Revisiting How Departments of Economics Evaluate Teaching

William E. Becker, William Bosshardt and Michael Watts

December 6, 2011

William E. Becker
Professor Emeritus of Economics
Indiana University
325 E Blue Ridge Drive
Bloomington IN 47408
Tel: 812-336-9119 Fax: 812-855-3736
beckerw@indiana.edu

William Bosshardt
Associate Professor of Economics
Florida Atlantic University
777 Glades Road
Boca Raton, FL 33431
Tel: 561-297-2908 Fax: 561-297-2542
wbosshar@fau.edu

Michael Watts
Professor of Economics
Purdue University
425 W. State St.
West Lafayette, IN 47907-2056
Tel: 765-494-8543 Fax: 765-496-6028

Abstract: Based on results from a 1999 national survey, William Becker and Michael Watts were the first to report that student evaluations of teaching (SET) were, by far, the most widely used and often the only method used by economics departments to evaluate teaching in undergraduate economics courses. To investigate whether departments of economics have moved beyond the use of SET, in 2011 the current authors conducted a national survey of departments based largely on questions used in the 1999 survey. The surveys included items on how courses and teaching are evaluated and on how that information is used in departmental promotion and salary decisions.

The authors thank Jane Anderson, Mara Gallo, and Chineze Christopher for help with the administration, data entry, and preliminary analysis of the survey.

Introduction

Becker and Watts (1999) provided the first information on the use of student evaluations of teaching (SET) in undergraduate economics courses, versus alternative and/or supplementary means of evaluating teaching, based on 330 responses to a national survey of Chairs of U.S. departments of economics. They found that SET were the almost exclusively used standard for evaluating teaching, with approximately 83 percent of all departments using both fixed-response (e.g., a discrete qualitative scale for which an arbitrary 1 to 5 measure is often imposed) and student free-response questions in their SET forms.

Despite that revealed preference for SET measures, which may well constitute a "bandwagon effect" that has the advantage of being "student centered" and therefore viewed as promoting a more "democratic" classroom setting, there are many alternative ways of evaluating teaching, some of which are arguably superior or valuable because they offer complementary extensions and insights. For example, Glen (2010b) recently summarized the major advantages and disadvantages of 22 different ways of evaluating courses and instructors. Here we briefly address the debate over using end-of-semester SET as an exclusive measure of teaching effectiveness. Then, to investigate whether departments of economics have moved beyond the use of SET, we report the results from our 2011 national survey of departments and compare those results to those reported for the 1999 Becker and Watts survey.

Evaluating Teaching

Walstad and Saunders (1998, p. 339) claimed that "the primary purpose of the common end-of-course evaluation form is to provide comparative data for administrators . . ." Becker

(2000) reviewed many reasons why an exclusive reliance on SET is troubling, but he also recognized that SET measures offer several key advantages to departments, administrators, and to faculty and students. First, and possibly most compelling, they are relatively inexpensive to administer and report. Second, they offer flexibility (AKA "wiggle room") to departments and administrators who want to meet a wide range of possible personnel goals and constraints – such as promoting a good researcher who is a mediocre (or worse) teacher, especially someone covering courses that are expensive and difficult to cover with new hires. The flexibility with SET measures exists because the simple numerical results can be interpreted more or less favorably by reviewers, and therefore weighted or discounted more heavily according to reviewers' discretion/preferences, based on a wide range of factors such as students' expected grades in a class, class meeting times, required vs. elective classes, or different class compositions (for example in terms of prior coursework in economics or different mixes of majors and class standing). Finally, SETs provide a visible means to mollify students, parents and others calling for accountability in the classroom.

The major shortcomings of student evaluations can be summarized in four points:[i]

First, although some education specialists such as Marsh, (2007) claim that SETs are valid and reliable because ratings are positively related to many criteria of teaching effectiveness—including ratings from former students, test scores or other measures of student achievement, teacher self-evaluations, and peer evaluation ratings or observations of trained observers—in fact, contrary to claims, there is little empirical support for the claim that SET ratings capture most of the elements of good teaching. The simple correlation coefficients

cited by Marsh and others are typically less than 0.7, implying that student evaluation scores explain less than 50 percent of the variability in these other teaching outcomes.

Second, departments often misuse these scores by comparing each instructor with numerical means or medians for all instructors in a course or in "similar" courses, which results in treating the scores as if they have far more precision than they actually do. Decisions are often made on small numerical differences that are unlikely to distinguish between competent and incompetent teachers. By implication these comparisons damn those below the arbitrary critical number regardless of its level.

Third, if administrators treat student evaluations of teaching as important, then teachers can be expected to react to them in ways that may be inappropriate. To instructors, generating positive student answers to questions about overall effectiveness and communication skills may smack of entertainment and dumbing down. To raise scores on the end-of-term entertainment quotient, teachers can be expected to modify student activities and grading; they can manipulate timing and procedures for student evaluations of teaching data collection; they can drive the unhappy out of the class, with no trace showing on end-of-semester student evaluations of teaching. To raise their scores on organizational questions, instructors may attempt to gain class sympathy by alleging that chaotic events are out of their control. Instructors facing the judgment of student evaluators may also avoid innovation. As McKeachie (1997, p. 1219) points out: "Many students prefer teaching that enables them to listen passively -- teaching that organizes the subject matter for them and that prepares them well for tests ... research, however, points to better retention, thinking, and motivational

effects when students are more actively involved in talking, writing, and doing ... Thus, some teachers get high ratings for teaching in less than ideal ways."

A fourth concern with the student evaluation forms used in economics courses is that they usually ask few questions that deal with what education specialists say is important: active student learning. Little attention is given to students' perception of what they were expected to learn versus what they believe they learned. Instead, the top four items on which student opinion is typically sought include the teacher's overall effectiveness, communication skills, organization and planning, and knowledge of material (Becker and Watts, 1999); yet students have little basis for judging an instructor's knowledge of the material, or alternative ways of organizing a course.

Weinberg, Fleisher and Hashimoto (2009) go beyond highlighting these shortcomings of end-of-term student evaluations and document that they are positively related to current grades but unrelated to learning once current grades are controlled. They offer evidence that the weak relationship between learning and student evaluations arises, in part, because students are unaware of how much they have actually learned in a course. They conclude by developing an original measure of learning for assessing teaching and learning, based on grades in subsequent courses. This work is impressive and should lead department heads and faculty members to rethink their sole use of end-of-semester student evaluations. But for that matter, the broader and longstanding concerns about SET measures being at best only partial and rather severely incomplete measures of teaching effectiveness should, in an ideal world, also lead to more systematic and comprehensive assessment procedures for evaluating teaching. It

is not at all clear that has happened in economics or other fields over the past decade, as shown by comparing the results of our 2011 survey to the results reported by Becker and Watts (1999).

The 2011 Survey, Response Rates, and Institutional Characteristics

In February 2011 we mailed 774 questionnaires to chairs of economics departments identified by the American Economic Association.[ii] We restricted our sample to U.S. institutions offering a baccalaureate degree or higher, as classified by the Carnegie Foundation. The chairs were given the choice of completing and returning a "hard copy" paper questionnaire or completing the form online. As seen in Table 0, we received 182 completed surveys – 129 paper and 53 online – for a response rate of 23.5%. The response rate was similar across institutions, ranging from 30.6% from Research I institutions to 21.6% from Masters Institutions. The response rate for this survey was lower than in 1999, with a drop in the response rate from Doctoral I institutions (59.5% to 30.6%) accounting for much of the drop in the overall response rate.

Table 1 describes the distribution of our institutions across the Carnegie classifications. In the presentation of our results, we combine Doctoral with Research I and Research II institutions because we received only 11 responses from Doctoral institutions. As in the 1999 study, Masters and Baccalaureate institutions form the other two categories. Administratively, economics departments at Research and Doctoral institutions tend to reside in liberal arts or arts and sciences. This is in stark contrast to economics departments at Masters institutions, which tend to reside in business colleges (100% in our survey). Economics departments at

Baccalaureate institutions, which include smaller liberal arts schools that often do not have business schools, are housed primarily in liberal arts or arts and sciences colleges. With the exception of Baccalaureate institutions, which include smaller liberal arts colleges, the vast majority (over 80%) of institutions offer degrees in business fields.

The characteristics of responding departments are shown in Table 2. Compared to the 1999 Becker and Watts study, the number of majors and students enrolled reported by respondents has increased substantially. For example, in 1999, Baccalaureate institutions had a median enrollment of 390 – today this is 673. For Masters institutions, the increase is from 900 to 1335 students. For Research institutions, the median rose from 3,847 to 4,939, although the mean remained about the same. Department size as measured by regular faculty increased by approximately 10% during the same time, less than the increase in median student enrollments. We did not ask about class size in the surveys, but we suspect that either class sizes have increased over this time or the use of adjunct faculty with higher teaching loads has increased – most likely both.

Table 2 also reports the weight departments place on teaching in determining annual raises and awarding tenure and promotions. Masters and Baccalaureate institutions place relatively more weight on teaching than Research and Doctoral institutions in evaluating candidates for promotion and tenure, but weight teaching less for annual raises than for promotion or tenure decisions. In comparing these results to our 1999 study, we find that Masters institutions are now placing less weight on teaching for promotion and tenure decisions, with the average weight falling by approximately 5 percentage points. Separating

out the Research institutions as in the 1999 study, we find that teaching has increased in importance for promotion and tenure by about 5 percentage points (25% to 30%). But despite this slight increase, no Research or Doctoral institution placed more than 50% weight on instruction in our current study.

The evaluation of teaching is widely mandated by a school's central administration or department, as seen in Table 3. Only 12 of the 182 reported no mandated evaluation of teaching, about the same percentage as in 1999. Despite the fact that instructional evaluation is not always mandated, only one institution indicated that evaluation by students was not typical.

Student Evaluation of Teaching (SET) Procedures

SETs are administered somewhat differently than in 1999. Overall, 35% of departments now administer SETs electronically/ on-line. The 1999 survey did include electronically as an explicit response option[iii], because the use of electronic SETs was seen as experimental at the time, which gives an idea of how quickly technology has moved in this application. Predictably, this has resulted in less frequent use of SET administrations by students, teachers, secretaries or other methods. However, in Table 4 it is shown that students are still used to administer the SET evaluations in half of the responding departments. With more SETs being given electronically, further research is needed to understand the potential biases that may be introduced compared to in-class hard-copy SETs. Avery et al. (2006) find a lower response rate, but no change in the average evaluation scores.

Practically all of the SET forms are composed of both Likert-scale style and written response questions (93%). Only 4% consisted of exclusively written responses, with the remaining 3% being solely Likert-scale items. Written responses are typically reviewed by the department in forming an evaluation of an instructor's teaching, with over 80% of departments reviewing responses. This is a surprising result given the amount of time involved in reading student comments and possible confidentiality issues, but students who take the time to write the evaluations may be pleased to know that they are read by others.

In terms of frequency of administration, SET evaluations continue to be administered almost exclusively as a one-time, end-of-term assessment, as seen in Table 5.

Department chairs were provided a list of commonly cited dimensions of teaching (see Marsh 1991 for examples), and asked to check the aspects of teaching students at their schools were asked to evaluate. These results are shown in Table 6. Organization, clarity, and instructor knowledge were the three most frequently mentioned aspects. Rapport with groups of students and the use of instructional technology were the two least mentioned. As was the case in 1999, the use of examples and applications does not even appear as an SET item in many evaluations, which we find surprising and disappointing. The use of two SET items differed across different Carnegie types of schools: rapport with individual students and grading. In both cases, Baccalaureate schools were more likely to include these items than Research and Doctoral schools.

The results of SET scores are typically given to instructors and chairpersons, although chairs are somewhat less likely to see them at Baccalaureate institutions. Deans and tenure

and promotion committees are also likely to see SET results. Beyond these groups, as shown in Table 7, the results of SETs are not widely available. Students at the university generally cannot see the evaluations – ranging from 15% of Research and Doctoral institutions to only 2% of Baccalaureate institutions. Less than 10% of chairs reported that scores are publicly available, and only 2% of Baccalaureate institutions report the results publicly. The percentage of SETs released to a broader audience seems relatively small given recent efforts to provide better "consumer" information to students and to carefully and publically evaluate university professors.

The importance of SETs as a component of the overall evaluation of teaching is relatively high[iv], accounting for 50% of an instructor's teaching evaluation. For 20% of departments, the weight was 75% or higher, and for 20% it was 25% or lower. For Research and Doctoral institutions the percentage is a bit higher, 53.1%, compared to the other institutions where the average weight was about 47%.

Peer Evaluation of Teaching

Despite complaints from faculty about SETs, alternative formal evaluation methods remain less used. Only about half of departments conduct peer reviews – slightly more than in 1999. Even for those departments that do peer evaluations, the frequency of peer reviews is relatively rare. Most of the time, particularly among Baccalaureate institutions, the review is conducted only for the purposes of promotion and tenure evaluation. Peer review each semester is rare, and annual reviews are conducted at only about 30% of the institutions that do peer reviews. Whether peer evaluation is done only for tenure and promotion or more

regularly also seems to change the nature and structure of the review. Specifically, the promotion and tenure peer reviews are more likely to be under the control of the university or some other broad processes (such as collective bargaining agreements). Promotion and tenure reviews are also more likely to involve external reviewers. Of the approximately one-third of peer review departments that use external reviewers, 75% said the frequency of review was only for promotion and tenure reviews.

Peer review can be an extensive and expensive process, with 90% of peer reviews involving direct classroom observation. Of the ten alternatives given in Table 9, on average department chairs checked three choices. Over 60% of peer reviews involved classroom observation and a syllabus review. Other common peer review methods include reviews of quizzes/exams or instructional materials. Reviews of grading distributions were included in about 30% of peer reviews. Examples from the "other" category include discussing SETs with the instructor. Reviewing student work was more common at the Baccalaureate institutions, and reviews of student drop rates were more common at Research and Doctoral institutions. Videotaping and student interviews are infrequently used.

Table 10 shows the criteria most often used in peer reviews, and can be compared to the criteria listed in Table 6, which lists items most often used on SET forms. For the most part, the frequency of a criteria's use in SETs generally matches the frequency of a criteria's use in peer evaluations. Significant differences[v] between the two types of evaluation emerged for four criteria. Course organization and planning appeared in a smaller proportion of peer reviews, though it still was present in 70% of peer review criteria. Grading was not a criteria

considered in most peer reviews, but it was an important part of SETs – perhaps signaling that the item is used more as a control in the context of SET than as an evaluation of teaching effectiveness, per se.[vi] Applications and examples were considered approximately 75% of the time in peer reviews, about twice as heavily weighed as the 40% use of this item for SET forms. Finally, workload and difficulty was less cited in peer reviews than in SET scores. As was with SET evaluations, teacher-student interactions on an individual basis were more frequently a criteria for Baccalaureate institutions than at to Research and Doctoral institutions.

Peer evaluations are not shared with students or the public, and in fact are less likely to be shared with the faculty being evaluated, showing that they are used as summative evaluations for promotion and tenure decisions more than as formative assessment or mentoring programs. The weight given to peer reviews, when peer reviews are conducted, is about 28%, which is considerably less than the weight given SETs in evaluating teaching.

Curriculum Development and Instructional Research

We asked department chairs for the percentage of faculty that engaged in the curriculum development and instructional research activities, and found the results listed in Table 12. The median answer for each item except that of non-departmental teaching awards was zero. Therefore, instead of reporting the mean of the reported percentage, we report the percentage of chairs that reported **any** faculty engaging in the activity over the last five years. Most commonly reported were winning non-departmental teaching awards and the publication of journal articles on instructional methods (about 45% of departments overall). The

publication of textbooks and workbooks were also commonly reported – about a third of the departments reported these two activities.

The weight placed on curriculum development and instructional research in forming the overall teaching evaluation averaged 8.7%. However, approximately 40% of departments place no weight on these activities. The average of the remaining departments placed a weight of 15% on these activities.

Overall, the SET has a higher weighting in composing an overall teaching evaluation than either peer evaluations or other instructional activities. Only 15 of the 182 departments or 8% reported a higher weight for peer evaluations than SETs (assuming a weight of zero for a department without peer evaluations). Only 7 of the 182 respondents reported a higher weight on curriculum development and instructional research than on SETs.

Conclusion

Not much has changed in departmental practices in evaluating teaching since 1999. With few exceptions, the numbers reported here are qualitatively the same as those found in 1999. The biggest change in the evaluation of teaching has been the advent of on-line SETs. This technology has grown from no use in 1999 to accounting for 35% of departments today. Beyond that technology change, we find that the evaluation of instruction tends to rely heavily and almost exclusively on SETs, with almost every department using them in their evaluations of faculty. Other techniques such as peer review are sometimes conducted, but tend to be used only for promotion and tenure reviews. In short, it appears the conclusions from the 1999 study remain – the relatively lower cost of SET data is sufficient to justify their nearly exclusive

use. With the growth of on-line administration of these evaluations, the cost of SET data is likely to decrease further, undoubtedly increasing their use in years to come even though there are serious (and some would say compelling) arguments against their sole use as THE measure of teaching.

References

Avery, R. J., W. K. Bryant, A.M.H. Kang, and D. Bell. (2006). Electronic Course Evaluations: Does an online delivery system influence student evaluation? Journal of Economic Education 37: 21-37.

Becker, W. E. (2000). Teaching Economics in the 21st Century. Journal of Economic Perspectives, 14(1), 109-119.

Becker, W.E. and Watts, W. (1999). How Departments of Economics Evaluate Teaching. *The American Economic Review*, Papers and Proceedings of the One Hundred Eleventh Annual Meeting of the American Economic Association, 89(2): 344-349.

Becker, W.E. and Sherwin Rosen. The Learning Effect of Assessment and Evaluation in High School. Economics of Education Review. 11, June 1992: 107-118.

Bosshardt, W. and Watts, W. (2001) Comparing Student and Instructor Evaluations of Teaching. *The Journal of Economic Education*, 32(1), 3-17.

Glenn, David (2010a). Rating Your Professors: Scholars Test Improved Course Evaluations, Chronicle of Higher Education, 25 April, 2010. http://chronicle.com/article/Evaluations-That-Make-the/65226/?sid=at&utm_source=at&utm_medium=en

Glenn, David (2010b). Is Your Psychology 102 Course Any Good? The Chronicle of Higher Education, December 12. < http://chronicle.com/article/Is-Your-Psychology-102-Course/125698/>

Marsh, H. W. (1991). Multidimensional students' evaluations of teaching effectiveness: A test of alternative higher-order structures. Journal of Educational Psychology, 83(2), 285-296.

Marsh, H. W. (2007). Students' evaluations of university teaching: A multidimensional perspective. In R. P. Perry & J C. Smart (Eds.), The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective (pp.319-384). New York: Springer. Reports that SETs

McKeachie, Wilbert (1997). Student Ratings: The Validity of Use. American Psychologist, 52(11, November): 1218-25.

Walstad, William B. and Phillip Saunders (1998). Using Student and Faculty Evaluations of Teaching to Improve Economics Instruction, in W. Walstad and P. Saunders, eds., Teaching Undergraduate Economics: A Handbook for Instructors. New York: McGraw-Hill: 337-55.

Weinberg, Bruce, Belton Fleisher and Masanori Hashimoto (2009). Evaluating Teaching in Higher Education," Journal of Economic Education. 40(3, Summer) : 227-261.

Tables

Table 0: Response Rates

	Associates	Bachelors	Masters	Doctoral	Research I	Research II	Totals (Not Assoc.)
Total mailed*	4	223	310	44	108	89	774
#returned	0	50	67	11	33	21	182
Paper	0	34	46	6	27	16	129
Electronically	0	16	21	5	6	5	53
Response rate	0	.224	.216	.25	.306	.236	.235

*****4 more had either bad addresses or no Carnegie classification and did not respond.

Table 1: Administrative Location and Business Fields

	Overall	**RI, RII & D**	**M I & II**	**B I & II**
Total Number of Institutions	182	65	67	50
Administrative Location of Economics Department (% of total departments)				
Business	37.9	32.3	61.2	14.0
Liberal Arts or Arts and Sciences	55.5	60.0	31.3	82.0
Other	11.0	15.4	9.0	8.0
Does Institution Offer Bachelors Degrees in Business Fields? (%of total departments)				
Yes	83.4	83.1	100.0	62.0

Note: Some respondents checked more than one location.

Table 2: Institutional Characteristics

	Overall	RI, RII & D	M I & II	B I & II
Characteristics of Department:				
Department Size	11 (12.9)	17 (20.3)	9 (9.7)	7 (7.6)
Majors	80 (172.2)	200 (339.0)	52.5 (81.6)	52 (74.9)
Enrollments	1500 (2417.6)	4339 (4429.1)	1335 (1596.45)	673.5 (807.7083)
Percentage of Courses Taught by Regular Faculty	90 (81.0)	70 (70.9)	90 (86.7)	93 (86.6)
Weight, in percent, of Teaching Importance for:				
Annual Raises	40 (33.7)	40 (30.7)	40 (32.6)	50 (38.9)
Tenure	50 (45.3)	38.75 (31.2)	50 (48.7)	55 (58.1)
Promotion	40 (41.9)	33 (30.7)	50 (43.2)	50 (54.0)

Medians, with means in parenthesis.

Table 3: Teaching Evaluation Requirements

	Overall	RI, RII & D	M I & II	B I & II
Evaluation of Teaching Mandated (vs. Optional)	93.4	95.4	92.5	92.0
Evaluation by Students Typical	99	100	100	98
Source of Mandate of Evaluation of Teaching				
School Administration	40.4	32.8	39.7	52.2
Central Administration	30.1	37.5	25.4	26.1
Board of Trustees	8.7	9.4	12.7	2.2
State Legislature	4.6	6.3	4.8	2.2

Numbers are presented as percentages.

Table 4: Method of Administering Student Evaluations of Instructors

	Overall	D I, II &R	M I & II	B I & II
Student Member	50.0	44.6	54.6	51.0
Student Nonmembers	0.0	0.0	0.0	0.0
Class Teacher	2.8	0.0	3.0	6.1
Other Teacher	3.3	1.5	6.1	2.0
Secretary	6.7	6.2	9.1	4.1
Electronically	35.0	46.1	24.2	34.7
Other	2.2	1.5	3.0	2.0

Numbers are presented as percentages.

Table 5: SET Response Types and Frequency

	Overall	RI, RII & D	M I & II	B I & II
Student Responses Ordinal	97.8	98.5	91.0	98.0
Student Responses Written	95.6	96.9	95.5	93.9
Written Responses Reviewed By Dept.	82.7	81.3	82.3	85.1
Student Evaluations Given Once at End of Term	97.8	95.4	98.5	100.0

Numbers are presented as percentages.

Table 6: SET Items

	Overall	RI, RII & D	M I & II	B I & II
Course Organization and Planning	96.1	93.9	95.5	100.0
Instructor Clarity and Communication	98.9	96.9	100.0	100.0
Teacher-Student Interaction and Rapport with Individual Students	72.9	64.6	70.2	87.8
Teacher-Student Interaction and Rapport with Groups of Students	28.7	23.1	31.3	32.7
Grading	70.7	55.4	77.6	81.6
Student Self-Ratings of Amount Learned	59.1	66.2	53.7	57.1
Instructor Enthusiasm	71.3	60.0	80.6	73.5
Instructor Knowledge of Course Content	86.8	90.8	83.6	86.0
Instructor Use of Applications and Examples	39.2	32.3	43.3	42.9
Course Workload and Difficulty	71.8	78.5	62.7	75.5
Use of Instructional Technology	23.2	18.5	25.4	26.5

Numbers are presented as percentages.

Table 7: Availability of SET Results

	Overall	RI, RII & D	M I & II	B I & II
Individual Instructors	93.9	95.4	94.0	91.8
Course Coordinators	4.9	10.8	3.0	0.0
Department Chairperson	84.5	95.4	83.6	71.4
Department Review Committee for Annual Raises	28.7	47.7	19.4	16.3
T&P Committees	74.6	78.5	68.7	77.6
Dean or Associate Dean	66.9	60.00	67.2	75.5
Students in the Course	2.2	6.2	0	0
Any Student in the University	7.7	15.4	4.5	2.0
Publicly Available	6.6	7.7	9.0	2.0
Other	3.3	3.1	1.5	6.1
Weight SETs Are Given for Raises, Promotion, Tenure	48.7 (24.0)	53.1 (23.1)	46.0 (25.2)	47.0 (23.2)

Numbers are presented as percentages, with last row being the mean and standard deviation in parenthesis.

Table 8: Peer Evaluation of Teaching

	Overall	RI, RII & D	M I & II	B I & II
Is Peer Review Conducted? (Yes)	54.4	46.1	59.7	58.0
Who is Under Control of Evaluations?				
Faculty Member Being Reviewed	26.0	16.7	36.8	21.4
Economics Department	58.3	80.0	44.7	53.6
University	11.5	0.0	13.2	21.4
Other	3.1	0.0	5.3	3.6
Who Selects Evaluator?				
Faculty Member Being Reviewed	22.7	13.3	35.9	14.3
Department Chairperson	45.4	46.7	38.5	53.6
Dean or Associate Dean	2.1	0.00%	2.56%	3.57%
Student Group	0.0	0.0	0.0	0.0
Other	28.9	36.7	23.1	28.6
External Evaluators Usage				
Used	36.4	22.6	42.5	42.9
Percentage of cases used	43.7	37.4	34	62.8
Frequency of Review:				
P&T Time	60.4	56.7	53.9	74.1
Annually	29.2	36.7	28.2	22.2
Once each semester	8.3	6.7	12.8	3.7
More than Once a Semester	2.0	0.0	5.1	0.0

Numbers are presented as percentages. 99 out of 182 chairs indicated peer evaluations were conducted.

Table 9: Method of Peer Evaluation

	Overall	RI, RII & D	M I & II	B I & II
Classroom Observation	88.9	86.67	90.0	89.6
Videotape Review	2.0	0.0	5.0	0.0
Syllabus Review	72.7	73.3	70.0	75.9
Quiz/Exam Review	43.4	43.3	37.5	51.7
Instructional Materials Review	56.6	60.0	52.5	58.6
Student Work Review	11.1	6.7	12.5	13.8
Grade Distribution Review	29.3	26.7	30.0	31.0
Student Drop Rate Review	7.0	10.0	7.5	3.5
Student Interviews	2.0	6.7	0.0	0.0
Other	6.1	3.3	5.0	10.3

Numbers are presented as percentages. 99 out of 182 chairs indicated peer evaluations were conducted.

Table 10: Peer Review Criteria

	Overall	RI, RII & D	M I & II	B I & II
Course Organization and Planning	75.8	70.0	77.5	79.3
Instructor Clarity and Communication	92.9	90.0	95.0	93.1
Teacher-Student Interaction and Rapport with Individual Students	73.5	56.7	79.5	82.8
Teacher-Student Interaction and Rapport with Groups of Students	38.4	43.3	27.5	48.3
Grading	17.2	16.7	12.5	24.1
Instructor Enthusiasm	67.7	63.3	67.5	72.4
Instructor Knowledge of Course Content	81.2	80.0	87.5	75.9
Instructor Use of Applications and Examples	77.8	76.7	72.5	86.2
Course Workload and Difficulty	43.4	43.3	40.0	48.3
Use of Instructional Technology	36.4	30.0	45.0	31.0
Classroom Decorum	61.6	53.3	65.0	65.5
Other	2.0	0.0	5.0	0.0

Numbers are presented as percentages. 99 out of 182 chairs indicated peer evaluations were conducted.

Table 11: Availability of Peer Review Results

	Overall	RI, RII & D	M I & II	B I & II
Individual Instructors	75.8	73.3	85.0	65.5
Course Coordinators	1.0	3.3	0.0	0.0
Department Chairperson	84.9	96.7	77.5	82.8
Dean or Associate Dean	55.6	50.0	52.5	65.5
Students in the Course	0.0	0.0	0.0	0.0
Any Student in the University	0.0	0.0	0.0	0.0
Publicly Available	0.0	0.0	0.0	0.0
Other	21.21	40.0	10.0	17.2
Weight Peer Reviews Are Given for Raises, Promotion, Tenure (St. Dev.)	28.7 (19.2)	27.1 (16.3)	27.0 (19.5)	32.5 (21.3)

Numbers are presented as percentages. 99 out of 182 chairs indicated peer evaluations were conducted.

Table 12: Recognition of Curriculum Development and Instructional Research

Percentage of Department Chairs That Reported Some Faculty Developing in the Last Five Years:	**Overall**	**RI, RII & D**	**M I & II**	**B I & II**
Textbooks*	34.1	38.5	32.8	30.0
Workbooks and Study Guides*	32.4	30.8	35.8	30.0
Test Banks*	17.6	16.9	20.9	14.0
Course Outlines*	1.7	1.5	3.0	0.0
Computer Software for Instructional Use*	7.1	12.3	3.0	6.0
Refereed Journal Articles on Instructional Methods	44.5	43.1	55.2	32.0
Refereed Journal Articles on Assessment and Evaluation of Instruction	19.8	9.2	29.9	20.0
Authored Published Books on Teaching Economics	3.3	3.08	6.0	0.0
Edited Published Books on Teaching Economics	5.0	6.2	6.0	2.0
Contributed Chapters to Edited Volumes on Teaching Economics	17.6	16.9	20.9	14.0
Non-departmental Teaching Awards	53.3	56.9	56.7	44.0
Weight Curriculum Development and Instructional Research Are Given for Raises, Promotion, Tenure (St. Dev.)	8.7 (11.6)	6.9 (8.5)	10.4 (9.7)	8.8 (16.3)

*The question specified commercially published.

Numbers are percentages, with the last number being the mean response with the standard deviation in parenthesis.

[i] D. Larry Crumbley, co-editor of *Measure Learning Rather than Satisfaction in Higher Education* (2009), was quoted by David Glenn (2010a) saying "They (student evaluations of teaching) should be outlawed . . . They have destroyed higher education." In terms of legislation and enabling action, just the opposite seems to be happening. In 2009, the Texas legislature enacted a law (HB 2504) that requires public colleges and universities to post on Web sites curricula details on faculty members, including student evaluation scores. At the same time, Michael McKinney, Chancellor of Texas A&M University, enacted a bonus program in which the highest rated 15 percent of professors on student evaluations receive up to $10,000. Ironically, it was also in Texas that felon Jeffery Skilling instituted the performance review committee (PRC) at ENRON. PRC resulted in the annual firing and replacing of 15 percent of employees receiving the worst rating while at the other end of the five-point rating scale immediate gratification was prized and rewarded above long-term potential. (http://www.journalofaccountancy.com/Issues/2002/Apr/TheRiseAndFallOfEnron.htm) Apparently, because of tenure restrictions, Chancellor McKinney could not fire the bottom 15 percent, as bankrupted ENRON CEO Skilling could, but McKinney thinks that he can make Texas A&M run more like a Texas business by rewarding the top 15 percent of those providing contemporaneous customer satisfaction. Encouragingly, there are some who disagree: "Professors Question Texas A&M's Plan to Award Bonuses on the Basis of Student Evaluations." (http://chronicle.com/article/Professors-Question-Texas/42221)

Crumbley, was also quoted saying "Students are the inventory . . . At General Motors, you don't ask the cars which factory workers are good at their jobs." Becker and Rosen (1992) made clear, however, that unlike the assigning of quality grades to fabric, steel and any other inanimate objects, students are not merely inert pieces of inventory and indifferent to the manner in which they are treated and graded. Students are both an input and output in the education production process. Becker and Rosen explicitly modeled how the educational environment, the type of grading process employed and the student's position in the grading distribution affects his or her behavior.

[ii] The total mailed was 782. However, four were sent to schools classified with associate degrees, two to schools with no Carnegie classification, and two to bad addresses. No responses were received.

[iii] "Electronically (e.g., via computers)" was the language used.

[iv] The exact wording for the question is: In evaluating teaching as one component of decisions concerning annual raises, promotion, and tenure, how much weight is given to student evaluations of

teaching? (E.g., if only student evaluations are used to evaluate teaching, the response would be 100%.)
_____%

[v] For differences in proportion at a 1% level.

[vi] Bosshardt and Watts (2001) noted that grading rigor was an important component in determining the overall evaluation by students for instructors whose first language was not English.