Student Grades and Course Evaluations in Engineering: What Makes a Difference

by

Kara Maria Kockelman

Assistant Professor of Civil Engineering
The University of Texas at Austin
6.9 E. Cockrell Jr. Hall
Austin, TX 78712-1076

kkockelm@mail.utexas.edu
Phone: 512-471-0210
FAX: 512-475-8744

Presented at the ASEE Annual Conference
Albuquerque (Paper #2793), New Mexico, June 2001,
and at the ASEE Gulf Section Annual Conference, College Station, Texas, March 2001

Abstract:

This research investigates the impact of different instructor, course, and student attributes on student grades and course evaluations. The data come from undergraduate courses given at the University of Texas at Austin during the 1992 through 1998 calendar years. Instructor experience, standing, and gender; course department and credit hours; and student classification, test scores, gender, and other variables are used to explain variation in both grades and evaluation scores.

The results of multivariate weighted-least-squares regressions of average grades given across a sample of over 2,500 courses suggest that the average male instructor assigns lower grades than female instructors, while lecturers and teaching assistants assign higher grades than full, associate, assistant, and adjunct faculty. Instructors teaching chemical, mechanical, and petroleum-and-gas engineering courses assign higher grades, on average, than those teaching aerospace, architectural, civil, and electrical engineering, and engineering mechanics. The results also indicate that non-Asian and non-foreign males taking lower-division courses for more credit hours receive lower grades, after controlling for student entrance-test scores and year in school.

Weighted-least-squares analyses of average evaluation scores given to instructors were conducted over five different qualities: course organization, instructor communication, instructor

teaching skill, the instructor overall, and the course overall. Evaluations from over 2,500 courses comprised the data set. In general, female students and African Americans rated their courses and instructors higher; and male faculty rated somewhat lower than female faculty. There are interesting gender interaction effects – between students and their instructors – evident as well. Instructors who had received their PhDs relatively long ago (which is expected to be highly correlated with instructor age and teaching experience) rated lower, except in the area of course organization. Senior lecturers consistently rated higher than full faculty, and assistant and adjunct faculty rated lower (on two and four of the five questions, respectively). Students in engineering mechanics and aerospace, architectural, civil, mechanical, and petroleum engineering rated their courses and instructors higher, on average, than did electrical and chemical engineering students.

Also of interest to educators are the consistently positive and statistically significant associations between student ratings of a course and student GPAs – and the lack of any statistically significant relation between evaluations and grades biases (as captured by an average grade minus average GPA variable). These results are apparent only after controlling for other factors, including instructor, course, and student attributes and suggest that educators need not be very concerned about the biasing effects of "easy grading" on instructor evaluations.

INTRODUCTION

Student grades and course evaluations are important descriptors of student and faculty performance. Student grades represent instructor evaluation of students and have been used pervasively for probably as long as there have been universities. In contrast, the acquisition and dissemination of student evaluations of their instructors and courses have arisen relatively recently, from student-based efforts in the 1960's. Many universities now incorporate evaluation results in faculty salary and promotion decisions, and nearly all major U.S. universities regularly collect such data. (Ory 1990, Seldin 1993) Students are probably the best resource universities have to assess instructor performance; they experience all aspects of many courses and thus can compare and contrast such experiences. Moreover, aggregation of their responses provides a large data set, helping to minimize any variation in estimation of average response. However, their use in merit, promotion, and other decisions engenders some controversy.

Student ratings of courses are not perfectly reflective of student learning. For example, laboratory studies have suggested that while instructor enthusiasm significantly impacts student ratings, it does not much affect student learning. In contrast, lecture content appears to have a much greater effect on student learning than on ratings. (Abrami, Leventhal, and Perry 1982) And correlations between average ratings and average learning (based on standardized test results across multiple course sections) generally fall well below 0.5. For example, Cohen's meta-analysis (1981) deduced that the highest correlations relate test performance to overall course and overall instructor ratings; these were estimated to be 0.47 and 0.43, respectively. In contrast, his correlation of performance and instructor-student interaction ratings was just 0.22. Therefore, in terms of a linear goodness of fit (R^2), course and instructor ratings explain just 22% and 18% of the variance associated with average student achievement, and instructor-student interaction explains less than 5%. After controlling for other variables (for example, student intelligence and course time of day), it is very possible that the explanatory contributions of such ratings will fall to even lower levels. Clearly, much more information on students, their instructors, and their courses is needed – in a single model.

While the relationships between student learning and ratings are not very strong, they are significant enough that many (e.g., Cohen 1990, Franklin and Theall 1990, Cuseo 2000, Wankat and Oreovicz 1993) place great value on their collection and use. Cohen's "research-based refutations" to myths concerning student ratings include the following: student ratings are reliable, stable, and "not unduly influenced by the grades" received; and "students are qualified to rate certain dimensions of teaching." (1990, p. 124) However, most researchers agree that ratings are just one component of a comprehensive assessment of faculty teaching (see, e.g., McKnight 1990).

The mechanisms linking student grades and instructor evaluation to student, instructor, and course characteristics are complex; but, thanks to a large set of detailed data, the investigation described here illuminates many such relationships. The work analyzes University of Texas at Austin grades given to undergraduate students and course evaluations received by their instructors in the College of Engineering. It relies on weighted-least-squares (WLS) regression models for estimates of different instructor, course, and student attribute effects on student grades and course/instructor evaluations. What follows is a discussion of the data sets, the models, and the analytic results.

DATA SETS

Instructor experience, standing, and gender, course department and credit hours, and student classification, test scores, gender, and other variables are used here to explain variation in grades and evaluation scores. A description of all variables used is provided in Table 1.

The primary data represent grades, evaluations, and student information collected and maintained by U.T. Austin's College of Engineering from the spring semester of 1992 through the fall semester of 1998 (21 semesters total). The instructor-attribute data set was produced only recently, based on personnel databases and publicly available lists maintained by the University. Tables 1 and 2 present the definitions and some basic descriptive statistics of the variables used.

Course and Instructor Evaluations

The initial course data set contained 8,458 records. However, instructor gender could not be reliably matched to approximately 2,100 of these cases, and the year of an instructor's PhD (or coming PhD, in the case of teaching assistants and some others) could not be computed for another 1,300 cases. The remaining course records were merged with student attribute data and information on grades given, and another set of records was removed (due to a lack of grades data and other, student attributes). In general, the removed course records were for very small courses and/or courses taught by temporary instructors whose information did not enter the University database. The resulting, complete data set had close to 2,700 observations, and these were used here for analysis.

The survey instrument and its administration are very important in producing reliable results. The course and instructor ratings data analyzed here permit only five responses, which, according to Cashin (1990), may be ideal for analytical distinction of student satisfaction and dissatisfaction. Moreover, a category for a response of "don't know" or "not sure" is not provided, helping avoid other issues (see, e.g., Arreola 1983). The five questions, as shown in Table 3, do not suffer from ambiguity, though one may argue that there is some ambiguity in the description of permitted responses, shown in Table 4. There is an opportunity for more open-ended responses on the right side of the Scanton survey sheets, where two questions are clearly posed: "What did you like most about this course?" and "How might this course be improved?" While helpful for overall assessment (Cuseo 2000), these two textual responses are non-numeric and are not analyzed here.

In terms of survey administration, instructors are told to have students acquire the survey forms from departmental administrative offices, distribute these to classmates, and provide standardized instructions to all classmates present for survey completion. These students are then to collect all responses and return these to administrative offices. The survey's administration takes approximately ten to fifteen minutes, and instructors are required to leave the classroom during this time. In general, such administrative methods comply with the core standards promoted in the student-evaluation literature (see, e.g., Cuseo 2000).

Student Grades

Instructor reporting of student grades is quite standardized and must be done within five calendar days of a course's final examination. Grades of "incomplete" and "no credit" are permitted, but these were not analyzed here. 131,071 records of student grades were assembled over a period of seven years. Almost 10,000 of these lacked a letter grade of A through F and were removed. When combined with the faculty and evaluations data, roughly 57,000 additional records were lost, because many matching instructor and evaluation records were missing. This left roughly 64,000 complete student-grade records for analysis. These were averaged across students in a single course to produce measures of average grade given and average student attributes (e.g., percent male and average verbal SAT score).

As a supplement to the average-grade-given-in-a-course model, an analysis of *individual* student GPAs was conducted. Only the latest GPA record of any specific student (during the data set's seven-year period) was used in this regression model, so almost 50% of the resulting 16,076 records came from senior-level students and no student was duplicated in the data set (thus avoiding inter-record correlations). Unfortunately, many student records did not contain achievement-test score information, so those records were not used in the full-model results (which are shown in Table 6 and are based on 9,297 complete records).

MODELS USED

The primary models formally presented here rely on weighted-least-squares (WLS) regressions of the response variables (*i.e.*, average course grade and average course evaluation scores) on student, instructor, and course attributes. Since the response variables are averages of largely independent variables (for example, the grades received/given in a single course probably do not much depend on one another, though they may be highly correlated), their variance is expected to vary inversely with the number of values averaged. Thus, the weighting variables – which characterize response value precision – are the number of students whose grades have been averaged (in the case of the grades models) and the number of evaluation forms whose scores have been averaged (in the case of the evaluations models). For further discussion of this method, one may review a number of statistical texts, including Greene (1993) and Rice (1995).

RESULTS

The WLS parameter estimates are shown in Tables 5 and 7 through 11. In each of these cases two models are run. The first incorporates (and thus controls for) all available variables simultaneously; the second retains only those variables that remain statistically significant at a level of 0.15 (p-value) through a series of step-wise deletions (where the least statistically significant explanatory variables are removed one at a time and regressions re-run, before the next variable is removed). Parameter estimates for the linear models are shown in the columns labeled "Coefficient" and their statistical significance can be deduced from the second and third columns, which contain the t-statistics and p-values for the tests of a null hypothesis that the true coefficient equals zero (and thus the associated variable does not contribute, in statistically significant way, to the model's prediction). The standard errors of parameter estimates can be deduced by simply dividing coefficients by their t-statistics.

Many explanatory indicator variables are used in these regressions, and many parameter estimates relate to certain reference variables. For example, the data distinguish gender by assigning values of one to an indicator variable for instructor gender (MALEPROF) when the instructor is male – and assigning values of zero, when the instructor is female. In addition, the fraction of a class that is represented by male students (MALE) is computed. Thus, the reference gender is female, and coefficients for female gender are not estimated. Instead, the results associated with the parameter estimates for the male-gender variables are to be appraised with respect to a female reference level of zero. Reference variables are omitted from the model because inclusion of all such categories in the presence of a constant/intercept term (the first parameter estimate) provokes perfect collinearity across categories (and the constant), resulting in a statistically unidentifiable/inestimable model. (See, e.g., Greene's [1993] discussion of this topic.) In the models examined here, the reference student is a male Caucasian freshman, the reference instructor is a female full professor, the reference course is a lower-division electrical engineering course taught in the spring semester.

Student Grades

WLS regression results of average grades data are shown in Table 5. In this model, the dependent variable is the average grade across students receiving a letter grade in each course record. Thanks to the Central Limit Theorem, the average of discrete individual grades (where an "A" counts as a 5 and an "F" counts as a 1) approaches a normal distribution as class sizes increase. The average class size in the data set is 23.1 students, with a standard deviation of 21.6 (as shown in Table 2). These class sizes are used to weight the individual course observations, so that those with less variable average grades (thanks to an averaging over larger class sizes) are weighted more. The adjusted goodness-of-fit (R_{adj}^2) for this model is 0.223, suggesting reasonable prediction.

Class size is positively associated with average grade given in a course; however, the effect is not very practically significant for most course sizes (just +0.107 grade for every 100 students added). The largest class size in the data set was 408 students, so this estimate could mean a sizable difference in average student grade received when compared to the average class size of 23 students.

The presence of certain ethnicities and gender appears to affect grades received. For example, males are associated with 0.288 lower grades than females, while Asian Americans are associated with 0.236 higher grades than non-foreign Caucasian students. International students also fare better than non-foreign Caucasian students, on average; they are estimated to receive a +0.177 higher grade. In contrast, Hispanic and African Americans are estimated to have no statistically significant influence.

The number of years "experience" (i.e., years since PhD) an instructor has does not offer predictive information in this model. However, Table 5's results suggest that male instructors assign -0.102 lower average grades than female instructors, and associate professors assign 0.121 lower average grades given than full professors. Senior lecturers, lecturers, TAs, and specialists are all estimated to grade "easier" than full professors, assigning, respectively, 0.126, 0.237, 0.292, and 0.386 higher grades, on average. It may be that certain types of instructors are assigned to certain courses where students tend to perform better or worse, but this proxying effect is probably unlikely after controlling for several course and student attributes. According to the estimates in Table 5, assistant instructors and adjunct and assistant faculty grade students consistently with full faculty. Male instructors show no special interaction with male students, in terms of grades given (via the MSMP variable).

Courses from several departments are associated with statistically different average student grades than courses from the reference department, electrical engineering. These are architectural engineering (+0.236), chemical engineering (+0.0837), mechanical engineering (+0.125), and petroleum and geotechnical engineering (+0.158).

Students in fall semester courses are graded consistently with those in spring semester courses, and average grades show no trend over the years (either up or down). Students in upper-division courses are graded higher than those in lower-division courses, and this effect is rather expected, given that students enrolled in courses more aligned with their major interest are more likely to have the interest and basis for strong performance. However, students in higher credit-hour courses tend to perform less well, receiving, on average, -0.23 grade points per extra credit hour. This may be expected if students tend to invest equal time in their courses, rather than spending proportionally more time on those courses with higher credit hours.

The initial average-grades regression model controls for four different college entrance-exam scores, but only the average verbal and quantitative SAT scores received by students in a course are statistically significant (in the presence of the other standardized test scores) and thus remain in the final model. 100 more points on the verbal SAT is estimated to improve a student's grade by 0.136 points, while 100 more points on the quantitative SAT suggests an improvement of 0.109 grade points.

Rather interestingly, Table 5's results suggest that engineering and math achievement-test scores (ENGACH and M1ACH) offer less valuable information about average course grades given than do SAT scores. However, this is not the case when analyzing GPAs of individual students; such a model was run in order to supplement the grades analysis, and these results are shown in Table 6. In this supplementary model (which lacks course and professor attribute data, due to the dependence of individual GPAs on a variety of courses taken), both the engineering and math achievement-test scores provide practically (and highly statistically) significant positive contributions to student GPA prediction. Though not shown here, the ENGACH scores explained more variation than most other variables in this model (as was seen through a process of stepwise variable additions and deletions). Rather unexpectedly and in contrast to the average-grade-given results (Table 5), Table 6 suggests that verbal SAT scores may contribute in a negative (but not very statistically significant) way to individual student GPAs. This result is reversed when the variable of ENGACH is excluded from the model; under those conditions the VSAT coefficient becomes highly statistically significant and positive. Thus, it is likely that collinearity between these VSAT and ENGACH (due, perhaps, to some shared information) is provoking this result. In any case, the contribution of engineering and math achievements scores to the model's prediction of student GPAs is substantial: 100 more points on the engineering achievement exam is predicted to produce a 0.135 higher GPA, while 100 points on the quantitative SAT is predicted to contribute less than half that level.

Average grade given in a course and individual GPAs are both important measures of student performance. Based on these results, it appears that Achievement test scores provide for better prediction at the individual student level, while SAT scores are more helpful at the class level. The course-level average-grade-given results (Table 5) are useful for assessing the impacts of class size, instructor attributes, and course characteristics on average grades given/received. And the student-level GPA results (Table 6) are most useful for assessing the impacts of student attributes alone. Viewed together, the two sets of results are consistent in their general prediction of quantitative test score contributions, student gender impacts (negative for males), level of student/level of course (increasing GPAs and grades with student year in school and upper division course status, respectively), and certain ethnicities (increasing grades and GPAs for Asian Americans and international students, and somewhat lower grades and GPAs for Native Americans). However, the fraction of African Americans in a course was estimated to not impact average grade received, at the course level, but was strongly linked to lower GPAs, at the individual student level (-0.25 GPA). One possible explanation for this distinction in model results is that such students take more difficult courses; it is difficult to be certain without a more controlled experiment and/or better data. The results of an ordered probit model for individual student grades received (see, e.g., Greene [1993] for a discussion of this model's construction) corroborate the negative sign associated with this ethnicity, at the individual student level.[1]

Course and Instructor Evaluations

Tables 7 through 11 show the results of the five course evaluation models. Among these, course organization ratings (Question 1, in Table 7) were the least predictable, having an adjusted R^2 (a measure of model fit) of just 0.053. The model of overall course rating enjoyed the highest explanatory power, with an R^2 of 0.124 (Table 11). Even though a variety of variables describing the "average" enrolled student, the instructor, and the course were available for these analyses and examined in the initial models in all cases, the low goodness of fits suggest that a great deal of course and instructor success lies with less obvious and probably much less quantifiable attributes. These attributes are likely to include instructor enthusiasm and proficiency in the specific subject, as well as student interest in the subject. Some incorporated variables (e.g., instructor level and years-since-PhD variables, and an upper-division indicator variable) may proxy for these more qualitative attributes, but they cannot duplicate them.

Average test scores aid in ratings prediction in only three of the five cases. Higher average verbal SAT scores (AVGVSAT) of enrolled students suggest lower ratings of instructors and their courses, overall (Tables 10 and 11, respectively), suggesting that such students may be more critical. However, the AVGVSAT effect is estimated to be positive in the evaluation of teaching skills, indicating that such students may judge *specific* skills less severely. The effects of this variable on overall instructor performance (Table 10) and teaching skills (Table 9) are mostly (but not entirely) negated by the effects of quantitative test scores (QSAT and QACH), and the effects are not of great practical significance (e.g., a +0.079 lower overall-instructor rating for every 100 verbal SAT points is the largest coefficient estimated for these standardized test scores). Ratings of class organization and instructor communication skill are found to be unrelated to entrance test scores.

Student ethnicities do not appear to play a role in overall class ratings, but Asian Americans appear to rate instructors more highly than non-foreign Caucasian students on the dimensions of course organization and instructor communication skill. African American students are associated with higher ratings of instructor communication skill and overall instruction, while Hispanic Americans appear to rate teaching skill more highly.

Gender plays an interesting role here, including gender interaction between students and instructors. Male students are estimated to rate all five elements of instruction more severely, particularly instructor communication skill, where a rating drop of 1.04 points is estimated for a 100-percent male class, relative to a 100-percent female class. Male instructors rate as highly as female faculty in the qualities of course organization and overall instruction (where no statistically significant difference between instructor genders was found). But, in the three other evaluation areas (i.e., communication skill, teaching skill, and overall instruction), female instructors appear to lead. For example, in terms of communication skill, a 0.67 reduction is predicted for a male instructor. However, due to gender interaction effects, this rather severe male-instructor reduction is offset. A positive male student-instructor interaction effect leaves the *net* male student-male instructor reduction at –0.75 (i.e., -1.04-0.67+0.9632, rather than -1.04-0.67, or –1.71).

Evidently, male students rate male faculty higher than they rate female faculty in four of the five questions. (No significant effect was found for question 1, regarding course organization.) However, this bias, if it in fact exists, is not enough to compensate for the lower overall ratings predicted for male instructors teaching male students. In general, given these results, one might conjecture that female students rate female instructors higher and male students rate male instructors higher, but, overall, a combination of female students and female instructors results in the highest evaluations. It may be that the average female instructor takes teaching more seriously and/or exhibits a student-preferred teaching style. These may be relatively innate gender distinctions, or they may arise from hiring more capable female instructors.

Rather interestingly, the longer an instructor has his/her PhD (a proxy for teaching experience), the lower his/her ratings on every dimension except course organization. It is not surprising that more experience translates to more organized courses (since practice makes perfect!), but none of the effects is very practically significant. For example, 20 more years of "experience" (since one's doctorate) predicts just a 0.037 higher organization rating and a 0.21 lower communication-skill rating.

Assistant professors score somewhat lower than full professors on course organization, communication skill, and the course overall, while senior lecturers score consistently higher, adjunct faculty score consistently lower, and associate professors are predicted to score no differently than full professors. The ratings performance of adjunct faculty – relative to full professors – is particularly poor; this may be due to lack of preparation and/or lesser administrative support of such faculty.

There is significant and rather consistent variation across courses given in different departments. Aerospace, engineering mechanics, and civil engineering performed best, in general, but courses in the departments of architectural engineering, mechanical engineering, and petroleum and geosystems engineering consistently "beat" courses from the model's reference department, electrical engineering (EE). Only courses in the Department of Chemical Engineering (ChemE) were estimated to receive lower evaluation scores than those in EE. There are a number of possible reasons for such ratings distinctions. For example, it may be that students in EE and ChemE are more discriminating, it may be that undergraduate EE and ChemE courses are not as well taught, or it may be that undergraduate EE and ChemE courses are evolving so rapidly (given changing technologies) that they are difficult to perfect. Further investigation may shed light on this result.

In all evaluation models except that of teaching skill, higher-unit courses are found to receive slightly lower ratings. This may be consistent with popular expectations, but it is in some contrast to work by Marsh (1987), Marsh and Dunkin (1992), and Sixbury and Cashin (1995) who concluded that course load and course difficulty do not produce negative ratings. The sheer size of the data sets used here (N_{obs} ~2,700) produces many statistically significant estimates, and their support of this popular expectation is rather consistent. However, the effect is relatively minor (on the order of 0.1 points for every additional credit hour), and the number of units or credit hours in a course is only a proxy for workload. A better test of this expectation/hypothesis would distinguish evaluations of comparable courses assigning different workloads, rather than try to relate evaluations of distinct courses.

Existing literature does suggest that larger class sizes and required courses are associated with lower ratings (Cashin 1988, Feldman 1984, Braskamp & Ory 1994, Marsh and Dunkin 1992). Here the variable of CLASSIZE was estimated to have a slightly negative effect on students' evaluations of course organization (-0.0007 per student, as shown in Table 7), teaching skill (-0.0009, as shown in Table 9), and overall instructor rating (-.0005 per student, as shown in Table 10). Thus, if one were to teach to 100 students, instead of a single student, one might expect reductions on the order of –0.05 to –0.09 in certain areas; these are not of much practical significance – though they are statistically significant. The suggestion that required courses are associated with lower ratings is consistent with results achieved here – and these results have exhibit some practical significance. The consistently positive sign on the variable UPPER supports the hypothesis that upper-division courses are preferred by the students enrolled in them; this result is to be expected given that many lower-division courses are required and general in discipline (in contrast to active student selection of major program of study and many upper-division courses). The effect of this variable is strongest (+0.17 points) for evaluation of instructor communication skill and weakest (+0.05) for teaching skill.

The data sets used here come exclusively from engineering courses, and it should be noted that there is good evidence that courses in engineering and other "hard sciences" rate lower than those in the humanities and social sciences (see, e.g., Feldman 1978 and Cashin 1990). This may be due to a variety of factors, including characteristics of the students, faculty, and subject matter. However, in general, research has concluded that "students are generous evaluators" (Wankat and Oreovicz, 1993, p. 312). In other words, the average score for courses and their instructors often suggest an "above average" rating. In the evaluations responses used here (Table 4), there is no "average performance" category, but the average response of 3.85, on a scale of 1 to 5, suggests that U.T. Austin students find their engineering courses and instructors to be "very good" – on average. It would be interesting to compare similar ratings across other universities that students have attended, since these results may reflect the College's top-10 ratings (according to *U.S. News and World Report* [2000]) in all 6 of its undergraduate departments.[2]

Evaluations and Grades

A possible weakness in university use of student evaluations is the favorable rating of instructors who "grade easy." Cohen (1981) did find some evidence of lower ratings when students' perceptions of their personal performance fell, but Theall *et al.* (1990) could not conclude that students receiving higher course grades rate such courses more highly. The anonymity of student evaluation forms precludes our linking course grades and evaluation responses of individual students; however, we are able to examine indications of this relation at the course level. For example, simple correlations of average grades given in a course and an average rating on each of the five survey questions are the following: +0.050 for course organization, +0.116 for instructor communication skill, +0.103 for teaching skill, +0.138 for the instructor overall, and +0.175 for the course overall. All of these are highly statistically significant and suggest that grades given are positively related to ratings – everything else constant. However, there are many other variables that may be proxied by grades given (most importantly, student ability). A multivariate analysis is needed.

In this multivariate analysis of evaluation scores, the relationship between ratings and grades was estimated to be positive (and highly statistically significant) in four of the five models (Tables 8 through 11). But a variable to assess grades biases, "GRD-GPA" (grade minus GPA), was found to be statistically insignificant for the prediction of all five areas of evaluation. The simultaneous control for these variables illuminates the deeper dependencies, hopefully minimizing any spurious correlations.

It should be mentioned that the GPA variable used (AVGGPA) contains not only prior but also the most recent grades given to students, including those grades given in each course under examination. So it is picking up some of the grades-given effect and provoking some undesirable collinearity in the explanatory variable set (which produces less precise estimation of parameters associated with these two variables). However, one grade – among a set of roughly 5 to 40 that comprise the AVGGPA variable – is unlikely to significantly impact this variable or prediction precision. Thus, one can draw some inferences.[3]

It is comforting to see that biased grading (in the form of average grades minus average GPAs) is having no estimable effect on evaluation scores, at least in this aggregate, course-based context. Students rarely know their final grades in advance of these evaluations, so the results do make sense. Moreover, positive faculty-student interaction may bring out better learning on the part of the students, and Cohen (1981) has found somewhat positive correlations between evaluations and better learning. Thus, the positive effect of grades given (on ratings) may be simply the result of better learning raising grades and evaluation scores. Biased (easy or hard) grading seems to have no effect.

CONCLUSIONS

This research investigated University of Texas at Austin College of Engineering multi-year data bases containing grades given to undergraduate College of Engineering students and course evaluations received by their instructors. A variety of variables were considered simultaneously, offering much more insight than the correlation analyses that characterize the great majority of the literature in this controversial area.

Based on the results of weighted-least-squares regressions of average grades given across a sample of over 2,500 courses, U.T. College of Engineering male faculty are likely to assign lower grades than female faculty, while full faculty, lecturers, and teaching assistants are found to assign higher grades than associate, assistant, and adjunct faculty. Controlling for student entrance-test scores and year in school, results suggest that non-Asian and non-foreign males taking high credit-hour lower-division courses receive lower grades. When compared to faculty teaching electrical engineering courses, those teaching chemical, mechanical, and petroleum-and-gas engineering courses appear to assign higher grades.

Based on the results of weighted-least-squares analyses of average evaluation given (on each of five different questions, in over 2,500 courses), female students and African Americans rated their courses and instructors higher, on average. Interesting gender interaction effects – between students and their instructors – were evident as well, suggesting that same-gender preferences/biases exist. Overall, male faculty tended to rate somewhat lower than female faculty, and the longer a faculty member had been teaching (since receiving her/his PhD) was associated with lower ratings. However, assistant and associate faculty also rated somewhat lower (on three of the five questions). Students in architectural, civil, aerospace, mechanical, and petroleum engineering rated their courses and instructors more highly, on average, than did electrical and chemical engineering students.

On average, higher grades given/received in a course were associated with more favorable evaluations. But grade biases (captured by a grade-minus-GPA variable) had no statistically significant effect. Thus, it seems that positive instructor-student interaction produces better student performance – and student evaluation of their instructors.

These and other results are of interest to engineering educators because they provide information for more careful appreciation (and comparison) of student and instructor performance. Such information is expected to aid in decisions of new-student admissions, instructor hiring and promotion, and mentorship of both instructors and students.

ACKNOWLEDGEMENTS

The author would like to thank Young-Jun Kweon, Margaret Stephens, Rick Moseley, and Lisa Barden for their acquisition and manipulation of data sets and several results. The author also is grateful to support provided by Alvin Meyer, the University of Texas at Austin's College of Engineering, and the Jeff and Carol Heller Engineering Excellence Fund.

REFERENCES

Abrami, P.C., L. Leventhal, and R.P. Perry. 1982. "Educational seduction." *Review of Educational Research* 52: 446-464.

Arreola, R.A. 1983. "Establishing successful faculty evaluation and development programs." In A. Smith (ed.), *Evaluation Faculty and Staff*. New Directions for Community Colleges 41. San Francisco: Jossey-Bass.

Braskamp, L.A., and J.C. Ory. 1994. *Assessing Faculty Work: Enhancing Individual and Institutional Performance*. San Francisco: Jossey-Bass.

Cashin, W.E. 1988. "Student ratings of teaching: A summary of the research." *IDEA Paper No. 20*. Manhattan, Kansas: Kansas State University, Center for Faculty Evaluation and Development.

Cashin, William E. "Students do rate different academic fields differently." *New Directions for Teaching and Learning* 43 (Fall): 113-121.

Cohen, Peter A. 1981. "Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies." *Review of Educational Research* 51: 281-309.

Cohen, Peter A. 1990. "Bringing research into practice." *New Directions for Teaching and Learning* 43 (Fall): 123-132.

Cuseo, Joe. 2000. "The case for student evaluation of college courses." Weblink at http://www.brevard.edu/fyc/FYA_contributions/CuseoLink.htm. Available 20 October 2000.

Feldman, K.A. 1978. "Course characteristics and college students' ratings of their teachers and courses: What we know and what we don't." *Research in Higher Education* 9: 199-242.

Feldman, K.A. 1984. "Class size and college students' evaluations of teachers and courses: A closer look." *Research in Higher Education* 30 (2): 137-194.

Greene, William H. 1993. *Econometric Analysis, Second Edition*. New York, NY: McMillan.

Marsh, H.W. 1987. "Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research." *International J of Educational Research* 11: 253-388.

Marsh, H.W., and M. Dunkin. 1992. "Students evaluations of university teaching: A multidimensional perspective." In J.C. Smart (ed.), *Higher Education: Handbook of Theory and Research* 8 (pp. 143-233). New York: Agathon Press.

McKnight, Philip. 1990. "The evaluation of postsecondary classroom teaching: A wicked problem?" *New Directions for Teaching and Learning* 43 (Fall): 57-62.

Ory, John C. 1990. "Student ratings of instruction: Ethics and practice." *New Directions for Teaching and Learning* 43 (Fall): 63-74.

Rice, John A. 1995. *Mathematical Statistics and Data Analysis, 2nd Edition*. Belmont, CA: Duxbury Press.

Seldin, P. 1993. “How colleges evaluation professors, 1983 vs. 1993.” *AAHE Bulletin* 46 (2): 6-8, 12.

Sixbury, G.R., and W.E. Cashin. 1995. “Description of database for the IDEA diagnostic form.” *IDEA Technical Report No. 9*. Manhattan, Kansas: Kansas State University, Center for Faculty Evaluation and Development.

Theall, M., J. Franklin, and L.H. Ludlow. 1990. “Attributions and retributions: Student ratings and the perceived causes of performance.” Paper presented at the annual meeting of the American Educational Research Association, Boston.

Theall, Michael, and Jennifer Franklin (eds.). 1990. *Students Ratings of Instruction: Issues for Improving Practice*. New Directions for Teaching and Learning 43 (Fall). San Francisco: Jossey-Bass.

U.S. News & World Report. 2000. “Best Undergraduate Engineering Programs.” Vol. 129 (10), September 11.

Wankat, Phillip C., and Frank S. Oreovicz. 1993. *Teaching Engineering*. Chapter 16. New York, NY: McGraw-Hill, Inc.

Table 1: Definition of Variables Used

Variable	Description
AVGGRADE	Average grade of students in a course (5=A, 4=B, 3=C, 2=D, 1=F)
AVGSCRQ1	Average score for evaluation question 1 in a course
AVGSCRQ2	Average score for evaluation question 2 in a course
AVGSCRQ3	Average score for evaluation question 3 in a course
AVGSCRQ4	Average score for evaluation question 4 in a course
AVGSCRQ5	Average score for evaluation question 5 in a course
AVGVSAT	Average Verbal SAT score of students in a course (as re-centered by the College Board in 1995)
AVGQSAT	Average Quantitative SAT score of students in a course (as re-centered by the College Board in 1995)
AVGENACH	Average English Achievement test score of students in a course
AVGM1ACH	Average Math1 Achievement test score of students in a course
AVGGPA	Average GPA of students in a course (for UT courses taken prior to & including the course semester)
AVGCLSS	Average Classification of students in a course (Freshman = 1, Sophomore = 2, Junior = 3, Senior = 4)
AVGMALE	Fraction of male students in the course
AVGCAUCA	Fraction of Caucasian students in the course
AVGNATAM	Fraction of Native American students in the course
AVGAFRAM	Fraction of African American students in the course
AVGASIAM	Fraction of Asian American students in the course
AVGHISAM	Fraction of Hispanic American students in the course
AVGINTNT	Fraction of non-US students in a course
MSMP	Fraction of male students * MALEPROF
MALEPROF	Instructor is male? (Male = 1, Female = 0)
YEAR_P	Course year minus PHDYEAR
FULLPROF	Instructor is a full professor (Yes = 1, No = 0)
ASSTPROF	Instructor is an assistant professor (Yes = 1, No = 0)
ASSOPROF	Instructor is an associate professor (Yes = 1, No = 0)
ADJUNCT	Instructor is an adjunct professor (Yes = 1, No = 0)
SENIRLEC	Instructor is a senior lecturer (Yes = 1, No = 0)
LECTURER	Instructor is a lecturer (Yes = 1, No = 0)
ASSTINST	Instructor is an assistant instructor [title given to PhD students in their final year] (Yes = 1, No = 0)
SPECIAL	Instructor is a specialist/non-adjunct outside instructor (Yes = 1, No = 0)
TA	Instructor is a TA (Yes = 1, No = 0)
ARE	Course is in Architectural Eng. (Yes = 1, No = 0)
ASE	Course is in Aerospace Eng. (Yes = 1, No = 0)
CE	Course is in Civil Eng. (Yes = 1, No = 0)
CHE	Course is in Chemical Eng. (Yes = 1, No = 0)
EE	Course is in Electrical Eng. (Yes = 1, No = 0)
EM	Course is in Engineering Mechanics (Yes = 1, No = 0)
GE	Course is in General Eng. [supplemental to math/science] (Yes = 1, No = 0)
ME	Course is in Mechanical Eng. (Yes = 1, No = 0)
PGE	Course is in Petroleum & Geosystems (Yes = 1, No = 0)
FALL	Course given in fall semester (Fall = 1, Spring = 0)
LOWER	Course classified as lower division (Yes = 1, No = 0)
UPPER	Course classified as upper division (Yes = 1, No = 0)
UNITS	Number of credit hours for the course
YEAR_C	Course year minus 1992

CLASSIZE	Same as "NUMSTUD"
NUMSTUD	Number of students in the course (weighting variable for grades data)
NUMANSW	Number of evaluations received for the course (weighting variable for evaluations data)

** Shaded rows are dependent variables. Italicized variables represent reference groups.*

Table 2: Data Statistics

Variables	Mean	SD	Minimum	Maximum	N
AVGGRADE	2.917	0.511	0	4	2748
AVGSCRQ1	3.891	0.529	1.7895	5	2748
AVGSCRQ2	3.866	0.624	1	5	2748
AVGSCRQ3	3.910	0.483	1	5	2748
AVGSCRQ4	3.927	0.595	1	5	2748
AVGSCRQ5	3.779	0.550	1	5	2748
AVGVSAT	584.68	41.96	230	800	2740
AVGQSAT	655.69	36.06	310	800	2740
AVGENACH	535.93	43.26	270	734	2712
AVGM1ACH	633.96	44.08	410	800	2677
AVGGPA	3.003	0.238	1.6610	3.9710	2748
AVGCLSS	3.452	0.692	1	4	2748
CLASSIZE	23.161	21.565	1	408	2748
AVGCAUCA	0.584	0.176	0	1	2748
AVGNATAM	0.003	0.015	0	0.25	2748
AVGAFRAM	0.029	0.053	0	1	2748
AVGASIAM	0.136	0.122	0	1	2748
AVGHISAM	0.127	0.106	0	1	2748
AVGINTNT	0.121	0.127	0	1	2748
AVGFRESH	0.058	0.171	0	1	2748
AVGSOPHO	0.083	0.153	0	1	2748
AVGJUNIO	0.207	0.207	0	1	2748
AVGSENIO	0.652	0.337	0	1	2748
AVGMALE	0.822	0.140	0	1	2748
MSMP	0.757	0.262	0	1	2748
MALEPROF	0.918	0.275	0	1	2748
YEAR_P	21.460	11.268	-5	52	2748
FULLPROF	0.530	0.499	0	1	2748
ASSTPROF	0.109	0.311	0	1	2748
ASSOPROF	0.138	0.345	0	1	2748
ADJUNCT	0.012	0.107	0	1	2748
SENIRLEC	0.097	0.296	0	1	2748
LECTURER	0.083	0.275	0	1	2748
TA	0.012	0.111	0	1	2748
SPECIAL	0.019	0.135	0	1	2748
ASSTINST	0.000	0.019	0	1	2748
ARE	0.052	0.221	0	1	2748
ASE	0.112	0.315	0	1	2748
CE	0.137	0.344	0	1	2748
CHE	0.113	0.316	0	1	2748
EE	0.291	0.455	0	1	2748
EM	0.045	0.207	0	1	2748
ME	0.231	0.421	0	1	2748
PGE	0.020	0.139	0	1	2748
FALL	0.466	0.499	0	1	2748

UPPER	0.791	0.407	0	1	2748
UNITS	2.864	0.543	1	4	2748
YEAR_C	3.215	1.968	0	6	2748

Table 3: Evaluation Questions

Question #	Question Text
1	The organization of the course was…
2	The instructor's skill in communicating information effectively was…
3	The instructor's skill in helping me think for myself in this course was…
4	Overall, this instructor was…
5	Overall, this course was…

Table 4: Evaluation Responses

Response Level	Definition of Response Level
1	Very Unsatisfactory
2	Unsatisfactory
3	Satisfactory
4	Very Good
5	Excellent

Variables	Initial Model			Final Model		
	Coefficients	t-stats	p-values	Coefficients	t-stats	p-values
(Constant)	2.14E+00	6.53	0.000	1.98E+00	9.34	0.000
AVGVSAT	1.54E-03	4.16	0.000	1.32E-03	4.23	0.000
AVGQSAT	8.02E-04	1.76	0.079	1.09E-03	3.04	0.002
AVGENACH	-2.84E-04	-0.92	0.357			
AVGM1ACH	3.03E-04	0.99	0.324			
AVGCLSS	5.49E-02	2.64	0.008	5.85E-02	2.96	0.003
CLASSIZE	1.14E-03	6.01	0.000	1.07E-03	6.03	0.000
AVGNATAM	-1.18E+00	-1.87	0.062	-1.11E+00	-1.77	0.077
AVGAFRAM	-2.15E-01	-1.08	0.282			
AVGASIAM	2.17E-01	2.23	0.026	2.36E-01	2.78	0.005
AVGHISAM	3.62E-02	0.33	0.739			
AVGINTNT	1.54E-01	1.48	0.139	1.77E-01	1.87	0.062
AVGMALE	-3.72E-01	-1.46	0.144	-2.88E-01	-3.60	0.000
MSMP	7.02E-02	0.27	0.787			
MALEPROF	-1.48E-01	-0.69	0.489	-1.02E-01	-3.34	0.001
YEAR_P	-1.35E-03	-1.41	0.158			
ASSTPROF	-2.54E-02	-0.79	0.427			
ASSOPROF	-1.30E-01	-4.62	0.000	-1.21E-01	-5.05	0.000
ADJUNCT	-1.04E-01	-1.30	0.194			
SENIRLEC	1.12E-01	4.10	0.000	1.26E-01	4.91	0.000
LECTURER	2.25E-01	7.49	0.000	2.37E-01	8.40	0.000
TA	2.66E-01	2.78	0.006	2.92E-01	3.26	0.001
SPECIAL	3.78E-01	4.88	0.000	3.86E-01	5.14	0.000
ASSTINST	2.67E-01	0.50	0.614			
ARE	2.10E-01	4.96	0.000	2.36E-01	6.55	0.000
ASE	-4.88E-02	-1.30	0.195			
CE	-1.38E-02	-0.40	0.692			
CHE	7.52E-02	2.35	0.019	8.37E-02	2.87	0.004
EM	-4.01E-02	-0.86	0.390			
ME	1.06E-01	3.73	0.000	1.25E-01	5.72	0.000
PGE	1.29E-01	1.75	0.080	1.58E-01	2.29	0.022
FALL	-3.89E-03	-0.24	0.811			
UPPER	8.07E-02	2.28	0.023	8.43E-02	2.68	0.007
UNITS	-2.33E-01	-14.35	0.000	-2.30E-01	-14.83	0.000
YEAR_C	1.02E-03	0.22	0.823			
R-sq	0.232			0.229		
Adj R-sq	0.222			0.223		
N	2666			2740		

Table 5: WLS Model of Average Grade of Students in Course

Variables	General Model		
	Coefficients	t-stats	p-values
(Constant)	0.6638	10.31	0.000
VSAT	-0.0002	-1.57	0.116
QSAT	0.0006	4.09	0.000
ENGACH	0.0013	12.74	0.000
M1ACH	0.0014	11.25	0.000
MALE	-0.2031	-12.92	0.000
NATAMR	-0.1022	-1.02	0.306
AFRAMR	-0.2557	-8.16	0.000
HISAMR	-0.0846	-4.53	0.000
ASIAMR	0.0475	2.50	0.012
INTNTL	0.2610	7.34	0.000
SOPHOR	0.3263	15.59	0.000
JUNIOR	0.4278	20.23	0.000
SENIOR	0.6071	35.41	0.000
R-sq	0.2583		
Adj R-sq	0.2573		
Nobs	9297		

Table 6: OLS Model of Individual Student GPAs

Variables	Initial Model			Final Model		
	Coefficients	t-stats	p-values	Coefficients	t-stats	p-values
(Constant)	3.9732	9.33	0.000	4.0845	37.38	0.000
AVGVSAT	-3.9125E-04	-0.82	0.414			
AVGQSAT	3.6050E-04	0.59	0.556			
AVGENACH	5.1098E-04	1.28	0.200			
AVGM1ACH	3.8565E-04	0.97	0.330			
AVGGRADE	3.2273E-02	1.18	0.238			
GRD_GPA	-8.4833E-02	-1.11	0.268			
AVGCLSS	-3.4669E-02	-1.28	0.199			
CLASSIZE	-6.1852E-04	-2.51	0.012	-6.7059E-04	-3.13	0.002
AVGNATAM	-3.7574E-02	-0.05	0.963			
AVGAFRAM	0.1202	0.47	0.642			
AVGASIAM	0.1957	1.56	0.120	0.1667	1.47	0.142
AVGHISAM	0.1035	0.74	0.458			
AVGINTNT	0.1065	0.79	0.427			
AVGMALE	-0.5206	-1.59	0.112	-0.2144	-2.15	0.032
MSMP	0.3223	0.96	0.335			
MALEPROF	-0.2451	-0.89	0.374			
YEAR_P	9.6736E-04	0.78	0.433	1.8588E-03	1.97	0.049
ASSTPROF	-3.5869E-02	-0.87	0.384			
ASSOPROF	-3.1682E-02	-0.87	0.385			
ADJUNCT	-7.8924E-02	-0.77	0.441			
SENIRLEC	0.1054	2.98	0.003	0.1117	3.40	0.001
LECTURER	-4.0482E-02	-1.04	0.301			
TA	-4.8747E-02	-0.40	0.692			
SPECIAL	9.2587E-02	0.92	0.355			
ASSTINST	9.9615E-02	0.15	0.884			
ARE	0.2569	4.63	0.000	0.2344	4.75	0.000
ASE	0.2742	5.65	0.000	0.2496	5.99	0.000
CE	0.2529	5.59	0.000	0.2227	5.82	0.000
CHE	-5.2927E-02	-1.28	0.200	-6.9602E-02	-1.79	0.074
EM	0.2960	4.93	0.000	0.2361	4.43	0.000
ME	0.1471	4.01	0.000	0.1305	4.09	0.000
PGE	0.1028	1.07	0.283			
FALL	3.0131E-03	0.14	0.886			
UPPER	0.1666	3.64	0.000	0.1068	4.06	0.000
UNITS	-9.4039E-02	-4.33	0.000	-0.1028	-5.21	0.000
YEAR_C	1.1370E-02	1.93	0.054	1.0072E-02	1.88	0.061
R-sq	0.063			0.058		
Adj R-sq	0.051			0.053		
N	2665			2747		

Table 7: WLS Model of Average Evaluation of Course Organization

Variables	Initial Model			Final Model		
	Coefficients	t-stats	p-values	Coefficients	t-stats	p-values
(Constant)	4.1594	8.70	0.000	4.3075	13.20	0.000
AVGVSAT	-3.5945E-04	-0.67	0.504			
AVGQSAT	9.6926E-04	1.41	0.159			
AVGENACH	-1.6490E-04	-0.37	0.712			
AVGM1ACH	-4.8665E-04	-1.10	0.273			
AVGGRADE	0.2710	3.43	0.001	0.1875	7.01	0.000
GRD_GPA	-9.8941E-02	-1.15	0.250			
AVGCLSS	-5.4920E-02	-1.81	0.070	-4.7086E-02	-1.69	0.091
CLASSIZE	4.1415E-05	0.15	0.881			
AVGNATAM	-0.2923	-0.32	0.750			
AVGAFRAM	0.5733	1.98	0.048	0.5153	1.86	0.063
AVGASIAM	0.2112	1.50	0.135	0.2751	2.10	0.036
AVGHISAM	0.2567	1.64	0.101	0.2359	1.59	0.112
AVGINTNT	0.1708	1.13	0.257	0.2383	1.83	0.067
AVGMALE	-1.1064	-3.01	0.003	-1.0452	-3.00	0.003
MSMP	1.0322	2.75	0.006	0.9632	2.68	0.007
MALEPROF	-0.7312	-2.36	0.018	-0.6688	-2.26	0.024
YEAR_P	-1.1476E-02	-8.28	0.000	-1.0722E-02	-8.69	0.000
ASSTPROF	-8.7799E-02	-1.90	0.058	-6.5026E-02	-1.58	0.115
ASSOPROF	-5.1578E-02	-1.26	0.208			
ADJUNCT	-0.4515	-3.93	0.000	-0.4359	-3.87	0.000
SENIRLEC	0.2446	6.15	0.000	0.2631	6.97	0.000
LECTURER	-4.4941E-02	-1.02	0.306			
TA	-0.2741	-1.98	0.047	-0.2293	-1.73	0.084
SPECIAL	0.1053	0.94	0.349			
ASSTINST	0.3934	0.51	0.607			
ARE	0.2289	3.67	0.000	0.2050	3.56	0.000
ASE	0.3189	5.85	0.000	0.3007	6.20	0.000
CE	0.3832	7.54	0.000	0.3646	8.06	0.000
CHE	-7.9003E-02	-1.70	0.089	-8.7081E-02	-1.95	0.051
EM	0.4592	6.81	0.000	0.4348	6.88	0.000
ME	1.2511E-01	3.03	0.002	0.1307	3.42	0.001
PGE	0.3072	2.86	0.004	0.2814	2.81	0.005
FALL	-2.5231E-02	-1.07	0.285			
UPPER	0.1801	3.50	0.000	0.1690	3.47	0.001
UNITS	-0.1006	-4.12	0.000	-9.4876E-02	-4.00	0.000
YEAR_C	1.7422E-03	0.26	0.793			
R-sq	0.129			0.125		
Adj R-sq	0.117			0.117		
N	2665			2747		

Table 8: WLS Model of Average Evaluation of Instructor’s Communication Skills

Variables	Initial Model			Final Model		
	Coefficients	t-stats	p-values	Coefficients	t-stats	p-values
(Constant)	3.7204	10.19	0.000	3.7340	11.43	0.000
AVGVSAT	-3.4475E-04	-0.84	0.402	7.4340E-04	1.92	0.055
AVGQSAT	8.1807E-04	1.56	0.119			
AVGENACH	-4.1581E-04	-1.22	0.224	-5.3833E-04	-1.88	0.060
AVGM1ACH	-6.7530E-05	-0.20	0.842			
AVGGRADE	0.2190	3.63	0.000	0.1531	7.64	0.000
GRD_GPA	-8.7254E-02	-1.33	0.184			
AVGCLSS	1.8654E-03	0.08	0.936			
CLASSIZE	-7.9804E-04	-3.77	0.000	-8.9643E-04	-4.67	0.000
AVGNATAM	-0.5394	-0.77	0.442			
AVGAFRAM	0.3247	1.47	0.142			
AVGASIAM	-2.8396E-02	-0.26	0.792			
AVGHISAM	0.2010	1.68	0.093	0.1715	1.53	0.126
AVGINTNT	-2.7700E-02	-0.24	0.810			
AVGMALE	-0.6579	-2.34	0.019	-0.6990	-2.61	0.009
MSMP	0.6998	2.44	0.015	0.7176	2.61	0.009
MALEPROF	-0.4800	-2.03	0.043	-0.4958	-2.19	0.028
YEAR_P	-6.8465E-03	-6.47	0.000	-6.1600E-03	-7.40	0.000
ASSTPROF	-0.0365	-1.03	0.301			
ASSOPROF	-1.4422E-02	-0.46	0.645			
ADJUNCT	-0.3418	-3.89	0.000	-0.3424	-3.96	0.000
SENIRLEC	6.8807E-02	2.26	0.024	6.9917E-02	2.43	0.015
LECTURER	-0.1212	-3.61	0.000	-0.1138	-3.61	0.000
TA	3.5824E-02	0.34	0.734			
SPECIAL	3.1584E-02	0.37	0.713			
ASSTINST	0.5382	0.92	0.357			
ARE	0.1252	2.63	0.009	0.1200	2.83	0.005
ASE	0.2556	6.14	0.000	0.2511	6.84	0.000
CE	0.2732	7.04	0.000	0.2614	7.65	0.000
CHE	-7.8281E-02	-2.21	0.027	-8.0087E-02	-2.36	0.018
EM	0.2677	5.20	0.000	0.2626	5.68	0.000
ME	6.0862E-02	1.93	0.053	6.7225E-02	2.44	0.015
PGE	0.1990	2.43	0.015	0.1909	2.47	0.014
FALL	-1.5004E-02	-0.83	0.405			
UPPER	4.8306E-02	1.23	0.219	4.9951E-02	2.23	0.026
UNITS	-2.5191E-02	-1.35	0.177			
YEAR_C	-1.5317E-03	-0.30	0.762			
R-sq	0.121			0.118		
Adj R-sq	0.109			0.112		
N	2665			2711		

Table 9: WLS Model of Average Evaluation of Instructor's Teaching Skill

Variables	Initial Model			Final Model		
	Coefficients	t-stats	p-values	Coefficients	t-stats	p-values
(Constant)	4.1761	9.23	0.000	3.7549	11.37	0.000
AVGVSAT	-5.8537E-04	-1.15	0.250	-7.8787E-04	-2.00	0.045
AVGQSAT	8.7795E-04	1.35	0.177	0.0007	1.52	0.128
AVGENACH	-3.3330E-04	-0.79	0.431			
AVGM1ACH	-4.0486E-04	-0.96	0.335			
AVGGRADE	0.2813	3.77	0.000	0.2167	8.47	0.000
GRD_GPA	-7.8522E-02	-0.96	0.335			
AVGCLSS	-6.4531E-02	-2.25	0.024	-5.9906E-02	-2.18	0.029
CLASSIZE	-4.0777E-04	-1.56	0.120	-4.6074E-04	-1.86	0.063
AVGNATAM	-0.2565	-0.30	0.768			
AVGAFRAM	0.4770	1.74	0.082	0.4342	1.64	0.101
AVGASIAM	0.0973	0.73	0.466			
AVGHISAM	0.1845	1.24	0.213			
AVGINTNT	0.0563	0.40	0.693			
AVGMALE	-0.7926	-2.28	0.023	-0.3502	-2.92	0.004
MSMP	0.6689	1.88	0.060	0.1722	3.35	0.001
MALEPROF	-0.4124	-1.41	0.159			
YEAR_P	-1.0704E-02	-8.17	0.000	-9.2110E-03	-8.87	0.000
ASSTPROF	-0.0763	-1.75	0.081			
ASSOPROF	-4.8645E-02	-1.26	0.209			
ADJUNCT	-0.4482	-4.12	0.000	-0.4279	-4.01	0.000
SENIRLEC	0.1183	3.14	0.002	0.1382	3.90	0.000
LECTURER	-1.0651E-02	-0.26	0.798			
TA	-8.7541E-02	-0.67	0.503			
SPECIAL	0.1556	1.46	0.143	0.1808	1.76	0.078
ASSTINST	0.5005	0.69	0.489			
ARE	0.1768	3.00	0.003	0.1545	2.93	0.003
ASE	0.3084	5.98	0.000	0.2992	6.57	0.000
CE	0.3289	6.85	0.000	0.3188	7.47	0.000
CHE	-0.1025	-2.34	0.020	-9.5430E-02	-2.24	0.025
EM	0.3790	5.94	0.000	0.3668	6.24	0.000
ME	7.5232E-02	1.93	0.054	6.6555E-02	1.95	0.051
PGE	0.2676	2.63	0.008	0.2522	2.65	0.008
FALL	-1.2204E-02	-0.55	0.584			
UPPER	0.1578	3.24	0.001	0.1559	3.34	0.001
UNITS	-7.0888E-02	-3.07	0.002	-6.5030E-02	-2.90	0.004
YEAR_C	2.8496E-03	0.45	0.649			
R-sq	0.127			0.123		
Adj R-sq	0.115			0.116		
N	2665			2739		

Table 10: WLS Model of Average Evaluation of Instructor Overall

Variables	Initial Model			Final Model		
	Coefficients	t-stats	p-values	Coefficients	t-stats	p-values
(Constant)	3.7563	9.10	0.000	3.9732	12.39	0.000
AVGVSAT	-6.3331E-04	-1.36	0.173	-7.4066E-04	-2.36	0.018
AVGQSAT	5.5544E-04	0.94	0.349			
AVGENACH	-2.1194E-04	-0.55	0.583			
AVGM1ACH	-2.6020E-04	-0.68	0.498			
AVGGRADE	0.2683	3.94	0.000	0.2289	9.87	0.000
GRD_GPA	-4.4448E-02	-0.60	0.550			
AVGCLSS	-2.1436E-02	-0.82	0.413			
CLASSIZE	-1.0445E-05	-0.04	0.965			
AVGNATAM	-0.1187	-0.15	0.881			
AVGAFRAM	0.3750	1.50	0.134			
AVGASIAM	0.1590	1.30	0.192			
AVGHISAM	0.2002	1.48	0.139			
AVGINTNT	0.1269	0.98	0.329			
AVGMALE	-0.7297	-2.30	0.022	-0.6566	-2.21	0.027
MSMP	0.6914	2.13	0.033	0.6320	2.07	0.039
MALEPROF	-0.4292	-1.60	0.109	-0.3791	-1.51	0.131
YEAR_P	-7.3084E-03	-6.11	0.000	-6.5942E-03	-6.24	0.000
ASSTPROF	-0.1170	-2.93	0.003	-9.0472E-02	-2.55	0.011
ASSOPROF	-5.6260E-02	-1.59	0.112			
ADJUNCT	-0.3762	-3.79	0.000	-0.3574	-3.67	0.000
SENIRLEC	9.7856E-02	2.85	0.004	0.1142	3.51	0.000
LECTURER	-4.1970E-02	-1.11	0.269			
TA	-4.2789E-02	-0.36	0.720			
SPECIAL	0.2000	2.06	0.040	0.2205	2.37	0.018
ASSTINST	0.5512	0.83	0.404			
ARE	0.2460	4.57	0.000	0.1981	4.29	0.000
ASE	0.3386	7.19	0.000	0.2952	8.17	0.000
CE	0.2972	6.78	0.000	0.2548	7.36	0.000
CHE	-0.1150	-2.87	0.004	-0.1126	-2.93	0.003
EM	0.3288	5.64	0.000	0.2787	5.40	0.000
ME	7.1953E-02	2.02	0.043	4.3440E-02	1.48	0.139
PGE	0.2587	2.79	0.005	0.2129	2.48	0.013
FALL	5.6863E-04	0.03	0.978			
UPPER	0.1413	3.18	0.001	0.1049	4.00	0.000
UNITS	-3.9017E-02	-1.85	0.064	-3.8731E-02	-1.91	0.056
YEAR_C	1.0016E-02	1.75	0.080	1.1905E-02	2.27	0.023
R-sq	0.136			0.131		
Adj R-sq	0.124			0.125		
N	2665			2739		

Table 11: WLS Model of Average Evaluation of Course Overall

ENDNOTES

[1] The results of the ordered probit model are not shown here because of the very low predictive power (pseudo-R^2 = 0.0291) at this level of student-course resolution. The predictive power rose substantially for this model (to a pseudo-R^2 of 0.187) when a control variable of student GPA was added, but this explanatory variable was felt to provide too much of an "assist" to the model, rendering the model unreasonable for assessment of underlying behavioral mechanisms.

[2] Architectural engineering and engineering mechanics courses are separated here, for purposes of analysis, but they are formally taught under the Departments of Civil and Aerospace Engineering, respectively. These "sub-departments" are not rated in *U.S. News and World Report* at the undergraduate level. The departments/disciplines of aerospace, chemical, civil, computer, electrical, environmental, mechanical, and petroleum engineering were rated 8th, 6th, 5th, 7th, 9th, 7th, 10th, and 2nd, respectively.

[3] Note that the variable of GPA by itself was not also included in these evaluations models, due to the perfect collinearity that would then arise with the model's constant term (and the AVGGRADE and GRD-GPA variables). Such a specification would render the model unidentifiable/inestimable.