possibly because students lack prior exposure or preparation in a subject). This suggestion would serve as a guideline, not policy, because there would, of course, be many exceptions to the rule. In any case, these guidelines might simply serve as a starting point for discussion among faculty and between faculty and their institution.

It is intriguing to find that neither LGPA nor LPW are significantly different for different instructor gender, job and qualification or year, semester and class time period. In other words, the seemingly real differences particularly inherent in different instructor gender, job and qualifications are being compensated for by each instructor to give consistent LGPAs and LPWs. If consistency is possible for these variables, then why is consistency not possible for the other variables?

It is also interesting to note that the models which best accounted for LGPA also best accounted for LPW, but, having said this, LGPA, more than LPW, is effected by the explanatory variables in the study. This implies there is a lot of similarity between LGPA and LPW in the sense that the institutional-academic type variables considered in this study all provide good descriptions of both LGPA and LPW. However, there appear to be other variables outside of the model, most likely socio-economic variables, playing a role in LPW. It appears a greater number of descriptive variables are necessary to fully describe LPW than LGPA, which implies the LPW variable is a more complicated than the LGPA variable.

Future Work

Future work might involve more carefully identifying the causes and effects of local grade inflation, of significant differences in LGPA. For example, one may measure how fair or impartial students felt a course was, to determine if this influences a student enough to either avoid or be attracted to a class. Course fairness could be found on student evaluations or with student surveys. Measuring avoidance or attraction of a course might involve tracking individual students to determine if they stuck with one academic program or changed between programs. It would be interesting how these variables influenced LPW as well.

Also, it would be interesting to investigate the extent to which differences in LGPA may be based upon differences in content mastery. This would be easiest to determine within a course, across instructors. An externally developed and validated assessment instrument could be uniformly administered by all instructors teaching a course, in a pre/post-test design. Ideally, LGPA for an instructor's class should be significantly positively correlated with the pre/post-test learning gain.

Other explanatory variables could be introduced in this study, including percentage of female/male students, and age of students or entrance GPA. Local grade inflation might be measured by variables other than LGPA, such as LGPM (local grade point median) or, possibly, LGPS (local grade point skew). Other possible future work might include: assessing the impact on results of excluding class sections of students with extreme LGPA = 0.00 or LGPA = 4.00 or LPW = 1.00; comparing groups of students with perfect LGPAs against those that had less than perfect LGPA; comparing no withdrawals (LPW = 0) against those where some withdrew (LPW > 0).

REFERENCES

Bean, J. P. (1980). Dropouts and turnover: The synthesis of a causal model of student attrition. *Research in Higher Education*. 12(2): 155-87.

Bean, J. P., and Metzner, B. S. (1985). A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research*. 55(4): 485-540.

Bar, T., and Zussman, A. (2005). Quest for knowledge and pursuit of grades: Information, course selection, and grade inflation at an ivy league school. *Working Papers.* Paper 53. Downloaded January 5, 2011 from http://digitalcommons.ilr.cornell.edu/workingpapers/53

Bar, T., Kadiyali, V., and Zussman, A. (Summer 2009). Grade information and grade inflation: The Cornell experiment. *Journal of Economic Perspectives* 23(3): 93-108

Boddy, C. R. (2010). A projective technique to help understand the non-rational aspects of withdrawal and undergraduate attrition. *The Journal of the Education Research Group of Adelaide*. 1(3): 11-20.

Braxton, J. M., Brier, E. M., and Hossier, D. (1988). The effect of student problems on student withdrawal decisions: An autopsy on "Autopsy" studies. *Research in Higher Education*. 28(3): 241-253.

Braxton, J. M., Milem, J. F., and Sullivan, A. S. (2000). The effect of active learning on the college student departure process. *The Journal of Higher Education*. 71(5): 569-90.

Braxton, J. M., Jones, W. A., Hirschy, A. S., and Hartley, H. V. (Autumn 2008). The role of active learning in college student persistence. *New Directions for Teaching and*

Learning. 115: 71-83.
Breslow, N. E. (July 1996). *Generalized Linear Models: Checking Assumptions and Strengthening Conclusions*. Prepared for the Congresso Nazionale Societa" Italiana di Biometria Centro Convegni S. Agostino, Cortona. Downloaded January 5, 2011 from http://faculty.washington.edu/norm/pubs/glm.pdf.
Driscoll, M.F. and Borror, C.M. (2000). Sums of squares and expected mean squares in SAS. *Quality and Reliability Engineering International.* 16: 423-433.
The Educational Policy Committee (February 2000). *Grade Inflation at UNC—Chapel Hill; A Report to the Faculty Council*. Chapel Hill, NC. Downloaded January 15, 2011 from http://www.unc.edu/faculty/faccoun/reports/1999-00/R2000EPCGrdInfl.PDF
Eiszler, C. F. (2002). College students' evaluations of teaching and grade inflation. *Research in Higher Education*. 43(4): 483-501.
Freund, R. J. (1980). The case of the missing cell. *The American Statistician*. 34(2): 94-98.
Garson, G. D. (2010). *Generalized Linear Models and Generalized Estimating Equations*. North Carolina State University, Raleigh, NC. Downloaded March 3, 2010 from http://faculty.chass.ncsu.edu/garson/PA765/gzlm_gee.htm
Garson, G. D. (2010). *Linear Mixed Models: Hierarchical Linear, Random Effects, Multilevel, Random Coefficients, and Repeated Measures Models*. North Carolina State University, Raleigh, NC. Downloaded March 3, 2010 from http://faculty.chass.ncsu.edu/garson/PA765/multilevel.htm
Germain, M-L. and Scandura, T. A. (2005). Grade inflation and student individual differences as systematic bias in faculty evaluations. *Journal of Instructional Psychology*. 32(1): 58-67.
Gose, B. (July 25, 1997). Efforts to curb grade inflation get an F from critics. *The Chronicle of Higher Education.* 43: A41-2.
Grove, W. A. and Wasserman, T. (2004). The life-cycle pattern of collegiate GPA: Longitudinal cohort analysis and grade inflation. *The Journal of Economic Education*. 35(2): 162-174.
Hausmann, L. R. M., Schofield, J. W., and Woods, R. L. (2007). Sense of belonging as a predictor of intentions to persist among african american and white first-year college students. *Research in Higher Education*. 48(7): 803-839.
Ishitani, T. (2006). Studying attrition and degree completion behavior and first-generation college students in the United States. *The Journal of Higher Education*. 77(5): 861- 885.
Johnson, V. E. (2003). *Grade Inflation: A Crisis in College Education*. New York: Springer-Verlag.
Kember, D., and Harper, G. (1987). Implications for instruction arising from the relationship between approaches to studying and academic outcomes. *Instructional Science*. 16: 35-46.
Kendall, M., Stuart, A. and Ord, J. K. (1983). *The Advanced Theory of Statistics, Volume 3, 4th Ed*. Charles Griffin & Company Limited, London & High Wycombe.
Kohn, A. (November 8, 2002). The dangerous myth of grade inflation. *The Chronicle of Higher Education*. 49(11): p B7-9.
Kolevzon, M. S. (1981). Grade inflation in higher education: A comparative study. *Research in Higher Education.* 15(3): 195-212.
Kuh, G. D., and Hu, S. (1999). Unraveling the complexity of the increase in college grades from the mid-1980s to the mid-1990s. *Educational Evaluation and Policy Analysis*. 21: 297-320.
Kuter, M. H., Nachtsheim, C. J., Neter, J. and Li, W. (1996). *Applied Linear Statistical Models.*

5th Ed. New York: McGraw-Hill Irwin.
Langsrud, O. (2003). ANOVA for unbalanced data: Use type II instead of Type III sums of squares. *Statistics and Computing*. 13: 163–167.
Littell, R. C., Freund, R. J. and Spector, P. C. (1991). *SAS System for Linear Models, 3rd Ed.* SAS Incorporated, Cary, NC.
Manhire, B. (2005). *Grade inflation in engineering education at Ohio University.* Proceedings of the 2005 American Society for Engineering Education Annual Conference & Exposition, Session 3460, Portland, OR. Downloaded January 5, 2011 from http://sdsu-physics.org/sdsu_per/articles/GradeInflation.pdf.
McSpirit, S., Jones, K., Chapman, A. D., and Kopacz, P. (June 2000a). Identifying grade inflation at an open-admissions institution. *College Student Journal*. 34(2): 228-35.
McSpirit, S., Jones, K., Chapman, A. D., and Kopacz, P. (June 2000b). Faculty ironies on grade inflation. *Journal of Instructional Psychology*. 27(2): 104-9
Miller, R. G. (1986). *Beyond ANOVA, Basics of Applied Statistics*. New York: John Wiley & Sons.
Napoli, A. R., and Wortman, P. M. (1998). Psychosocial factors related to retention and early departure of two-year community college students. *Research in Higher Education*. 39(4): 419-455.
Nelson, J. P., and Lynch, K. A. (1984). Grade inflation, real income, simultaneity, and teaching evaluations. *The Journal of Economic Education*. 15(1): 21-37.
Pascarella, E. T. (1980). Student-faculty informal contact and college outcomes. *Review of Educational Research.* 50(4): 545-595.
Pascarella, E. T., Duby, P. B., and Iverson, B. K. (1983). A text and reconceptualization of a theoretical model of college withdrawal in a commuter institution setting. *American Sociological Association*. 56(2): 88-100.
Pascarella, E. T., Terenzini, P. T., and Wolfle, L. M. (1986). Orientation to college and freshman year persistence/withdrawal decisions. *The Journal of Higher Education*. 57(2): 155-175.
Pendleton, O. J., Tress, M. V. and Bremer, R. (1986). Interpretation of the four types of analysis of variance tables in SAS. *Communications in Statistics - Theory and Methods*. 15(9): 2785-2808.
Perry, R. P., Hladkyj, S., Pekrun, R. H., Clifton, R. A., and Chipperfield, J. G. (2005). Perceived academic control and failure in college students. *Research in Higher Education*. 46(5): 535-569.
Prather, J. E., Smith, G., and Kodras, J. E. (1979). A longitudinal study of grades in 144 undergraduate courses. *Research in Higher Education*. 10(1): 11-24.
Remmers, H. H. (1930). To what extent do grades affect student ratings of instructors? *Journal of Educational Research*. 21: 314-316.
Rosovsky, H., and Hartley, M. (2002). *Evaluation and The Academy: Are We Doing the Right Thing? Grade Inflation and Letters of Recommendation*. American Academy of Arts and Sciences, 136 Irving Street, Cambridge, MA 02138. Downloaded January 5, 2011, from http://www.amacad.org/publications/monographs/Evaluation_and_the_Academy.pdf
Sabot, R., and Wakeman-Linn, J. (Winter 1991). Grade inflation and course choice. *Journal of Economic Perspectives.* 5(1): 159-170.
Searle, S. R. (1994). Analysis of variance computing package output for unbalanced data from fixed-effects models with nested factors. *The American Statistician*. 48(2): 148-153.

Shoichet, C. E. (July 12, 2002). Reports of grade inflation may be inflated, study finds. *The Chronicle of Higher Education.* 48(44).

Simmons-Welburn, J., and Welburn, W. C. (2001). Cultivating partnerships/realizing diversity. *Journal of Library Administration*. 33(1/2): 5-19.

Spady, W. G. (1970). Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange*. 1: 64-85.

SPSS Advanced Statistics 17.0. (2007). Chicago, IL. Downloaded January 7th, 2011 from http://www.hks.harvard.edu/fs/pnorris/Classes/A%20SPSS%20Manuals/SPSS%20Advanced%20Statistics%2017.0.pdf.

Stone, J. E. (1995). Inflated grades, inflated enrollment, and inflated budgets: An analysis and call for review at the state level. *Education Policy Analysis Archives*. 3(11): 1-30.

Stupnisky, R. H., Renaud, R. D., Perry, R. P., Ruthig, J. C., Haynes, T. L., and Clifton, R. A. (2007). Comparing self-esteem and perceived control as predictors of first-year college students' academic achievement. *Social Psychology of Education*. 10: 303-330.

Thomas, S. L. (2000). Ties that bind: A social network approach to understanding student integration and persistence. *The Journal of Higher Education*. 71(5): 591-615.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research* 45: 89-125.

Yates, F. (1934). The analysis of multiple classifications with unequal numbers in the different classes. *Journal of the American Statistical Association*. 29(185): 51-66.

Walton, I., Cousin, A., Davison, D., Echevarri, A., Koller, B., Mohon, R., Murillo, A., and Takahashi, J. (Spring 2008). *Promoting Thoughtful Faculty Conversations About Grade Distributions*. The Academic Senate for California Community Colleges. Sacramento, CA. Downloaded January 7, 2011, from http://www.asccc.org/node/175018.

APPENDIX

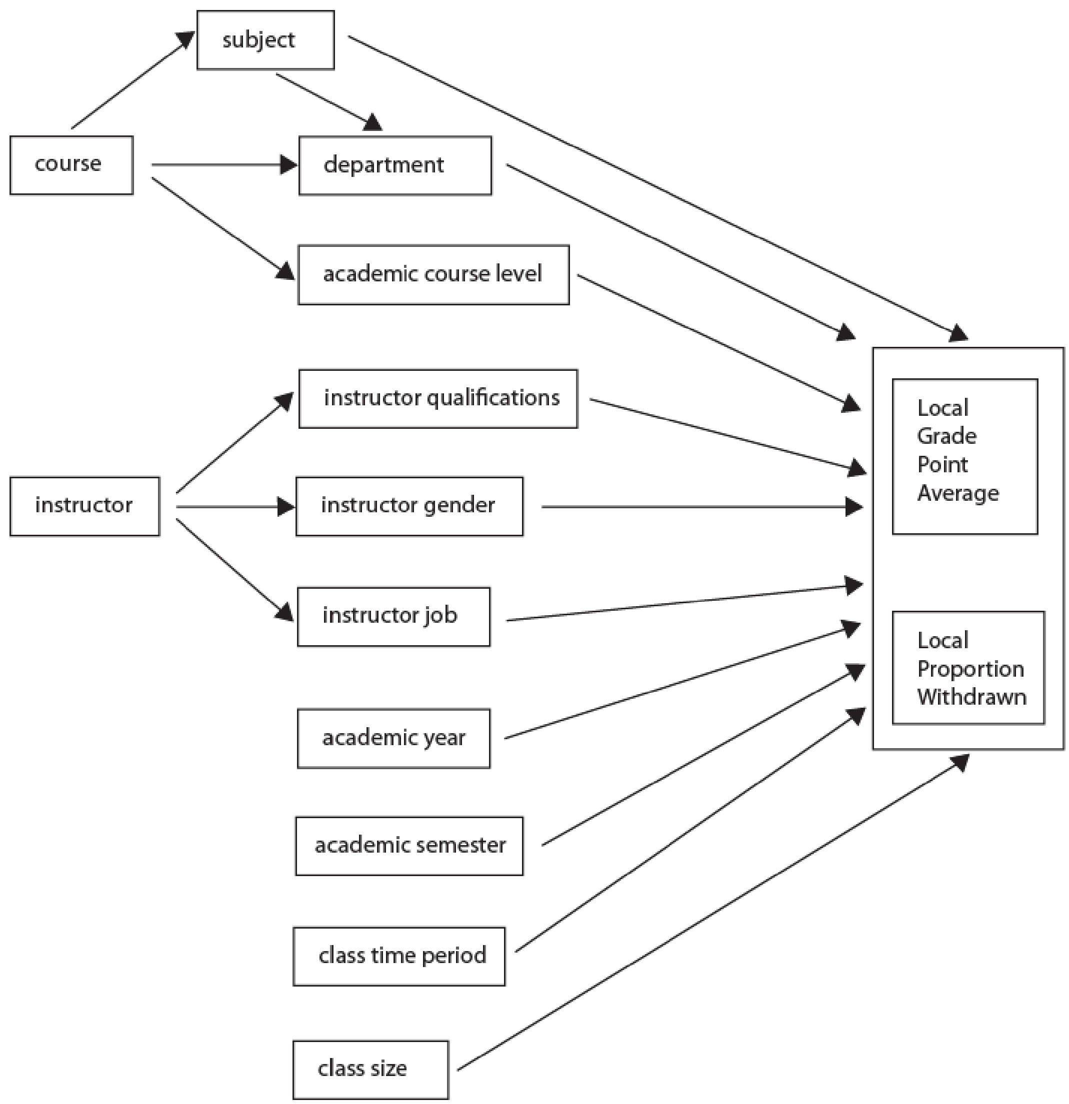

FIGURE 1. Flow chart of variables for model of grade distribution records

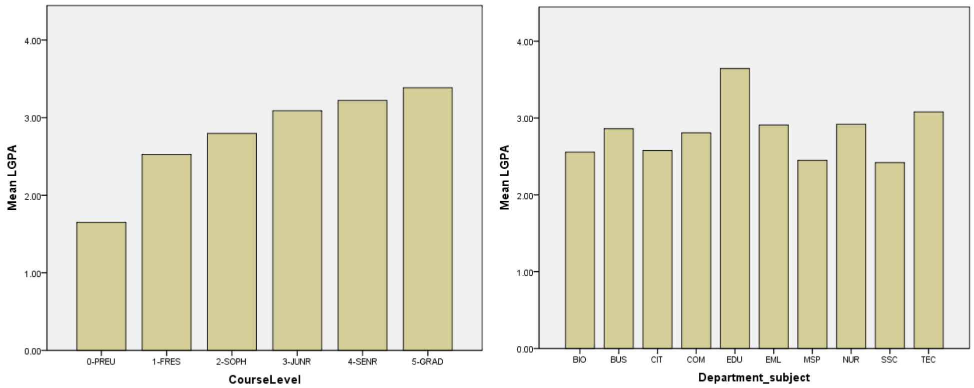

FIGURE 2. LGPAs for course level and department, weighted by class size without Ws or Is

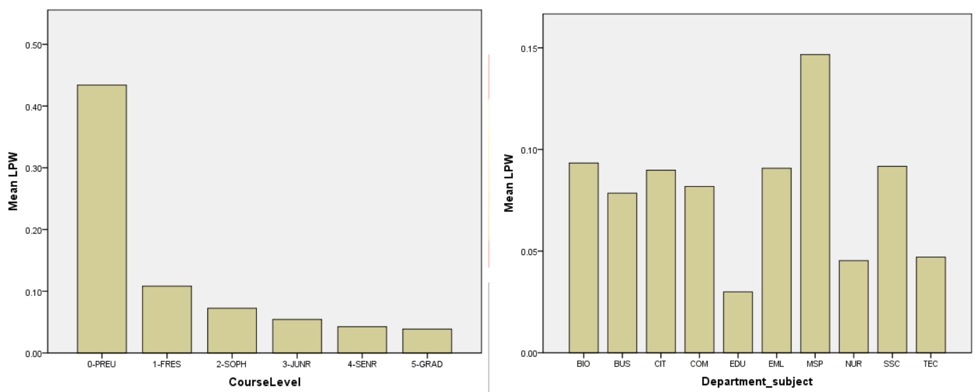

FIGURE 3. LPWs for course level and department, weighted by class size with Ws

Rec	Instr	Course	Sec	Subj	Depart	Year	Sem	A	B	C	D	F	No	W	I	No+W	LGPA	LPW
1	111	MA 223	001	Math	MSP	05-06	F	3	11	13	1	0	28	1	0	29	2.57	0.03
2	093	MA 223	002	Math	MSP	05-06	F	3	5	5	5	0	18	6	0	24	2.33	0.25
3	093	MA 223	003	Math	MSP	05-06	F	1	5	4	3	1	14	3	0	17	2.14	0.18
4	111	STAT 213	001	Stat	MSP	05-06	F	4	11	12	1	1	29	3	1	32	2.55	0.09
5	111	MA 223	001	Math	MSP	05-06	S	5	6	9	1	1	22	0	0	22	2.59	0.00
6	005	STAT 301	001	Stat	MSP	05-06	S	7	8	6	4	3	28	3	0	31	2.43	0.10
7	056	STAT 301	002	Stat	MSP	05-06	S	7	7	7	4	1	26	3	0	29	2.58	0.10
8	031	PSY 201	001	Psy	SSC	05-06	S	3	5	7	1	0	16	1	0	17	2.63	0.06

TABLE 1. Example grade distribution records data of eight class sections

Dependent → Lambda, λ Cramer's V Independent ↓	I	G	J	Q		C	S	D	L		Y	Sm	P
Instructor (I)		1 1	1 1	1 1		0.419 0.575	0.833 0.739	0.986 0.991	0.440 0.594		0.102 0.299	0.051 0.160	0.336 0.521
Gender (G)	0.011 1		0.000 0.160	0.026 0.186		0.000 0.750	0.038 0.574	0.027 0.447	0.000 0.097		0.000 0.069	0.000 0.000	0.000 0.107
Job (J)	0.050 1	0.017 0.160		0.298 0.382		0.022 0.695	0.047 0.454	0.059 0.248	0.020 0.160		0.018 0.132	0.000 0.030	0.041 0.132
Qualifications (Q)	0.032 1	0.011 0.186	0.133 0.382			0.013 0.689	0.026 0.447	0.053 0.283	0.006 0.127		0.000 0.056	0.000 0.010	0.013 0.128
Course (C)	0.537 0.575	0.576 0.750	0.549 0.695	0.579 0.689			1 1	1 1	1 1		0.093 0.296	0.288 0.478	0.316 0.562
Subject (S)	0.268 0.739	0.370 0.574	0.267 0.454	0.348 0.447		0.334 1		1 1	0.364 0.582		0.026 0.146	0.013 0.092	0.169 0.313
Department (D)	0.090 0.991	0.267 0.447	0.110 0.248	0.207 0.283		0.131 1	0.454 1		0.183 0.303		0.001 0.043	0.000 0.037	0.115 0.221
Academic Course Level (L)	0.024 0.594	0.000 0.097	0.055 0.160	0.070 0.127		0.044 1	0.078 0.582	0.126 0.303			0.003 0.046	0.000 0.034	0.013 0.078
Year (Y)	0.005 0.299	0.000 0.069	0.057 0.132	0.001 0.056		0.002 0.296	0.006 0.146	0.001 0.043	0.000 0.046			0.035 0.265	0.017 0.062
Semester (Sm)	0.000 0.160	0.000 0.000	0.000 0.030	0.000 0.010		0.000 0.478	0.000 0.092	0.000 0.037	0.000 0.034		0.000 0.265		0.000 0.011
Period (P)	0.009 0.521	0.000 0.107	0.044 0.132	0.023 0.128		0.000 0.562	0.014 0.313	0.008 0.221	0.000 0.078		0.000 0.062	0.000 0.011	

TABLE 2. (Symmetric) Cramer's V and (dependent) Lambda measures of association between explanatory variables

Dependent —› R^2 F, degrees of freedom	LGPA Type III and IV	LPW Type III and IV
Instructor	0.550 25.57*, 337, 7043	0.326 10.18*, 340, 7152
Instructor Gender	0.012 87.44*, 1, 7379	0.002 13.88*, 1, 7491
Instructor Job	0.011 16.78*, 5, 7375	0.012 18.00*, 5, 748
Instructor Qualifications	0.013 33.23*, 3, 7377	0.001 3.44*, 3, 7489
Course	0.615 16.05*, 667, 6713	0.500 10.14*, 672, 6820
Subject	0.373 64.86*, 67, 7313	0.224 31.55*, 68, 7424
Department	0.217 226.75, 9, 7371	0.095 87.71*, 9, 7483
Academic Course Level	0.144 248.76*, 5, 7375	0.153 269.96*, 5, 7487
Year	0.023 19.03*, 9, 7371	0.014 11.83*, 9, 7483
Semester	0.001 7.38*, 1, 7379	0.000 2.12, 1, 7491
Period	0.017 63.83*, 2, 7380	0.005 17.40*, 2, 7490

TABLE 3. Various R^2 and F values (F significant, p-value < 0.01, if starred*) with degrees of freedom, weighted by sample size, for type III (and IV) ANOVA model analyses

LGPA → R^2 model fit III A,B,AB IV A,B,AB LPW ↓	I	G	J	Q	C	S	D	L	Y	Sm	P
Instructor (I)		0.550 I(G),G	0.550 I(J),J	0.550 I(Q),Q	0.784 I,C,IxC I,C,IxC	0.610 I,S,IxS I,S,IxS	0.560 I,D,- I,D,-	0.691 I,L,IxL I,L,IxL	0.679 I,Y,IxY I,Y,IxY	0.571 I,-,-	0.633 I,P,IxP I,P,IxP
Instructor Gender (G)	0.326 -,I(J)		0.029 G.J,GxJ	0.025 -,Q,GxQ	0.519 -,C,GxC -,C,GxC	0.245 -,S,GxS -,S,GxS	0.128 G,D,GxD	0.217 G,L,GxL	0.035 G,Y,-	0.013 G,-,-	0.034 G,P,GxP
Instructor Job (J)	0.326 J,I(J)	0.016 J,G,JxG		0.046 J,Q,JxQ J,Q,JxQ	0.561 -,C,JxC -,C,JxC	0.403 J,S,JxS J,S,JxS	0.127 J,D,JxD J,D,JxD	0.289 J,L,JxL -,L,JxL	0.040 J,Y,- J,Y,-	0.013 J,Sm,-	0.039 J,P,JxP
Instructor Qualifications (Q)	0.326 -,I(J)	0.007 -,-,QxG	0.025 -,J,QxJ -,J,QxJ		0.527 -,C,QxC -,C,QxC	0.269 -,S,QxS -,S,QxS	0.110 Q,D,QxD Q,D,QxD	0.179 Q,L,QxL Q,L,QxL	0.044 Q,Y,QxY	0.015 Q,-,-	0.030 Q,P,QxP
Course (C)	0.629 C,I,CxI C,I,CxI	0.637 C,G,-	0.670 C,J,CxJ C,-,CxJ	0.660 C,Q,CxQ C,Q,CxQ		0.614 C(S),S	0.615 C(D),D	0.615 C(L),L	0.756 C,Y,- C,-,-	0.632 C,-,- C,-,-	0.657 C,-,CxP C,P,CxP
Subject (S)	0.512 S,I,SxI S,I,SxI	0.391 S,G,- S,G,-	0.455 S,-,SxJ S,J,SxJ	0.421 S,-,SxQ S,-,SxQ	0.500 S,C(S)		0.373 S(D),D	0.464 S,L,SxL S,L,SxL	0.437 S,-,SxY S,-,SxY	0.378 S,-,- S,-,-	0.408 S,-,SxP S,-,SxP
Department (D)	0.337 D,I,- D,I,-	0.231 D,-,DG	0.288 D,J,DxJ D,J,DxJ	0.248 D,Q,DxQ D,Q,DxQ	0.500 D,C(D)	0.224 D,S(D)		0.332 D,L,DxL D,L,DxL	0.248 D,Y,DxY	0.219 D,S,DxSm	0.245 D,P,DxP
Academic Course Level (L)	0.523 L,I,LxI -,I,LxI	0.154 L,-,- L,-,-	0.183 L,J,LxJ L,J.LxJ	0.174 L,-,LxQ L,-,LxQ	0.500 L,C(L)	0.271 L,S,- -,S,-	0.214 L,D,LxD L,D,LxD		0.165 L,-,-	0.146 L,-,-	0.154 L,-,LxP L,-,LxP
Year (Y)	0.482 Y,I,YxI Y,I,YxI	0.017 Y,G,-	0.031 Y,J,-	0.019 Y,-,-	0.646 -,C,- -,C,-	0.279 Y,S,- Y,S,-	0.121 Y,D,YD	0.169 -,L,-		0.025 Y,Sm,- Y,Sm,-	0.042 Y,P,-
Semester (Sm)	0.361 -,I,SmxI -,I,SmxI	0.002 -,G,-	0.014 -,J,-	0.002 -,Q,-	0.520 -,C,- -,C,-	0.239 -,S,SmxS -.S,SmxS	0.100 -,D,SmxD	0.157 Sm,L,SmxL	0.018 Sm,Y,- Sm,Y,-		0.018 Sm,-,-
Period (P)	0.379 -,I,PxI -,I,PxI	0.007 P,G,-	0.022 P,J,PxJ	0.008 P,Q,-	0.541 -,C,- -,C,-	0.242 -,S,PxS -,S,PxS	0.103 P,D,PxD	0.188 P,L,PxL	0.019 P,Y,-	0.005 -,Sm,-	

TABLE 4. Type III and IV ANOVA of two-variable model analyses, weighted by sample size, for LGPA and LPW

LGPA Models										
1	C(S)**	S*	I(J)**	J	Y	Sm	P			
	CxJ(S)	CxY(S)	CxSm(S)	CxP(S)	IxS(J)	IxY(J)**	IxSm(J)	IxP(J)**	CxI(SxJ)	
	SxJ	SxY	SxSm	SxP	JxY	JxSm	JxP	YxSm	YxP**	SmxP
2	C(S)**	S*	I(Q)**	Q	Y	Sm	P			
	CxQ(S)	CxY(S)	CxSm(S)	CxP(S)	IxS(Q)	IxY(Q)**	IxSm(Q)	IxP(Q)**	CxI(SxQ)	
	SxQ	SxY	SxSm	SxP	QxY	QxSm	QxP	YxSm	YxP**	SmxP
3	C(S)**	S*	I(G)**	G	Y	Sm	P			
	CxG(S)	CxY(S)	CxSm(S)	CxP(S)	IxS(G)	IxY(G)**	IxSm(G)	IxP(G)**	CxI(SxG)	
	SxG	SxY	SxSm	SxP	GxY	GxSm	GxP	YxSm	YxP**	SmxP
4	C(D)**	D*	I(J)**	J	Y	Sm	P			
	CxJ(D)	CxY(D)	CxSm(D)	CxP(D)	IxD(J)	IxY(J)**	IxSm(J)	IxP(J)**	CxI(DxJ)	
	DxJ	DxY	DxSm	DxP	JxY	JxSm	JxP	YxSm	YxP**	SmxP
5	C(D)**	D*	I(Q)**	Q	Y	Sm	P			
	CxQ(D)	CxY(D)	CxSm(D)	CxP(D)	IxD(Q)	IxY(Q)**	IxSm(Q)	IxP(Q)**	CxI(DxQ)	
	DxQ	DxY	DxSm	DxP	QxY	QxSm	QxP	YxSm	YxP**	SmxP
6	C(D)**	D*	I(G)**	G	Y	Sm	P			
	CxG(D)	CxY(D)	CxSm(D)	CxP(D)	IxD(G)	IxY(G)**	IxSm(G)	IxP(G)**	CxI(DxG)	
	DxG	DxY	DxSm	DxP	GxY	GxSm	GxP	YxSm	YxP**	SmxP
7	C(L)	L*	I(J)**	J	Y	Sm	P			
	CxJ(L)	CxY(L)	CxSm(L)	CxP(L)	IxL(J)	IxY(J)**	IxSm(J)	IxP(J)**	CxI(LxJ)	
	LxJ	LxY	LxSm	LxP	JxY	JxSm	JxP	YxSm	YxP**	SmxP
8	C(L)	L*	I(Q)**	Q	Y	Sm	P			
	CxQ(L)	CxY(L)	CxSm(L)	CxP(L)	IxL(Q)	IxY(Q)**	IxSm(Q)	IxP(Q)**	CxI(LxQ)	
	LxQ	LxY	LxSm	LxP	QxY	QxSm	QxP	YxSm	YxP**	SmxP
9	C(L)	L*	I(G)**	G	Y	Sm	P			
	CxG(L)	CxY(L)	CxSm(L)	CxP(L)	IxL(G)	IxY(G)**	IxSm(G)	IxP(G)**	CxI(LxG)	
	LxG	LxY	LxSm	LxP	GxY	GxSm	GxP	YxSm	YxP**	SmxP

TABLE 5. Type III and IV ANOVA of 26-variable linear models with dependent variable LGPA and listed explanatory variables, weighted by class size (without Ws or Is)

LPW Models										
10	C(S)	S*	I(J)**	J	Y	Sm	P			
	CxJ(S)	CxY(S)	CxSm(S)	CxP(S)	IxS(J)	IxY(J)**	IxSm(J)	IxP(J)	CxI(SxJ)	
	SxJ	SxY	SxSm	SxP	JxY	JxSm#	JxP	YxSm**	YxP	SmxP
11	C(S)	S*	I(Q)**	Q	Y	Sm#	P			
	CxQ(S)	CxY(S)	CxSm(S)	CxP(S)	IxS(Q)	IxY(Q)**	IxSm(Q)	IxP(Q)	CxI(SxQ)**	
	SxQ	SxY	SxSm	SxP	QxY	QxSm#	QxP	YxSm**	YxP	SmxP
12	C(S)	S*	I(G)**	G	Y	Sm	P			
	CxG(S)	CxY(S)	CxSm(S)	CxP(S)	IxS(G)	IxY(G)**	IxSm(G)	IxP(G)	CxI(SxG)	
	SxG**	SxY	SxSm	SxP	GxY	GxSm	GxP	YxSm**	YxP	SmxP
13	C(D)	D*	I(J)**	J	Y	Sm	P			
	CxJ(D)	CxY(D)	CxSm(D)	CxP(D)	IxD(J)	IxY(J)**	IxSm(J)	IxP(J)	CxI(DxJ)	
	DxJ?	DxY*	DxSm**	DxP#	JxY	JxSm#	JxP	YxSm**	YxP	SmxP
14	C(D)	D*	I(Q)**	Q	Y	Sm#	P			
	CxQ(D)	CxY(D)	CxSm(D)	CxP(D)	IxD(Q)?	IxY(Q)**	IxSm(Q)	IxP(Q)	CxI(DxQ)**	
	DxQ	DxY	DxSm**	DxP**	QxY	QxSm#	QxP	YxSm**	YxP	SmxP
15	C(D)*	D*	I(G)**	G	Y	Sm	P			
	CxG(D)	CxY(D)	CxSm(D)	CxP(D)	IxD(G)	IxY(G)**	IxSm(G)	IxP(G)	CxI(DxG)**	
	DxG	DxY	DxSm**	DxP**	GxY	GxSm	GxP	YxSm**	YxP	SmxP
16	C(L)	L*	I(J)**	J	Y	Sm	P			
	CxJ(L)	CxY(L)	CxSm(L)	CxP(L)	IxL(J)	IxY(J)**	IxSm(J)	IxP(J)	CxI(LxJ)	
	LxJ	LxY	LxSm	LxP#	JxY	JxSm#	JxP	YxSm**	YxP	SmxP
17	C(L)	L*	I(Q)**	Q	Y	Sm#	P			
	CxQ(L)	CxY(L)	CxSm(L)	CxP(L)	IxL(Q)	IxY(Q)**	IxSm(Q)	IxP(Q)	CxI(LxQ)	
	LxQ	LxY	LxSm	LxP#	QxY	QxSm#	QxP	YxSm**	YxP	SmxP
18	C(L)	L*	I(G)**	G	Y	Sm	P			
	CxG(L)	CxY(L)	CxSm(L)	CxP(L)	IxL(G)	IxY(G)**	IxSm(G)	IxP(G)	CxI(LxG)	
	LxG	LxY	LxSm	LxP#	GxY	GxSm	GxP	YxSm**	YxP	SmxP

TABLE 6. Type III and IV ANOVA of 26-variable models with dependent variable LPW and listed explanatory variables, weighted by class size (with Ws)