Correlation between GPA and faculty evaluations at UIUC

Patrick Huang[1] and Nickvash Kani[1]

[1]University of Illinois Urbana-Champaign

1 Introduction

Course grades and student evaluations of teaching (SETs) are factors important for the career of students and professors respectively [1, 3, 6]. Grade Point Average (GPA) is used as a measure of student academic performance. Professor evaluations are commonly used to determine tenure and salary [6]. When faculty evaluations are determined by students, the integrity of the system can be compromised. Studies have found that teaching quality decreases as a result [6].

At the University of Illinois Urbana-Champaign (UIUC), faculty evaluations are determined in end-of-term surveys — a practice common in higher education [1, 6]. Because these are desirable measures mutually given to each other [6], we hypothesize that there is a correlation between the two. Although raw results from SETs are not released, teaching honors are awarded publicly at UIUC. Faculty can receive "excellent" or "outstanding" honors based solely on SETs [5]. This, combined with public course GPA data, can be used to determine the correlation between average course GPA and faculty rating. Our dataset is available for investigations of other factors that may influence end-of-semester teaching evaluations [4].

We find that there is a significant positive correlation between GPA and teaching evaluations, and compute the strength of correlation in different departments.

2 Data

We use data sources for faculty rating and course GPA. This study considers courses between 2010 and 2024 (inclusive), in the fall and spring terms.

2.1 Faculty rating

UIUC does not release the exact SET metrics. Our requests for this data via the Freedom of Information Act (FOIA) were declined. However, every semester, lists of "excellent" and "outstanding" faculty are published based on SETs. Ratings for professors and teaching assistants are released as a categorical variable [5]. If an individual is evaluated particularly well, they receive an "excellent" rating. If they pass an even higher benchmark, they get the more prestigious "outstanding" rating. If neither an "excellent" nor "outstanding" rating is earned, we refer to their rating as "not recognized". Ratings

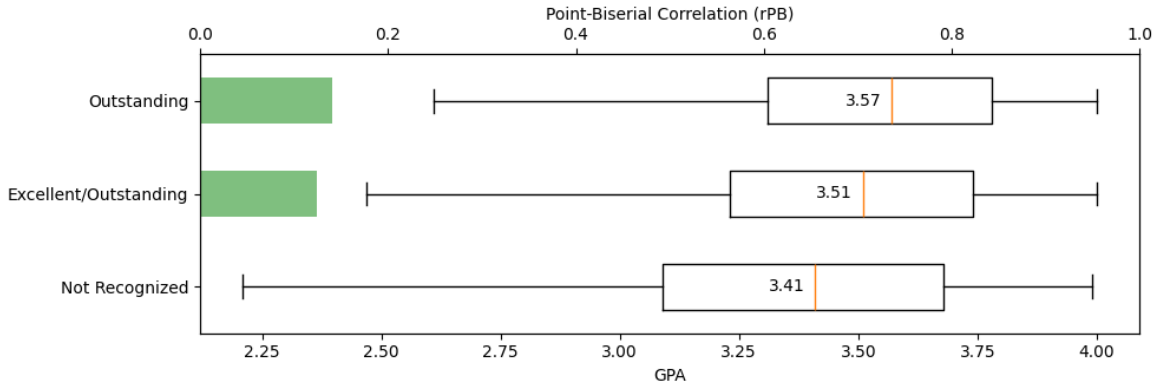

Figure 1: GPA distribution by faculty rating (outliers removed), and Point-Biserial correlation.

are given on a per-class basis: An individual receives a rating associated with a course, and may earn ratings for multiple classes in one term. We provide more details in the supplemental report [7].

2.2 Course GPA

Average course GPA at UIUC is released under FOIA. We use a collated dataset of historical UIUC course GPAs [8]. GPA data are only released when doing so maintains student confidentiality. Two criteria restrict the release of data: 1) sections with fewer than 20 students, or 2) sections in which all students received the same grade. Therefore, public GPA data only cover around 27 percent of all courses historically offered. The number of students in a course affects the class dynamic [3], and thus part of the student's perception of the professor. Therefore, we cannot circumvent criteria 1, and limit the applicability of our results to courses with more than 20 students. Faculty evaluations at UIUC are a survey that students individually complete, so the grades of peers should not significantly affect how individuals respond. Therefore, we hypothesize that criteria 2 does not affect our conclusions.

3 Results

We present two figures of interest.

Figure 1 shows GPA distributions separated by faculty rating. The bottom distribution plots the GPAs of faculty "not recognized". In the top two distributions, we consider "outstanding"; and at least "excellent" (i.e. "excellent" or "outstanding"). We compute the correlation between the bottom distribution with each of the top distributions.

The Point-Biserial correlation coefficient (r_{PB}) has been used in this context [2]. It quantifies the association between a binary categorical variable (whether the faculty received teaching honors), and a numerical variable (course GPA). In Figure 1, we compute r_{PB} of the course GPAs of faculty with high ratings compared to courses where faculty were not recognized. These produce correlations of 0.140 and 0.125, respectively. This relatively low correlation suggests that the difference in GPA between those that did and did not receive honors is not particularly large; smaller than the expected variation within a single distribution. We show that the difference in the means is statistically significant.

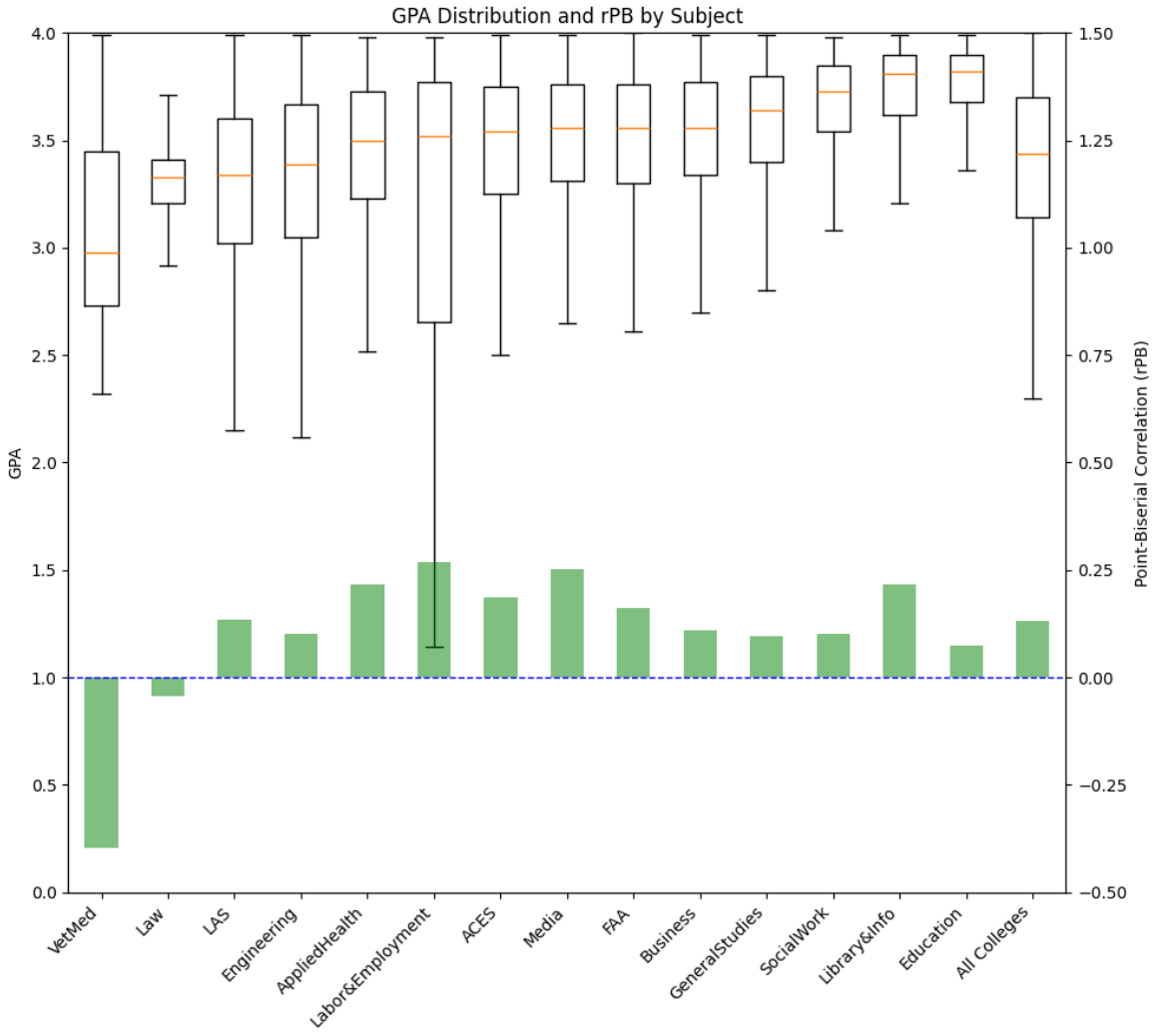

Figure 2: GPA distribution (outliers removed) and Point-Biserial correlation by college.

Details are in our supplemental report [7].

Figure 2 shows the correlation and GPA distribution separated by college. Colleges are sorted by median GPA from lowest to highest, on a standard 0 to 4 point scale (left axis). We plot the correlation coefficient on an arbitrary scale (right axis) for visual clarity, with zero marked.

The GPA distribution of all courses in a college, regardless of faculty rating, are shown as a single box plot. The correlation of GPAs between at least "excellent" and "not recognized" is shown as a bar graph. We sort the colleges by increasing median GPA. However, there is no obvious pattern to the correlations, which varies non-monotonically. Additionally, the two colleges with the lowest median GPA — Veterinary Medicine and Law — also have a negative correlation. More analysis is required to determine traits of these colleges that cause such patterns.

3.1 Limitations

Because the available data are averaged on a per-class basis, we cannot account for individual student characteristics. While studies notice significance in following students over time [3], or considering student gender [1], we can only present results regarding professor characteristics.

A categorical faculty rating limits our analysis. We hypothesize that a continuous measure of teaching performance — the raw SET scores used to generate the teaching honors, for example — would yield a stronger correlation.

As mentioned previously, not all GPA data are released, in order to protect the privacy of students. Therefore, our statistical results only pertain to courses of at least 20 students.

4 Conclusions

In this study, we collected data on faculty evaluations and the GPA of their associated courses, using data unique to the University of Illinois Urbana-Champaign. We determined that the two measures are significantly correlated. While further work is needed to determine the impact of this association on education and university quality, our results suggest that some degree of variation in student evaluations of teaching is affected by the grades students receive.

References

[1] S. Basow. "Student evaluations of college professors: When gender matters." In: *Journal of Educational Psychology* 87.4 (1995), pp. 656–665. DOI: https://doi.org/10.1037/0022-0663.87.4.656.

[2] R. Barker Bausell et al. "Expected Grade in a Course, Grade Point Average, and Student Ratings of the Course and the Instructor". In: *Educational and Psychological Measurement* 32.4 (1972). DOI: https://doi.org/10.1177/001316447203200415.

[3] S. Carrell et al. "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors". In: *Journal of Political Economy* 118.3 (2010), pp. 409–432. DOI: https://doi.org/10.1086/653808.

[4] *IcesEval*. https://github.com/phuang1024/IcesEval. Accessed: 2025-8-20.

[5] *List of Teachers Ranked as Excellent by Their Students*. https://citl.illinois.edu/citl-101/measurement-evaluation/teaching-evaluation/teaching-evaluations-(ices)/teachers-ranked-as-excellent. Accessed: 2025-8-18.

[6] W. Stroebe. "Student Evaluations of Teaching Encourages Poor Teaching and Contributes to Grade Inflation: A Theoretical and Empirical Analysis". In: *Basic and Applied Social Psychology* 42.4 (2020), pp. 276–294. DOI: https://doi.org/10.1080/01973533.2020.1756817.

[7] *Supplemental Report*. https://github.com/phuang1024/IcesEval/blob/master/supplement.pdf. Accessed: 2025-8-20.

[8] *wadefagen's Useful Datasets*. https://github.com/wadefagen/datasets. Accessed: 2025-8-18.