Approaches to Comprehensive and Systematic Faculty Evaluation

Steven H. Smartt
Southern Regional Education Board
Jon F. Wergin
Medical College of Virginia
Albert B. Smith
University of Florida
George E. Rolle
Southern Association of Colleges and Schools

It is ironic that faculty evaluation has been a long-standing problem in a profession which has as one of its primary activities the evaluation of others' performance. Until recently, however, this situation was not sufficiently critical to require a change in the methods colleges and universities used to upgrade their faculties or weigh personnel decisions. While enrollments and programs were expanding, new faculty and professional mobility provided a steady influx of new blood, and high tenure ratios were an unlikely concern. But in a period of little growth, such as we now are experiencing, the attitudes of constituencies have changed. Governing boards and institutional administrators recognize the importance of making valid and equitable personnel decisions. Faculty themselves feel that evaluation is often lacking in substance and fairness. And, increasingly, evaluation is recognized as a first step in the process of faculty development, particularly instructional improvement.

The Southern Regional Education Board (SREB) conducted a survey in 1975 to determine the characteristics of faculty

evaluation practices in colleges and universities in a fourteen-state region. This survey and subsequent case studies showed, among other things, that evaluation tended to concentrate on individual components or separate procedures rather than on comprehensive approaches, and evaluation data or evidence was not gathered systematically or consistently.[1,2] In the fall of 1977, SREB, with support from the Fund for the Improvement of Postsecondary Education and the Carnegie Corporation of New York, undertook a two-year demonstration project to promote the improvement of faculty evaluation, emphasizing two principles: comprehensive scope and systematic application.

Thirty institutions, representing an array of types and levels, were chosen from among some 60 applicants to participate in the project and revise their evaluation approaches based on these two principles. Each college was required to appoint a special committee (a project team) to serve as the central campus group to work toward improving evaluation. These teams, consisting of at least one academic administrator and two faculty members, attended three workshops sponsored by the project, and hosted up to an equal number of visits by project consultants to assist them with planning on their campuses. The workshop and consultation experiences did not encourage the schools to adopt a particular system of faculty assessment. Instead, they promoted locally-developed approaches based on awareness of a number of important considerations identified by the previous SREB research.

Characteristics of Improved Faculty Evaluation

The thrity institutions represented not only an assortment

of types but also a variety of goals for their participation in the project. The participants could be categorized into three groups, according to their original objectives. Half of the institutions had as their objective to develop a completely revised faculty evaluation system. Nine others sought to make improvements in an existing program (e.g., working on a new student rating form). The remaining six set out to analyze their current programs and concentrated on developing more consistent policies and procedures.

Progress toward these self-defined goals was good. A team of project evaluators observed all project activities, carefully examined all relevant materials, and conducted intensive site visits at half of the institutions, randomly selected on a stratified basis (by type of institution). The evaluators concluded that observable progress toward goal accomplishment was evident in all but four of the institutions. Progress was most impressive among schools in the first group, i.e., those starting from scratch. Among these 15 institutions, five had developed and implemented a comprehensive faculty evaluation program, and all but two of the other ten had made significant advances.

Except in the case of two relatively new institutions, at the outset of the project all participating schools were already evaluating faculty in one manner or another. The typical procedure consisted of a form for student ratings of instruction. By the end of the project, after giving deliberate attention to making their programs more comprehensive

and systematic, the evaluation program of most of the institutions had the following characteristics:[3]

A clear statement of purpose. Some programs were intended to serve only as a basis for faculty development, and their subsequent design was based on this function. Other programs were designed for both faculty development and personnel decisions.

Multiple faculty activities evaluated from multiple sources of evidence. While instruction was considered the primary faculty function to be rated, others were included, such as research and professional activities, public service, counseling and advising, teaching improvement activities, and syllabus materials. Sources of data included students, peers, self, department heads and deans.[4]

Publicly stated, regular procedures. A well-conceived, documented, and respected sequence of activities was viewed as critical to the equitable and useful application of the program. Procedures sometimes included description of an appeals process.

Flexibility to allow for individual differences. Many of the new systems provided for a variable rating scheme which required that the department chairperson and faculty member discuss in advance the individual's assignments for the coming year, and negotiate the weight each would receive in the evaluation.

Critical Factors in Progress and Permanence

Given the varied approaches to developing improved evaluation programs that were found among these 30 institutions, the project sought to identify those variables which distinguished the more successful from the less successful. Further, an attempt was made to speculate about the expected mortality of the new programs. Probabilities of permanent impact were defined in terms of relative evidence of progress, support, and follow through. After sorting the institutions according to the likelihood of permanence of their revised evaluation systems, it was concluded that neither the type of institution nor the nature of their goals were significant predictors of long-term impact.

The project did identify, however, several factors which seemed to characterize institutions which made more progress and had higher likelihood of sustained impact. These characteristics are as follows, in approximate descending order of significance:

active support and involvement of top-level administrators[5]

faculty involvement throughout the project

faculty trust in the administration

faculty dissatisfaction with the status quo

historical acceptance of faculty evaluation at the institution

presence of an institutional statement on the philosophy and uses of faculty evaluation

centralized institutional decision making

Student Learning as a Measure of Faculty Effectiveness

As an extension of the philosophical question, "If the student hasn't learned, has the teacher in fact taught?", some have advanced the notion that measures of student learning should be used as indicators of teaching effectiveness. The SREB project attempted to promote a related position, namely, that improved faculty evaluation contributes to improved instruction, which in turn leads to better learning. What is uncertain is whether better learning means more learning, faster learning, or some other variety of learning.

Although others have offered approaches to faculty evaluation based on student achievement and other measurement procedures,[6] it is of interest that attention to this means of assessing faculty performance received consideration by only a few participants in the SREB project. One large university undertook an experiment to obtain an assortment of test scores and other indicators both at the outset and the conclusion of an economics course, while varying the teaching approaches among the course sections. A community college pursued the idea of giving standard examinations to students upon completion of selected courses, regardless of instructor or section. The most significant aspect of these attempts--and the most encouraging--is that these measures were intended to be used in conjunction with other, more traditional means of faculty evaluation. While most of the

institutions voiced an appreciation for the importance of the relationship between faculty evaluation and student learning, their experiences have not yet contributed to our understanding of how to use one as a direct measure of the other.

Implications for Other Institutions

From assisting this group of institutions in the process of improving evaluation and attempting to identify the variables which seemed to have a causal effect on the results of their efforts, a number of conclusions are possible. Institutions desiring to make significant and lasting improvements in their evaluation programs should be attentive to four points in particular:

...strong administrative support. The presence of strong support will not be sufficient to bring about a successful program in the absence of other factors, but the lack of top-level support will almost certainly render a program ineffective.

...full and extensive faculty involvement. The most progress in the SREB project was evident in institutions where the project team expended its base of input and support.

...expertise. Consultants provided the strongest boost to institutional teams in several instances. There are many competent resource people who have a sense of what works and what does not. Especially needed was

assistance in designing and validating evaluation instruments.

...recognized need for improvement. Institutional change must always confront problems posed by institutional inertia. General dissatisfaction with the status quo is a strong ally in any change effort.

In addition to these points, several lessons were reported by project participants. They offer useful information to others who pursue a similar course.

1. The process is more time-consuming than anyone might expect. It was initially expected that most of the participants would design and implement a new program within 18 months. Only five of the 30 were able to do so, and even those had begun planning before their involvement with the project.

2. A statement of purpose should be considered "square one." When this step was overlooked or consensus was lacking, teams found they had to return to this basic requirement. Many design principles are dependent on the intended purpose of the evaluation program.

3. The incremental approach is advisable. Starting with one purpose (faculty development) is easier than two (development and personnel matters). A single rating procedure can be expanded to several over time.

4. Implementation can be more difficult than design.

5. It is likely that not all of one's assumptions about existing evaluation programs or campus attitudes are correct. Many teams found their "foregone conclusions" to

be myths on closer examination, e.g., faculty disdain for evaluating others or themselves being evaluated.

A Multi-Institutional Strategy for Change

The project design was predicated on several assumptions: (a) institutions could benefit from collectively addressing the same issues within the context of their own settings; (b) a select group of local campus leaders should be appointed to direct the efforts under special mandate or authority; (c) plans and goals should be recorded and regularly monitored for progress and revision; (d) a series of activities and externally imposed timeframes would help campus teams devote the necessary time and effort; and (e) expertise not readily available internally would have to be made accessible. Although a classical test of these assumptions by contrast with a control group was not part of the project design, these assumptions were generally supported by the project's outcomes.

Reactions to the workshops suggested that the benefits of convening participants for interaction and exchange of ideas were secondary to the opportunity provided for uninterrupted blocks of time for teams to work independently and develop rapport with potential consultants. Many participants did report, however, a feeling of stronger self-confidence after having a forum for discussing solutions and finding that others shared similar problems.

The project evaluators concluded that this kind of regional, multi-institutional approach to addressing a

problem is worthy of consideration for dealing with other issues in higher education. This conclusion is supported by SREB's experience in similar projects, as well. It should be noted that some of the project's high achievers (i.e., those which demonstrated good progress and likelihood of extended impact) probably could have made significant advancement even without assistance from this project because they exhibited strengths in the critical factors associated with successful change. There were also a few institutions which did not make much progress because of a prevailing lack of commitment, and the project made little noticeable difference in the state of faculty evaluation on these campuses. Institutions with a middle range or moderate degree of interest benefitted most from this collective design. Participation by these institutions resulted in commitment and involvement of their faculties and administrations.

In conclusion, it is obvious that effecting change in faculty evaluation is characteristic of other attempts at educational improvement. Most parties are initially concerned with questions of who is likely to gain and who might lose. Although an administrative directive can serve as an impetus for revising an evaluation system, a useful program will more likely result from long-term deliberations involving several participants. Observation of thirty institutions in such a process suggests that colleges and universities are

capable of establishing a clear purpose for faculty evaluation, identifying several faculty functions for assessment by several sources of information, providing for flexible weighting of criteria, and designing an orderly procedure for the overall process.

Notes and References

1. Boyd, James E., and E. F. Schietinger, Faculty Evaluation Procedures in Southern Colleges and Universities, Southern Regional Education Board, 1976.
2. Faculty Evaluation for Improved Learning, Southern Regional Education Board, 1977.
3. A further description of the project's design and activities, and the typical programs in the institutions before and after the project, can be found in Improving Faculty Evaluation: A Trial in Strategy, by William R. O'Connell, Jr. and Steven H. Smartt, Southern Regional Education Board, 1979.
4. An excellent discussion of the uses of student ratings, peer ratings, and self-evaluation is found in Grace French-Lazovik's paper, "Evaluation of College Teaching: Guidelines for Summative and Formative Procedures," Association of American Colleges, N.D.
5. The importance of support by top-level administrators has been identified by many studies. See Mayhew, Lew B., How Colleges Change: Approaches to Academic Reform, ERIC, 1976.
6. For a recent discussion, see Clark, John L. D., "Measures of Student Learning," in Centra, John A., Determining Faculty Effectiveness, Jossey-Bass, 1979.