

TEACHING & COURSE EVALUATIONS

Eiszler, C.F. (2002). College students' evaluations of teaching and grade inflation. *Research in Higher Education*, 43(4), 483-501.

This study examined the relationship between grades and students' composite ratings of teaching effectiveness over a 20-year period at a comprehensive, public university in the upper Midwest. Between the spring of 1980 and the fall of 1999, students in more than 37,000 thousand course sections submitted 983,491 teaching evaluations using the Student Opinion Survey: A Survey of Teaching Characteristics (SOS). As a first step in the analysis, linear and polynomial regression lines were fitted to the variables of central interest: the percentage of students expecting grades of A or A- and the composite teaching effectiveness rating for all instructors. In a second stage, calculation of the correlation across all semesters between expected grades and ratings of teaching effectiveness and the partial correlation between these variables. The data showed significant time related trends for both expected grades and teaching effectiveness ratings. The predictive relationship between student ratings of teaching and expected grades was significant. These results suggested that college students' ratings of instruction may be used in ways that raise questions of consequential validity, specifically by encouraging grade inflation.

Harrison, P.D., Douglas, D.K., & Burdsal, C.A. (2004). The relative merits of different types of overall evaluations of teaching effectiveness. *Research in Higher Education*, 45(3), 311-323.

The major objective of this article was to compare the usefulness of different types of overall evaluations of teaching effectiveness. For the Fall 2001 semester, a random sample of 265 sections was drawn from the pool of 770 sections requesting the administration of the SPTE (Student Perceptions of Teaching Effectiveness). The SPTE is a norm-based instrument that includes 37 evaluative items, two validity items, four items measuring the extent to which the student wished to take the course, and four research questions. Factor analysis was used to identify six first-order factors and two second-order factors. Four of the six first-order factors are: Course Organization and Design, Rapport with Students, Grading Quality, and Course Value. These four factors loaded on the second-order factor Perceived Quality Index (PQI), which proxies as an overall evaluation. The other two first-order factors are Course Difficulty and Workload, and these two factors load on the second-order factor Perceived Course Demands. One interesting finding is the extremely high inter-correlations that exist between the different factors. These inter-correlations are all so high that they virtually measure the same thing. With the SPTE, the PQI is doing a good job of serving as an overall evaluation, evidenced by the high inter-correlations between the PQI and the other overall evaluations made or computed in this study.

Nasser, F., & Hagtvet, K.A. (2006). Multilevel analysis of the effects of student and instructor/course characteristics on student ratings. *Research in Higher Education*, 47(5), 559-590.

The primary issue addressed in this study concerns the effects of student, instructor and course characteristics on student ratings. This study examined seven instructor/course variables that included instructor gender, status (tenured, non-tenured), workload, degree, pedagogical training, teaching experience, and class size. The study was conducted at a large teacher training college in Israel. Data was gathered from 1867 students enrolled in 117 courses during the first term of 2001 academic year. Student ratings of one course taught by each instructor were included in the data. Student ratings were collected using an evaluation questionnaire with 16 items measuring three dimensions pertaining to instruction quality: course content, instructor's planning and teaching strategies, and instructor's behavior relating to students. A multilevel SEM method was used to analyze the data. The results revealed that expected grade, interest in the course subject, and student's age have significant positive correlations with student ratings indicating that the higher the expected grade, the greater the interest in the course subject, and the older the student—the higher are the ratings. The results of the current study provide evidence against a spurious relationship between the two variables as asserted by Marsh and Rochi (2000).

Pan, D., Tan, G.S.H., Ragupathi, K., Booluck, K., Roop, R., & Ip, Y.K. (2009). Profiling teacher/teaching using descriptors derived from qualitative feedback: Formative and summative applications. *Research in Higher Education*, 50(1), 73-100.

This study attempts to devise a method for 'quantifying' students' comments to increase their usefulness in complementing/confirming ratings. Through this study, the researchers attempted to generate quantitative data from qualitative research. They obtained student feedback data, both qualitative and quantitative, for the highest 20% cohort (N = 278) and the lowest 20% cohort (N = 278) of faculty members based on student ratings on overall teaching effectiveness for module levels 1,000–3,000 (undergraduate), 4,000 (honors) and 5,000 (graduate). The difference between two means was compared by Students' t-test (two-tailed). The difference between multiple means was compared by one-way analysis of variance (ANOVA) followed by Bonferroni multiple range test. . The findings of this study support the proposition made by Ory and Ryan (2001) that there is no strong relationship between teacher characteristics and student ratings.

Renaud, R.D., & Murray, H.G. (2005). Factorial validity of student ratings of instruction. *Research in Higher Education*, 46(8), 929-953.

The main purpose of this study was to test the systematic distortion hypothesis (SDH) by examining the factorial validity of student ratings of teaching. The systematic distortion hypothesis (SDH) stems from implicit personality theory (Cronbach, 1958), which explains how traits can be judged as correlated when in reality they correlate little or not at all. This test compared the structure of actual teaching behaviors measured by direct observation, and the structure of conceptual associations. Two different teacher rating forms (TRFs) were used in the ratings component of the study. Teacher Behaviors Inventory (TBI) developed by Murray (1983). The second was a newly developed 9-item form consisting of eight high-inference items corresponding to the eight underlying factors or dimensions of the TBI, plus a ninth item assessing overall teaching effectiveness. Ratings and actual behavior counts were analyzed in a oneway analysis of variance (ANOVA) paradigm where instructors served as “treatments” and raters or observers served as “subjects”. It was found, first, that the structure of student ratings showed a moderately strong relation to the structure of actual behaviors, and a somewhat stronger relation to the structure of conceptual associations; and second, that the effects of systematic distortion were more pronounced for low-inference student ratings than for high-inference ratings.