

**Predicting Student Evaluation of Instructor and
Course: Revisiting the Relationship among Course Grades, Improving Teaching,
and Summative Evaluation**

EXECUTIVE SUMMARY

C. S. Gaede, M. D. Svinicki, and D. M. Zimmaro

Historically student ratings of instruction have been used for three purposes: personnel or administrative decisions, improving teaching, and guiding student selection of courses. The purpose of this paper is to examine the relationship between items on the rating form used at UT Austin that could inform each of these three purposes and the overall ratings of teaching effectiveness, along with some of the assumptions about other variables that might be affecting student ratings.

There were 5,077 cases in the data set for all instructor ranks, professor, associate professor, assistant professor, instructor, and lecturer. Correlations, regression equations, ANOVA, and descriptive statistics were computed in the analysis. The study produced the following major findings.

1. The CIS Basic Form items form clusters that represent each of the three purposes of the CIS (improving teaching, administrative decisions, and course selection), the six dimensions of improving teaching reported in literature (skill, rapport, structure, difficulty, interaction, and feedback), and self-reported student achievement.
2. Correlations between the student reported overall grade point average (GPA) are low across all items.
3. Correlations between the student predicted probable course grade and mean course grade with the items in the improving teaching and administrative decision clusters are low to moderate.

4. Moderate to high correlations exist between the items in the improving teaching and the administrative decision clusters.
5. When compared to the overall group mean, the assistant professor rank has the strongest positive effect size across all clusters and the professor rank is below the group mean.
6. When compared to the rank of professor, the assistant professor rank has the largest positive effect size across all three clusters.
7. The best predictors of the overall instructor rating are the “communication” and “interest” items from the improving teaching cluster. “Communication” is more influential than “overall course” in predicting overall instructor.
8. The best predictors of the overall course rating are the “value” and “communication” items from the improving teaching cluster. “Overall instructor” is more influential than “value” in predicting “overall course.”
9. Mean course grade point average is not a significant predictor in any equation.
10. Indications of nonlinearity exist for responses to the evaluation of “workload,” so that students with the highest mean grade point averages rate the “workload” insufficient, and students with the lowest mean grade point averages rate the “workload” excessive.
11. Across instructor ranks, lecturers received the highest percentage of ratings in all three workload categories, insufficient, average, and excessive, followed by the professor rank.
12. The preponderance of evidence indicates that students are making ratings of instruction that are unbiased by grading leniency.