

The Relationship Between Grades and Academic Program Satisfaction Over Four Years of Study

J. PAUL GRAYSON

York University

ABSTRACT

It is frequently assumed that the student experience, and, by implication, student program satisfaction, improves over the course of a university education. A four-year panel study of students at a large commuter university indicates some improvements in assessments of professor performance and GPA between first and fourth year; however, satisfaction with academic programs remains more or less the same across all four years of study.

Structural equation modelling was employed to estimate the relationships among professor performance, GPA, and program satisfaction within, and between, each of the four years of study. Contrary to expectations based on some conventional models, it was found that students' assessments of professors were not affected by GPA; conversely, professor performance had little impact on GPA. By contrast, student satisfaction was related to both GPA and professor performance. The greatest predictor of students' program satisfaction, however, was neither GPA nor professor performance, but program satisfaction in the previous year. This finding suggests that underlying

personality characteristics likely are more responsible for expressions of program satisfaction than either GPA or professor performance.

RÉSUMÉ

On suppose souvent que l'expérience de l'étudiant et dès lors, sa satisfaction à l'égard du programme, s'améliore au cours de ses études universitaires. Sur une période de quatre, l'étude réalisée auprès d'un groupe d'étudiants d'une grande université à population majoritairement externe montre une certaine amélioration de l'évaluation de la performance des professeurs et de la MPC entre la première et la quatrième année, mais la satisfaction à l'égard du programme universitaire reste à peu près la même pendant les quatre années de l'étude.

Nous avons utilisé la modélisation par équation structurelle pour estimer le rapport entre la performance du professeur, la MPC et la satisfaction à l'égard du programme dans les quatre années d'études et d'une année à l'autre. Contrairement aux attentes fondées sur quelques modèles conventionnels, nous avons constaté que la MPC n'influçait pas la cote donnée au professeur par l'étudiant; réciproquement, la performance du professeur avait peu d'effet sur la MPC. La satisfaction de l'étudiant, en revanche, était liée à la fois à la MPC et à la performane du professeur. Le meilleur prédicteur de la satisfaction d'un étudiant à l'égard du programme, toutefois, n'était ni la MPC ni la performance du professeur, mais la satisfaction exprimée à l'égard du programme l'année précédente. Cette conclusion suggère que les traits de personnalité sous-jacents sont probablement plus responsables de la satisfaction exprimée à l'égard du programme que la MPC et la performance du professeur.

INTRODUCTION

It is a common assumption in many universities that the quality of the student experience improves between first year and graduation. For one thing, because of decreasing student faculty ratios, the possibility of having small classes and contact with professors increases as students make their way up the academic ladder. For another, students become increasingly more knowledgeable of university folkways and, as a result, can make their way through the system with increasing competence. The limited amount of research that has been carried out on changes in students' experiences gives cautious support to assumptions such as these (Terenzini & Wright, 1987). As a result, it is reasonable to expect changes in students' satisfaction with various aspects of their education between first and final year.

While student satisfaction can be examined at many levels (Beltyukova & Fox, 2002), the focus of this article will be on satisfaction with *academic programs* in a large commuter university. In commuter settings, students are less likely than those in residential universities to be involved in various informal academic and social activities that contribute to the realization of desired educational outcomes. As a result, what goes on in the classroom is more important in commuter than in residential universities for the attainment of educational outcomes, and, potentially, for program satisfaction. Bearing these possibilities in mind, the focus of this study is on changes in program satisfaction between first and fourth years, and on the extent to which an integration of various models of the relationship between academic achievement and teaching effectiveness contributes to an understanding of academic program satisfaction.

INTEGRATION / INVOLVEMENT

Various studies have established that educational outcomes can be related to what occurs within the classroom as well as to academic and social integration/involvement in the wider university (Astin, 1993; Grayson, 1995, 2003; Pascarella, Edison, Nora, Hagedorn, & Terenzini, 1996; Terenzini, Springer, Pascarella, & Nora, 1995; Terenzini, Springer, Yaeger, Pascarella, & Nora, 1996; Tinto, 1993). In commuter institutions, however, academic integration, particularly in the classroom, makes a greater contribution than social integration to desired outcomes (Grayson, 1997, 1998). Consistent with this observation, Tinto (1997, p. 559) noted that, "for students who commute to college, especially those who have multiple obligations outside the college, the classroom may be the only place where students and faculty meet, where education in the formal sense is experienced."

While relatively little time has been spent on examinations of the degree to which academic and social integration/involvement have consequences for student satisfaction, one examination of undergraduate students that utilized some integration/involvement variables found that the degree of "fit" between the student and the university, academic integration, program utility, and having a positive social life all contributed to student satisfaction (Bean & Bradley, 1986). A study of changes in satisfaction over an academic term also discovered that students who were most involved in the college environment were more satisfied than others (Pennington, Zvonkovic, & Wilson, 1989). In addition to experiences that can be classified as forms of academic and social integration/involvement, studies have found that what goes on in broader aspects of students' lives, such as their romantic involvements and living arrangements, have consequences for general satisfaction levels (Benjamin & Hollings, 1995, 1997). In addition to the possibility that students' satisfaction with various aspects of their university experiences can be linked to more general theories focusing on the consequences of academic and

social involvement, it is important to note that there are “discrepancy models,” “cognitive models,” and “legitimation models” of student satisfaction per se (Benjamin & Hollings, 1997).

MODELS

There are three general models focusing on the relationship between teaching and grades that are relevant to the current analysis of grades and academic program satisfaction. In the “grading leniency bias model” it is assumed that students who get high grades give high ratings to the performance of their professors on teaching evaluations. As a result, instructors who give unwarrantedly high grades receive unwarrantedly high evaluations (Krautmann & Sander, 1999; Greenwald, 1995). By extension, students who get high grades also should be satisfied with their academic programs. Because of this possibility, teaching evaluations are viewed as potentially biased, and invalid measures of teaching effectiveness. The “teaching effectiveness model” assumes that students who learn more get high grades and give high evaluations to their instructors. By extension, they will also express satisfaction with academic programs. Within this model, teaching evaluations are viewed as valid measures of teaching performance.

In the “student characteristics model” it is postulated that certain student characteristics, like high motivation, result in greater learning and, as a result, high evaluations of teacher performance (Cashin, 1995; Howard & Maxwell, 1980; Marsh & Dunkin, 1992) and program satisfaction. In addition to student motivation, reasons for taking the course and expected grades are among student characteristics that can affect teaching evaluations. In general, higher ratings are given by students taking courses out of general interest or as an elective. Relatively low ratings are received in courses taken as a major requirement or in fulfilment of general education requirements. Although there are exceptions, studies have found that a student’s age and gender have no affect on teaching evaluations (Cashin, 1995).

Personality also falls within the student characteristics category; however, it is important to note that some researchers have found it to be unrelated to student ratings (Abrami, Perry, & Leventhal, 1982). In contrast to environmental variables, other researchers have found personality to explain relatively little of the variance in student satisfaction (Witt & Handal, 1984). By comparison, personality has been related to satisfaction with relatively unstructured course contexts. Students most satisfied with an unstructured learning environment are relatively forceful, persevering, dependable, conscientious, adaptable, and curious (Strom & Hocevar, 1982). Among students enrolled in telecourses it has been found that those most satisfied with instruction are “mature and humble, yet venturesome and outgoing” (Biner et al., 1997, p. 29).

Although student-centered inquiries suggest limited connections between personality and teaching evaluations and student satisfaction, research into the link between personality and job satisfaction has yielded findings that might be applicable to student satisfaction with various aspects of the university experience. For example, one study found that “core evaluations” of the self have effects on job satisfaction that are independent of the objective nature of the job. In essence, some people put the best light on their jobs, independent of the objective nature of their work conditions, whereas other individuals do the reverse. It is reasoned that people who believe themselves to be worthy and who are capable of coping with life’s problems bring a “positive frame” to their daily lives, including their jobs, whereas those viewing themselves as unworthy apply a “negative frame” to the same situations. The personality characteristics contributing most to positive evaluations of the environment are self-esteem and self-efficacy (Judge, Kluger, Locke, & Durham, 1998). Other research has shown that “negative affectivity” can have implications for expressions of job satisfaction (Abraham, 1999).

In view of the research on the relationship between personality and job satisfaction, it is reasonable to postulate that students with a

positive frame would be more inclined to give positive evaluations to their professors, and be more satisfied than others with various aspects of the university experience, than students with a negative frame. In essence, all else being equal, because of their personalities, some students would be more positive about their professors and more satisfied with their university experiences, than other students.

As the vast majority of studies based on the three models outlined above assume the validity of teaching evaluations, it is important to indicate that a considerable body of evidence supports this assumption (Abrami, Cohen, & d'Apollonia, 1988; Cohen, 1981; Marsh & Dunkin, 1992). It is also important to note conflicting findings on the relationship between year of study and teaching evaluations (Cashin, 1995; Feldman, 1978; Murray, Rushton, & Paunonen, 1990; Marsh & Dunkin, 1992). It is generally found, however, that evaluations in science are low (Marsh & Dunkin, 1992; Feldman, 1978). In terms of satisfaction, some evidence indicates that despite the possibility of smaller classes and more rewarding experiences, upperclassmen express less satisfaction than students early in their careers. The explanation given for this possibility is that as they proceed toward graduation, students become more fixed on their futures than in the early years of study. As a result, their focus on satisfaction with the present declines (Bean & Bradley, 1986).

Hypothesized Model

Although the grading leniency bias, teaching effectiveness, and student characteristics models were not designed for examinations of program satisfaction, in this analysis, using structural equation modeling, the assumptions underlying the models will be used in a longitudinal examination of the degree to which program satisfaction is related to professors' classroom performance and GPA. More specifically, the assumptions of each of the grading leniency bias and teaching effectiveness models will be combined with assumptions underlying the student characteristics model for independent samples of Arts and Pure and Applied Science Students.

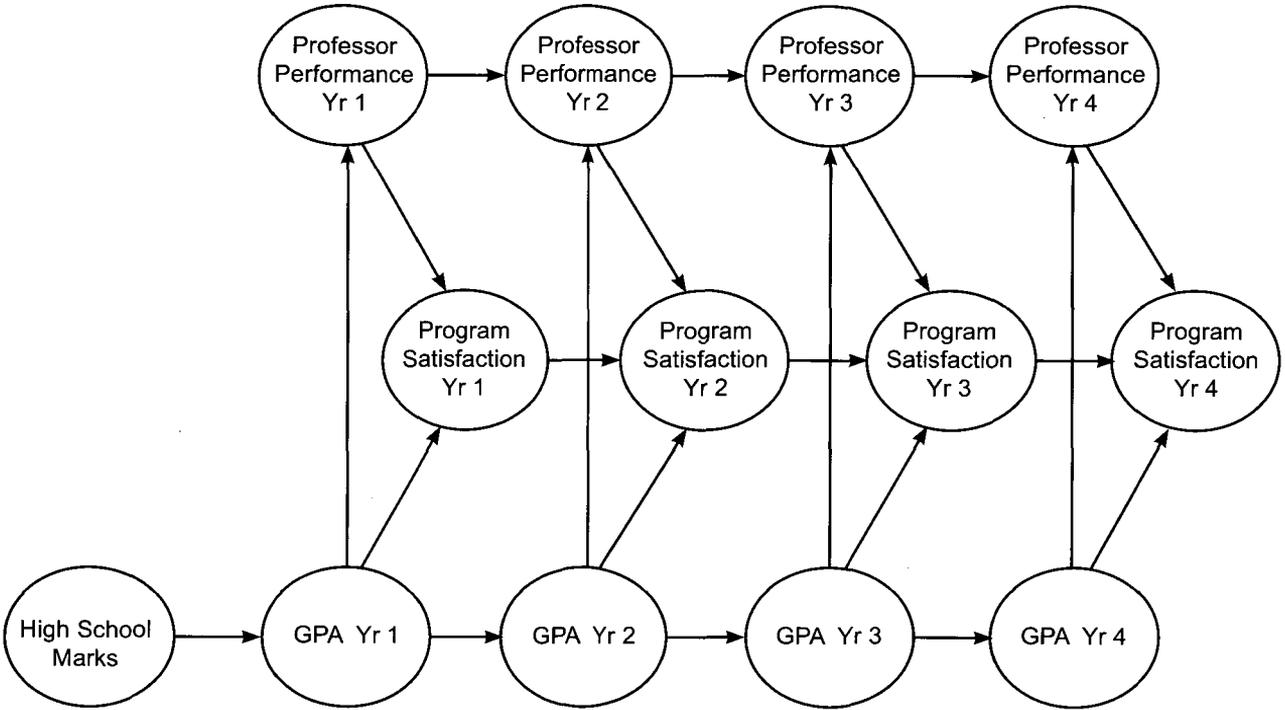


Diagram 1. Hypothesized Model of the Relationship among Program Satisfaction, Professor Performance, and GPA

Grading Leniency Bias and Student Characteristics Model

In the grading leniency bias model, it is postulated that high grades result in positive evaluations of professors' performance, and by extension, positive assessments of program satisfaction. Diagram 1 outlines the hypothesized relationships among variables from the grading leniency bias model when combined with assumptions of the student characteristics model over four years of study.

There is considerable research demonstrating that high school performance is a good predictor of first year grades. In addition, in Ontario, research shows that the variance in first year grades explained by high school grades is as high as the variance explained in some studies in the United States by both high school grades and the results of standardized tests, such as SATs. In turn, grades in early years of university are good predictors of grades in later years (Grayson, 1996). As a result of findings such as these, in the proposed model, it is hypothesized that across four years of study, previous grades will predict future grades. Consistent with the grading leniency bias model, within each year, it is expected that grades will predict both evaluations of professors' classroom performance and overall program satisfaction. The higher the grades, the more favourable the assessment of professors' performance and the higher the program satisfaction. As in a commuter university the classroom is the arena in which students most directly experience the benefits and short-comings of their overall academic programs, it also can be expected that positive evaluations of professors' performance in the classroom will predict program satisfaction.

As noted earlier in discussion, on the basis of the job satisfaction literature there is reason to believe that some students will interpret their university experiences within positive frames while others will bring negative frames to their interpretation of experiences. Consistent with the student characteristics model, in the current context this possibility means that we can reasonably expect some students to

consistently rate the classroom performance of their professors and their program satisfaction in a relatively positive way. Others will consistently do the opposite. If this is a valid assumption, it can be expected that assessments of professors' performance will be good predictors of performance in future years. Similarly, program satisfaction in one year will predict satisfaction in later years.

Teaching Effectiveness and Student Characteristics Models

In the teaching effectiveness model it is postulated that high caliber classroom performance by professors (i.e., effective teaching) will result in both good grades and, by extension, high levels of satisfaction with academic programs. A model expressing these possibilities would be exactly the same as depicted in Diagram 1 with the exception that the direction of the path in each year from GPA to professor performance would be reversed. Predictions resulting from assumptions underlying the student characteristics model would remain the same as in Diagram 1. Although the results of the combined teaching effectiveness and student characteristics models will be analyzed, because the visual changes to the model are limited, a drawing of the new model is not necessary.

SAMPLE

At the end of their first year of study in 1995, a mail-out questionnaire was sent to a random sample of 2,195 students who had entered York University directly from high school in 1994. Completed useable questionnaires were received from 1,865 students for a response rate of 64%. Exactly the same questionnaire was mailed to this original group of respondents at the end of 1996, 1997, and 1998. By the final year, 513 students (or 28% of the original sample) had responded to each wave of the study. When adjustments are made for the fact that in the intervening years students had left the university either before or after degree completion, the 513 students

who completed the final questionnaire represent 55% of those who had responded to the original survey and who were still enrolled in the university (Grayson, 1999). In a longitudinal study of this nature, this is an excellent retention rate (Dey, 1997). Overall, the sample on which the current study is based can be viewed as representative of students who entered York after high school in 1994 and who remained in their programs until 1998.

Students who remained in the sample entered the University with an average grade of 79% in their final year of high school. Sixty seven percent (67%) of the sample were female, 69% spoke English in their families when growing up, and 73% were of European origin. Forty four percent (44%) were the first in their families to go to university and 49% of the sample stated that their family's income was less than \$50,000 per year. In the final year of the study, students reported spending 22 hours a week outside of classes on their academic work and another 12 hours in paid employment.

The type of experiences students had over the four years of their education are analyzed elsewhere (Grayson, 1999). Suffice it to say that over the four years, on a number of dimensions, there was a general improvement in students' in-class experiences. Between first and fourth years, however, students became less involved in many out-of-class activities, but maintained a constant amount of interaction with friends. In general, the experiences of South Asian and Chinese origin students were relatively negative, but there were no differences in experiences based on family income and gender; however, experiences varied somewhat from one faculty to the next.

Of the final 513 who completed all phases of the study, 177 were in the Faculty of Arts, 104 in the Faculty of Pure and Applied Science, 84 in the Faculty of Fine Arts, 71 in Glendon College, 46 in the Schulich School of Business, and 31 in Environmental Studies. In the following analyses, given the large number of parameters to be estimated in the models, and in recognition of the fact that the student experience varies by faculty, it was decided to

restrict analyses to students in the faculties of Arts and Pure and Applied Science. Students in the Faculty of Arts alone account for approximately 51% of the university's enrolment.

MEASURES

Information on high school grades and grades for each year of study was obtained from the university's administrative records and merged with survey data.

Satisfaction was measured in two ways. First, students were asked the global question: "All things considered, how satisfied or dissatisfied are you with your academic program at York?" Response options ranged from 1 meaning very dissatisfied to 5 indicating very satisfied. Second, an index was constructed of questions with similar response options asking students how satisfied they were with various aspects of their academic programs: course content, quality of instruction, amount of required work, departmental administrative procedures, class size, grades, the extent to which they had developed intellectually, and the degree to which their knowledge of subject matter had increased over the previous academic year. Cronbach's alpha for each of the four years of study ranged from .71 to .75. Correlations between the index and responses to the global question ranged from .52 to .59.

Models were estimated using both measures for two reasons. First, it has been argued that for administrative purposes (e.g. tenure and promotion decisions) single item measures of teaching effectiveness are more useful than multiple measures. While satisfaction with academic program rather than teaching is under consideration, it makes sense that in processes such as program reviews single item measures may be more useful to assessors than multiple items. Second, as with teaching evaluations, multiple measures may be more useful in terms of feedback so that improvements can be made (Abrami et al., 1997; Marsh & Dunkin, 1992). Third, in studies of job satisfaction it has been found that single item measures are

more robust than scale measures and that single item measures are particularly useful in examining change in job satisfaction (Wanous, Reichers, & Hudy, 1997). Fourth, it made sense to see if using one measure or the other increased the fit of the proposed model.

A review of the literature reveals that teaching effectiveness has been operationalized in many different ways. One characteristic of many of these operationalizations is that they are not inspired by theories of teaching and learning (Marsh & Dunkin, 1992, p. 153). Moreover, there is reason to believe that what students view as the characteristics of effective teachers is related to what is being assessed: quality of instruction, quality of lectures, or the value of a course (Mason, Steagall, & Fabritius, 1995). A review of the literature conducted in the 1970s indicated that specific teaching dimensions having the highest correlations with overall teaching evaluations are the stimulation of interest, clarity of explanation, the provision of intellectual challenge, sensitivity to class level, and being prepared (Feldman, 1976). Somewhat consistent with these findings, more recent research has found that a sample of graduates identified effective communication, organization, and preparation as the most important characteristics of effective teachers (Rice, Stewart, & Hujber, 2000).

In the current study, questions focussing on exemplary performance by professors were derived from a study of students at the University of Guelph in which participants kept diaries of their first-year experiences and participated in interviews with researchers. The aspects of classroom performance by professors that were identified as exemplary were: having adequate technical expertise with regard to teaching; having knowledge of subject matter; being responsive to the class; caring about students in the class; having a sense of humour; and being well organized (Benjamin, 1990). In the current study, students were asked how many of the instructors in the courses in which they were currently enrolled had each of the foregoing characteristics. Using the total number of professors reported by the student as a base, a calculation was then made of the

percentage of professors with each of the characteristics. An average professor performance score was then calculated.

Cronbach's alpha was calculated for professors' performance scores in each year of the study. In all cases, the value of alpha exceeded .90.

CHANGE OVER TIME

As seen earlier, there is some reason to believe that certain aspects of the student experience may improve over time. For example, as they progress through the academic ranks, students may be exposed to smaller and smaller classes. As a result, satisfaction with academic programs may increase as time goes on.

Some preliminary information on this possibility for students in Arts and Pure and Applied Science is presented in Table 1. Professor performance and high school marks are expressed in percentages. GPA has high and low values of 9 and 0 respectively. Using the single item measure, program satisfaction ranges from 1 to 5 with 5 meaning very satisfied.

Data in Table 1 show that there is a monotonic increase in professor performance over the four years of study. In year one, students say that 72.1% of their professors display exemplary classroom behaviour. In year four the figure is 77.8%. It is important to note that for each year the standard deviation is 17% or 18% indicating considerable difference in assessments of professor performance. There is an irregular increase in GPA from 5.5 to 5.9, with a standard deviation between 1.5 and 1.7. Levels of program satisfaction are more or less the same across all years of study with a standard deviation of 0.8 or 0.9.

More light is shed on change in professor performance, GPA, and program satisfaction by repeated analyses of variance. The results of these analyses are presented in Figures 1 to 4. In the figures, mean scores for professor performance have been divided by 10 to facilitate graphing.

Table 1

Means and Standard Deviations for Variables in Analysis

	Year 1	Year 2	Year 3	Year 4
Professor Performance				
Mean	72.1	74.3	75.7	77.8
SD	17.8	17.3	18.0	18.4
GPA				
Mean	5.5	5.6	6.1	5.9
SD	1.6	1.6	1.5	1.7
Program Satisfaction				
Mean	3.9	3.8	3.8	3.7
SD	0.8	0.8	0.9	0.9
High School Marks				
Mean	79.3			
SD	5.8			

Figure 1 includes information on professor performance, GPA, and program satisfaction for all students included in the study (i.e., not just students from the faculties of Arts and Pure and Applied Science). It also includes a measure of average class section size in 1999 weighted for the reality that faculties have differing numbers of students. Overall, information in Figure 1 indicates that between years one and four, average class section size decreases drastically. Despite the considerable decrease in class section size, satisfaction with academic programs is relatively constant over four years

of study and there is only a slight increase in sessional GPA and assessments of professors' performance over the same time period. In essence, while in some respects, as indicated by decreasing class section size, the learning environment may improve between years one and four, the correspondence between this improvement and assessments of professor performance, GPA, and program satisfaction is very small.

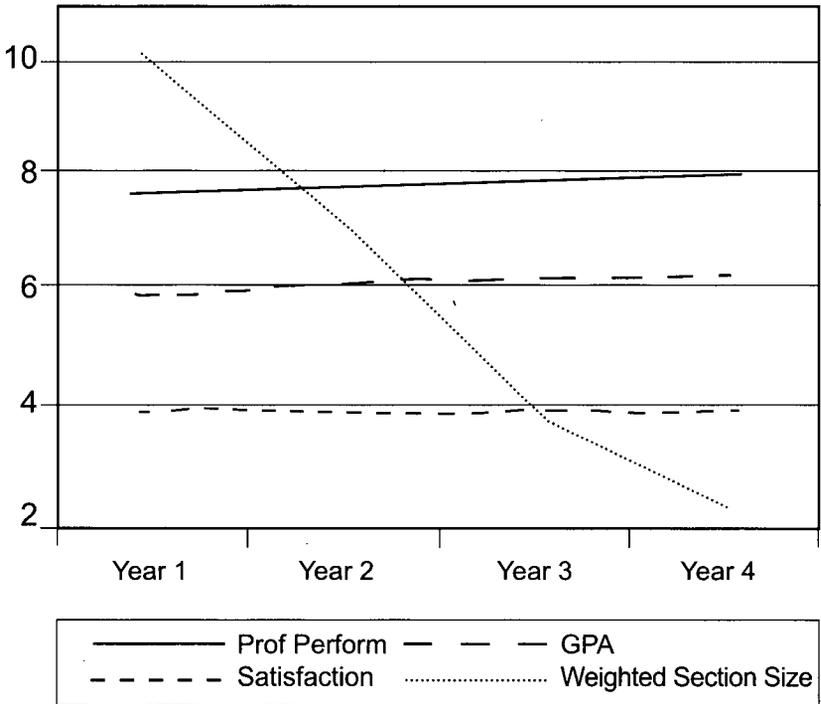


Figure 1. Professor Performance, GPA, Program Satisfaction, and Weighted Section Size by Year of Study

The information in Figures 2 to 4 focuses on differences between students in Arts and Pure and Applied Science. Data presented in Figure 2 show that for both Arts and Science students assessments of the percentage of professors displaying exemplary performance

were fairly high in each year of the study. For Arts students, there was an increase in the percentage of professors demonstrating behaviours that enhance learning from 74% in first year to 80% in fourth year. For Science students overall increases were from only 70% in first year to 74% in third and fourth year.

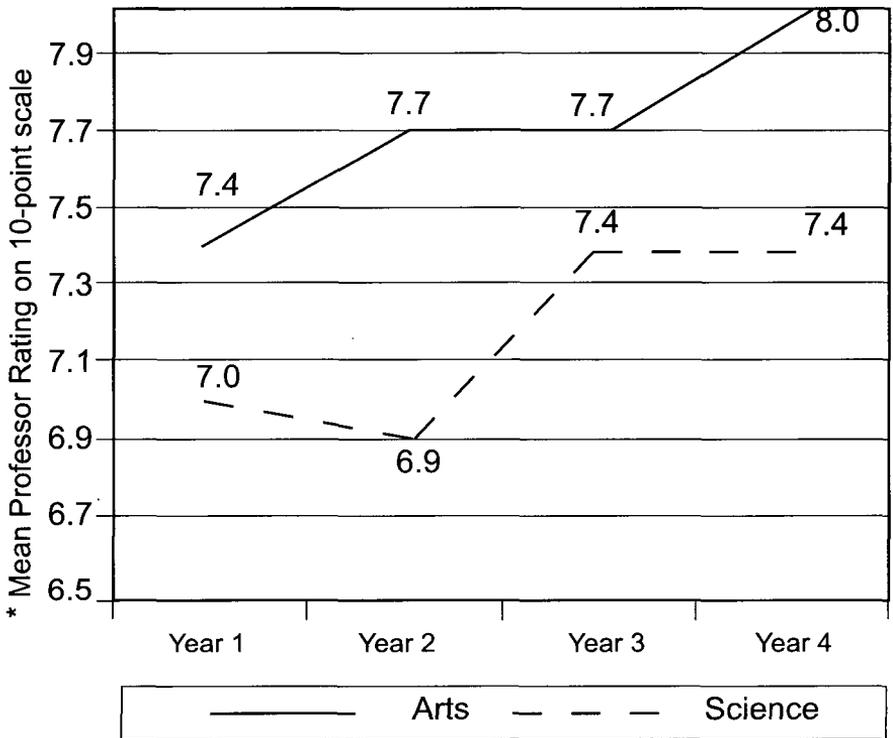


Figure 2. Professor Performance by Faculty

* The original 100-point professor evaluation scale is divided by 10 to produce a 10-point scale.

A repeated analysis of variance indicated that for Arts and Science students, increases in the percentage of professors exhibiting model performance over the four years were statistically significant (data not presented in tabular form). Differences between Arts and Science

students were also statistically significant. In other words, the perceived performance of professors increased over time, but Arts students viewed more improvement in teaching performance than Science students. This finding is consistent with the idea that the student experience improves over the academic career of the student.

Information presented in Figure 3 indicates that there is an overall increase in grades as students progress through the ranks in both Arts and Science.

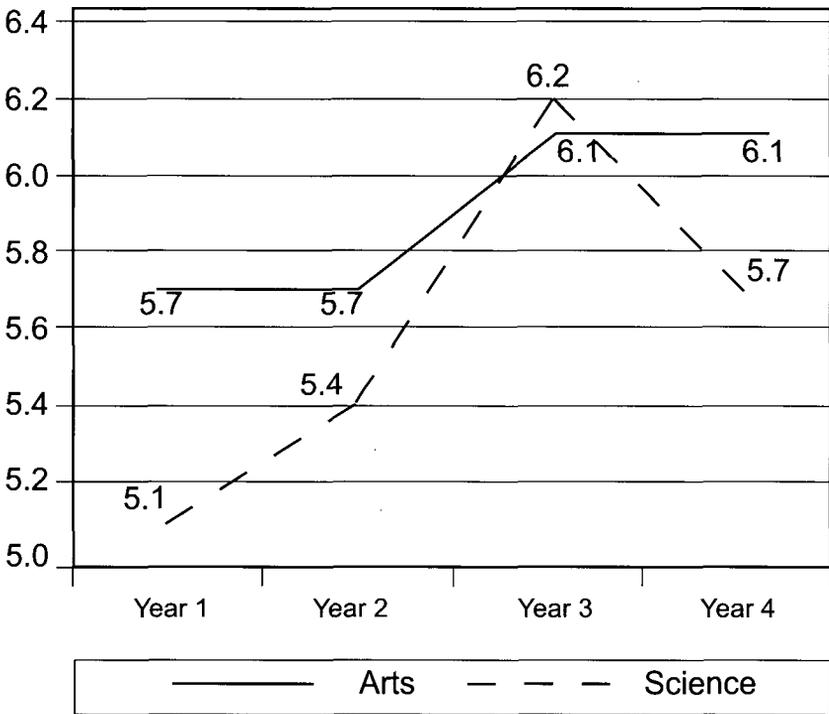


Figure 3. GPA by Faculty

A repeated analysis of variance reveals that for Arts and Science students increases between first and fourth year are statistically significant, but differences between students in the two faculties are not statistically significant. In short, an overall increase in GPA for both Arts and Science students occurs over the course of the

academic career and differences between students in the two faculties are not statistically significant. This finding is also consistent with the notion that the student experience improves over time.

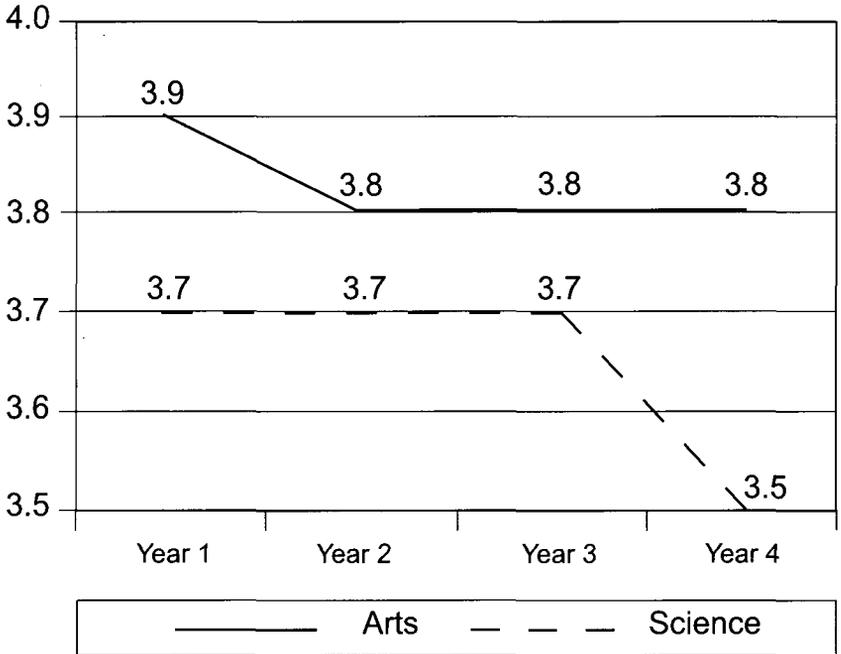


Figure 4. Program Satisfaction by Faculty

The data in Figure 4 show that for each year of the study the mean level of program satisfaction is greater for Arts than for Science students and a repeated analysis of variance indicates that differences between Arts and Science students are statistically significant; however, changes within Arts and Science over the four year period are not statistically significant. Students do not become increasingly more satisfied with their academic programs. Overall, while information on professor performance and GPA indicates improvement in the student’s experience between first and fourth years, the current finding indicates that such improvement is not paralleled by an increase in satisfaction with academic programs.

Taken collectively, findings expressed in Figures 2 to 4 indicate that for both Arts and Science students, assessments of professor performance increase over time, as do GPAs; however, change in these variables is not proportionate to the decrease in class section size as indicated in Figure 1. Changes in academic program satisfaction between first and fourth year are not statistically significant.

The fact that students perceive better performance on the part of their professors as time moves on, and that grades increase over time, is more or less consistent with findings reported earlier in this article. A perception by Arts students of relatively better performance by their professors, and greater academic program satisfaction than Science students, is also consistent with previous findings.

MODEL FITTING

Grading Leniency Bias and Student Characteristics Model

While repeated analyses of variance are appropriate in examining change in professors' performance, GPA, and student program satisfaction, structural equation modelling allows an examination of the relationships among these variables within and between years. Structural equation model fitting was accomplished by using AMOS 4. Variables used in the analysis are normally distributed. The amount of missing data ranged from a low of 0% to a high of 17% with an average of 5% per variable. Missing data were estimated using the maximum likelihood method.

As a first step in analysis, estimates were made of the model in Diagram 1 for Arts ($n = 177$) and Pure and Applied Science students ($n = 104$) separately. In evaluating the models, a RMSEA value up to .050 indicated a good fit and values more than .050 but less than .080 represented a reasonable fit (Browne & Cudeck, 1993). Values between .080 and .100 indicated a mediocre fit and values of .100 or more indicated a poor fit (MacCallum, Browne, & Sugawara, 1996). It must be stressed that these critical values are not accepted by all.

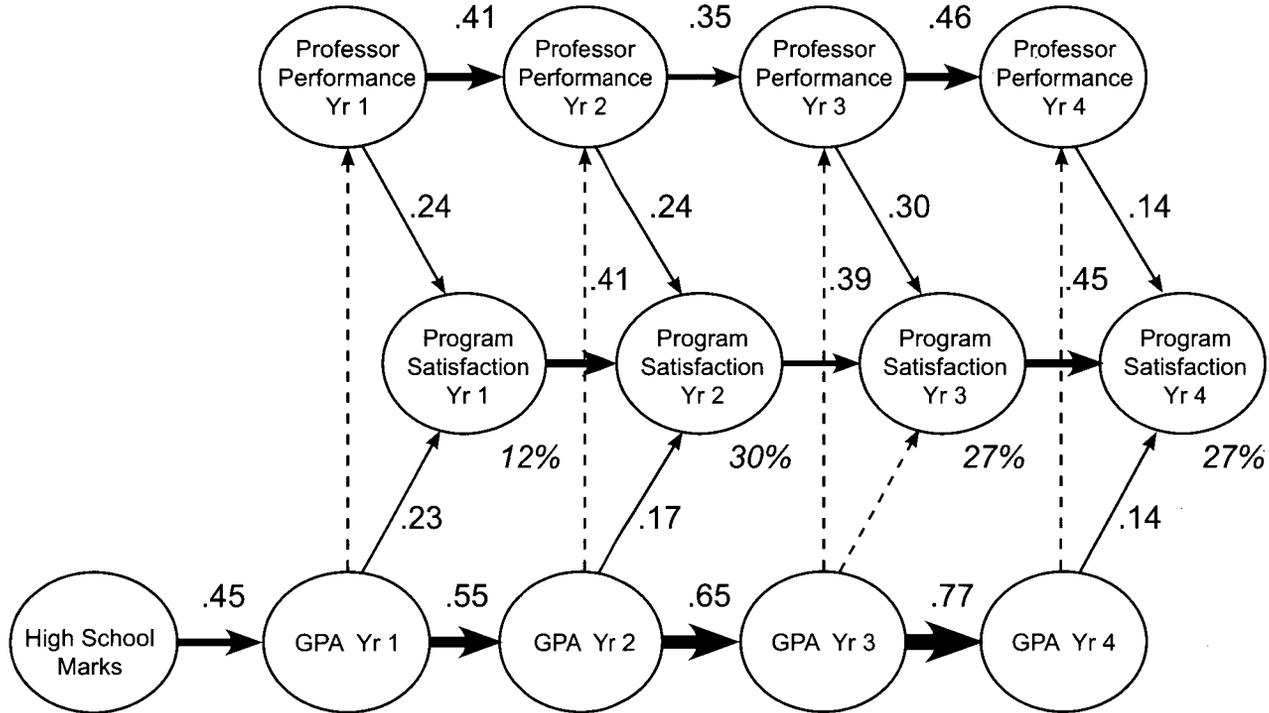
For the Science model, chi-square was 84.9 with 56 degrees of freedom. RMSEA equaled .070 with a 90% confidence interval between .036 and .100. The p test for close fit was .141 and CFI was .993. Despite the low p test, these values suggest a reasonable fit of the Science model.

The value of chi-square for the Arts model was 107.4 with 56 degrees of freedom. RMSEA was .072 with a 90% confidence interval between .051 and .093. The p test for close fit was .041 and CFI equaled .993. Again, despite the low p test value, these figures indicate a reasonable fit.

Because of acceptable fits of each model separately, as a second step, in a simultaneous two group analysis the results of the unconstrained model as found in Diagram 1 were compared to the results of a fully constrained model. The value of chi-square for the unconstrained model was 191.9 with 112 degrees of freedom. The value of RMSEA was .051 with a 90% confidence interval between .038 and .063. The p test for closeness of fit was .454 with a CFI of .993. These values indicate a good model fit. The chi-square value for the fully constrained model was 226.4 with 134 degrees of freedom. RMSEA had a value of .050 with a 90% confidence interval between .038 and .061. The p test of closeness of fit was .503, and the CFI .992. The chi-square difference between the unconstrained and constrained models of 34.5 with 22 degrees of freedom is statistically significant at the .05 level; therefore, we can conclude that differences exist between students in Arts and students in Pure and Applied Science.

Results of the simultaneous two group unconstrained analysis are presented in Diagrams 2 and 3. Dashed lines indicate no statistically significant relationship. Standardized regression coefficients are on, or near to, arrows. The thickness of arrows is in rough proportion to the size of the standardized regression coefficient. Explained variance in program satisfaction is in italics.

Information on Arts students is presented in Diagram 2. Given that the hypothesized model was based on the assumptions of



For two unconstrained models: Chi-square = 191.9; df = 112; RMSEA = .051 (lower bound = .038; upper bound = .063); P test for close fit = .454; CFI = .993

Diagram 2. Final Model of the Relationship Among Program Satisfaction, Professor Performance, and GPA for Arts Students over Four Years of Study

the grading leniency bias model and the student characteristics models, it is important to note that for no year is there a statistically significant relationship between sessional grades and assessments of professor performance. In other words, for Arts students, there is no support for the basic assumption of the grading leniency bias model: assessments of professor performance in all years of study are made independent of grades. This finding contributes to the body of evidence suggesting that teaching evaluations are not biased by grading practices.

The second observation that can be made is that while grades do not affect assessments of professors' performance, in all years but the third they make a small contribution to students' satisfaction with academic programs as measured by the global question: students who get good grades are more satisfied with their academic programs than students who get poor grades. Similarly, judging from the moderate to small size of the regression coefficients, for all years of study, good teaching results in slightly enhanced program satisfaction.

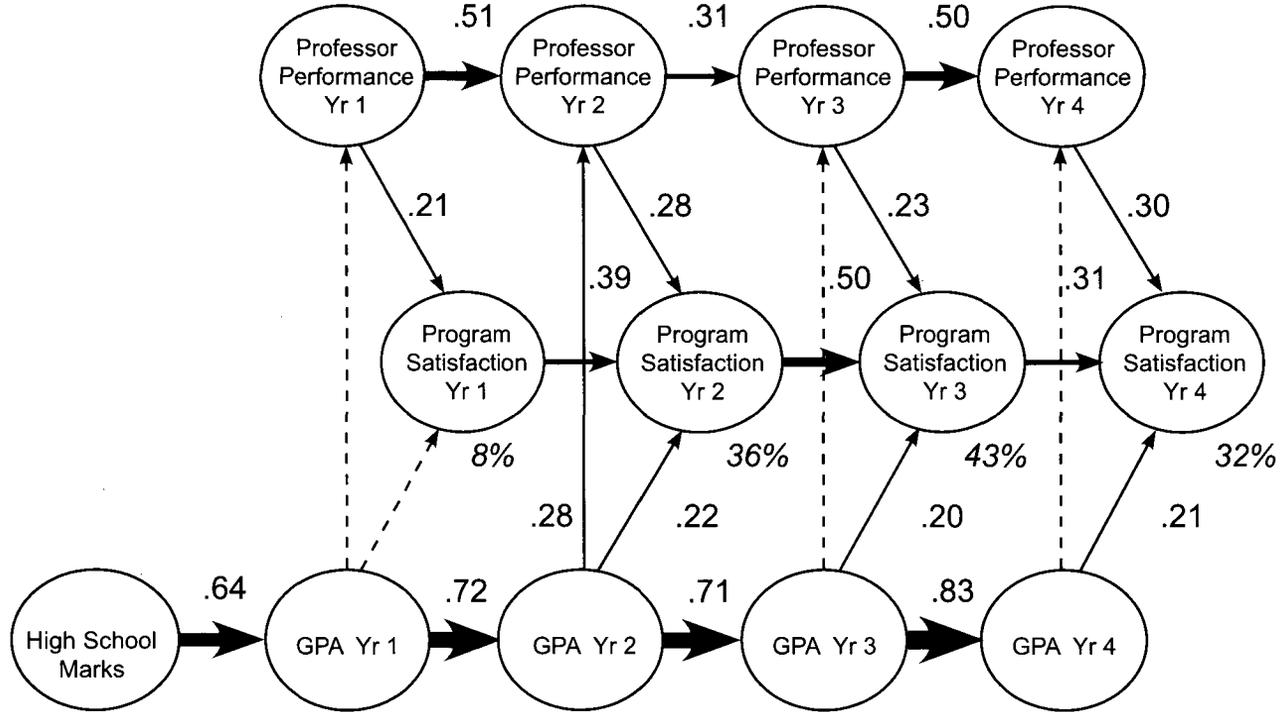
While no support is provided for assumptions underlying the grading leniency bias model, a lot of support is given to assumptions derived from the student characteristics model. For GPA, assessments of program satisfaction, and professor performance, the single best predictor in any one year is the score in the previous year. The effect is greatest for GPA suggesting an accumulation of advantage for students as they progress over their academic careers. For example, there is a coefficient of .55 for the effect of GPA in year one on GPA in year two. The coefficient increases to .77 for the effect of GPA in year three on GPA in year four. Figures such as these indicate the cumulative effect of learning in Arts.

The sizes of the coefficients between professor performance in one year and the next and between program satisfaction in one year and the next are smaller than for GPA, but they are still substantial and generally larger than the coefficients between GPA and program satisfaction, and professor performance and program satisfaction. Put differently, students are most likely to positively evaluate their

professors' performance if they have made positive evaluations in the past. Similarly, in any given year, students are most likely to be satisfied with their academic programs if they have been satisfied in the past. In other words, independent of GPA, even though in each year students may experience different courses and professors, some students are more inclined than others to have positive regard for their professors' performance and for their academic programs. This finding is consistent with research reported earlier in which it was found that individuals with particular personality characteristics are more likely than others to be satisfied with the same jobs. Judging from the current findings, although we cannot say with certainty, personality may have a greater impact on assessments of professor performance and satisfaction with academic programs than either GPA or the overall performance of professors.

In the Faculty of Arts the amount of variance explained in program satisfaction by GPA and professor performance in year one is 12%. The figures for years two, three, and four are 30%, 27%, and 27% respectively.

Information on students in Pure and Applied Science is found in Diagram 3. Although the differences between the models for Arts and Pure and Applied Science students are statistically significant, the differences between the two are minor. While no paths between GPA and professor performance are statistically significant for Arts students, the path between GPA for year two and professor performance in year two is significant, but moderate, for Science students. In addition, while the only path between GPA and program satisfaction that is not statistically significant for Arts students is between GPA year three and program satisfaction year three, for Science students only the path for GPA in year one and program satisfaction in the same year is not statistically significant. Apart from this, the general magnitude of the paths among variables is comparable. As a result, the overall conclusions that derive from an examination of the Science model are similar to those based on



For two unconstrained models: Chi-square = 218.9; df = 112; RMSEA = .058 (lower bound = .047; upper bound = .070); P test for close fit = .110; CFI = .992

Diagram 3. Final Model of the Relationship Among Program Satisfaction, Professor Performance, and GPA for Science Students over Four Years of Study

the Arts model. GPA is of little consequence for assessments of professors' performance, students who get high grades are somewhat more satisfied with their academic programs than students receiving low grades, and good teaching results in somewhat higher satisfaction with program. In any one year, the best predictor of GPA is GPA in the prior year, and students who assess their professors' performance positively in one year are likely to do the same in the next. The same applies to assessments of academic program satisfaction. In essence, it is reasonable to postulate that certain personality characteristics incline science students to positively view both their professors and their academic programs.

For Science students, the amounts of variance in program satisfaction explained by GPA and professor performance are 8%, 36%, 43%, and 32% in years one, two, three, and four respectively.

A separate model was estimated in which the index of program satisfaction described earlier was substituted for the global measure of satisfaction. The chi-square value for the unconstrained model was 218.9 with 112 degrees of freedom. The value of RMSEA was .058 with a 90% confidence interval between .047 and .070. The p test of close fit was .110, and CFI equaled .992. While the p test is low, RMSEA and CFI are acceptable.

Given equal degrees of freedom, and in both cases roughly comparable and acceptable RMSEA and CFI indexes, in determining whether this or the model with the global measure of academic program satisfaction is a better fit, it is appropriate to compare the Akaike information criteria (AIC). For the unconstrained model with the global measure AIC = 383.9; for the current model AIC = 410.9. As a result, we can conclude that the model with the global measure is more parsimonious, and therefore better, than the one using the satisfaction index. Moreover, despite some minor differences in the sizes of the regression coefficients, an examination of the model that includes the satisfaction index results in conclusions exactly the same as the model employing the global measure. In essence, nothing is to be gained by employing the satisfaction index rather

than the global measure of academic program satisfaction. This suggests that particularly for administrative purposes, the global measure is as good as the index.

Teaching Effectiveness and Student Characteristics Model

As noted earlier, a model based on the assumptions of the teaching effectiveness and student characteristics models would be the same as the hypothesized model represented in Diagram 1 with the direction of the paths from GPA to professor performance reversed. This change would indicate the expectation that in any year, good teaching increases both GPA and satisfaction with academic programs. Support for the assumptions of the student characteristics model would be evident in a retention of all other links in the model.

The model estimated by reversing the direction of causation outlined above results in a model (not shown) with a chi-square of 198.5 and 112 degrees of freedom. RMSEA is .053 with a 90% confidence interval between .040 and .064. The p test for close fit is .346 and the CFI equals .993. These figures indicate a generally good fit. The Akaike information criterion of 390.5 is slightly higher than the 383.9 AIC of the unconstrained grading leniency bias and student characteristics model. In essence, the models share a more or less equal fit of the data, although the latter is slightly more parsimonious than the former.

When the specific paths in the Arts teaching effectiveness and student characteristics model are examined (not shown), in no year are the paths from professor performance to GPA statistically significant. In essence, for Arts students, good teaching does not result in increased grades. As a result, the model does not support the fundamental assumption of the teaching effectiveness model. The nature and magnitude of all other links in the model are the same as in the model based on the assumptions of the grading leniency bias and student characteristics model tested earlier.

When the model is applied to the Pure and Applied Science data (not shown) in second year, good teaching results in better grades. For no other year is the path from professor performance to GPA statistically significant. In essence, for Science students, the basic assumption of the teaching effectiveness component of the model receives very weak support. The values of the standardized coefficients of all other paths in the model are the same as in the grading leniency bias and student characteristics model. Overall, for both Arts and Pure and Applied Science students, teaching effectiveness assumptions of the model receive virtually no support. Although both professor performance and GPA enhance student satisfaction with their programs, the best predictors of professor performance and program satisfaction are scores on these variables in the preceding year. Again, this finding suggests the presence of underlying personality characteristics that incline students to positively evaluate both their professors' performance and their academic programs. In essence, the student characteristics components of the model have more support than the teaching effectiveness components.

From the foregoing analyses, it is possible to conclude that in the samples of Arts and Pure and Applied Science students under consideration, the assumptions underlying neither the grading leniency model nor the teaching effectiveness model receive much support. In both Arts and Pure and Applied Science high grades do not necessarily translate into positive evaluations of teaching, nor does being exposed to good teaching necessarily result in good grades. By contrast, good teaching and getting high grades does lead to increased program satisfaction.

In both faculties the best predictor of high grades is past achievement. This is to be expected. To the degree that high grades suggest intellectual ability and/or high motivation, it can be argued that academic achievement has more to do with particular student characteristics than with good teaching. *This does not mean that students would do as well without instructors. It likely indicates that*

even poor teachers provide sufficient guidance and/or that students work around them.

In both faculties the best predictors of both positive assessments of professor performance and program satisfaction are previous positive assessments. As a result, we can posit the existence of other student characteristics (certain personality traits) that contribute to both favourable assessments of professors and academic programs. This interpretation is consistent with interpretations of why certain individuals are satisfied with their jobs, somewhat independently of the objective circumstances in which they work.

Despite these conclusions, it must be stressed that the current study was carried out in one large commuter university. It is likely that similar results would be found on small and/or residential campuses; however, further research is needed to confirm this assumption.

CONCLUSIONS

It is frequently assumed that the quality of the undergraduate experience improves over the course of a university education. Data on a panel of students followed over four years of study analyzed in this article suggest that for students in the faculties of Arts and Pure and Applied Science there is somewhat of an improvement in professors' performance over time. This increased appreciation of professors' performance may be related to the fact that upper level courses' section sizes were much smaller than in first and second year, to the possibility that in upper level courses students feel relatively confident, or to the likelihood that some professors relate best to students who have acquired some disciplinary knowledge. There is also an increase in GPA as students progress through the academic ranks. This may result from the fact that as they mature academically, students are more able to grasp the nuances of their disciplines or from the possibility that in upper level courses marking is more lenient than in earlier years. Whatever the reason,

improvements in assessments of professor performance and GPA are not paralleled by increases in academic program satisfaction. On this measure, there is little change between first and fourth years.

In addition to examining changes in professor performance, GPA, and program satisfaction, the current study focused on the dynamics among these three variables within each year of study, and over the four years of the student experience. In contrast to the assumptions of both the student leniency bias model, and the teaching effectiveness model, little support was found for the ideas that positive assessments of professor performance were a reflection of grades or that, in comparison, grades are a result of good teaching. Instead, it is likely that some students are more likely than others to put a positive light on their experiences and to give high ratings to their professors. While I have suggested that this tendency is probably related to underlying personality characteristics, further research is needed to test this possibility. This does not mean that professors are irrelevant to learning. It may mean that even relatively poor instruction as operationalized here is sufficient for student learning.

While there may be little relationship between professors' performance and GPA, it is clear that good teaching results in enhanced program satisfaction. In addition, good grades result in increased program satisfaction. As was the case for assessments of professor performance, however, program satisfaction likely is most directly a result of certain personality characteristics that predispose students to evaluate their experiences in a positive manner.

IMPLICATIONS

There are two main implications of the foregoing analysis. First, a single item measure of academic program satisfaction has as much utility as a multi-trait index of the same concept.

As a result, in evaluations of program satisfaction where space constraints may prevent the inclusion of questions dealing with various aspects of students' programs, a global measure is adequate for most administrative purposes, such as program reviews. The same argument has been made for single item measures of teaching effectiveness (Abrami et al., 1997).

Second, good teaching does not necessarily translate into high grades; however, it does increase student satisfaction. As a result, if the promotion of learning is the benchmark, committees assessing the quality of various programs should be cautious when they interpret the results of teaching evaluations. Positive evaluations may indicate that students in the program are satisfied with their courses, but not necessarily that they have learned a lot because of the performance of their professors. This said, student satisfaction should be viewed as a legitimate end in itself, and should be an important criterion in program reviews. ❁

References

- Abraham, R. (1999). Negative affectivity: Moderator or confound in emotional dissonance-outcome relationships? *The Journal of Psychology*, 133(1), 61–72.
- Abrami, P.C., Cohen, P.A., & d'Apollonia, S. (1988). Implementation problems in meta-analysis. *Review of Educational Research*, 58, 151–179.
- Abrami, P.C., d'Apollonia, S., & Rosenfield, S. (1997). The dimensionality of student ratings of instruction: What we know and what we do not. In R.P. Perry & J.C. Smart (Eds.), *Effective teaching in higher education: Research and practice*. New York: Agathon Press.
- Abrami, P.C., Perry, R.P., & Leventhal, L. (1982). The relationship between student personality characteristics, teacher ratings, and student achievement. *Journal of Educational Psychology*, 74, 111–125.
- Astin, A. (1993). *What matters in college?* San Francisco: Jossey Bass.
- Bean, J.P., & Bradley, R.K. (1986). Untangling the satisfaction-performance relationship for college students. *Journal of Higher Education*, 57(4), 393–412.

- Belyukova, S.A., & Fox, C.M. (2002). Student satisfaction as a measure of student development: Towards a universal metric. *Journal of College Student Development*, 43(2), 161–172.
- Benjamin, M. (1990). *Freshman daily experiences: Implications for policy, research and theory*. Guelph: University of Guelph Student–Environment Group.
- Benjamin, M., & Hollings, A.E. (1995). Toward a theory of student satisfaction: An exploratory study of the “quality of student life.” *Journal of College Student Development*, 36(4), 574–586.
- Benjamin, M., & Hollings, A.E. (1997). Student satisfaction: Test of an ecological model. *Journal of College Student Development*, 38(3), 213–228.
- Biner, P.M., Summers, M., Dean, R.S., Bink, M.L., Anderson, J.L., & Gelder, B.C. (1997). Personality characteristics predicting continuing education student satisfaction with interactive telecourses. *Journal of Continuing Higher Education*, 45(3), 22–32.
- Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models*. Newbury Park, CA: Sage.
- Cashin, W.E. (1995). *Student ratings of teaching: A summary of the research*. Idea paper no. 20: ERIC NO: ED302567.
- Cohen, P.A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, 51(3), 281–309.
- Dey, E.L. (1997). Working with low survey response rates: The efficacy of weighting adjustments. *Research in Higher Education*, 38(2), 215–227.
- Feldman, K.A. (1976). The superior college teacher from the students’ view. *Research in Higher Education*, 5(3), 243–248.
- Feldman, K.A. (1978). Course characteristics and college students’ ratings of their teachers: What we know and what we don’t. *Research in Higher Education*, 9, 199–242.
- Grayson, J.P. (1995). Does race matter? – Outcomes of the first year experience in a Canadian university. *The Canadian Journal of Higher Education*, 25(2), 79–109.
- Grayson, J.P. (1996). *Under- and over-achievement in first year*. Toronto: Institute for Social Research, York University.

- Grayson, J.P. (1997). Place of residence, student involvement, and first year marks. *The Canadian Journal of Higher Education*, 27(1), 1–23.
- Grayson, J.P. (1998). Race and retention in a Canadian university. *Higher Education*, 36(3), 323–352.
- Grayson, J.P. (1999). *The student experience at York University: The effects of income, race, and gender over four years*. Toronto: Institute for Social Research, York University.
- Grayson, J.P. (2003). The consequences of early adjustment to university. *Higher Education*, 46, 411–429.
- Greenwald, A.G. (1995). *Applying social psychology to reveal a major (but correctable) flaw in student evaluations of teaching*. Paper presented at the 103rd Meeting of the American Psychological Association, New York.
- Howard, G.S., & Maxwell, S.E. (1980). Correlations between student satisfaction and grades: A case of mistaken causation? *The Journal of Educational Psychology*, 72(6), 810–820.
- Judge, T.A., Kluger, A.N., Locke, E.A., & Durham, C.C. (1998). Dispositional effects on job and life satisfaction: The role of core evaluations. *Journal of Applied Psychology*, 83(1), 17–34.
- Krautmann, A.C., & Sander, W. (1999). Grades and student evaluations of teachers. *Economics of Education Review*, 18, 59–63.
- MacCallum, R.C., Browne, M.W., & Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130–149.
- Marsh, H., & Dunkin, M.J. (1992). Students' evaluations of university teaching: A multidimensional perspective. *Higher Education: Handbook of Theory and Research*, 8, 143–233.
- Mason, P.M., Steagall, J.W., & Fabritius, M.M. (1995). Student evaluations of faculty: A new procedure for using aggregate measures of performance. *Economics of Education Review*, 14(4), 403–416.
- Murray, H.G., Rushton, P., & Paunonen, S.V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *The Journal of Educational Psychology*, 82(2), 250–261.
- Pascarella, E., Edison, M., Nora, A., Hagedorn, L., & Terenzini, P. (1996). Influence on students' openness to diversity and challenge in the first year of college. *Journal of Higher Education*, 67(2), 174–195.

- Pennington, D.C., Zvonkovic, A.M., & Wilson, S.L. (1989). Changes in college satisfaction across an academic term. *Journal of College Student Development, 30*, 528–535.
- Rice, R.E., Stewart, L.P., & Hujber, M. (2000). Extending the domain of instructional effectiveness assessment in student evaluations of communication courses. *Communication Education, 49*(3), 253–266.
- Strom, B., & Hocevar, D. (1982). Course structure and student satisfaction: An attribute-treatment interaction analysis. *Educational Research Quarterly, 7*(1), 21–30.
- Terenzini, P., Springer, L., Pascarella, E., & Nora, A. (1995). Influences affecting the development of students' critical thinking skills. *Research in Higher Education, 36*(1), 23–39.
- Terenzini, P., Springer, L., Yaeger, P.M., Pascarella, E., & Nora, A. (1996). First-generation college students: Characteristics, experiences, and cognitive development. *Research in Higher Education, 37*(1), 1–22.
- Terenzini, P., & Wright, T.M. (1987). Influences on students' academic growth during four years of college. *Research in Higher Education, 26*(2), 161–179.
- Tinto, V. (1993). *Leaving college*. Chicago: University of Chicago Press.
- Tinto, V. (1997). Classrooms as communities: Exploring the educational character of students' persistence. *Journal of Higher Education, 68*(6), 599–623.
- Wanous, J.P., Reichers, A.E., & Hudy, M. (1997). Overall job satisfaction: How good are single-item measures. *Journal of Applied Psychology, 82*(2), 247–252.
- Witt, P.H., & Handal, P.J. (1984). Person-environment fit: Is satisfaction predicted by congruency, environment, or personality? *Journal of College Student Development, 25*, 503–508.