

Teaching Assessment in Plant and Soil Science and Agricultural Economics Departments

Victoria Mundy,* Larry J. Grabau, and M. Scott Smith

ABSTRACT

The process of evaluating teaching should involve students, teachers, and teachers' colleagues. In 1997 we used a mail survey and phone interviews to investigate teaching evaluation methods in departments of plant and soil science and agricultural economics at land-grant universities in the USA. These teaching evaluation programs were not always carefully planned, and 48% of respondents believed that their teaching evaluation systems needed improvement. Of 97 departments, only 36 used data from students, peers, and teachers themselves. Seventeen departments used data from only one source. Student rating forms were very important to 90% of the departments, peer review of teaching was very important to 59%, and teaching portfolios were very important to 50%. Random student comments, student exit interviews, student achievement, and in-class visits were of much less importance. Credit toward promotion and tenure was the most common reward for good teaching, but the practical implications of such credit depended on the institution's overall attitude toward teaching.

THE stated mission of any university or college describes undergraduate classroom teaching as an activity of primary importance. Most professors value and enjoy teaching, advising, and other contact with undergraduates (Edgerton, 1993). Furthermore, classroom teaching is consistently reported as being important in "overall performance evaluation of faculty," although teaching is rewarded less tangibly and less frequently than research (Seldin, 1993; Austin, 1992).

Perhaps one reason that classroom teaching takes second place to research in the university reward structure is that teaching quality and quantity seem more difficult to evaluate than research quality and quantity. Seldin (1998) observed a 10-yr trend in liberal arts institutions toward structured data-gathering from many sources to evaluate teaching as a factor in personnel decisions. Other recent papers report that teaching is often evaluated with general, superficial information rather than with direct observation of teaching skills or careful analysis of course content and materials (Massy et al., 1994; Way, 1992).

Teaching is a human endeavor that cannot, and perhaps should not, be measured against generic criteria and summarized with a few numbers. But rewards in academia depend on thorough evaluation of scholarship; consequently, teaching must be evaluated if it is to be rewarded. Fair, valid, and reliable evaluation of classroom teaching requires the involvement of at least three parties: students, teachers, and teachers' disciplinary peers and other colleagues.

Dep. of Agronomy, N122 Agricultural Sciences North, Univ. of Kentucky, Lexington, KY 40546-0091. Corresponding author (toria@pop.uky.edu).

Published in *J. Nat. Resour. Life Sci. Educ.* 28:26-30 (1999).
<http://www.JNRLSE.org>

Furthermore, many different data sources should be used to evaluate teaching fairly and effectively, either for teaching improvement (formative evaluation) or for personnel decisions (summative evaluation) (Arreola, 1995).

The objective of this study was to investigate and describe methods of teaching evaluation used by departments of plant and soil science (PS) and agricultural economics (AE) at land-grant institutions in the USA.

METHODS

In June 1997 we mailed a survey to chairs of PS and AE departments at all 1862 and 1890 land-grant institutions in the USA. We considered PS departments and AE departments to compare *hard* science and *social* science in an agricultural context. Mailing lists came from the American Society of Agronomy and the American Agricultural Economics Association. Administrators who did not return the survey received a reminder memo. If need be, a new copy of the survey was sent. We held follow-up phone interviews with seven department chairs or undergraduate coordinators from AE and eight from PS who were willing to discuss their teaching evaluation (approximately 15% of respondents).

Our survey assumed that *teaching* meant classroom teaching, except for one question that specifically mentioned advising and service on departmental teaching committees. Respondents rated their summative evaluation procedures for teaching as completely inadequate, needing improvement, acceptable, or completely adequate, compared with summative evaluation of research in their departments. Respondents described departmental systems for calculating distribution of effort in teaching. Respondents ranked the sources of information used for teaching evaluation and ranked the importance of input from various peers or administrators. Checklists requested information on rewards for teaching efforts, opportunities for professional development in teaching, and teaching portfolios. Respondents assigned value to each of several aspects of teaching. Demographic information such as number of faculty was requested.

In our results and discussion, we present survey data and interview data together, using the two data streams to clarify one another. We used chi-square frequency analysis, Duncan's Multiple Range Test, and Pearson-Spearman dependence and correlation procedures to analyze quantitative survey data. Qualitative results from surveys and interviews were analyzed by pattern coding (Yin, 1997).

RESULTS AND DISCUSSION

Of 63 surveys mailed to agronomy or plant and soil science (PS) related departments, 48 were returned. Of 67 sur-

Abbreviations: PS, plant and soil science; AE, agricultural economics; DOE, distribution of effort.

Table 1. Adequacy of teaching evaluation, compared with effectiveness of research evaluation, in departments of PS and AE.†

Rating	Frequency of response‡			
	PS		AE	
	No.	% of total	No.	% of total
Completely inadequate	0	0	0	0
Needs improvement	27	56.3	20	40.8
Acceptable	16	33.3	25	51.0
Completely adequate	5	10.4	4	8.2

† PS, plant and soil science; AE, agricultural economics.

‡ Chi-square frequency analysis revealed no significant differences between PS and AE departments.

veys mailed to agricultural economics (AE) related departments, 49 were returned. The total return rate was 75%. Department chairs responded to all but 11 surveys. Undergraduate education coordinators completed those 11.

Most respondents perceived teaching evaluation strategies to be adequate or only slightly lacking in effectiveness compared with research evaluation (Table 1). The survey described *effective* evaluation as fair, reliable, and using all relevant information. We asked survey respondents to compare teaching evaluation with research evaluation, because the value of teaching is often compared with that of research when faculty and administrators make decisions about time and money. Therefore, it was reasonable for us to ask if the two activities are evaluated fairly in reference to one another. Although chi-square frequency analysis did not reveal statistically significant differences ($P > 0.1$), AE departments generally seemed more satisfied with teaching evaluation than PS departments. Respondents wrote qualitative comments about the reasons for their ratings. We grouped these responses into the categories presented in Table 2.

We found that teaching evaluation processes were deliberately and carefully planned in only 5 of the 15 departments we interviewed. In the other 10 departments, administrators either inherited review policies, or the procedures were informal and varied from professor to professor, or both. This lack of planning and consistency in evaluation may be one reason why <20% of the respondents felt that their departmental teaching evaluation was entirely satisfactory.

Summative teaching evaluation should, ideally, take several sources of information into account (Arreola, 1995). In order of importance, respondents ranked the sources of information used to evaluate instructors for merit or promotion and tenure decisions. We have reported only promotion and tenure results in this article, because merit is usually a much less rigorous decision than promotion and tenure.

Rankings went from 1 (most important) to 9 (least important or not used). Table 3 reports the frequencies of rankings, pooled across AE and PS departments because chi-square frequency analysis did not reveal differences between the disciplines. Totals for each column of Table 3 are >100% because some respondents used each ranking more than once. For example, some respondents ranked both student evaluations and peer review as 1.

Student Rating Forms

Student rating forms were clearly the most significant data in teaching evaluation. More than 90% of the respons-

Table 2. Comments concerning teaching evaluation programs.

Comment	No. of similar comments
Dissatisfaction with Current Teaching Evaluation	
We rely too much on student rating forms.	20
Evaluation is subjective and inconsistent (e.g., "popularity contest.")	15
We lack serious peer evaluation.	10
We lack "hard facts" for teaching evaluation, compared with numbers of research publications, grants, etc.	6
Evaluating teaching adequately takes too much time.	3
Satisfaction with Current Teaching Evaluation	
A range of data comes from students, peers, and administrators.	10
Student evaluations or "bits and pieces of information" are adequate.	5
It is fair, reliable, and objective (no further explanation).	3
It's been fine in the past.	3
No system is perfect.	3
Other Issues Mentioned	
We need to include more data.	5
Research evaluation isn't good either.	4
We need to evaluate learning outcomes instead of teaching techniques.	2

Table 3. Importance of various data sources in summative teaching evaluation.

Source of information‡	Frequency of response†		
	Importance of data source		
	Very important	Somewhat important	Unimportant
Student rating forms	90.7	5.2	4.1
Peer review	58.9	4.2	36.9
Teaching portfolios	49.9	13.8	36.2
Casual student comment	26.3	29.5	44.3
Student exit interviews	30.5	20.0	47.4
In-class visits	20.8	13.5	64.6
Student achievement	11.6	29.5	59.0
Videotaping	0	7.4	92.7

† Frequency of response is expressed in each category for each data source as the percentage of the total number of responses for that data source.

‡ Sources of information were ranked from 1 (most important) to 9 (unimportant or not used). Very important includes ranks 1 to 3. Somewhat important includes ranks 4 to 6. Unimportant includes ranks 7 to 9.

es placed student ratings as one of the three most important data sources. More than 58% of the responses described student ratings as the data source of first importance. But the perception that student ratings were overused and inadequate is clear from the comments in Table 2.

Peer Review

Peer review of teaching was very important to 59% of the departments we surveyed (Table 3). On the other hand, 37% of the departments did not consider peer review of teaching at all important in summative or judgemental evaluation.

Peer review of teaching is not a clearly defined term. Peer review may include first-hand observation and analysis of a professor's teaching by his or her peers. Peer review can also connote decisions made by faculty review committees based on consideration of second-hand data such as student evaluations and hallway conversations. In seven of nine interviews, we found that *peer review* implied that faculty colleagues gathered data by scrutinizing course materials, visiting classes, and talking with teachers. The intensity of data-gathering varied from department to department.

Eighteen departments reported that peer review was actually the most important single data source in teaching eval-

uation, even more important than student ratings. None of these 18 reported that peer reviewers were trained to evaluate teaching. Training is highly recommended to improve the reliability of peer review (Keig and Waggoner, 1994).

Four of these 18 departments had no formal system in which faculty peers supported one another in teaching development by visiting classes or consulting about classroom materials. Peer review in those four departments may have connoted only decision-making by peers. The other 14 departments reported a range of completeness in teaching development activities; 13 had in-class visits but may or may not have used pre- and postvisit consultation, review of materials, or student interviews.

Peer review has become an important movement in higher education in the last 10 yr (Edgerton, 1993), partly because institutions are under pressure to become more accountable for the quantity and quality of faculty work in the classroom (Seldin, 1993). Also, many professors desire that teaching should be accorded the level of prestige and reward that research enjoys. Such prestige requires changing teaching from a private enterprise to one openly discussed by the academic community (Shulman, 1993). Several of our interviewees mentioned that their departments were considering peer review. Their interest in peer review was not surprising, given the strength of the general educational movement toward better evaluation and reward for teaching.

Teaching Portfolios

The use of and value placed on teaching portfolios was not clear. Almost 50% of the departments reported that teaching portfolios were very important. Thirty-six percent reported that portfolios were not at all important. Three departments commented in writing that portfolios were not used but they reported in the ranking of data sources that portfolios were the most important or second most important sources. Many departments reported that portfolios were not used, but then later reported the required features of their portfolios.

Confusion about teaching portfolios also appeared in portfolio features that departments described. Our survey showed that student evaluations were the only data included in 100% of the teaching portfolios. Syllabi were the second most consistent feature at 90%, followed by personal teaching statements at 85%. Chair reports appeared in approximately 70% of portfolios, and teaching improvement strategies were found in 68%. Peer reviewer reports were included in 58%. Less than 50% of portfolios contained actual teaching artifacts such as assignments and tests. Murray's (1995) review of assessment literature indicates that portfolios should contain all of the above information, carefully selected to represent the range of teachers' activities and skills.

Completeness of Teaching Evaluation

We expected that departments in which many professors have relatively heavy teaching loads would have more comprehensive teaching evaluation systems. We calculated the percentage of teachers in each department who had teaching loads of >35%. Then we assigned an *index of completeness*

to each department, based on the three most important data sources used to evaluate teaching in that department. Our intent was to sort teaching evaluation systems that were clearly limited in scope from those systems that used more diverse data sources.

We referred to a *complete* teaching evaluation program as one that used data from students, from peers including administrators, and from faculty themselves (assumed to be teaching portfolios) as the three most important data sources. Thirty-six departments had complete programs. An *incomplete* program used data from two groups; 43 departments fell in this category. Seventeen departments used data from only one group, so we referred to them as *one-dimensional* programs. Most one-dimensional teaching evaluation programs used student rating forms, random student comment, and student exit interviews as the three most important data points. One department did not respond to the appropriate question. Chi-square frequency analysis revealed no differences in teaching evaluation completeness between AE and PS departments.

We found no correlation, according to Pearson-Spearman procedures, between the percentage of heavy teaching loads in the department and the comprehensiveness of teaching evaluation as described by our definitions. Departments that had a large percentage of faculty members who had >35% teaching appointments were no more or less likely to have comprehensive teaching evaluation than departments with few teachers with heavy teaching loads.

We found no correlation between reported satisfaction with teaching evaluation and the completeness of teaching evaluation as described by our definitions. Some departments that reported very complete teaching evaluations were in fact entirely satisfied. But some departments had well-rounded programs to evaluate teaching and were still dissatisfied. The latter situation seems to point out the difficulty of teaching evaluation and may indicate a serious commitment to the process of teaching and teaching evaluation and reward.

Some departments reported incomplete teaching evaluation and rated themselves as needing improvement. Others had very limited teaching evaluation and reported that their evaluation systems were adequate or completely satisfactory. In the latter departments, formal teaching evaluation may not have been a priority for some reason. One respondent, for example, reported that his department was so small and faculty talked among themselves so much that formal teaching evaluation was not felt to be necessary.

Attributes of Teaching

Teaching evaluation methods should be appropriate for the teaching characteristics that departments value. Respondents gave a score of importance to each of several different aspects or attributes of teaching. Respondents did not compare the aspects of teaching to one another but rated each aspect separately. A score of 10 indicates a very important aspect of teaching, while a score of 1 indicates an aspect of low importance. Responses were quite variable. Table 4 presents the data, pooled across AE and PS departments; chi-square frequency analysis did not reveal significant differences.

Table 4. Mean ratings by chairs or undergraduate coordinators of importance of aspects of teaching.

Aspect of teaching	Score of importance†
Mastery of content	8.9 a*
Commitment to teaching	8.1 b
Positive interaction with students	7.9 b
Stimulation of student interest	7.5 bc
Effective instructional materials	7.5 bc
Selection of course content	6.7 cd
Student assessment devices	6.5 de
Appropriateness of course objectives	6.2 de
Appropriate methodology for specific content	6.1 de
Innovative instructional materials	6.0 de
Advising	5.4 e
Service on departmental committees for instruction	4.6 f

* Means with the same letter are not significantly different ($P < 0.05$), Duncan's multiple range test.

† Each aspect of teaching was rated from 10 (very important) to 1 (unimportant).

The responses indicated that stimulation of student interest, effective materials, and appropriate methodology for teaching specific subjects were considered less valuable than mastery of content when teaching was evaluated summatively. Many departments did not use appropriate methods to judge the aspect of teaching they considered most important. Students are not suitable judges of a professor's mastery of content, yet student evaluations were the most commonly used and important data for teaching evaluation. The discrepancy between what departments value in teaching and their data for evaluating it should, and did, concern the departmental decision-makers we spoke with.

Distribution of Effort

Since deciding how to spend time is a primary concern of faculty, we asked respondents to describe how their departments calculate distribution of effort (DOE) for teaching. Reported methods of calculation varied tremendously, as did explanations of those methods. Forty-one respondents reported that a full-time teaching load was equivalent or nearly equivalent to 12 credit hours per semester. Seventeen respondents stated that full-time teaching was >12 credit hours per semester. Advising activities may or may not have counted in these calculations.

Twenty respondents reported that full-time teaching was <12 credit hours per semester; advising did not usually count toward the DOE. Thirteen respondents reported either that no formal DOE system existed in their departments or that the formal departmental system had no relationship to how faculty actually spend time. Five respondents reported that DOE was calculated by formula but did not describe the formulas.

Even when clear formulas for teaching DOE existed, faculty might not have been evaluated on the basis of their formal responsibilities. Only two respondents specifically wrote that their teaching evaluation systems deliberately took time commitment to teaching into account. Another person wrote that his department had been directed, apparently by some level of administration, to place no more than "one-third weight on teaching in retention decisions." One department chair said that even teachers with very heavy teaching loads were evaluated much more upon research productivity than upon teaching effectiveness, but if a person was not serious about teaching, then he or she would be

Table 5. Importance of different evaluators of teaching in decision making.

Evaluator	Frequency of response†				
	Importance of evaluator‡				
	1	2	3	4	5
Faculty peers	53	33	6	1	7
Department chair or head	46	46	4	1	2
Deans	8	16	44	2	30
Other administrators	1	2	11	32	55

† Frequency of response is expressed in each category for each evaluator as the percentage of the total number of responses for that evaluator.

‡ Decision-makers are ranked from 1 (very important) to 5 (unimportant) in teaching evaluation.

fired. A second chair said, "Teaching only counts if it's really lousy." But a third chair said, "The days of the top-notch researcher being a bad teacher are gone."

Since the DOE calculations are usually confusing to begin with, and discrepancy may exist between what people *do* and what they are *evaluated* upon, we are not surprised that many faculty members, particularly untenured professors, wonder how they should spend their time. They might not be rewarded, and may even be harmed professionally, if they choose wrongly. Teaching improvement would be easier if departments and institutions formally and clearly communicated expectations for both teaching and research responsibilities, then followed up by rewarding people who met those expectations.

Rewarding Teaching

Assigning meaningful, tangible rewards to teaching is challenging for many institutions because evaluating teaching equitably is so difficult. In our survey, 72 departments reported giving merit or promotion credit for efforts to improve teaching. The second most common reward for good teaching was teaching awards, reported by 19 departments. Ten departments reported the contradictory practice of releasing good teachers from teaching in order to do research, while five departments sometimes released teachers from research to do more teaching.

The administrators we spoke to clearly tried to reward teaching. But credit for merit pay and promotion decisions, the most common and probably most necessary reward for good teaching, was unevenly assigned. One interviewee said that his institution valued teaching over research in assigning credit for pay or promotion. A second interviewee reported that his department and the college of agriculture recommended a candidate for promotion based on teaching, but the new university provost, who preferred research, overturned the decision. The practical implications of merit and promotion credit depend on a department's, a college's, and perhaps even a university's attitude toward teaching.

Whatever the institutional attitude toward teaching, someone evaluates teaching for decisions such as tenure, promotion, or teaching awards. Respondents ranked four groups of evaluators in order of importance. Faculty peers and department chairs were equally important evaluators of teaching quality in all departments we surveyed (Table 5). Data are pooled across AE and PS departments. Faculty peers and chairs were much more important than deans or other administrators such as provosts. One department

reported that student advisory groups were somewhat important in teaching evaluation.

Departmental Attitudes Toward Teaching

In our interviews, two extreme types of departmental attitudes emerged. Most departments' attitudes fell between the two poles. At one extreme, teaching was valued only as the transmission of subject matter knowledge. This mindset was exemplified by one department in which a faculty member who had an 80% teaching appointment would still be evaluated twice as heavily on research productivity as on teaching effort. In this department, content expertise was the most necessary ingredient for good teaching. "Competent [disciplinary experts] are good teachers," the department chair remarked.

At the other end of the philosophic scale, teaching was valued as both a transmission of knowledge and as an exercise in human thought and relationships. This pole was exemplified by a department in which credit was given for interest and involvement in teaching. Performance evaluation was based on an annually written job description. This department believed that stimulation of student interest and concern for students were the most necessary ingredients for good teaching—twice as important, in fact, as mastery of content. "Teaching's what we do. If it weren't for teaching, we wouldn't be here," said this department chair.

CONCLUSION

Teaching evaluation is a continuing challenge, even for departments whose evaluation programs are carefully designed and use various data sources effectively. Student evaluations were by far the most commonly used data for teaching evaluation in both AE and PS departments. Peer review of teaching and teaching portfolios were distant but important seconds to student evaluations. Other evaluation methods such as class visits, student interviews, student achievement, student comment, and videotaping were not used very much.

We found that slightly more than half of the departments were content, if not completely satisfied, with their efforts in teaching evaluation. We had expected much more dissatisfaction, especially since student evaluations were so prevalent but so often remarked upon as inadequate. Student evaluations are especially inadequate because students are not appropriate judges of mastery of content, the trait that many departments most value in their teachers.

Part of the general satisfaction with teaching evaluation may have come from the relatively high use of peer review.

More than half the departments we surveyed used some form of peer review of teaching. The seriousness and depth of peer review varied from department to department. Our own experience with peer review suggests that the usefulness of peer review depends greatly on the willingness of faculty to participate thoughtfully and give honest, in-depth feedback. The usefulness of peer review may have varied in the departments we surveyed.

We found that departments with heavy teaching responsibilities were no more or less likely to have well-rounded, complete teaching evaluation programs than departments in which teaching responsibilities were lighter. We had expected that departments whose primary commitment was teaching would tend to take teaching evaluation more seriously than other departments.

Distribution of effort calculations were usually complicated and were sometimes unimportant in how faculty decided to spend their time or how they were rewarded. The most common reward for efforts in teaching was credit toward merit or promotion and tenure. Career credit seemed to depend on the institution's attitude toward teaching. Departments may be able to encourage good teaching by having well-communicated expectations for teaching and following through by rewarding faculty who meet those expectations.

REFERENCES

- Arreola, R.A. 1995. Developing a comprehensive faculty evaluation system. Anker Publ. Co., Bolton, MA.
- Austin, A.E. 1992. Supporting junior faculty through a teaching fellows program. *New Dir. Teach. Learn.* 50:73-86.
- Edgerton, R. 1993. The re-examination of faculty priorities. *Change* 25:10-25.
- Keig, L., and M.D. Waggoner. 1994. Collaborative peer review: The role of faculty in improving college teaching. ASHE-ERIC Higher Education Rep. 2. George Washington Univ. Grad. School of Educ. and Human Devel., Washington, DC.
- Massy, W.F., A.K. Wilger, and C. Colbeck. 1994. Overcoming "hollowed" collegiality. *Change* 26:11-20.
- Murray, J.P. 1995. Successful faculty development and education: The complete teaching portfolio. ASHE-ERIC Higher Educ. Rep. 8. George Washington Univ. Grad. School of Educ. and Human Devel., Washington, DC.
- Seldin, P. 1993. How colleges evaluate professors: 1983 vs. 1993. *AAHE Bull.* October 1993:6-12.
- Seldin, P. 1998. How colleges evaluate teaching: 1988 vs. 1998. *AAHE Bull.* March 1993:3-7.
- Shulman, L.S. 1993. Teaching as community property: Putting an end to pedagogical solitude. *Change* 25:6-7.
- Way, D. 1992. What tenure files can reveal to us about evaluation of teaching practices: Implications for instructional/faculty developers. *To Impr. Academy* 11:71-85.
- Yin, R.K. 1997. Case study research design and methods. Sage Publ., London.