

Use of Student Evaluations in Promotion and/or Tenure:
An Annotated Bibliography
Compiled by Lisa LeVasseur, Institutional Effectiveness, WMU

Becker, W. E., & Watts, M. (1999). The state of economic education: How departments of economics evaluate teaching. *The American Economic Review*, 89(2), 344-349.

Abstract summary: Findings are reported from a national survey of *economics departments* concerning the use of student evaluations of *teaching* (SET) and alternative or supplementary means of evaluating *teaching*, such as peer review. The use of SET instruments is controversial in many respects, but has become commonplace at US universities in virtually all disciplines. The basic question addressed is how *departments of economics* are currently evaluating *teaching*. Specifically, how widespread and extensive is departmental use of SET ratings, peer evaluation and other measures of *teaching* effectiveness and scholarship?

Cutler, A. (2010). It's not just you. *Journal of College Science Teaching*, 39(4), 6-7.

Abstract summary: Cutler discusses the issue on faculty governance structure in relation to promotion and tenure processes. She cites the Carnegie studies by Youn and Price which indicate, among other things, the affirmation of faculty and administrators to the statement, "It is difficult for a person to receive tenure and promotion if he/she does not publish." She comments that the current scenario does form a scramble for status wherein the prestige of a faculty is being parameterized as to the number of his/her publications. She commends the senior faculty and administrators on her campus who make appropriate decisions on tenure and promotion, and suggests that every committee from other institutions should likewise be mindful of the broad situation and issues that are truly campus-centered.

d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *The American Psychologist*, 52(11), 1198-1208.

Abstract summary: Many colleges and universities have adopted the use of student ratings of instruction as one (often the most influential) measure of instructional effectiveness. In this article, the authors present evidence that although effective instruction may be multidimensional, student ratings of instruction measure general instructional skill, which is a composite of three subskills: delivering instruction, facilitating interactions, and evaluating student learning. The authors subsequently report the results of a metaanalysis of the multisection validity studies that indicate that student ratings are moderately valid; however, administrative, instructor, and course characteristics influence student ratings of instruction.

DeLorme, C. D., Hill, R. C., & Wood, N. J. (1979). Analysis of a quantitative method of determining faculty salaries. *Journal of Economic Education*, 11(1), 20–25.

Abstract summary: Several researchers have systematically examined the factors that relate to faculty promotions and salary increases. In this paper, the authors report on a faculty-evaluation system used at the College of Business Administration of the University of Georgia. Using a linear regression model to analyze the data, they considered the effects of publications, teaching experience, departmental affiliation, school from which Ph. D. was granted, and teaching effectiveness on the faculty member's salary. They found that variables not included in the quantitative reward system were statistically significant. The impact of teaching effectiveness (as measured by student evaluations of faculty) varied over the four-year period covered, the quality of teaching actually having a negative influence during some years.

Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction. *The American Psychologist*, 52(11), 1182-1186.

Abstract summary: The validity of student rating measures of instructional quality was severely questioned in the 1970s. By the early 1980s, however, most expert opinion viewed student rating measures as valid and as worthy of widespread use. In retrospect, older discriminant-validity concerns were not so much resolved as they were displaced from research attention by accumulating evidence for convergent validity. This article introduces a Current Issues section that gives new attention to validity concerns associated with student ratings. The section's 4 articles deal, respectively, with (a) conceptual structure (are student ratings unidimensional or multidimensional?), (b) convergent validity (how well do ratings correlate with other indicators of effective teaching?), (c) discriminant validity (are ratings influenced by factors other than teaching effectiveness?), and (d) consequential validity (are ratings used effectively in personnel development and evaluation?). Although all 4 articles favor the use of ratings, they disagree on controversial points associated with interpretation and use of ratings data.

Hamermesh, D. S., & Parker, A. (2005). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review*, 24(4), 369-76

Abstract summary: Adjusted for many other determinants, beauty affects earnings; but does it lead directly to the differences in productivity that we believe generate earnings differences? We take a large sample of student instructional ratings for a group of university teachers and acquire six independent measures of their beauty, and a number of other descriptors of them and their classes. Instructors who are viewed as better looking receive higher instructional ratings, with the impact of a move from the 10th to the 90th percentile of beauty being substantial. This impact exists within university departments and even within particular courses, and is larger for male than for female instructors. Disentangling whether this outcome represents productivity or discrimination is, as with the issue generally, probably impossible.

Kulik, J. A. (2001). Student ratings: Validity, utility, and controversy. *New Directions for Institutional Research*, 27(5), 9-25.

Abstract summary: This chapter reviews the conclusions on which most experts agree, cites some of the main sources of support for these conclusions, and discusses some dissenting opinions and the research support for those opinions.

Marsh, H. W., & Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American Educational Research Journal*, 30(1), 217-251.

Abstract summary: Abstract currently unavailable.

Smith, B. (2009). Student ratings of teaching effectiveness for faculty groups based on race and gender. *Education*, 129(4), 615-624.

Abstract summary: The purpose of this study was to describe student ratings of teaching effectiveness in a College of Education at a Southern Research I institution for faculty groups based on race and gender. Student ratings for undergraduate, graduate, and undergraduate and graduate courses combined, were used to determine the teaching effectiveness for the 190 tenure-track faculty in six groups of faculty (White males, White females, Black males, Black females, male faculty from "Other" racial groups, and female faculty from "Other" racial groups). Of the 36 items on the end-of-course evaluation form, the two global items - overall value of course and overall teaching ability - are reported. On overall value of course, student ratings were very good for four of the six faculty groups and good for the remaining two groups. On overall teaching ability, student ratings were very good for half of the faculty groups and good for the remaining half. Very good indicates a rating above 4.0 on a 5.0 scale while good indicates a rating below 4.0, but higher than 3.0 on a 5.0 scale.

Trout, P. A. (1997). What the numbers mean. *Change*, 29(5), 24-30.

Abstract summary: The context for numerical student evaluations of courses, which are used in about 80% of postsecondary institutions, are discussed. Trout examines the reliability and validity of the scores.

Yagoda, B. (2010, March 28). Why I hate annual evaluations. *The Chronicle of Higher Education*. Retrieved from <http://proquest.umi.com>

Abstract summary: American colleges and universities still view faculty members as a group of scholars and teachers devoted to and bound by self-instilled standards of excellence. Tenure, as long as it continues to exist, must and does require evaluation. But that evaluation and judgment are performed not by the chair but by one's peers.