

Grading And Assessments: Correlations Of Variables Affecting Teaching And Course Assessments

Joyce L. Ellis, National University, USA
Thomas J. Francl, National University, USA

ABSTRACT

Faculty assessments are a hot topic in academia. Mostly considered an unproven statistic holding the professor hostage for a good grade, the benefits are still in discussion. At course end, we expect students to have the ability to analyze what we think they should have learned. In reality, most students do not have the maturity to realize correlation for what you put into your education is what you get out. Consensus is that it does not happen the way we plan. Adding to the assessment issue is the clear fact that today most universities are adjunct professor-driven. Given these facts, the authors have statistically studied teaching assessments and associated grade point averages (GPA) representing academic rigor at a major non-profit university over the past three years. The authors' hypothesis is that there is no correlation between the two statistics. This paper reviews the literature, provides the study methodology, and presents the findings.

Keywords: Faculty Assessments; Teaching Assessments; Academic Rigor

INTRODUCTION

Issues surrounding faculty and student assessment processes have long been controversial throughout most universities. Most mandate evaluations of faculty through end-of-course student-answered questionnaires seeking student ratings on teaching effectiveness, while collecting information is utilized to determine professor career implications. This implication creates criticism from faculty and questions validity, reliability of the surveys, and the degree to which student opinions should influence the career of academics.

Other variables applied to the assessment composition question faculty rigor within analysis of course grade point averages (GPAs) compared with faculty student assessments. This study examines faculty evaluations, assessment of faculty rigor established from course GPAs, and the potential correlation regarding a grade-rating relationship. Traditional assessment questionnaires distributed through end-of-course surveys by the university's Office of Institutional Research and Assessments (OIRA) department in a non-profit university are the basis of the three-year study. Corresponding class GPA statistics contribute in statistical input for comparison purposes.

Research questions include: Do students rate professors higher, believing they will receive higher grades or because they believe the courses widened their intellectual capacity and critical thinking, or is the relationship between grades and teaching evaluation a complex system of relationships including student and professor motivation, instructor experience and education, as well as student response to instructor pedagogical choices?

The purpose of this research is to examine the relationship between student GPAs and professor assessments, questioning if there is statistical significance between the dependent variable - professor assessment scores - and the independent variable - student GPA. As a result of this research question, the null hypothesis tested in this study is:

H0: There is no correlation between the student GPAs and their assigned assessment scores for their professor.

LITERATURE REVIEW

Studies reviewed from experienced faculty and deans across the United States, with respect to validity of student professor assessments and correlations with professor grading reference the research questions of this study. Does a correlation exist between student assessment of professors and class grading (GPA)? In addition, should the results of these statistics be used for professor salary adjustment and ranking? Wilson (1998) referred to her studies stating, “Studies raise questions about validity of student rating of teaching and the tendency of professor to ‘teach to the evaluations’”. Her research found “professors who want high ratings have learned to dumb down material, inflate grades and keep students entertained. The ratings can make or break a professor’s career even though they do not always accurately measure teaching skills” (p. A12).

References include adjunct, full-time professors, tenured and non-tenured professors in analysis of professor grading and teaching assessments. Sonner (2000) found “adjunct professors, hired on a term-by-term basis, are easily replaced; thus, most face serious pressure to earn good evaluations from students. Keeping students happy may mean giving higher, potentially inflated grades” (p.1). Their two-year studying concluded that adjunct instructors do give higher grades than full-time faculty. Filetti, Wright, and King (2010) concluded:

Tenured professors tend to award fewer A's than tenure-track professors. The authors posit that at universities where emphasis is placed upon teaching, two specific patterns emerge: reciprocity may be an unspoken agreement between some faculty and students with regard to the exchange of good grades for good evaluations, or with experience comes rigor. (p.1)

In today’s academic world, adjuncts carry a majority of the teaching load at major universities. This study includes comparisons in statistical output between full-time and adjunct professors from testing the hypothesis that there is no correlation between the student GPAs and their assigned assessment scores for their professors.

Further literature addresses student bias, suggesting necessity of multiple methods and elements for professor evaluations. Baldwin and Blattner (2003) address guarding against potential bias in student evaluations, suggesting alternative methods to using single evaluations from students. They suggest multiple methods for teaching evaluations, including “putting results and comments into context, collecting formal and informal feedback from students in a variety of ways, and inviting others to evaluate your teaching (pp.30-31). Johnson (2002) collected student data analyzing teaching assessments with grade scores both before and after students received their final grades. Her analysis concludes “evidence of a biasing effect of students’ grades on student evaluations of teaching” (p.16). Universities relying on student teaching evaluations in administrative decisions effecting professors’ futures provide incentives for manipulation of grades by faculty. Consequences from faculty manipulation could be degradation of education quality.

Some literature questions student maturity in assessments and emphasize professor focus on high quality deliverables and academic rigor. Heller (2010) questions who professors’ customers are and who should evaluate them, stating:

Students are the ultimate consumer of the service, but I don’t believe professors should consider them the customer. Why? Because the vast majority of students do not have the maturity and life experiences to understand what is best for them. The objective of what is best for them should simply be delivering the highest capable students upon graduation to the business workplace. (p. 2)

In an updated study, Heller (2013) states, “Academic rigor is a necessary component in creating an elevated education experience. So how does the combination of both online and university professor rating systems result in applied pressure for professors to decrease academic rigor...most students reward professors that have easy course requirements and those that are liberal graders” (p. 1).

With academic rigor at stake and student immaturity in question, should student evaluations be the single factor in professor assessments? Carrell and West (2012) studied student evaluations concluding, “students confuse course grades with long-term learning and reward those professors who hand out the A’s”. They noted developing trends speculating “young, inexperienced faculty members were teaching more directly to the content on course exams, while more experienced faculty members taught in a way that promoted deep learning...if long-term learning is the mark of higher-quality faculty members, our research provides ample evidence that student evaluations do not reliably measure professor quality” (pp. 4-7). In that same year, Decatur (2012) noted, “Experiencing the anxiety of evaluation day...student evaluations should only be a part of the overall evaluation of an instructor’s effectiveness, and they’re best read in the context of peer observation” (p. 4). McCulloch-Lovell (2013) state, “Student evaluations can be effective if useful information prompts reflection from professors...at larger institutions, teachers often focus only on scores, which can serve as a rough numeric...this is a less effective model than providing observations of teaching styles and course objectives” (p. 4).

Some research reviewed indicates definitive opposing opinions on student assessment of their professors. Davidovitch and Soem (2009) conclude “correlation between the students’ grades and the lecturers are non-existent. This is nothing but a myth amongst the academic body and is not based on reality...the research still points out, just as one breaks or dismantles a myth, that there are some additional elements which are beyond the efficiency of teaching as we touch different levels of interaction between student and lecturer” (pp. 46-47). Sandefer (2012) does not completely agree with his colleagues stating, “Do student evaluations work? Absolutely. As long as you provide clear promises to students and make sure everyone’s incentives are aligned to the same learning goals” (p.4). Addressing evaluations, Sandefer pays close attention to students as customers, tying student assessments to professor bonuses, using a forced curve for student grades and students receiving their grades prior to professors’ evaluations. Studies have addressed grades and ranking of professors as they correlate with student grades and assessments. Rojstaczer (2012) states, “Student evaluations can be useful when they are divorced from tenure, retention and promotion evaluations. Student evaluations are a poor indicator of professor performance. He suggests, “measuring how students perform, on average, on the next sequence in a study area and inferences made on instructor quality from the outcomes of graduates in the major area of study” (p.4).

Professors are paying close attention to their student assessments, as well as their course GPA, as related to pedagogy rigor. Students’ assessments and course rigor reflected in GPA all affect a professor’s academic career. As a result of this interest, this study collected data during a three-year period in an attempt to show a correlation between the assessments and the GPA. Without significant association, one can question if supplements to this study variables should be included in assessment of faculty.

THE STUDY

This research collected data over a three-year period - 2011 through 2013 - from the School of Business and Management of a major Southern California university. For years the data were collected at the end of each course session by the university’s Office of Institutional Research and Assessments (OIRA), calculating student assessments of professors and class overall GPA.

The original study population totaled 3,476 class sessions with 61,781 students enrolled. Statistics from pass/fail courses were dropped from the study because the course grade cannot be combined with a normal grading scheme. Further, because the return of the student assessment forms was optional, some class sessions had none returned. Most classes did not have 100% student participation of professor assessment. The final sample totaled 2,975 class sessions with 54,424 enrolled students. This sample had 35,891 assessment forms returned and processed by OIRA. GPA scores were based upon a 4.0 scale where a grade of A was assigned 4.00 points, a grade of A- was assigned 3.7 points, a grade of B+ was assigned 3.3 points, etc. University guidelines specified GPA targets to avoid grade inflation. The target for undergraduate classes is 2.75 and 3.25 for graduate classes. If the instructor exceeded those targets by more than 0.25 points, the school dean questioned the rigor of the course. Repeated variances resulted in disciplinary action. Teaching assessments measured 12 criteria for an assessment of the instructor’s performance:

1. Instruction was well organized
2. Class time was used effectively
3. Content areas described in the course outline were covered
4. Method of assigning grades was clear
5. Instructor gave clear explanations
6. Instructor was receptive to questions
7. Instructor stimulated critical thinking
8. Instructor encouraged students to think independently
9. Instructor was available for assistance
10. Instructor provided timely feedback on my work
11. Instructor provided useful comments on my work
12. Overall, the instructor was an effective teacher

Each assessment was assigned a score of 1 to 5. The form also captured student assessment of course content and their self-assessment of learning, but only the teaching assessment scores were included in this study. The university also established guideline scores for teaching assessments. An assessment score below 4.00 was perceived in a negative light by the university administration. In fact, adjunct instructors have been terminated because of scores below that target of 4.00 and full-time faculty have suffered reduced salary increases and promotion delays.

DATA ANALYSIS METHODOLOGY

The SPSS Linear Regression method was used to determine the correlation between student grades and teaching assessment. A bivariate correlation analysis determined the relationship. A perfect correlation of 1.0 would indicate that the student's grade is a perfect predictor of their assessment of the teacher. It should be noted that the reverse - a high student assessment would result in a higher grade - does not follow.

The SPSS model produced a correlation coefficient, the amount of significance that can be placed on that result, and the amount of variance (r^2) in the dependent variable (teaching) that can be explained by the independent variable (GPA).

Independent of the SPSS model, weighted averages were calculated. The average GPA was calculated by multiplying the number of students in each class by the average GPA of that class. These products were summed and divided by the total number of students. Likewise, the average teaching assessment was calculated by multiplying the number of assessment forms returned by the average assessment for that class. These products were summed and divided by the total number of forms returned.

DATA ANALYSIS RESULTS

The SPSS correlation coefficient was reported to be 0.018 with an r^2 of 0.000. The standard error of the estimate was 0.445. The resulting correlation formula was:

$$\text{Assessment} = 4.324 + 0.019 \times \text{GPA}$$

The result can be interpreted to reflect that there is no correlation between the two variables. Using the resulting equation, the teacher assessment for a student earning a D grade would be 4.343 while the assessment for a student earning an A grade would be 4.400. Computing the result utilizing the weighted average GPA for all students of 3.06, the formula produces an Assessment of 4.382 compared to the assessment weighted average of 4.392.

Ten assessments were considered outliers but were not eliminated from the population as they had little effect on the other 35,881 data points. Microsoft Excel was also used as a confirmation of the results and to plot the data points in a graphical format, as presented in Figure 1, with regression analysis for all classes resulting in a correlation coefficient of 0.018 with $R^2 = 0.000$. The Y-axis intercept was 4.324 which closely approximated the weighted average of the sample of 4.392.

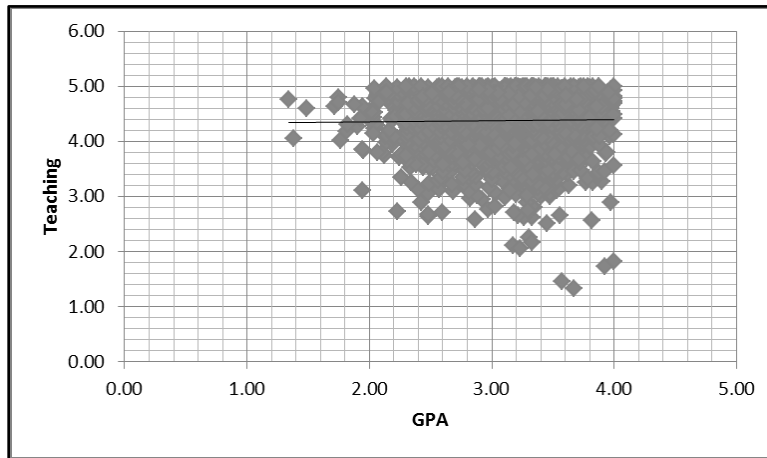


Figure 1: Combined Sample

An additional test of the relations was performed by separating full-time faculty and adjunct faculty. One might assume that results might be different because of teaching experience and expertise of the subject matter. However, the same SPSS formulae produced a correlation coefficient of 0.001 and an r^2 of 0.000 for full-time faculty and a correlation coefficient of 0.021 and an r^2 of 0.000 for adjunct faculty. These additional SPSS calculations repeated the finding, concluding that there is no correlation in these two data subsets. The Excel graph for adjunct faculty is presented in Figure 2 with the Y-axis intercept at 4.317 which closely approximated the weighted average of the sample, 4.393.

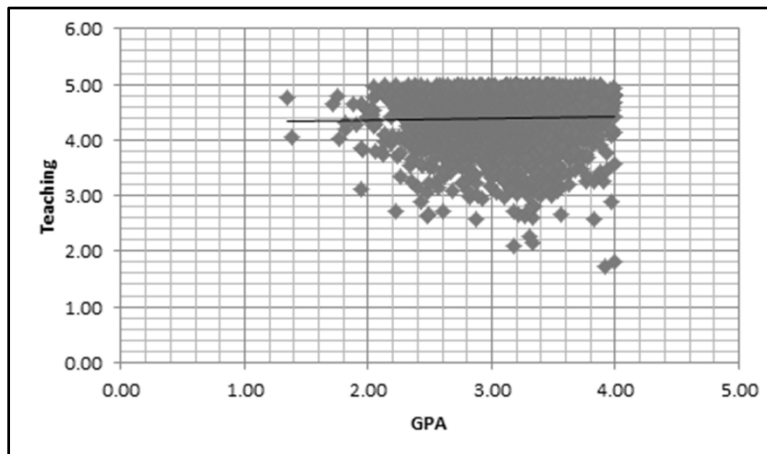


Figure 2: Full-Time Faculty Only

Figure 3 - the graph for full-time faculty - shows little difference. The Y-axis intercept is 4.372, which closely approximates the weighted average of the sample, 4.387, and the adjunct assessment of 4.393. The weighted averages of the Grade Point Averages were also similar - 3.04 in classes taught by full-time faculty versus 3.07 in adjunct-taught classes.

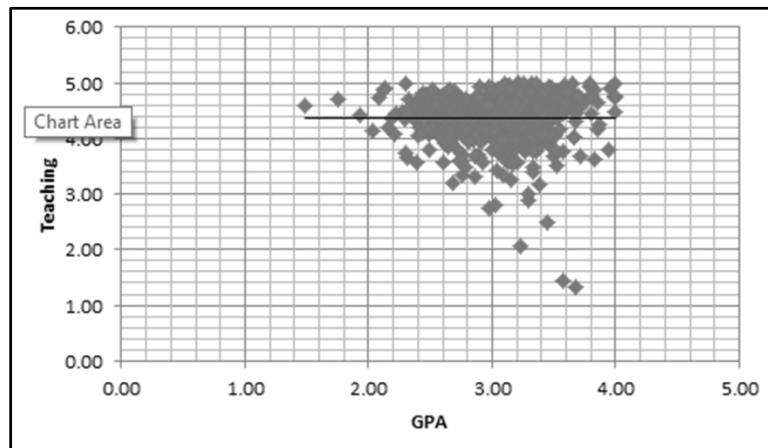


Figure 3: Adjunct Faculty Only

CONCLUSION AND FUTURE CONSIDERATIONS

All three calculations produce a zero or near zero correlation interpreted to mean that there is no correlation between the student's grade and his/her assessment of the teaching effectiveness. Students expecting an A grade often gave below-average assessments while students expecting a C grade often gave above-average assessments. The null hypothesis is supported:

H₀: There is no correlation between the students' GPAs and their assigned assessment scores for their professors.

This study contributes to existing literature and supports Davidovitch and Some (2009) stating a potential myth of correlation between GPAs and professor assessments. Another observation about this conclusion and the study, in general, indicate that the instructor can expect an assessment score of between 4.3 and 4.4 regardless of whether he or she is perceived as an "Easy A" or a "Tough B". Further, instructors need to know that an attempt to buy a higher assessment score with higher grades is misguided.

While this study was conducted within the School of Business and Management of a major Southern California University, additional studies conducted in other schools and universities may prove insightful in researching the hypothesis in this study. Future consideration suggests using a wider range of disciplines, including:

- The sciences
- Health sciences
- Teacher education
- Arts and humanities

Further research may also consider:

- Separating graduate from undergraduate students
- Distinguishing between on-site and online classes
- Additional variables including, but not limited to, professor assessment adjustments for circumstances found within biases of student ratings, peer-review ratings, structure, and rigor of courses

This study provoked several questions addressed in academia. The result of the study supporting the null hypothesis indicates that further research is necessary. If we cannot rely upon student assessment to correlate with their grade, then another model must be researched and suggested. Faculty has their careers at stake and students need assurance that upon graduation they have received the highest quality education.

AUTHOR INFORMATION

Dr. Joyce L. Ellis is an Associate Professor of Accounting at National University in LaJolla, California. She received her B.S. in accounting from Canisius College at Buffalo, New York, her Masters in Accounting and Financial Management from Keller Graduate School of Management, and her DBA from Argosy University. She is a licensed certified public accountant (CPA) in California. E-mail: jellis@nu.edu.

Thomas J. Francl is Associate Faculty of Accounting at National University in LaJolla, California. He received his B.S. in mathematics at Loyola University, Chicago, Illinois, his Masters in Business Administration from the University of Southern California, and a Certificate in Management Accounting (CMA). E-mail: tfrancl@nu.edu.

REFERENCES

- Baldwin, T., & Blattner, N. (2003). Guarding Against Potential Bias in Student Evaluations. *College Teaching*, 51(1), 27-32.
- What we saw. Retrieved February 16, 2014, from <http://www.nytimes.com/roomfordebate/2012/09/17/professors-and-the...>
- Davidovitch, N., & Soem, D. (2009). Myths and Facts about Student Surveys of Teaching the Links between Students' Evaluations of Faculty and course Grades. *Journal of College Teaching & Learning*, 6(7), 41-49.
- Decatur, S. (2012, September 17). Student Evaluations and Other Insights about Professors. *NY Times*, p.4.
- Filetti, J., Wright, M., & King, W. (2010). Grades and Ranking: When tenure affect assessment. *Practical Assessment, research & Evaluation*, 15(14).
- Heller, P. (2010, November 26). College Business Professor: Who is their customer and who should evaluate them? Retrieved February 13, 2014, from <http://Atlanta Social Media>.
- Heller, P. (2013, February). College professor evaluation systems penalize good instructors. Retrieved January 22, 2014, from <http://Genesis Framework>.
- Johnson, V. (2002, July 16). *Teacher Course Evaluation and Student Grades: An Academic Tango*. Retrieved April 7, 2014, from <http://www.isds.duke.edu/~dalene/chance/chanceweb/153.johnson.pdf>.
- As long as it's about more than scores. Retrieved February 16, 2014, from <http://www.nytimes.com/roomfordebate/2012/09/17/professors-and-the...>
- . Retrieved February 16, 2014, from <http://www.nytimes.com/roomfordebate/2012/09/17/professors-and-the...>
- Sandefer, J. (2012, September 18). Our Motto: Give the customer what they want. *New York Times*.
- Sonner, B. (2000). A Is for "Adjunct": Examining Grade Inflation in Higher Education. *Journal of Education for Business*, 76(1), 5.
- Wilson, R. (1998, January 16). New Research Casts Doubt on Value of Student Evaluations of Professors. *The Chronicle*, p.A12.

NOTES