Issues in Developing a Faculty Evaluation System

Raoul A. Arreola

Key Words: education, higher • teaching

The increasing demands for accountability in higher education are resulting in calls for important personnel decisions—such as promotion, tenure, pay, and continuation—to be based directly on the outcomes of systematic faculty evaluations. This article provides a step-by-step procedure for developing a fair and meaningful faculty evaluation system on which such personnel decisions can be based. The procedure systematically involves faculty and administrators in the design and development of a faculty evaluation program that reflects the unique values, priorities, and heritage of an institution. The resultant faculty evaluation system integrates data from students, peers, and administrators to provide meaningful evaluative information for both faculty use in self-improvement efforts and administrative use in making personnel decisions that are based on a valid and reliable faculty performance record.

long and has six matching chairs). But one couple evaluates the table as being good and the other evaluates it as being poor. The only thing that differs in this evaluation is the value system that each couple brings to the interpretation of the measurement. Perhaps the second couple was looking for a teak table of Scandinavian design, or maybe they did not like cherry wood, or maybe they wanted a rectangular table. The point is that the same measurement can result in either a positive or a negative evaluation, depending on the values and context in which the measurement is being interpreted.

The same is true of the evaluation of faculty performance. In designing a faculty evaluation system, we should make our measurements as objective as possible. However, instead of striving toward the illusion of an objective faculty evaluation system, we should recognize that our interpretations (evaluations) of those measurements will be based on subjective values and perspectives. What we should strive for is “controlled subjectivity,” that is, an agreed on set of values that will be used consistently in interpreting all measurements of faculty performance. The important principle here is that two persons may interpret the same measurement as indicative of either good or poor performance, depending on the value system by which the data are interpreted. The key to a good faculty evaluation system is to clearly identify and define an agreed on set of values and priorities that will be consistently used to interpret the performance measures.

The principle of controlled subjectivity, then, is based on the consistent use of an agreed on set of values. Observations are made of the performance of interest, and a judgment is made about whether that performance conforms to the set of values held by those interpreting the data resulting from the observations. If there is a good match between observed performance measures and values held, such performance is generally given a positive or good evaluation. If there is a poor match or a discrepancy between what is observed and what is held to be of value, such performance is generally given a negative or poor evaluation. Before an evaluation system can be built, the values of those who intend to use it must be determined. To develop a faculty evaluation system that correctly reflects the values of the institution, these values must not only be determined and be clearly in mind, but also must be expressed in such a way that enables evaluators to apply them consistently in the evaluation process.

Systematic Faculty Involvement

The strategy for developing a faculty evaluation system described here assumes that there is no one best system that could be successfully applied to all faculty groups. Developing a successful faculty evaluation system depends on the planned and systematic inclusion of the faculty’s values—the values that form the basis of the evaluation system. My experience has shown that the best approach to developing a faculty evaluation system is to appoint a committee composed primarily of faculty members, a few key administrators, and perhaps a student or two. This committee is responsible for gathering the information and following the steps described in this article. Thus, various steps in the process will refer to the faculty evaluation committee as the operational entity carrying out the process. It is important to secure faculty input as an integral part of the system’s design and construction.

Evaluation for Self-Improvement and Personnel Decisions

The results of any faculty evaluation system must serve two seemingly contradictory purposes: (a) providing feedback information for faculty member growth and development and (b) providing evaluative information on which to base personnel decisions such as promotion, tenure, continuation, and merit raise determination. The key to constructing a system that serves these differing purposes is in the policies determining the distribution of the information gathered. The general principle to be followed is that detailed information from questionnaires or other forms should be given exclusively to the faculty member for use in professional growth. However, aggregate data that summarize and reflect the overall pattern of performance over time of a faculty member can and should be used for personnel decisions.

Planning for the Development of a Comprehensive Faculty Evaluation System

Faculty Evaluation and Faculty Development: Two Sides of the Same Coin

Ideally, faculty evaluation programs and faculty development programs should work hand in hand. The operational rule of thumb assumed here is that if some aspect of faculty performance is to be evaluated, then there should exist resources or opportunities that enable faculty members to develop that performance.

Faculty evaluation systems that are implemented without reference to faculty development programs, no matter how well the evaluation systems are designed, are inevitably viewed by faculty members as being primarily punitive in intent. On the other hand, faculty development programs that are implemented without clear reference to information generated by faculty evaluation systems tend to be disappointing in their effect, no matter how well they are designed and funded. Without reference to a faculty evaluation system, faculty development programs tend to attract primarily those faculty members who are already motivated to seek out resources and opportunities to improve their performance. To achieve the greatest benefit from faculty evaluation and faculty development programs, the two types of programs must be linked. The faculty evaluation program should provide diagnostic information on strengths
and weaknesses of a faculty member, whereas the faculty development program should provide resources to enable the faculty member to enhance his or her strengths and overcome detected weaknesses.

**Obstacles to Establishing Successful Faculty Evaluation Programs**

Administrative commitment is necessary for establishing successful faculty evaluation and faculty development programs but so is faculty member acceptance. Resistance to being evaluated appears to grow out of three basic concerns: (a) resentment of the inferred assumption that faculty members may be incompetent in their subject area, (b) suspicion that they will be evaluated by unqualified persons, and (c) faculty members’ anxiety that they will be held accountable for performance in an area in which they have little or no training or interest. This last anxiety is not unusual, even though most faculty members may attribute most of their concern to the second factor. Milton and Shoben (1968) pointed out the basis for this anxiety when they stated that “college teaching is probably the only profession in the world for which no specific training is required. The profession of scholarship is rich in prerequisites for entry, but not that of instruction” (p. xvii).

Several publications have addressed the issue of overcoming faculty resistance to evaluation programs (Arreola, 1983, 1995; Grasha, 1977; O’Connell & Smartt, 1979; Seldin, 1980). The underlying premise for developing a comprehensive faculty evaluation system is the careful and deliberate preclusion of faculty resistance by systematically providing opportunities for faculty members to have a say in the development of their own evaluation system.

**Steps in Developing a Comprehensive Faculty Evaluation System**

**Step 1: Determining the Faculty Role Model**

The first step in developing a comprehensive faculty evaluation system is to determine which faculty activities to evaluate; that is, the faculty role model for the institution must be determined. A common description of what constitutes the faculty role model for an institution is teaching, research, and service (service including clinical service in health professions programs). However, faculty members must engage in a wide variety of activities in complex academic programs. In addition to teaching, conducting research, and performing various service activities, faculty members also advise students, publish articles and books, serve on committees, administer programs, and perform many other essential duties. All of these activities must be considered for inclusion in a faculty evaluation system if it is to be fair. Thus, the traditional faculty role model, consisting of teaching, research, and service, may be insufficient to adequately encompass the full range of the faculty member’s professional activities. Arreola (1995) and Miller (1972) provided more comprehensive treatments of the activities that define possible faculty roles in an academic institution.

**Step 2: Determining Faculty Role Model Parameter Values**

Within any institution, a wide variety of opinions or positions on the relative value of faculty roles will exist. Some faculty members will hold teaching to be of primary importance, others will hold faculty service to be of greatest importance, and still others will maintain that scholarly research and creative endeavors are the most important functions of the institution. To reach a consensus in developing a value structure for the evaluation system, the parameter values for each faculty role should be established; that is, the maximum and minimum weights that could be assigned to a role are determined.

Table 1 shows the parameter values for a hypothetical faculty role model. Notice that teaching ranges in weight from a minimum of 50% to a maximum of 85%. This would be interpreted to mean that in the evaluation of a faculty member’s overall performance, teaching performance would count no less than 50% of the overall evaluation and no more than 85%. This does not mean that a faculty member may have a 50% to 85% teaching load; rather, these numbers are an expression of how much weight will be given to performance of teaching in the faculty member’s overall evaluation. A teacher may or may not have a full-time teaching load, but the value associated with teaching performance in this evaluation system would range from 50% to 85%.

Determining the parameter values for a faculty role model is actually a political process that involves consensus building between faculty and administration. The minimum and maximum weights should reflect both the values and priorities of the administration and the general sentiment of the faculty. This is best accomplished by having faculty members respond to a questionnaire that asks them to indicate their values or priorities relative to the roles in the faculty role model and using that data to determine consensus parameter values for the faculty role model.

**Step 3: Defining Roles in the Faculty Role Model**

The third step in developing a comprehensive faculty evaluation system should be carried out concurrently with Step 2. As noted earlier, it is assumed that a faculty evaluation committee or similar committee will coordinate the detail work associated with this project. Step 3 involves coming to a consensus about how each identified role is defined. For example, teaching as a role will readily be agreed on. However, different faculty members mean different things when they use the word teaching. It is important to agree on a consistent definition of teaching if teaching is to be evaluated fairly for all faculty members.

The teaching role may be defined as being made up of the following four components:
1. Content expertise: the body of skills, competencies, and knowledge in a specific content area in which the faculty member has advanced experience, training, or education.

2. Instructional delivery skills: those human interactive skills and characteristics that (a) ensure clear communication of information, concepts, and attitudes and that (b) promote or facilitate learning by creating an appropriate effective learning environment.

3. Instructional design skills: those technical skills in (a) designing, sequencing, and presenting experiences that induce student learning and in (b) designing, developing, and implementing tools and procedures for assessing student learning outcomes.

4. Course management skills: those bureaucratic skills involved in operating and managing a course, including but not limited to (a) timely grading of examinations, (b) timely completion of drop-add and incomplete grade forms, (c) maintenance of published office hours, and (d) generally arranging for facilities and resources required in the teaching of a course.

By defining the total teaching act in terms of these four broad components or dimensions, it becomes clear that the evaluation of teaching cannot be accomplished neither with only one questionnaire or form nor on the basis of the judgment of one administrator or peer committee after only a few classroom visits. No one person or group can have a sufficiently detailed and complete view of the entire teaching process. A more accurate and valid perception of teaching performance would, of necessity, involve information from (a) students on their opinions and reactions to the teacher’s instructional delivery skills and characteristics; (b) peers, and perhaps informed experts, on the teacher’s instructional design skills; (c) peers and department heads on the teacher’s content expertise (if such information was required); and (d) the department head, or perhaps even the departmental secretary, on the teacher’s course-management skills. Additionally, information is desired from students about the teacher’s apparent content expertise and students’ reactions to several aspects of the course operation from which inferences could be made about the teacher’s instructional design skills. Thus, the key to more effective evaluation of teaching is to carefully take all the parts of this mosaic and put them together in such a way that the faculty member’s overall teaching competence is accurately reflected. The rest of the steps outline the procedure for doing this.

**Step 4: Determining Role Component Weights**

After developing definitions for the various roles in the faculty role model and determining the relative impact or parameter values that the different roles can take in the overall evaluation of a faculty member, the weight or relative importance the various components of each role should have in the overall evaluation of that specific role must be considered and determined. In the example in Step 3, teaching was defined as involving four components: instructional delivery skills, instructional design skills, content expertise, and course management. The issue now is to determine how much weight or relative importance each of these four defining teaching role components should have.

The Source Impact Matrix is a tool that controls the effect of the subjective data gathered as part of the overall evaluative process. Figure 1 shows an example of a partially completed Source Impact Matrix for teaching. Note that the sources of information have not yet been determined and are simply left blank. Similar matrices will have to be constructed for the other faculty roles (e.g., research, clinical service). In Figure 1, the instructional delivery skills component is weighted as 30%, the instructional design skills component as 40%, the content expertise component as 25%, and the course management component as 5%.

These weights reflect the relative importance that the various defining components of the teaching role hold for the faculty members in the hypothetical institution. Thus, whatever the rating or evaluation outcome is for the instructional delivery skills component for a given faculty member, that rating will count only 30% of the total evaluation of the teaching role. Likewise, the rating or evaluative outcome of the instructional design skills component will count 40%, and so on. The weights used in this example are entirely subjective. Your institution may establish different weights.

### Step 5: Determining Appropriate Sources of Information

Your next step is to agree on who should provide the information on which the evaluations will be based. Too frequently, students are automatically selected as the sole or primary source of the information used in a faculty evaluation system. Students are appropriate sources of information for certain kinds of activities, but they are not the best source of information for all the activities in which faculty members engage and on which they may be evaluated. The most important principle in identifying and selecting sources of information is to make certain that the source identified has firsthand knowledge of the performance being evaluated. Figure 2 shows a Source Identification Matrix completed appropriately for the teaching role as previously defined.

In completing this step, yes or no decisions are made
about whether a particular source of information should be tapped for the role component in question. There are no right or wrong answers. The purpose of this matrix is to provide a structure for the discussions of what sources should provide the data to be used in the evaluation of teaching. A similar matrix would be used for each of the other roles (e.g., research, service, clinical practice). Proceeding in this way through each role's defining components, it is possible to make rational decisions and determinations about what sources of information would be appropriate and acceptable to the faculty.

Step 6: Determining Information Source Weights

In any faculty evaluation system, judgments and evaluations will be based on information derived from a number of sources. This information will concern various elements or components of the roles being evaluated. The issue of the appropriateness of those sources was addressed in Step 5. After determining where the information to be used in the evaluation system will come from, the credibility of those sources is addressed. This is accomplished by the Weighted Source by Role Component Matrix (see Figure 3).

In this hypothetical situation, following the yes and no decisions made in completing the Source Identification Matrix (Figure 2), the faculty evaluation committee has determined that information provided by students on instructional delivery will be all the information that will be considered in the evaluation of that component. This results from the earlier decision that neither peers nor the department head would be sources of information for this component. Thus, the total weight for student input on the instructional delivery skills component is 100%.

To this point, we have specified the credibility of the various sources of information and the relative importance of each defining component for every role by determining the associated weights. This information enables us to complete the Source Impact Matrix and control the impact of information from each source on the overall evaluation of a faculty member.

Figure 4 shows an example of a completed Source Impact Matrix for the teaching role. In this example, the information recorded on the Weighted Source by Role Component Matrix for teaching has been entered into the small boxes in the corresponding cells of the Source Impact Matrix for the teaching role (see Figure 4). By multiplying the source weights for each cell by the row weight for that role component, it is possible to obtain an indication of the impact any one source of information will have on the overall evaluation of a role.

Note that both the sum of the component weights and the sum of the source impact weights must each equal 100%. In examining the matrix in Figure 4, we can see that...
the system we are designing will give 40% of the impact of the evaluation of teaching to what students say, 50% to what peers say, and 10% to what the department head says. If there is agreement with this value structure, the faculty evaluation committee can proceed. If there is disagreement, then the decisions that led to this value system must be revisited. For example, suppose we wished the department head input to have 25% of the impact on the evaluation of teaching. We must either increase the number of areas in which the department head provides information (with all that implies) or increase the weight given to those areas for which the department head already provides information. Appropriate adjustments would also have to be made in reducing the weights and activities of the other sources.

Step 7: Determining How Information Should Be Gathered

Because this step involves the more technical area of measurement, enlist the aid of those faculty members whose area of expertise is tests and measurement would be beneficial. Such faculty members may be placed on the faculty evaluation committee in the first place.

This step determines how the information specified in the role definitions is to be gathered from the identified sources. This step requires a careful review of the roles and the development of an operational plan for the final faculty evaluation system. The Data Gathering Tool Specification Matrix is completed for the teaching role (see Figure 5). Note that the cells completed in Step 6 containing zeros (0%) in the previous Source Impact Matrix (Figure 4) are blanked out. Because data will not be gathered from those sources for these elements, the tools or means for doing so need not be specified. The remaining cells are used to represent how information will be gathered from students, peers, and the department head. Again, the matrix is used simply as a tool to structure and guide the discussions of the best way to gather the information desired from each identified source. It should be clear by this step that there will be a variety of questionnaires, rating forms, and other data collection strategies used in the entire faculty evaluation system.

Step 8: Completing the System—Selecting or Designing Forms and Protocols

It is recommended that a small team of faculty members with expertise in tests and measurement design a final set of forms rather than the task being taken on by the entire faculty evaluation committee. The full committee has determined all the specifications about what is to be measured and who is to be tapped for the information. This specification provides sufficient directions for the technical team to follow in developing the questionnaires, protocols, and so forth. My experience has shown that if the entire committee takes on this task, previous agreements can unravel after the item-by-item determination of the forms and questionnaires is under way. Faculty members unfamiliar with the principles of psychological measurement are likely to overlook the fact that well-designed questionnaires may include questions that, taken in isolation, may be argued but, taken in the aggregate, provide valid and reliable measures of the characteristic or role component in question.

Timetable for Developing a Comprehensive Faculty Evaluation System

After the technical team designs the questionnaires, protocols, checklists, and other forms, the faculty evaluation system is ready to be implemented. The development of a comprehensive faculty evaluation system, using the strategy outlined in this article, generally takes about 2 years to complete. Guideposts during the process include the following:

- Month 1: Appoint faculty evaluation committee, familiarize committee with system development procedure, and hold general faculty meeting, sponsored by the committee, where procedure is presented and explained.
- Month 2 to 6: Committee distributes questionnaires to faculty to develop the faculty role model, weights for the roles, definitions of roles, sources of information, and weights for each source.
- Month 7: Committee reports to the general faculty the total value structure and role definitions as determined by their input.
- Month 7 to 12: System forms and protocols are designed, selected, and developed. Policy decisions on confidentiality and the use of the information in promotion, tenure, and merit pay decisions are finalized.
- Month 12 to 24: Trial run of system. Time of stress because decisions on promotion, tenure, and merit pay during this time will still have to be based on old system. Debug system, make adjustments.
- Month 25: Full implementation of system.

Using the System: Combining Data To Generate an Overall Composite Rating

The task now is to combine all the data gathered using the system in a usable form, and, to do this, all information...
gathered from each source must be reported on a common scale. That is, regardless of whether a questionnaire, an interview schedule, or some other technique has been used in gathering evaluative information from the various sources identified, those data will be reported on the same scale. In the examples that follow, a scale of 1 to 4 will be used, where 1 is the lowest rating and 4 is the highest. In actual practice, the precise meaning of a rating of 1, 2, 3, or 4 will depend on the particular form being used. A 4 on a student rating form focusing on instructional delivery skills may have a different specific meaning than a 4 on a peer rating of the course design. However, in all such forms a 1 will constitute the lowest rating, and a 4 will constitute the highest.

Computing the Composite Role Rating

After determining and specifying the weights to be assigned to various activities and sources in the overall faculty evaluation system, it is possible to compute an overall rating for each role reflective of a consensus value system. This rating is referred to as the “composite role rating” because it is derived from various sources, with each source providing information on various components of each role and with the information from each source and component weighted in ways that reflect the collective value structure of the institution. The following is an example of how the composite role rating for teaching is computed.

Figure 4 shows that the information that students provide about the faculty member’s instructional delivery skills will affect the overall rating of the teaching role by 30%. Likewise, student information about the instructional design skills component will count 10% and peer information 30%. Further, peer input on content expertise will count 20%, and department head input will count 5%. Finally, it was determined that department head input on course management will count 5% of the overall rating on teaching. Table 2 shows these weights along with the rating (shown in brackets) each source has given each role component.

Note that the composite role rating of 3.45 was not assigned by any one student, peer, or administrator. Rather, this value represents an aggregate of information on activities the faculty members agreed should be evaluated, collected from sources that were agreed to be appropriate, and weighted to reflect both the credibility of the sources and the relative importance of each component of the entire role. Although the composite role rating may involve objective measurement data in its computation, it does not represent an objective measure in and of itself. Because values have been introduced into the computation, the resultant composite role rating contains an element of subjectivity; however, the subjectivity involved in computing it has been carefully controlled and prescribed by the values assigned to the sources and role components as agreed on in the design of the system. A similar procedure is followed in determining the composite role ratings for the other roles (e.g., research, faculty service, community service).

Individualizing the Evaluations

Assume that a faculty member has received the composite role ratings shown in Table 3 (computed as shown in Table 2); that is, each composite role rating is the result of gathering specific information from specified sources, and weighted in ways that reflect the value system of the faculty and the institution. Suppose this faculty member has an assignment as reflected in Table 4. Recall that the faculty role model for our hypothetical institution (Table 1) allowed a minimum of 50% weight on teaching and a maximum of 85%. Likewise, the minimum and maximum weights were 0% and 35% for scholarly research, 10% and 25% for faculty service, and 5% and 15% for community service. To combine the faculty member’s several separate composite role ratings into an overall composite rating, each composite role rating is multiplied by the assignment weights shown in Table 5 and the total is computed.

Note that the faculty member’s overall composite rating of 3.34 was not determined by any one individual or group. Rather, the overall composite rating can be thought of as a singular “index of success” because it was assembled by gathering information from various sources, weighted in ways that reflect the credibility of those sources, and further weighted by the assignment emphasis for this faculty member. That is, given the particular assignment this faculty member had this year, the various appropriate sources provide a mosaic of information, which is expressed in the overall composite rating. The computation of an individualized overall composite rating (OCR) results in what can be correctly characterized as an index of perceived success, or this faculty member’s individual and unique set of professional responsibilities. The OCR is a tool that may be appropriately applied to decisions on promotion, tenure, continuation, and merit pay.

Table 2

<table>
<thead>
<tr>
<th>Teaching</th>
<th>Students</th>
<th>Peers</th>
<th>Department Head</th>
<th>Weighted Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructional delivery</td>
<td>30% x [4]</td>
<td></td>
<td></td>
<td>1.20</td>
</tr>
<tr>
<td>Instructional design</td>
<td>10% x [4]</td>
<td>30% x [3]</td>
<td></td>
<td>1.20</td>
</tr>
<tr>
<td>Content expertise</td>
<td>20% x [4]</td>
<td></td>
<td>5% x [3]</td>
<td>0.95</td>
</tr>
<tr>
<td>Course management</td>
<td></td>
<td>5% x [2]</td>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td>Composite role rating</td>
<td></td>
<td></td>
<td></td>
<td>3.45</td>
</tr>
</tbody>
</table>
Conclusion

By maintaining a record of a person's OCR over a course of years, it is relatively easy to track his or her success in accomplishing each year's unique set of professional responsibilities. With such data, tenure or promotion decisions can be made based on the basis of a reliable track record. Additionally, the OCR resulting from the current year's evaluation provides appropriate data on which to base merit pay increases or other performance pay raises fairly and more consistently. In serving as a consultant to more than 100 institutions, including universities, community colleges, technical institutes, and 4-year liberal arts colleges that have used this approach in developing their faculty evaluation systems, I have found that no two institutions will develop identical systems. However, each system developed will reflect the values and priorities of the institution's faculty and administration and will thus have the best chance of successfully enhancing faculty performance and administrative decision making.

Table 3
Hypothetical Composite Role Ratings

<table>
<thead>
<tr>
<th>Role</th>
<th>Composite Role Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching</td>
<td>3.45</td>
</tr>
<tr>
<td>Research</td>
<td>3.20</td>
</tr>
<tr>
<td>Faculty service</td>
<td>3.60</td>
</tr>
<tr>
<td>Community service</td>
<td>2.60</td>
</tr>
</tbody>
</table>

Table 4
Hypothetical Assigned Role Weights

<table>
<thead>
<tr>
<th>Role</th>
<th>Assigned Role Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching</td>
<td>50</td>
</tr>
<tr>
<td>Research</td>
<td>35</td>
</tr>
<tr>
<td>Faculty service</td>
<td>10</td>
</tr>
<tr>
<td>Community service</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 5
Computation of Overall Hypothetical Composite Rating

<table>
<thead>
<tr>
<th>Role</th>
<th>Assigned Weight (%)</th>
<th>Composite Role Rating</th>
<th>Weighted Composite Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching</td>
<td>50</td>
<td>3.45</td>
<td>1.73</td>
</tr>
<tr>
<td>Research</td>
<td>35</td>
<td>3.20</td>
<td>1.12</td>
</tr>
<tr>
<td>Faculty service</td>
<td>10</td>
<td>3.60</td>
<td>0.36</td>
</tr>
<tr>
<td>Community service</td>
<td>5</td>
<td>2.60</td>
<td>0.13</td>
</tr>
<tr>
<td>Overall composite rating</td>
<td>50</td>
<td>3.34</td>
<td></td>
</tr>
</tbody>
</table>

References


