



THE ROLE OF STUDENT RATINGS OF INSTRUCTION IN ABET ACCREDITED INFORMATION SYSTEMS PROGRAMS

Gerard Morris, Wayne Haga and Abel Moreno

Metropolitan State University of Denver, USA

The use of student evaluations in retention, tenure, promotion, and post-tenure review has always been controversial. At the authors' institution, the importance of student evaluations has gradually increased over the last 15 years from a very small portion of the faculty evaluation to easily the most critical factor. At the same time as the percentage that the evaluations counts towards the overall evaluation has increased, the minimum expected average score on the evaluation instrument for a successful review has also gradually increased. To see if the current proposed guidelines for the Computer Information Systems department at the authors' institution are in line with expectations at peer institutions, 34 of the 39 Information Systems programs accredited by the Accreditation Board for Engineering and Technology (ABET) were surveyed to see how student evaluations are being used in tenure and post-tenure review decisions. Results of this survey are presented and analyzed.

Keywords: SRIs, Evaluation of teaching, ABET-accredited, CIS programs.

Literature Review

The role of Student Ratings of Instruction (SRI), also known as Student Evaluations of Teaching (SET) for faculty evaluation is very controversial. Feldman noted that SRIs can be used for three purposes: 1) to help students select classes; 2) to provide feedback to faculty; and 3) "to supply information for administrators and personnel committees in their deliberations on the promotion and tenure of individual faculty members" [1]. It is this last use that is most controversial and despite being the subject of extensive research for decades there is no consensus on the appropriateness of using these evaluations for performance evaluation. Feldman noted that SRIs are "hardly universally accepted" with some faculty against their use, some holding the opposite view and others holding an opinion in the middle. Faculty that oppose the use of SRIs consider SRIs to be "not reliable, valid, or useful, and may even be harmful" [1]. Zabaleta stated "since the relationship between student evaluations and the actual merits of teaching performance has not been clearly identified, numerical values of those evaluations should not be used in critical personnel decisions such as retention, tenure and promotion of faculty, unless they are properly interpreted within a sound theory of teaching effectiveness" [2].

Berk describe global items used for summative decisions as providing "general broad-stroke, summary index of teaching performance or course quality". Examples include questions such as "Overall, this instructor was an excellent teacher" and "Overall, this course was excellent". He concluded "Global items provide the illusion of (1) simplicity, (2) accurate and reliable information, and (3) the pinpoint precision needed for summative decisions about faculty. Unfortunately, the single rating of a global item can be (a) unreliable, (b) unrepresentative of the domain of teaching behaviors it was intended to measure,

and (c) inappropriate for personnel decisions according to U.S. professional and legal standards” and he stated that administrators “should be urged to ‘cease and desist’ before he or she is ordered legally to do so” in terms of using global items for faculty performance evaluation [3].

Part of the complexity in the use of SRIs is the number of variables involved, as attested to by the plethora of research on these items. Variables such as race, grades, class size, discipline, classroom versus online courses, and gender of professors and students have been investigated. Carle studied SRIs using a variety of variables including instruction mode online versus classroom and found that “in face-to-face classes only, students consistently and stably gave lower SETEs to minority professors. They did not differentially rate professors as a function of the professor’s sex, tenure status, discipline, or level of the course the professor taught.” No such effect was observed in online classes [4]. Basow found students gave female professors lower ratings compared to male professors [5]. Smith looked at SRIs for both undergraduate and graduate courses and found “that for both global items, overall value of course and overall teaching ability, male faculty identified as “Other” received the highest mean score while Black male faculty received the lowest mean score; Black female faculty received the second lowest mean scores” [6]. Basow investigated the effect of race and gender and related it to student learning by have students evaluate a computer-animated professor and then take a quiz on the content in the engineering presentation. The study found “that student evaluations may not be a good indicator of actual teaching effectiveness. Students may rate professors highly even when they do not seem to learn from them, as suggested by the higher student evaluations received by African American professors but the lower quiz performance of their students” [7]. Student major is another variable and Basow found that on “all measures, engineering majors gave the most negative evaluations, and humanities majors gave the most positive” [5]. Clayson states that “due to discipline differences the meta-analysis indicates that the academic discipline area is an important variable” [8].

The link between grades and SRIs is probably the most controversial item. Many faculty firmly believe that giving higher grades will result in higher SRIs. Langbein found that actual and expected course grades affect SRIs and warned that “faculty have a clear incentive to give higher grades to students” [9]. Feldman found “In all, currently available evidence cannot be taken as definitely establishing a bias in teacher evaluation due to the grades students receive or expect to receive in their courses, but neither is it presently possible to rule out such bias” [10]. Zabaleta found “a moderate correlation between low grades and low evaluations, but no correlation between high grades and high evaluations when all cases are considered together” [2].

Interestingly Feldman compared dimensions of instruction and ranked them on the correlation shown with student achievement and student evaluations and then grouped them into items of high importance, moderate importance, etc. High importance items included “Clarity and Understandableness”, “Teacher’s Stimulation of Interest in the Course and Its Perceived Subject Matter” and “Teacher’s Preparation; Organization of the Course” [1]. Many faculty would argue their performance should be evaluated on a measure of all these items, not just the usual two global items.

History of SRIs at MSU Denver

Faculty are evaluated initially when a decision is being made whether to retain a non-tenured faculty member. Thereafter they are evaluated for tenure and promotion decisions. Evaluation of faculty at MSU Denver has three main components: teaching, professional development, and service. A regular teaching load is eight three-credit courses per academic year. The requirements for a faculty member to get an acceptable score in the teaching component has continually changed over the years with an ever-increasing importance given to the numerical SRIs plus a concomitant increase in the numerical value needed for an acceptable rating. The continually changing requirements create a confusing and stressful environment but even more concerning is the increasing value of the required SRIs.

As well as the aforementioned changes there have been changes in the number of ratings from six in 2003 to three from 2008 on. What classes could be included/excluded has also changed. Table 1 summarizes the main changes [11].

Table 1. Changes in the Evaluation of Teaching

Year*	Ratings					% of Teaching Evaluation
2003	Excellent 4.71-6	Very Good 4.3-4.7	Good 3.85-4.29	Satisfactory 3.4-3.84	Unsatisfactory 0-3.39	7.5% 12.5%
2004	A 4.76-6.0	B 4.2-4.75	C 3.2-4.19	D 3.0-3.19	F 0-2.99	15%
2005	A 4.76-6.0	B 4.2-4.75	C 3.2-4.19	D 3.0-3.19	F 0-2.99	15%
2007	A 4.6-6.0	B 4.0-4.6	C 3.2-4.0	D 3—3.19	F 0-2.99	5%-10%
2008	Exceeds Standards 4.6-6.0 None below 3.75	Meets Standards 4-4.6 None below 3.25	Needs Improvement 0-3.99 OR 1 score below 3.25			7.5%-12.5%
2012	Exceeds Standards 4.65-6.0 May qualify the Faculty	Meets Standards No more than .2 below the department mean	Needs Improvement More than .2 below department mean for tenure/promotion. Average above 4.0 for successful post tenure review			100%

* Note no changes in 2006, 2009, 2010, 2011, 2013, and 2014.

2003: Faculty were allowed to choose to count SRIs between 7.5% and 12.5% of the Teaching evaluation. The minimum average score for a satisfactory evaluation was 3.4 out of 6.

2004: SRIs were increased to count for 15%, but the minimum score was dropped to 3.2 out of 6. Faculty were also allowed to just count the best 5 classes out of 8 classes taught.

2005: As online courses started to become prevalent, a clause was added that expected SRIs in online classes would be reduced by 1 full point.

2007: SRIs in all 8 classes were required, and online classes could only be adjusted by ½ point. Faculty were again allowed to choose weight for SRIs, from 5%-10% of Teaching.

2008: SRIs could be counted from 7.5%-12.5% of the evaluation. The evaluation could be dropped if the response rate was less than 10. Online scores can be adjusted by .25 only if the college-wide mean for online classes is more than .5 lower than classroom. The minimum acceptable score was raised to 4.0, with none of the 8 classes falling below 3.25.

2012: No adjustment for online classes, and no option to drop low response classes. SRI's need to be no more than .2 below the department mean.

Motivation for this Study

MSU Denver uses two global items from the SRIs to evaluate the teaching component:

1. Course as a whole
2. Contribution to the course

As noted in the literature review, using only global items for evaluation purposes has been seen as less than robust. In addition, the institution's administration is currently suggesting that faculty, both tenured and tenure-track, should be able to receive scores of five on a six-point scale for successful teaching evaluations. The authors wanted to investigate how the other IS programs that are accredited by ABET approach this matter.

Methodology

The chair of the CIS Department emailed 34 chairs of ABET-accredited CIS programs and encouraged them to fill out the survey which they would be receiving shortly. The next day a survey created in Qualtrics was sent to these department chairs. Given the target population was so small, a second email from the chair encouraging participation in the survey followed.

Results

Fourteen of the 34 ABET-accredited CIS departments surveyed completed the survey for a response rate of 41%. One response was unusable as the department chair submitted negative numbers and argued the questions asked were invalid as the score to determine whether a faculty member scores were high enough for tenure or post tenure depended on trend and relationship to the department average.

In addition to the fourteen responses above, three department chairs started the survey, but did not complete it – possibly because they also do not use a point system and were unable to answer the questions, or possibly they did not know the information without taking a lot of time to search for it. It should be noted that another chair responded via email that they use a qualitative evaluation. The 13 detailed responses are shown in Tables 2 (tenure-track faculty) and Table 3 (post-tenure faculty).

Table 2. Detailed Responses – Tenure-Track Faculty Evaluations

Percentage student evaluations count towards the evaluation of teaching for tenure-track faculty	Minimum Expected Average Score	Maximum Score	Equivalent minimum score on a six-point scale used at MSU - Denver
26-50%	80	100	4.8
26-50%	3.5	5	4.2
11-25%	3	4.5	4
51-100%	3	5	3.6
26-50%	3.5	5	4.2
11-25%	3.5	5	4.2
11-25%	3	4	4.5
1-5%	2	4	3
51-100%	3.5	5	4.2
missing	3	5	3.6
26-50%	3.2	5	3.84
11-25%	3.4	5	4.08
6-10%	2	4	3
Average	3.94		

Table 3. Detailed Responses – Post-Tenure Faculty Evaluations

Percentage student evaluations count towards the evaluation of teaching for post-tenure review	Minimum Expected Average Score	Maximum Score	Equivalent minimum score on a six-point scale used at MSU - Denver
26-50%	80	100	4.8
26-50%	3.5	5	4.2
11-25%	3	4.5	4
51-100%	3	5	3.6
26-50%	3.5	5	4.2
11-25%	3.5	5	4.2
11-25%	3	4	4.5
1-5%	1	4	1.5
51-100%	3.5	5	4.2
missing	3	5	3.6
26-50%	3.2	5	3.84
11-25%	3.4	5	4.08
6-10%	1	4	1.5
Average	3.71		

For comparison purposes, minimum expected scores were converted to the six- point scale used at the authors' institution. The majority of institutions use a five-point scale for SRIs. One institution indicated an unusual 100-point scale.

The percentages that SRIs counted towards the teaching evaluation at peer programs was identical for tenure-track and tenured faculty – see Table 4.

Table 4. SRI Weights in Evaluation of Teaching

Percentage that student evaluations count towards the evaluation of teaching	Tenure-track Faculty – Number of Departments	Post-Tenure Faculty – Number of Departments
1-5%	1	1
6-10%	1	1
11-25%	4	4
26-50%	4	4
51-100%	2	2

The results in table 4 show there is a huge variation in how much faith departments put in SRIs, with some institutions counting them for 5% or less and others counting them for 51-100%. There was one missing response to the question. Six of the 12 (50%) indicated the SRIs counted for 25% or less. Only two of the 12 (16.7%) counted the SRIs for more than 50% of the evaluation.

The average minimum expected score for tenure-track faculty was 3.94, with a standard deviation of .53. The average minimum expected score for post-tenure review was 3.71, with a standard deviation of 1.03. The expected scores were identical for tenure-track and post-tenure review at 11 of the institutions.

Two of the institutions had essentially no minimum score required for post tenure-review (required one out of four, and it was assumed one is the minimum score a student can give). This accounts for the lower average and higher standard deviation for post-tenure review. It is interesting to note that the two institutions that require the lowest minimum scores, also count the SRIs for the smallest percentage of the evaluation (less than 5%). The two institutions that count the SRIs for more than 50% have relatively reasonable expectations (3.6 and 4.2 respectively).

The survey also asked if student comments became part of the faculty permanent record. Four out of thirteen respondents indicated that they did. At the authors' institution this is a controversial issue that was opposed by faculty. Student comments on the evaluations are uploaded to Digital Measures, a web-based faculty reporting system. These comments are then available to numerous faculty serving on the Department, College, or University retention, tenure, promotion (RTP) and post tenure review (PTR) committees, as well as the various levels of administrators.

Conclusion

Based on the results obtained from the institutions participating in this study, a score in teaching performance of 4.0 on a 1-6 scale where 6 represents the highest mark, appears to be a reasonable expectation for faculty teaching computer information systems courses at ABET-accredited institutions. Further, the survey results indicate that while the weight assigned to those scores varies from institution to institution, SRIs are not the sole criterion considered when evaluating the teaching performance of a faculty member either going up for tenure or undergoing post-tenure review. The high standards and challenging student learning outcomes upheld by ABET accredited programs likely have an effect on the rigor and on the students' work expectations for the course, and its impact on SRIs should be considered in the overall evaluation of a faculty member's performance in teaching.

References

1. K. Feldman, "Identifying Exemplary Teachers and Teaching: Evidence from Student Ratings", *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, pp. 93-143, Springer 2007.
2. F. Zabaleta, "The use and misuse of student evaluations of teaching", *Teaching in Higher Education*, vol. 12, no. 1, pp. 55-76, February 2007.
3. R. Berk, "Should Global Items on Student Rating Scales Be Used for Summative Decisions?", *Journal of Faculty Development*, vol. 27, no. 1, pp. 63-67, January 2013.
4. A. Carle, "Evaluating college students' evaluations of a professor's teaching effectiveness across time and instruction mode (online vs. face-to-face) using a multilevel growth modeling approach", *Computers & Education*, vol. 53, issue 2, pp. 429-435, September 2009.
5. S. Basow, N. Silberg, "Student Evaluations of College Professors: Are Female and Male Professors Rated Differently?", *Journal of Educational Psychology*, vol. 79, no. 3, pp. 308-314, 1987.
6. B. Smith, "Student Ratings of Teaching Effectiveness for Faculty Groups Based on Race and Gender" *Education*, vol. 129 no. 4, pp. 615-624, Summer 2009.
7. S. Basow, S. Codos, J. Martin, "The Effects of Professors' Race and Gender on Student Evaluations and Performance", *College Student Journal*, vol. 47, no. 2, pp. 352-363, June 2013.
8. D. Clayson, "Student Evaluations of Teaching: Are They Related to What Students Learn? A Meta-Analysis and Review of the Literature", *Journal of Marketing Education*, vol. 31, no. 1, pp. 16-30, April 2009.
9. L. Langbein, "Management by Results: Student Evaluation of Faculty Teaching and the Mismeasurement of Performance", *Economics of Education Review*, vol. 27, issue 4, pp. 417-428, 2008.
10. K. Feldman, "Grades and college students' evaluations of their courses and teachers", *Research in Higher Education*, vol. 4, no. 1, pp. 69-111, 1976.
11. W. Haga, G. Morris, and J. Fustos, "A Database Management System for Tracking Changes to Faculty Evaluation Guidelines" *Journal of Teaching and Education*, 2(2), pp. 485-491, 2013.