DOCUMENT RESUME

ED 149 683 HE 009 597

TITLE	Faculty Evaluation for Improved Learning.
INSTITUTION	Southern Regional Education Board, Atlanta, Ga.
SPONS AGENCY	Carnegie Corp. of New York, N.Y.; Fund for the Improvement of Postsecondary Education (DHEW), Washington, D.C.
PUB DATE	77
GRANT	G007701123
NOTE	66p.
AVAILABLE FROM	Southern Regional Education Board, 130 Sixth Street, NW, Atlanta, Georgia 30313
EDRS PRICE	MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS	Administrative Policy; *College Faculty; College Role; *Evaluation Criteria; *Evaluation Methods; Faculty Development; *Faculty Evaluation; Faculty Promotion; Higher Education; Rewards; *Southern States; Teacher Salaries

ABSTRACT

Summarized in the first part of the report are the deliberations of the Southern Regional Education Board's Task Force on Faculty Evaluation and Institutional Reward Structures during 1976-77. The group assisted the SREB staff to analyze procedures currently used by institutions to judge faculty effectiveness, drew conclusions, and arrived at a set of recommendations addressed to institutions willing to consider changing their evaluation systems. The second part of the report, by W. Edmund Mocnaw, reviews the results of a study on current faculty evaluation practices in the southern states. (MSE)

Faculty Evaluation For Improved Learning

Southern Regional Education Board

Faculty Evaluation For Improved Learning

Southern Regional Education Board
130 Sixth Street, N.W. Atlanta, Georgia 30313 • 1977

This report was made possible in part by funds granted by Carnegie Corporation of New York. The statements made and views expressed are solely the responsibility of the authors.

The publication of this report has been supported in part by funds granted by the Fund for the Improvement of Postsecondary Education under grant number G00 7701123, 1977-78.

Foreword

The community of scholars which makes up higher education consists primarily of students who seek knowledge and teachers who impart knowledge, and it may be assumed that good teachers are an essential ingredient for producing good students.

SREB research has indicated that faculty evaluation programs of Southern colleges and universities, by and large, are conducted to improve the effectiveness of their faculties. It has also shown, however, that many evaluation programs emphasize decision making about retention, tenure, advancement and termination of staff, while largely neglecting the widely espoused formative or developmental objectives aimed at improvement of faculty effectiveness.

This publication reviews recent SREB studies which document these and related conclusions and looks toward institutional and regional programs which can be organized for contributing to the resolution of the apparent contrast in objectives and outcomes of faculty evaluation. In a word, the Board is concerned that faculty evaluation at colleges and universities be organized and operated for maximum enhancement of the student learning environment which faculty can provide. How this can be accomplished remains a major problem in the path toward quality education.

This report presents the initial conclusions of a Task Force on Faculty Evaluation and Institutional Reward Structures convened as part of SREB's Undergraduate Education Reform project and a summary of SREB's latest research efforts in this area. It is published with joint support from the Carnegie Corporation and the Fund for the Improvement of Postsecondary Education, in connection with a two-year SREB effort for "improving programs, personnel and instruction through systematic faculty evaluation" in the region, beginning in 1977.

Winfred L. Godwin
President

SREB Task Force on Faculty Evaluation and Institutional Reward Structures

Martha Church, President
Hood College

Joseph Hammock, Vice Chancellor for Academic Development,
University System of Georgia
(now Professor of Psychology, University of Georgia)

Jan LeCroy, Vice Chancellor for Academic Affairs
Dallas County Community College System

Herman Spivey, Professor of English
University of Tennessee at Knoxville

Andrew Torrence, Provost and Executive Vice President
Tuskegee Institute

Loren Williams, Director
Educational Planning and Development
Medical College of Virginia
Virginia Commonwealth University

Jeaninne Webb, Director
Office of Instructional Resources
University of Florida

SREB Staff

W. Edmund Moomaw
Staff Associate, Research
(Now Vice President and Dean of Faculty, Birmingham-Southern College)

William R. O'Connell, Jr.
Project Director, Undergraduate Education Reform and Faculty Evaluation

E. F. Schietinger
Director of Research

Steven H. Smartt
Associate Project Director, Faculty Evaluation

Table of Contents

A Regional Review of Faculty Evaluation

Evaluating faculty performance for purposes of promotion, tenure and salary increases is of singular importance today because of leveling and declining student enrollments, lack of faculty mobility and increasing financial pressures on institutions. It is vital that evaluation procedures be suitable for helping faculty improve their performance and that the relationship between faculty evaluation and development, institutional goals and the institutional reward structure be recognized.

As part of the SREB Undergraduate Education Reform Project, which is supported by a grant from the Carnegie Corporation, a Task Force was formed to consider ways for institutions to improve their evaluation of faculty in order to assist faculty enhance the quality of their work. These seven persons, who have agreed to work together over a three-year period, represent public and private higher education and several types of institutions.

As a starting point, this group spent its first year assisting the SREB staff analyze procedures used by institutions to judge faculty effectiveness, drawing conclusions about those procedures and arriving at a set of recommendations addressed to institutions willing to consider changing their evaluation systems.

What follows is a summary of the deliberations of the SREB Task Force on Faculty Evaluation and Institutional Reward Structures during 1976-77.

Rationale for Evaluation

The desire to achieve equity in making administrative decisions about faculty has been a major stimulus for increased faculty assessment activities in U.S. colleges and universities in the 1970's. An examination of the assessment instruments and procedures, sometimes quite elaborate, which are used by some institutions indicates progress toward attaining this goal.

Equal progress toward assessing individual faculty contributions in improving the quality of student intellectual and moral growth has not been achieved, however. Yet improvement in the quality of the college experience, and hence improvement in the quality of life generally, may be the chief challenge and opportunity facing the American institution of higher learning in the remainder of this century. Institutions must find a reliable way to assess faculty contributions to student growth in knowledge, basic skills, depth of understanding, discriminating ability, motivation, expanding interests and breadth of awareness if the undergraduate years are to be truly adventures of the mind for most students. Such student growth is facilitated by the continuing concern and growing ability of the faculty to stimulate and assist in this expanded intellectual development. Current faculty assessment efforts do not seem to be designed, or used, to provide an effective means of faculty self-assessment and/or self-improvement in trying to meet this enormous challenge, yet faculty competence seems to have a significant impact on student achievement.

Efforts to assess faculty achievements in research and in contributions to institutional and community improvement seem more nearly adequate, although institutional approaches to evaluating the quality and significance of research may not be completely satisfactory. It is clear that the quality of life has been and will continue to be positively affected by the contributions to applied research which universities have made. Two of the big challenges to future university research are ways to protect and improve the environment and acceptable ways to meet the energy needs of a sophisticated technological society. Therefore, universities must continue to promote and to assess continual improvement in research and the dissemination of its results.

The expansion of community colleges has greatly enlarged the scope of public service contributions which institutions of higher education are able to make, however, assessment and development of faculty efforts may need further refinement to permit meaningful evaluation of this aspect of institutional responsibility.

In summary, the main purposes of a faculty evaluation program are to make possible equitable administrative decisions affecting faculty status, to assure growth in the ability of the faculty to contribute to the intellectual, aesthetic and moral achievement of students (and hence to improvement in the quality of life), to promote expansion of the scope and quality of basic and applied faculty research, and to keep alive a sensitivity to the needs of the local, state, and national community. A good faculty evaluation and development program, in other words, can help a college or a university to remain dynamic and useful to the society of which it is a part.

Conclusions about Evaluation

Based on a review of the information collected and analyzed by SREB, as well as their own collective experiences in the area of evaluation, the Task Force members reached a number of conclusions about the current state of faculty evaluation in the Southern region.

In developing and operating any system of evaluation, astute and sensitive leadership is an essential ingredient. The appropriate leadership can come from administrative ranks or from faculty ranks, but it is most desirable that it develop in both areas. In addition to enhancing the quality of leadership, involving general faculty early in the development of a program and often in its operation makes a system for evaluating faculty stronger and more acceptable participants.

For any program to be successful, the purposes for which it is established and operated must be clearly and publicly stated. This principle especially applies to faculty evaluation programs. Institutions enter into evaluation programs for a variety of purposes, although they usually seem to be initiated in response to some specific need that has arisen. All too often the impetus for the system is not clearly recognized or articulated, and frequently the real reasons for evaluation are hidden under other publicly announced purposes. Seldom do institutions identify broader and long-range purposes for establishing an evaluation procedure. Programs now in operation serve varied and sometimes multiple purposes which include: making decisions about salary, promotion and tenure of faculty; instructional development and improvement; and institutional development, presumably through the collective improvement of the faculty.

In relation to stated objectives for evaluation and desirable goals for such programs, several general deficiencies have been observed by the Task Force. Two areas of faculty work which are generally regarded as important to the institution are given low status and little emphasis in most faculty evaluation programs: academic advising and public service. Likewise, no evaluation program examined indicates the existence of specific approaches for judging the administrative and institutional service activities which make up a large portion of the job of many faculty members. And most striking, it also has been noted that faculty evaluation programs generally use techniques with no demonstrated validity for measuring progress and success in the area of student learning, even though the area of instruction is reported to be of highest priority for evaluation.

One of the greatest problems in establishing, operating, and especially describing for others, a program of evaluation relates to language. The

most commonly used terms raising semantic problems are: a) comprehensive, b) equitable, c) systematic evaluation—subjective and objective, d) summative and formative evaluation, e) assessment, f) measurement, and g) faculty development. These terms are frequently found to be included in functioning programs and too often seem to have special definitions in particular settings.

An evaluation system must be designed to suit the particular institution in which it will function. In other words, it should be tailor-made for the institution by those who will be involved in it. One institution cannot adopt a system from another institution, although some successful components might be appropriately adapted if they relate to the purposes and desired outcomes of an institutional program. One of the chief reasons a system cannot be transferred from one institution to another, aside from the fact that it may not be feasible to do so, is that no system remains static once designed and approved, but continues to evolve to suit changing needs and times. This is a desirable condition and should be both expected and encouraged.

One of the reasons for greater attention to the ways in which faculty are evaluated is the desirability for more open approaches that are clearly fair to all involved. Programs vary greatly in the extent to which procedures, criteria, and standards are made explicit for all of the various areas being judged. For some areas there is a great deal of information about how data are to be collected, about what is to be considered, and who will be involved in judging. For other areas in the same system, frequently little or no information is provided, indicating the need for equal explicitness across all areas to be evaluated. However, in examining the need for spelling out what is to be judged and by whom, institutions should be aware that there is a tendency to place great emphasis on examining quantity items when there should be equal emphasis on examining indicators of quality.

In reviewing currently operating evaluation programs, it is clear that the most organized place a greater emphasis on procedures than on other elements of the system. This is not surprising since procedural steps obviously are easier to develop than are clearly delineated and stated purposes, criteria and standards. It is also true that if overemphasized, procedures can become so complex that they may subordinate reasons for the system's existence and overshadow other elements of the program.

Not all aspects of an evaluation system can consist of totally objective measurement; many aspects legitimately rely on subjective views and professional judgments, ideally based on specific criteria. Systems with components that seem to be working well use a multiple set of sources for gathering data to be so judged.

The Task Force sees little evidence that institutions have been able to specify the overall value of having an evaluation program, even though there seem to be values which can be identified. In spite of the belief of many institutions that faculty evaluation programs will stimulate

improvement, there is little or no indication that the data are so used. Likewise, no information has been found to indicate that a system of evaluation, by stimulating faculty improvement, will have a positive effect on student growth—an implied assumption of many programs.

In the absence of research and study to indicate specific and verifiable outcomes, there is the possibility, however, that the institutional ambiance, or climate which may emerge from a "systematic, comprehensive, equitable" evaluation program, and even from informal understandings about its purposes, can lead individual faculty to seek opportunities for enhancing their growth, competence, and general well-being.

It is clear that in recent years institutions of higher education have enlarged their efforts toward effective faculty evaluation and development, but this area of educational endeavor remains a frontier inviting more effective exploration and experimentation, especially in the area of judging improved teaching and learning.

Recommendations

The SREB Task Force on faculty evaluation, after review of the current state of the art of institutional evaluation of faculty, offers several recommendations for institutions ready to begin or revise a system for evaluating faculty.

The Task Force recommends that:

The purposes of a faculty evaluation program be comprehensive, clearly articulated and publicly stated.

Faculty members and administrators be involved in determining comprehensive purposes and be committed to using the results of evaluation for meeting all purposes, including those that relate to faculty growth and instructional improvement.

Procedures to be employed in the program's operation be suitable to achieve the stated purposes.

Faculty members be involved in developing the program and continually in its operation and in ways that will stimulate and enhance strong and astute faculty leadership.

The evaluation of faculty be part of a larger system which also focuses on administrators, chairpersons, and other staff, and be designed to reflect local circumstances, both inside the institution and in its supporting environment.

The system be designed to include a feedback mechanism for regular review through which the system can evolve and improve with experience.

Appropriate flexibility be built into the evaluation system, with provisions for differentiated criteria to serve varying professional roles of faculty and for differentiated criteria to be used in judging varying approaches to instruction.

Activity areas such as advising, public and institutional service, personal attributes, administrative involvement and others which are frequently considered in evaluation, be explicitly identified if they are to be included as part of the system.

A system for evaluation include provisions for collecting data from many sources and recommendations from multiple participants, since decisions made even in the most carefully conceived systems of evaluation will still largely depend upon a collection of subjective judgments.

Evaluation systems place emphasis on improving ways of assessing student learning and growth as a means of determining instructional effectiveness.

Evaluation programs be designed and data collected which can be used to further faculty growth.

Section 1

Current Faculty Evaluation Practices in the South

All colleges and universities evaluate their faculties. They all hire, promote, terminate and, from time to time, provide rewards; therefore, they all necessarily make evaluative judgments. At some institutions, evaluation procedures are formally established, regularized and clearly communicated to all concerned. At other institutions, practices are so informal they are almost indiscernible, even within the institutions themselves.

There are increasing institutional concerns about ways to effectively judge faculty performance, and there is a need for more information and for assistance on this subject. For these reasons, the Southern Regional Education Board (SREB), over the past several years, has studied faculty evaluation practices in institutions in the 11 states it serves. The study has had two major components: a survey questionnaire directed to all 843 postsecondary institutions in the region in 1975, and a number of in-depth case studies of individual faculty evaluation programs during 1976-77.

From the survey 536 institutions submitted usable responses to the questionnaires, representing an overall response rate of 63.6 percent. The response rate ranged from 80.5 percent of the doctoral level institutions to 58.5 percent of the two-year institutions. The questionnaires were completed by institutional administrators, primarily academic vice presidents, although presidents and provosts also completed a substantial number.

Institutions were chosen for the in-depth case studies on the basis of survey responses indicating operation of a systematic approach to overall faculty evaluation. Case studies were developed from detailed interviews with persons at all levels of the evaluation process—presidents, deans, department chairpersons, faculty members. In addition to the visits for formal case study, a number of more informal visits to a variety of institutions were made to learn about faculty evaluation practices.

These studies have shown that in the Southern region many institutions are making serious efforts to develop sound and systematic evaluation practices to assure fair and appropriate personnel decisions and to provide assistance to faculty for development and improvement.

The study shows that Southern institutions use a great diversity of approaches to faculty evaluation. It also shows unevenness in the validity of practices used, in the effectiveness of some programs in achieving their purposes, and in the consequences various evaluation practices have on institutions using them.

The purpose of this section of the report is to provide an overview of the characteristics of faculty evaluation practices in the South by summarizing the results of these studies and to discuss observable consequences of current practices.

Characteristics of Current Practices

1. Detailed and systematic evaluation practices are most likely to exist in large, doctoral level institutions.

Detailed and systematic procedures which specify criteria and standards, use multiple sources of evidence, require examination of evidence through institution-wide procedures, and include avenues of appeal are most likely to exist at doctoral level institutions. The master's and bachelor's level and two-year colleges generally tend to have more informal faculty evaluation procedures, not systematically communicated within the institutions. One reason may be that most non-doctoral level institutions have not felt the same legal pressures—institutional or otherwise—which have caused the large institutions to move toward more formal programs. But pressures for due process and fair and systematic practices seem to be increasing at all institutional levels.

On the whole, few non-doctoral level institutions seem to have moved toward very formal systematic programs of evaluation. Many colleges tend to use fairly informal procedures which are not very well communicated, probably as a result of the traditions of collegiality associated with many of these institutions. However, one newer more formal approach to a combination of faculty evaluation and development is the "growth contract" being developed primarily at private liberal arts colleges, with a few examples at community colleges and in at least one experimental college which is part of a major public university.

Two-year colleges also have generally maintained fairly informal procedures, but a number are moving toward newer types of formal systems. Several two-year colleges and one state-wide community college system are considering a "Management by Objectives" approach to evaluation and development, while another statewide community college system is designing a competency-based approach to faculty evaluation.

2. Faculty evaluation tends most often to be used as a basis for personnel decisions and for management purposes rather than for faculty development and improvement.

Both the survey and the case studies conducted by SREB confirm today's widely held belief that faculty evaluation has two primary purposes: providing data with which to make decisions regarding tenure, promotion, and salary increases; and providing information on which faculty development and improvement can be based (called "summative" and "formative" evaluation, respectively, in much of the current literature). A majority (54 percent) of the total respondents to the survey questionnaire stated that the main purpose of faculty evaluation is "to provide a basis for general faculty development and improvement." Slightly less than a majority (46 percent) reported the primary purpose is "to provide each faculty member with diagnostic information concerning his instructional behavior and effectiveness." Only 36 percent said their evaluation procedures are used mostly "to provide information needed in making decisions on salary, promotion and tenure." Indeed, 40 percent of the respondents did not rank the

Table 1[1]

Percentage of Institutions Assigning First Rank to Respective Reasons for Evaluation by Institutional Level

Reasons for Evaluation	Percentage of Institutions Assigning First Rank[2]				
	All Institutions	Doctoral	Master's	Bachelor's	Two-Year
Information for making personnel decisions	36%	62%	50%	11%	18%
A basis for general faculty development and improvement	54	38	40	56	66
Provide each instructor with diagnostic information on teaching effectiveness	46	33	52	47	53
Assure equity of employment practices	3	6	1	5	1
Research data	1	0	0	1	2

[1]This and the other statistical tables in this publication first appeared in Boyd and Schietinger (1976).

[2]Columns may total more than 100 because respondents were allowed to assign a given rank to more than one choice.

summative purpose as being either first or second in importance. The survey results are somewhat different, however, when viewed by institutional type, as Table 1 shows.

Table 1 makes it clear that the administrators who completed the questionnaires considered the summative function of the evaluation most important in the doctoral level institutions, the formative function most important in the bachelor's and two-year institutions, and a more-or-less even split between the two in the master's level institutions. A mere 18 percent of the two-year institutions consider the summative purpose most important.

While a majority of all survey respondents stated that promoting faculty development and improvement is the primary reason for evaluation, few examples of institutions that in fact use the results of evaluation for that purpose could be found in the cases studied or in the informal institutional visits. Fewest examples were found in the bachelor's level and two-year institutions, which claim faculty development and improvement as the main purpose of th evaluation practices. Of these two types of institutions, two-year institutions are more likely to have programs that serve these purposes than are bachelor's level institutions. Campus faculty development centers are most likely to appear in the doctoral and master's level institutions, where administrators state that the primary purpose of evaluation programs is for making personnel decisions. It should be noted, however, that faculty development centers at these institutions are not part of evaluation programs, and they neither have access to data nor opportunity to make use of the data that evaluation generates. It should be noted also that, while a majority of administrators who completed the survey questionnaire stated that the primary purpose was providing a basis for faculty development and improvement, most faculty interviewed in the case studies stated they thought the primary purpose was providing data for personnel decisions. These results suggest a lack of clarity in evaluation objectives.

It seems clear that, while many may think the results of faculty evaluation programs should be used as a basis for faculty development and improvement, the results are used instead primarily for making personnel decisions. Little evidence could be found that current evaluation practices do promote development and improvement or that they include specific procedures likely to have that effect. It is possible therefore to conclude that if the primary purpose of current faculty evaluation practices is to promote faculty development and improvement, then that purpose is poorly communicated and seldom achieved. While specific procedures for promoting development and improvement rarely exist, procedures for using data for personnel decisions are more common. Persons in some institutions believe that these summative procedures also often have formative effects. It is thought by some, for example, that when faculty become aware that certain activities are being examined closely for purposes of personnel decisions, they try to improve performance in these areas. Some

administrators in doctoral level institutions state they have observed this phenomenon, particularly in publications.

Some institutions gather data on faculty performance and simply return it directly and solely to the individual faculty members. Little evidence could be found, however, that this practice resulted in development and improvement. Through the case studies it was found that when this practice was followed, rarely did any discussion concerning the evaluation results take place even among faculty, less with administrators.

Sometimes there is difficulty in distinguishing clearly between summative and formative evaluation. Persons active in faculty development use the two terms separately and usually think of them as leading to separate acitivities. "Formative evaluation" is considered to be evaluation that can be used to guide a set of activities specifically designed to assist a faculty member with development and improvement. "Summative evaluation" is thought of as a process which can provide information for making personnel decisions unrelated to helping faculty members change or improve. In the institutions studied, however, it was clear that this use of these terms is not universal. Administrators in a number of doctoral and master's level institutions interpret the personnel decision making process itself as resulting ideally in overall faculty improvement. One such administrator stated the overall objective of his university's evaluation program this way: "... to improve the professional competence and performance of the faculty collectively by improvement in the quality of individuals retained and dismissal of those judged to be below the institution's standards." For this administrator and others, summative and formative evaluation are in reality one. A properly administered summative process is thought to improve the overall quality of the total faculty by identifying, rewarding and promoting the top performers, while terminating low performers. This use of the terminology may explain why a high number of survey respondents stated that their purpose was faculty development and improvement. At any rate, the number of institutions with formal faculty development programs is much smaller than the number stating that general faculty development and improvement is the main purpose of their faculty evaluation activities.

The case studies revealed that faculty evaluation programs also are used by some institutions to serve a management function. At these institutions each department's work load is divided among the department's faculty at the beginning of each year, and then each faculty member's work distribution agreement identifies the evaluation areas and the criteria for his or her evaluation. Administrators in these institutions state that this practice allows them to ensure that all the department's responsibilities are covered, that the work load is divided equally and fairly, that faculty know clearly what their responsibilities are and precisely on what activities they will be judged. This practice also seems to provide a degree of individualization of faculty evaluation.

3. Faculty evaluation practices usually are initiated and carried out by administrators, with little faculty involvement.

Evaluation is by nature a threatening activity, so it should come as no surprise that faculty rarely take the initiative in organizing faculty evaluation programs. But what may be surprising is that, after programs are initiated, faculty rarely play a substantial role in the functioning of a program.

The survey results demonstrate conclusively that the academic dean and department chairman are the two most important officials in any faculty development program, whether for formative or summative purposes. Table 2 shows that for summative decisions, the academic dean is the principal decision maker in master's, bachelor's and two-year institutions, and department chairmen are the principal decision makers in doctoral level institutions. Table 2 shows also that faculty members individually, and faculty committees, are more likely to participate in the evaluation process in doctoral level institutions than elsewhere. However, only 13 percent of the doctoral level

Table 2

Assignment of Principal Evaluation Responsibility for Decisions on Salary, Promotion and Tenure, by Institutional Type, Percentages

Responsible Person(s)	Percentages Reporting Respective Sources as Principal			
	Doctoral	Master's	Bachelor's	Two-Year
Department Chairman	36.7%	30.6%	24 1%	27.8%
Academic Dean or Vice President	30.4	32.0	39.6	35.7
Faculty Committee	13.3	8.1	9.3	1.7
President or Provost	4.4	7.5	12.2	11.7
Colleagues	3.3	2.2	0.3	0.0
Peers (other institutions)	2.3	0.0	0.0	0.0
Students	1.3	1.9	1.3	2.5
Others*	0.4	1 1	1.2	0.8
No Response	8.0	16.7	12.0	20.1

*Including alumni, joint student-faculty groups, self.

institutions assign faculty committees a major role. It was found in the case study institutions that faculty are used in doctoral level institutions primarily to make judgments on the quality of research and publications, which can explain why faculty do not have major roles outside the doctoral level institutions. It also is of interest to note in Table 2 that students are more likely to be assigned a major role in the two-year institutions than in others. Indeed, students are assigned a greater role in these institutions than are either individual faculty or faculty committees.

When the focus shifts from summative to formative evaluation (Table 3), similar results are shown although faculty committees play an even smaller role. The major difference between Table 3 and Table 2 is that students are assigned a greater role in formative than in summative evaluation. In other words, student ratings of instruction are used more often for development and improvement than for personnel decisions. This may also suggest that student ratings of instruction are the only developmental activity taking place at some institutions.

Table 3

Assignment of Principal Evaluation Responsibility for Faculty Development, by Institutional Type, Percentages

Source of Principal Responsibility	Percentages Reporting Respective Sources as Principal			
	Doctoral	Master's	Bachelor's	Two-Year
Department Chairman	35.5%	36.3%	26.0%	30.1%
Academic Dean or Vice President	18.7	25.3	32.8	36.9
Students	13.9	9.8	12.9	12.5
Faculty Committee	5.1	3.9	1.8	0.8
Self	4.4	4.7	3.4	3.7
President or Provost	4.0	1.9	2.7	0.6
Colleagues	2.0	2.0	2.6	0.6
Others*	3.1	1.3	1.9	1.4
No Response	13.3	14.8	16.0	13.4

*Including alumni, joint student-faculty groups, peers from other institutions.

The overwhelming importance of the roles of academic deans and department chairmen raises two concerns: (1) Should the faculty itself have a greater role in the evaluation process? and (2) Are the academic deans and department chairmen prepared to carry out this important responsibility? The case studies suggest that institutions which assign a greater role to faculty in the evaluation process tend to have higher faculty morale and greater faculty acceptance of evaluation and its results than institutions which do not include as much faculty participation. A common complaint of top administrators interviewed for the case studies was that department chairmen do not play their roles in the evaluation process properly. A number of academic deans and department chairmen interviewed in the institutions studied indicated the need for training, especially in using evaluation for developmental purposes. Some deans and chairmen seem to find assisting individual faculty with development and improvement a difficult and awkward role for which they have little training, or perhaps understanding. While some institutions have developed special offices with staffs to serve this role, budget restrictions have prevented the establishment of such offices in a large number of institutions.

4. Administrators say that instruction is the most important evaluation area at their institutions, but, in fact, procedures for evaluating instruction are generally poorly developed.

Respondents to the survey were asked to indicate the relative importance of nine evaluation areas in making personnel decisions at their institutions. Table 1 shows the results of this ranking process. "Instructional activities" (defined as including classroom teaching, laboratory supervision, thesis direction, and course preparation) were ranked the most heavily weighted area for overall faculty evaluation in all types of institutions. Respondents were permitted to assign a given rank to more than one area if several areas were equally important, and some doctoral level institutions ranked instruction equal in importance to research and publication. Each category of institutions, taken as a group, ranked instruction above everything else, however. The survey results show conclusively that administrators consider instruction the most important activity for evaluation in all of the responding institutions.

But while administrators reported in the survey that instruction is the most important area for faculty evaluation, it was found in the case studies that procedures for evaluating instruction are not as well developed as are procedures in other areas, particularly research and publication. It may be that administrators feel instruction *should* be the most important area both for faculty activity and for faculty evaluation but this attitude has not yet been translated into practice.

Table 4

Rank Order of Evaluation Areas Considered in Evaluation for Advancement, by Type of Institution

Evaluation Area	Rank Order of Evaluation Areas						
	All Institutions	Doctoral	Master	Bachelor's	Two-Year	Below 5,000	Above 5,000
Instructional Activities	1	1	1	1	1	1	1
Student Advising	2	4	2	2	2	2	3
Administrative Activities	3	5	3	4	2	3	3
Personal Attributes	4	6	5	3	4	4	6
Research	5	2	4	5	7	5	2
Publications	6	3	6	6	9	7	5
Activity in Professional Societies	6	7	7	6	5	6	7
Public Service	8	8	8	8	8	9	7
Civic Activities	8	9	9	8	5	8	9

5. Data or evidence on which judgments are made, particularly in evaluating instruction, are not gathered systematically or consistently.

The survey results demonstrate that not only are administrators the main decision makers in current faculty evaluation practices, they also are the main sources of information for these evaluations. Table 5 shows sources of information for use in faculty evaluation. Tables 6 and 7 provide the same information according to the purpose of evaluation; Table 6 is concerned with summative evaluation, Table 7 with formative evaluation.

Table 5 shows that the three most frequently used sources of information are the academic vice president or dean, the department chairman and students. Department chairmen are the single most frequent source of information in doctoral and master's level institutions, and the academic vice president or dean is most often used in bachelor's level and two-year institutions. Tables 6 and 7 show the dominance of the academic vice president and the department chairman, regardless of the purpose of evaluation.

Table 5

Sources of Information for Overall Use by Control, Type, and Enrollment, Percentage Utilizing

Source of Information	Control		Type of Institution				Enrollment		
	Public	Private	Doctoral	Master's	Bachelor's	Two-Year	Below 1,000	1,000-5,000	Above 5,000
Academic Dean or V.P.	88%	81%	84%	83%	96%	93%	92%	92%	84%
Alumni	13	24	20	18	22	14	23	16	13
Colleagues	45	43	65	50	48	32	38	42	61
Department Chairman	94	84	93	92	92	87	78	95	100
Faculty Committee	37	43	69	58	41	17	38	34	53
Joint Committee (faculty/student)	11	9	18	7	10	9	8	10	15
Peers (other institutions)	9	9	35	7	5	4	5	7	23
President or Provost	43	53	39	48	57	42	51	46	43
Self-Evaluation	63	46	46	53	52	61	52	55	65
Students	88	88	84	87	87	91	88	91	82

It seems significant to note as shown in Tables 6 and 7 that students are used more often as sources of information for formative than for summative purposes. Fewer than half of the responding institutions report that they use information from students for making personnel decisions, but 88 percent of the respondents report using students as sources of information for overall use although we do not know precisely how their opinions are used or to what extent they are valued. An analysis of the case studies showed that few institutions include students as sources in a way that results in reliable, consistent, or comparable data. Many institutions employ forms for student evaluation of instruction, but not systematically. In a surprising number of cases, students seem to be used in a casual and sometimes even gossipy way as sources of information. A large number of department or division chairmen seem to rely heavily on informal student comments or complaints as valid sources of information about teaching quality.

Many chairmen and deans also tend to rely heavily on informal comments from faculty colleagues. The case study results indicate that

Table 6

Use of Sources for Decisions on Reappointment, Promotion and Tenure by Various Degree-Type Institutions

Source of Information	Percentage Use by Level			
	Doctoral	Master's	Bachelor's	Two-Year
Academic Dean or V.P.	81%	79%	90%	75%
Alumni	0	7	8	2
Colleagues	54	40	32	13
Department Chairman	86	86	88	73
Faculty Committee	61	50	35	12
Joint Committee (faculty/student)	11	2	3	4
Peers (other institutions)	32	4	1	2
President or Provost	38	46	56	36
Self-Evaluation	32	38	32	32
Students	47	51	46	41

Table 7

Use of Sources of Information for Faculty Development by Institutional Level, Percentage Utilizing

Source of Information	Percentage Use by Level			
	Doctoral	Master's	Bachelor's	Two-Year
Academic Dean or V.P.	59%	75%	88%	86%
Alumni	15	12	16	12
Colleagues	43	39	35	27
Department Chairman	82	85	84	82
Faculty Committee	41	34	24	10
Joint Committee (faculty/student)	11	7	7	7
Peers (other institutions)	14	2	5	2
President or Provost	27	36	45	32
Self-Evaluation	39	47	45	61
Students	72	75	79	84

chairmen and deans who do rely on these informal comments are for the most part frustrated over the inadequacy of these sources, although a few thought they were quite adequate and accurate. Those who expressed frustration felt that any systematic information gathering about instruction would be resisted by their faculties.

Nevertheless, some institutions do gather systematic and reliable information about instructional effectiveness. Some use student evaluation forms, student committees to evaluate instruction, formal colleague classroom visitation and reporting, or formal examination of teaching materials. In addition, there are departments which have appointed interdepartmental committees to make formal evaluations of instruction for the department chairman to use in making recommendations.

One effective and unusual use of students as a source of information was found at a doctoral level institution and in a professional school. Student committees appointed at the departmental level conduct a thorough investigation of teaching effectiveness and write a detailed report of their findings. In the doctoral level institution the committee consists of both graduate and undergraduate students. That committee conducts random interviews with students and other instructors; examines teaching materials, including examinations, and interviews the instructor. Several department chairmen called these reports the best sources of information ever received.

The general absence of reliable or systematic evidence is a crucial factor in the deliberations of the department or division chairman or the dean because it is at these levels that the evidence is first considered and the quality of performance initially designated. The chairman or dean not only initiates the evaluation process but also makes the original qualitative judgment. In the evaluation steps that follow, the evaluators see primarily what the chairman has said, rather than the raw data, if there were any. When there is no systematic information for the chairman to use, he is left to his own devices. Thus, many evaluation systems that appear to be objective actually begin with a judgment based on little evidence. As that value judgment goes up the institutional ladder, it is treated increasingly as if it were an objective assessment, and at the end of the process the determinations made by the chairman often are used as if they were entirely reliable and accurate data.

The case studies demonstrated that, with regard to evaluating instruction, department and division chairmen often tend to give the benefit of the doubt to the instructor and give him or her an adequate rating unless some overwhelming evidence to the contrary has come to their attention. In short, in many evaluation systems, instruction is presumed to be satisfactory unless shown otherwise. As several administrators pointed out in the interviews, the result is that most teaching evaluation does identify the very unsatisfactory performers, but does not indicate which instructors belong in various categories above the merely satisfactory.

The gathering of information on performance in the research and publication area seems to be much more thorough and reliable than in other evaluation areas. Of course publications can easily be counted, but there also are effective ways of judging quality in the research and publication area which make use of colleagues in the same institution and of peers at other institutions. Also, various journals are judged by their different reputations for quality, and other sources, such as citation indexes, are used widely in doctoral level institutions.

For areas other than those of instruction, research and publication, the gathering of data or evidence on which to judge performance levels is less thorough and not as sophisticated. In fact few institutions gather any data at all on performance in areas such as student advising, administrative activities, or personal attributes. There are examples of institutions which use data-gathering forms in the areas of student advising and administrative activities but, for the most part, information on faculty performance in these areas remains very informal.

6. Current faculty evaluation practices tend to concentrate on individual components or separate procedures rather than on comprehensive approaches.

It was found in the case studies that almost all institutions use one or more individual components of faculty evaluation, but that few have developed a total comprehensive program. Faculty evaluation remains informal at many institutions. Doctoral level institutions are most likely to have an elaborate procedure for making decisions, but many of them continue to be vague about precise criteria, standards, and evidence to be used. In non-doctoral level institutions, where all academic affairs are more likely to be dominated by the administration, even the procedure by which decisions are made is often unclear.

Perhaps the single faculty evaluation activity most widely used is the gathering of evidence through student evaluation of instruction forms. The use of these forms has become very popular over the past several years, with many institutions using forms of their own devising. A smaller number use professionally designed instruments available through various national organizations or institutions. It appears that in many institutions, however, little or nothing is done with the data collected from these forms. The survey results suggest that about half the institutions simply return the data to each individual instructor. It is unclear what this practice accomplishes. Other institutions use the results heavily for personnel decision making on the grounds that this is the only information available in the instructional area. A few administrators suggested that they use the student forms primarily to give the students a feeling of involvement and responsibility.

It appears that few institutions have taken a comprehensive approach to developing a faculty evaluation program which begins with clearly

stated purposes, and then provides a program likely to bring about accomplishment of those purposes. Instead, current faculty evaluation activities seem more likely to have been designed in a piecemeal fashion without a clear vision of either the purpose to be achieved or of the components necessary for a comprehensive program. Finally, in a great many institutions faculty evaluation seems still to be a vague, unclear process, even to administrators most intimately involved.

Consequences of Current Practices

The key issue in considering the consequences of faculty evaluation seems to be whether or not current practices achieve the purposes desired. A second issue concerns the effects which evaluation practices have on the institutions using them.

Do present practices achieve their objectives ?

The objectives of encouraging faculty development and improvement and of making fair and appropriate personnel decisions should satisfy almost all concerned. A glaring problem with present practices is that often there is little evidence that they effectively stimulate or assist with faculty development and improvement. This statement assumes that the term faculty development and improvement refers to a set of practices or activities which provide positive assistance for change and improvement of the faculty. While a small number of institutions in the region do have organized practices of this kind, few are directly associated with the results of evaluation. It seems fair to conclude that if the chief purpose of evaluation is the promotion of development and improvement, then some link between objectives and practices would be essential.

On the other hand, it is clear that some administrators use the term development and improvement to refer to the end results of a personnel decision making process that retains and rewards top performers and terminates the remainder. There is some evidence that this purpose is achieved in a number of institutions - primarily doctoral and master's level - that have a systematic faculty evaluation program in operation. However, as described in more detail below, some institutions achieve this result at considerable expense to institutional morale and perhaps to overall institutional effectiveness.

The findings of the survey and the case studies indicate a fairly strong belief among adfministrators that sufficient data are being generated upon which to make personnel decisions. The problem is, however, that when one looks into the data, grave weaknesses are apparent. Casual comments by students and faculty members are relied on very heavily—a practice that appears to contribute to a general

distrust of evaluation systems. Few institutions outside the doctoral and master's level gather evidence on faculty performance in systematic ways, causing one to question the validity of the ultimate decisions. Those institutions which have comprehensive evaluation programs including systematic approaches to data gathering and judgment making appear to arrive at fair and appropriate personnel decisions and achieve general acceptance of the process by the institutional community. Components of some of these programs are described more fully in Section 2.

What are the effects of current practices?

The final question on the survey questionnaire asked the respondents: *Please describe briefly what you think have been the major (good or bad) consequences of your faculty evaluation program, e.g., changes in faculty morale, in the instructional process, in the quality of teaching.* Only slightly more than half of the institutions responded to this question and sixty-three others stated that such assessment of their programs would be premature. The 388 responding institutions listed 645 consequences, of which the "good" outnumbered the "bad" about four to one. Judging by the small number of responses to this question and the vagueness of the responses themselves, it is clear that one of the greatest unknowns about faculty evaluation is the effect it has upon the institution as a whole or upon the faculty. "Good" effects which were reported by respondents, in order of frequency, were improvement of teaching and/or counseling, improvement of morale, improved faculty-administrative relations, avoidance of inequities, involvement of students, improvement of faculty motivation, improvement of relations between faculty and students, and facilitation of administration. The most frequently reported "bad" effects were negative effect on morale, ineffectiveness of the evaluation system, faculty objections, and misuse by faculty. The difficulty with these findings is that we are not sure which type evaluation programs have what effects.

An attempt was made in the case studies to distinguish the consequences of faculty evaluation activities as they differ between institutions with informal approaches and institutions with more systematic, comprehensive approaches. Institutions which have informal, seemingly vague approaches to faculty evaluation tend to have less faculty turnover than those institutions with systematic approaches and faculty personnel decisions appear to be very difficult to make and are usually put off until the last possible opportunity. Satisfactory faculty performance seems to be assumed in these institutions, with personnel action being taken primarily when major adverse circumstances arise. The criteria and standards for promotion are obscure, and many promotions seem to be automatic when a certain amount of time has passed. Salary increases rarely are given for merit, but rather on an across-the-board basis. The decision making power in these institutions tends to rest exclusively with the department chairmen and the academic deans.

Institutions with informal approaches usually encounter major problems when it becomes necessary to make a negative personnel decision, especially if the faculty member involved wishes to contest the decision. Often the institution has no criteria or standards by which to define acceptable and unacceptable performance. Evidence of a faculty member's performance is usually acquired through informal procedures or from sources which the institution often is unwilling to reveal. In addition, these procedures do not result in evidence which permits formal comparisons because the approach has been so informal and possibly different for each faculty member. The institution is likely to have difficulty demonstrating that it has used fair procedures or allowed for due process, especially if the department chairman and dean simply meet together and discuss each situation. The result often seems to be that negative personnel decisions are so unpleasant and difficult in these institutions that they either are postponed to the last possible moment or avoided altogether.

On the other hand, there are those who find the congenial and collegial atmosphere of evaluation at these institutions more effective and more likely to generate development and improvement. One academic vice president of a master's level institution which recently adopted a systematic approach to evaluation would like to return to an informal, collegial approach. He states his views as follows:

> I'd scrap the whole thing and go back to the days when chairmen and deans talked in detail with faculty members about performance, expectations, and processes for improving performance; although our [current] process would, I'm sure, pass muster in the eyes of today's academic world. Indeed, I would postulate that it would be described as exemplary. Nevertheless, I would opt for old-fashioned, less stratified, less ritualized processes. I believe results would be as good and trauma would be greatly lessened.

Visited institutions with systematic approaches to faculty evaluation reported both positive and negative consequences. The major positive consequence reported was that better personnel decisions are being made, with the result that marginal faculty members are being terminated sooner, strong faculty members are rewarded for their strengths, and all are stimulated toward improved performance. Administrators at several doctoral level institutions reported that they believe the overall quality of their faculty has improved, and that faculty tenured before adoption of the present system perform better now than before. Concrete evidence for increased productivity in the area of research and publications is available in some of these institutions. Evidence for improved instruction usually is not available, but many administrators have the general "feeling" that instruction has improved as well. Using the results of systematic evaluation to assist individual faculty improve their performance or correct certain mistakes is unusual, however. Most often, faculty simply receive their evaluation results each year and are left on their own to take what

actions they wish. Conversations about performance usually must be initiated by the faculty member.

Additional positive consequences reported by administrators in institutions with systematic evaluation programs are that the program provides them a data base not only on which personnel decisions can be made, but also on which they can be justified, if need be, to courts and/or governmental agencies. Administrators in large universities also feel their programs provide additional information about the strengths and weaknesses of individual departments and programs which they otherwise might not receive.

The major negative consequence of systematic faculty evaluation programs revealed through the case studies was that of increased faculty anxiety, lower faculty morale, and in several cases, complete faculty rejection of the evaluation program. Realizing that evaluation is by nature threatening, some institutions have taken steps to minimize the anxiety level and thereby to increase acceptance. The most successful such step appears to be to increase the role of the faculty in operating the evaluation program. In the programs with the lowest faculty anxiety level, faculty have been involved in the design of the program, in its formal adoption, at key levels in the process, and in the evaluation of the program itself. More specific examples of the roles faculty play in evaluation programs are given in Section 2.

Another practice which seems to lower the faculty anxiety level is to have the evaluation program as open as possible. One institution had a major problem, for example, when it kept the membership of one of the faculty review committees secret. Not all institutions with systematic approaches allow faculty access to all written information. Institutions which do not allow access to all written materials provide summaries of the materials, either in written or oral form. In no case was it found that negative comments are kept from faculty, but in some cases the name of the person making the comments is withheld. Another institution which allows open access to all materials has found that the number of persons promoted and tenured has increased under the current system. Several administrators pointed out that a major problem in implementing a systematic evaluation program, which requires written evaluations based on specified criteria, standards, and evidence, is getting department chairmen to do their jobs adequately.

It is clear that there are both positive and negative consequences to both informal and systematic approaches to faculty evaluation. Neither approach is likely to automatically have only positive outcomes. The end result of a particular approach appears to depend to a large extent on how the program is adopted in the beginning and the extent of faculty involvement and responsibility for program operations. Section 3 will deal with strategies for initiating, developing, and operating systematic faculty evaluation programs. But first, Section 2 will describe a framework for such programs.

Section 2

Components of a Systematic Faculty Evaluation Program

It was pointed out in Section 1 that two key characteristics of current faculty evaluation practices are (1) that there is little evidence that current practices achieve the objectives set for them, and (2) that most evaluation practices are not organized into comprehensive programs. Analysis of the SREB case studies of faculty evaluation suggests these two characteristics may be closely related. Institutional evaluation programs appearing most nearly to achieve their summative and/or formative objectives were those which had identified and fully developed various components which make up a comprehensive evaluation approach. In many institutions it seems that a large amount of data is collected regarding faculty performance, with little consideration given to why these data are being gathered and how they will be used to achieve desired objectives. In short, few institutions have approached faculty evaluation in a systematic and comprehensive way. As a result, many institutions carry out a number of isolated practices with little overall impact.

One outcome of the SREB study of current practices has been the identification of components which make up a systematic approach to faculty evaluation. As a result, we have developed a framework for a systematic faculty evaluation program which will be presented in the remainder of this section of the report. We hope this framework will serve both to clarify the concept of systematic faculty evaluation and to provide guidelines institutions can use in developing programs of their own. This framework is intended to offer an approach which includes those components common to all evaluation programs as well as one that allows institutions to design the details of each evaluation component in accordance with their own individual needs and circumstances.

Framework

Four separate components of systematic faculty evaluation programs were identified through the case studies. The four components, in what appears the logical order for development, are as follows:

1. Purpose—The objectives, or desired outcomes, of the evaluation program clearly stated can guide to further development and operation of the program.

2. Areas—The areas for evaluation are those functions or attributes that are to be examined in determining faculty effectiveness. Traditionally, the three main areas of faculty activity have been assumed to be teaching, research, and service. However, additional areas have been identified.

3. Elements—The three essential elements of any evaluation program are criteria, standards, and evidence. Criteria are specific measurable attainments subsumed under each evaluation area Standards are the levels of attainment which are expected or defined for each criterion. Evidence is data or information gathered as the basis for determining performance levels for each criterion. Evidence also includes the methodology used in gathering the data.

4. Procedures—Procedures are the steps designated by a flow chart or other device for examining and applying the evidence to determine level of achievement for each criterion. Clear procedures are needed in making personnel decisions and in providing assistance for development and improvement, or for achieving objectives of the program.

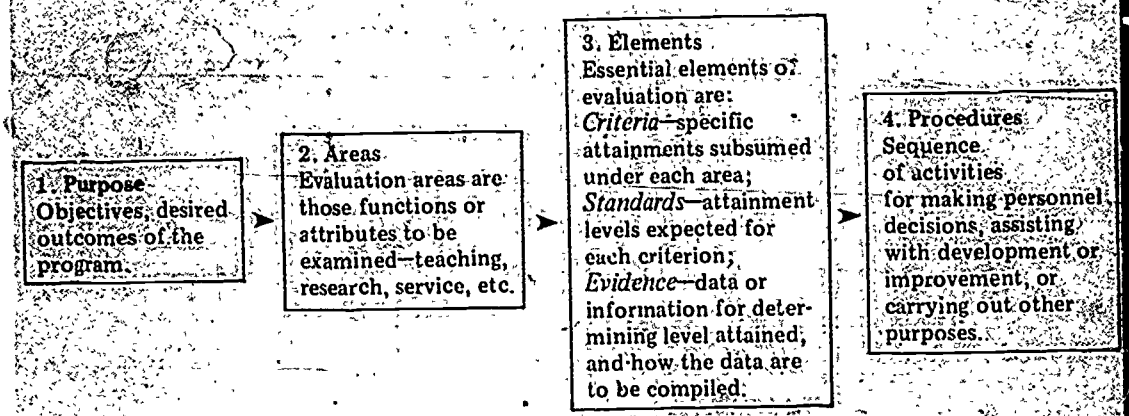
Figure 1, shows these four components in graphic form.

The figure suggests an order in which the four components of a faculty evaluation program may logically be developed by an institution. While few institutions were found to have followed this step-by-step process precisely, a larger number of institutions which clearly have all four of these components as part of their faculty evaluation programs was identified. In fact, it is undoubtedly true that all faculty evaluation programs or activities include parts of all four of these components, although in some cases it may be difficult to locate them all.

Examples drawn from the SREB case studies of current practices will be used to describe these four components of faculty evaluation.

Figure 1

Framework for Developing the Components of a Systematic Faculty Evaluation Program



1. Purpose of Evaluation

The literature of higher education generally confirms the SREB survey finding that there are two main purposes of current faculty evaluation practices. (1) providing data for making fair and appropriate personnel decisions, and (2) providing assistance or stimulation for faculty development and improvement. As pointed out in Section 1, 54 percent of the respondents to the SREB survey stated that the formative purpose was the most important in their institutions. However, the results of the case studies of current practices led to the conclusion that the summative function is more important in practice.

There seems little point in quibbling over which of these two purposes should be more important as institutions go about developing sound, systematic faculty evaluation practices since a good argument might be made that both of these purposes are indispensable. Certainly all institutions must make personnel decisions, and they ought to devise a process by which those decisions can be made in a fair, equitable, timely, and acceptable manner. By the same token, it is difficult to imagine an institution that does not wish to provide assistance and stimulation for faculty development and improvement. So it would seem appropriate for institutions to consider both of these purposes important—perhaps equally important. As pointed out earlier, some institutions do indeed find it difficult to separate the summative and formative purposes. Some institutions consider these two purposes to be one. This is especially true in institutions where development and

improvement are looked upon as applying to the faculty collectively rather than individually.

Table 1, in the preceding section, indicated importance attributed to additional purposes of faculty evaluation that were listed on the survey questionnaire. The categories *provide each instructor with diagnostic information on teaching effectiveness* and *a basis for general faculty development and improvement* are sufficiently similar that both can be considered formative purposes. It was found in the case study interviews, in fact, that these two concepts often are used interchangeably by administrators and faculty. Table 1 also shows that the purposes entitled *assure equity of employment practices* and *research data* are not considered significant in current practice. Assuring equity of employment practices probably is considered subsumed under the summative function by most administrators. It appears, therefore, that the two categories of summative and formative purposes cover rather completely the responses to the SREB survey.

An additional purpose of faculty evaluation programs identified through the case studies and noted earlier was that of facilitating the management process. A number of institutions begin each academic year by having each faculty member prepare a "distribution of effort" agreement with his or her department chairman, setting forth explicitly the faculty member's responsibilities for the year. This agreement specifies the evaluation areas and the criteria for that faculty member's evaluation. This process provides a means for assuring that all departmental responsibilities are covered, and for dividing the work load in accordance with the strengths and desires of individual faculty. It also provides a process for individualizing evaluation. Each member is evaluated on the basis of his or her own responsibilities, rather than on general expectations for all faculty. This procedure allows faculty to concentrate on different activities in different years. For example, in some years the agreement may provide that a certain professor concentrates on teaching a greater than average load, with an agreement that he or she will not be expected to prepare publications that year. In this case, that faculty member would not be evaluated on publications.

The purposes of faculty evaluation programs are sometimes stated clearly, sometimes vaguely. Several examples of evaluation program purposes taken from the institutions studied are listed below:

> *"The primary purpose of the annual performance evaluation of all faculty is to promote individual and institutional self-improvement."*
>
> *"Evaluation for promotion, tenure and termination has as its sole purpose continually upgrading the quality and performance of the faculty."*
>
> *"The purpose of the present evaluation and management system is to assist the faculty member in promoting student*

learning, and to reward each faculty member in proportion to the extent that he promotes such student learning."

"The overall objective is to provide the fullest and fairest possible evaluation to insure:

- *a. that the faculty member has optimum assistance to help him realize his maximum possible potential;*
- *b. that we are as fair as possible in extending rewards, e.g. merit increments, promotion, tenure;*
- *c. that we are able to eliminate those who have failed to remedy shortcomings of major moment in teaching, research, and service."*

"The primary goal is faculty development, the improvement of teaching through feedback. There are secondary goals involving the pay, promotion, and tenure process."

"The overall objective is to get the best possible professional judgment made and to ensure that recommendations are not capricious or arbitrary."

"Objectives of our evaluation programs are:

- *a. To keep and maintain a history of the performance of an individual throughout his employment at . . . College;*
- *b. To identify superior performance and to identify marginal or unsatisfactory performance;*
- *c. To give recognition and motivation to those individuals who are performing in an outstanding manner."*

It appears that clarifying purpose, objectives or expected outcomes, and achieving institutional agreement on them, is the first step in developing a faculty evaluation program. Taking this step provides the foundation on which the program can be developed. Without taking this step first, institutions are likely to have no clear guidance on what procedures to adopt and how those procedures should be used. The way the purpose of the evaluation program is determined may be as important as the substance of the decision itself. Without agreement on the program's purpose by all affected constituencies, the likelihood of its achievement is small. One possible explanation for the finding that current practices do not generally achieve their formative purpose may be that this purpose was stated after the program had been established, and that the purpose was not agreed to or recognized by all institutional constituencies.

2. Areas for Evaluation

Evaluation areas are those broad aspects of faculty functions or attributes which are to be examined in determining a faculty member's effectiveness. Traditionally, faculty activities have been referred to as falling into three areas—teaching, research, and service. Table 8 shows the percentage of institutions which assigned certain ranks to various evaluation areas named in the survey questionnaire. Table 4 in Section 1 gave this same information by institutional type.

Table 8

Ranking of Evaluation Areas for Faculty Evaluation

Evaluation Area	Percentage of Institutions Assigning Respective Ranking[1]				
	1st	2nd	(1st to 3rd)	(4th to 6th)	(7th to 9th)
Activity in Professional Societies	2%	10%	30%	44%	11%
Administrative Activities	8	33	69	25	2
Civic Activities	1	4	19	42	20
Instructional Activities	96	3	99	0	0
Personal Attributes	11	32	67	22	4
Publications	6	14	33	31	16
Public Service	2	6	22	33	24
Research	9	18	41	29	12
Student Advising	9	44	75	17	2

[1] Columns may total more than 100% because respondents were allowed to assign a given rank to more than one choice.

These tables show that for all institutions the four most important evaluation areas in current practice are instruction, student advising, administrative activities (including committee work), and personal attributes. Doctoral institutions rank research and publications second and third, and student advising fourth. A similar ranking was found by Seldin (1975) in a survey of liberal arts colleges. He found that the four most important evaluation areas were classroom teaching, student advising, length of service in rank, and personal attributes. Committee work ranked fifth in Seldin's survey. Astin and Lee's report on their 1963 survey (1967) shows the same results as Seldin's. There is little

mystery, therefore, about what the major faculty activity areas are. It is clear that different institutions have different ideas about which areas are most important, and what weight should be assigned the various areas. Institutions have different missions, and thus it is appropriate that each institution determine the importance of faculty activity areas in accordance with institutional purposes.

The "distribution of effort" agreements described above provide a means for weighting faculty activity areas in accordance with the desires of the institution and also in accordance with the specific responsibilities of each individual faculty member. Faculty members not having student advisees, for example, cannot be evaluated in the advisement area even though the institution may look upon that area as an important faculty activity. It is important, therefore, to look upon identification of evaluation areas as a decision to be made for the institution as a whole, separately for units within the institution, and for faculty members individually. This is particularly important for large institutions which have faculty performing many different sets of activities. Several doctoral level institutions which were studied allow each department to establish as well as to set the weightings for the areas to be evaluated. In another doctoral level institution the various colleges have responsibility for establishing the evaluation areas and weightings for their respective faculties.

In addition to those stated in Tables 4 and 8, other subjects of evaluation used by institutions in the region include: academic training, experience, length of service in rank, competing job offers, and consultation.

3. Evaluation Elements

Determining the areas of faculty activity that are to be examined in an evaluation program is not sufficient for evaluating those areas. Making a judgment about the level of performance of a faculty member in a certain activity area requires that one also have (1) *criteria* on which the area can be judged, (2) *standards* against which the extent or degree of achievement can be determined and (3) *evidence* on which to base the determination of the standard or level of achievement. Criteria, standards, and evidence are needed for each area to be evaluated. Whether or not an institution can develop criteria, standards, and evidence for an evaluation area may help determine whether that area is appropriate for evaluation.

Criteria.

Specific behaviors, features, measures or indicators to be examined for each evaluation area are the criteria for evaluation. It was found in the case studies that some institutions have difficulty generating

appropriate criteria for each evaluation area, and that some generate so many criteria that the system becomes unmanageable. One test for valid and workable criteria is whether or not it is possible to develop both standards and evidence for each.

An example of one way to begin the process of developing criteria is to ask the question, "What do we need to examine in order to determine a faculty member's performance level in instruction?" Other evaluation areas, such as student advising, committee work, research, publication, etc., can be substituted for the word instruction. Answers to the above question can be generated from various institutional constituencies and a common, acceptable core of criteria can result. Clearly there are no established complete sets of criteria available for use in various evaluation areas. Therefore it seems the best course for an institution to follow is to develop its own criteria by involving all constituencies within the institution. This process seems most likely to assure acceptance of the criteria, and later the success of the program.

Institutions studied use both quantitative and qualitative criteria. Some refer to their criteria as objective and subjective, rather than quantitative and qualitative, but these two sets of terms seem to have similar meanings. Most institutions use a combination of quantitative and qualitative criteria, although some use a quantitative approach to all criteria, standards and evidence.

Examples of quantitative criteria for instruction used by institutions studied are the following:

- enrollment trends
- number of students who become majors
- number of students who withdraw

The following are examples of qualitative criteria for the instructional area:

- extent of student learning
- student understanding of course objectives
- current and relevant syllabus and other course materials
- performance of students in higher level courses
- clarity of classroom presentations
- success of students after graduation

Examples of combined quantitative and qualitative criteria in the research and publication area are:

- number of publications
- quality of publications
- number of research grants
- citations in other publications
- speaking or consulting invitations
- author's reputation in the discipline

Some evaluation areas, such as that of personal attributes, may present more difficult problems in developing criteria. Examples of criteria that are used in this area by some institutions include the following: loyalty to the institution, honesty, punctuality, and neatness. The areas of student advising and administrative activity also

give difficulty to some institutions. Examples of criteria used in the student advising area are: student satisfaction, number of students advised, number of students who change advisors, and number of advisees who experience difficulties. In the area of administrative activities, criteria include: number of committee or other administrative assignments, effectiveness on committees, completion of administrative work, and colleague satisfaction.

Another approach to developing criteria for evaluating the area of instruction has been proposed by Meeth (1976). Basing his proposal on work in another context by Robert Thorndike, Meeth suggests three categories of criteria—ultimate, intermediate, and immediate—that can be applied to teaching. Ultimate criteria are based on student learning gains, intermediate criteria on the process or methodology of teaching, and immediate criteria on the sense of the learning experience.

In this approach, immediate criteria are judged by student satisfaction and the evidence used is primarily student ratings of instruction. Some examples of Meeth's intermediate and ultimate criteria are as follows:

Intermediate Criteria

Students were motivated to learn.

The structure of the learning experience was determined by the goals of the experience.

The content was well ordered, comprehensive, and appropriate to the abilities of the learners.

Rewards and sanctions were appropriate to the goals of the learning experience.

Goals and/or outcomes were clearly specified.

Evaluation criteria, standards, and methodologies were clear and appropriate to the goals of the experience.

Methodology was appropriate to the goals of the experience and the abilities of the learners.

Ultimate Criteria

The students learned what the instructor was trying to teach
- in cognitive, affective, and/or psychomotor development
- in rate and or absolute achievement

Students retained what was learned.

Teacher goals and or outcomes for the learning experience were met.

Student goals and or outcomes for the learning experience were met.

The learning experience related to other learning experiences students might have had (congruence, continuity, sequence): prior learning was capitalized upon; learning increased in other formal experiences the students had at the same time; learning improved in the rest of a sequential series the students had afterward.

Enrollment levels were sustained or increased in subsequent offerings of the learning experience.

Whatever the approach used to develop criteria, the best recommendation probably still is for each institution to go through the process for itself. On the other hand, there has been considerable research into the reliability of various criteria, particularly in the area of instruction. Institutions may wish to review this literature in advance to help avoid possible mistakes. Centra (1976), Hildebrand, *et. al.* (1971), and Smith (1977) are good places to begin.

Standards.

The reason for setting standards in the evaluation process is so performance can be evaluated only against stated expectations. Levels of expectation are always present in evaluation situations, however they often are so vague that it is almost impossible to know what they are. Faculty members continually use standards in assigning examination and course grades, even though they rarely are stated publicly. By the same token, standards are applied when institutions make personnel decisions even though those expectations may not be articulated.

The term "standard" may be defined as any definite rule, principle or measure to which something can be compared. In the academic world, when the word standard is used, it usually is assumed that somewhere there is a reference group or reference point to which a person's performance is compared. However, recent discussions have dealt with two kinds of standards: norm-referenced and criterion-referenced. When norm-referenced standards are used, a faculty member's performance is compared against the performance of his or her peers. When criterion-referenced standards are used, a faculty member's performance is measured against pre-set levels. In the framework that is being proposed here, it is necessary to have clear standards of some type for each criterion. The use of both norm- and criterion-referenced standards was found in the case studies.

Norm-referenced standards were found to be used in circumstances where there are scarce resources to distribute, such as promotions, tenure, and salary increases. In some institutions for example, only a certain percentage of assistant professors may be promoted to associate professors, or only a certain percentage of the non-tenured faculty may be granted tenure in a given year. In such circumstances, individual faculty members' performances are compared to each other, and only those in a certain percentile are chosen. Norm-referenced standards also are used when an institution finds itself in a retrenchment situation

which makes it necessary to reduce faculty. Norm-referenced standards might be compared to the practice of "grading on the curve."

Criterion-referenced standards are precisely set and stated in advance of their application. Faculty performance is measured against these pre-set standards without regard to how many faculty fall within, or above, or below the standards. Most institutions studied claim to have criterion-referenced rather than norm-referenced standards, but rarely are these standards precisely stated. Terms, such as "outstanding," "satisfactory," "unsatisfactory," often are used with little explanation.

Criterion-referenced standards can be stated either qualitatively or quantatively. Some institutions use a totally quantitative approach, not only to setting standards, but also to defining criteria and evidence. an approach recommended by Miller (1972). One institution has developed a numerical rating system of 1 to 5 for all criteria. This institution also gives a qualitative adjective and a short descriptive explanation for each numerical rating, as follows:

Rating	Quality	Explanation
1	Poor	The faculty member's performance is generally unsatisfactory or inadequate in this area.
2	Fair	The faculty member's performance is not entirely satisfactory and needs to be improved. Improvement may come with experience, but increased effort may be needed.
3	Good	This rating means the faculty member's performance in this area is generally adequate and definitely acceptable. It should be given when a faculty member performs at a level which corresponds with expectations for most faculty members.
4	Very Good	This rating should be given if the faculty member is above average with respect to the trait or standard. It should be given when a faculty member demonstrates above average talent and effort.
5	Outstanding	This rating should be given only in *exceptional* cases. It should be given only when a faculty member demonstrates *rare talent* and *performance* with respect to the area or category. A rating of superior should be supported with a statement documenting the basis for the rating, giving concrete examples or evidence, especially if more than 2 or 3 categories or items are assigned this rating.

Some feel that one problem with a quantitative approach to standards is that much academic work cannot be quantified. Another problem identified is that average or composite scores are computed in most such systems and therefore one has to be able to accept the assumption that all rated criteria are equally important. Moreover, the same assumption is said to be applicable to the questions on various forms, such as student and colleague evaluation forms, which are used. If a composite or average score is to be computed and used, one must assume, the argument goes, that all questions have the same weight. These assumptions sometimes are difficult to make.

Another institution states that it makes both a "quantitative assessment and qualitative judgment of (each faculty member's) activities. . . ." The quantitative assessment is made by each faculty member on an annual Academic Personnel Report. The qualitative assessment is made by following a procedural flow chart which this institution uses. The qualitative assessment terms used at this institution are "marginal," "professional," and "exceptional achievement." Some departments use a numerical ranking system of 1 to 5 in making the qualitative judgments. The dean at this institution has made the following statement regarding his understanding of the rankings of 1 to 5:

> The mid-rating, or 3, on this continuum seems the critical one since it connotes a satisfactory level of performance. So far as the Dean's office is concerned, this rating indicates that during the past year one has done reasonably well those teaching, research, and service activities that he or she is employed to do. This means that an occasional publication, a satisfactory evaluation for teaching, and a not unusually heavy service contribution, might very well characterize the 3 rating. Indeed, if these are the expectations of the position, it is difficult to conceive of a slight departure from that expectation warranting either a rating of excellence or extreme marginality.

It is of interest to note that at this institution, nearly 90 percent of the total faculty scored 3 or above.

Another institution uses a numerical system for applying standards for merit raises and also provides a descriptive explanation of each rating. This example represents a combination of norm- and criterion-referenced standards. Categories for merit raises are as follows:

> *Category I*—Faculty member doing a truly outstanding job in all five (5) areas. This should be limited to a maximum of 20% of your faculty unless you and the dean agree that there are some unusual circumstances in your department.
>
> *Category II*—Faculty member doing a good, above average, meritorious job when judged against the standards of what is

considered an ideal college professor. This category signifies that the administration of the university is satisfied with the manner in which the faculty member is carrying out his/her job. (No percentage limits on this category.)

Category III—Faculty member performing a less than adequate job and needs to improve his/her performance. This raise is used to signify to the faculty member that he/she needs to improve his/her performance. (No percentage limits on this category.)

Category IV—Faculty member not performing the job expected of the faculty. This category signifies a "zero" raise. (No percentage limits on this category.)

Some institutions also state standards for each academic rank. These standards may be stated quantitatively or qualitatively as well. Below are two examples of stating standards qualitatively for each rank:

Example 1

Professor. To be eligible for the rank of professor, a faculty member must have a record of outstanding performance normally involving both teaching and research or creativity or performance in the arts, or recognized professional contributions. As a general guideline, the faculty member is expected to hold the earned doctor's degree and to have at least nine years of effective and relevant experience.

Associate Professor. To be eligible for the rank of associate professor, a faculty member must have a good record of effective performance over a probationary period of time usually involving both teaching and research, or creativity or performance in the arts, or recognized professional contributions. He must possess strong potential for further development as a teacher and as a scholar. It is normally expected that he will hold the earned doctor's degree.

Assistant Professor. To be eligible for the rank of assistant professor, a faculty member must possess strong potential for development as a teacher and as a scholar. He will normally be expected to hold the earned doctor's degree or its equivalent.

Instructor. To be eligible for the rank of instructor a faculty member must normally hold the master's degree or bachelor's degree plus substantial additional graduate study—such as evidence of having fulfilled the requirements for admission to candidacy for the doctor's degree.

Example 2

Professor. A promotion to the rank of full professor is an indication that in the opinion of colleagues, this individual is outstanding in teaching and in research or other creative productivity and has earned national and perhaps international recognition. It should be further stressed that this rank is a recognition of attainment rather than of length of service.

Associate Professor. The promotion to associate professor should be made only after an indication of continuous improvement and contribution of the individual both in teaching and research or other creative productivity. Furthermore, the individual should have earned some regional recognition for excellence in his field.

Assistant Professor. Appointment or promotion to the rank of assistant professor should be made when it has been determined that the individual has a current capability for good teaching, research and university service and a potential for significant growth in these areas, and the terminal degree appropriate to his field.

Standards may be the most difficult element to develop and put into effect at the institutional level since many institutions seem accustomed to stating standards broadly, as in the two examples above. It is important in this framework, however, that standards be stated for each criterion in each area being evaluated. If standards for each criterion cannot be stated or agreed upon, it may perhaps indicate an inappropriate criterion.

Evidence.

Evidence is the information or data compiled for determining the level of attainment for each criterion. This element also includes the methodology designed for collecting the evidence. Most institutions have begun gathering various types of data on faculty performance, although usually not in a consistent and reliable way, and often not applied to specific criteria. Many institutions seem to treat the collection of evidence as if the accumulation of data alone constituted a comprehensive faculty evaluation and development program. However, in the framework presented here the gathering of data must be seen as simply one evaluation element which accomplishes little if not seen in its appropriate context.

Once criteria and standards have been determined for each evaluation area, the next step is to establish the evidence that can be compiled to judge the achievement of those criteria. As bei , an effective test of whether specific criteria and standards are valid is whether or not it is possible to compile evidence or d ta for each. It is important that

criteria and standards be determined first, and then evidence, because evidence alone appears to have little usefulness.

The technique to be used for gathering evidence will depend not only on the criteria and standards but also on the purpose or objectives of the evaluation program. A variety of ways of compiling evidence was found in the institutions studied. Tables 5, 6, and 7 in Section 1 show the evaluation techniques used by different types of institutions. Table 9 gives a summary of these sources and procedures according to the purpose of the evaluation program.

Table 9

Sources and Selected Uses of Information for Faculty Evaluation, All Reporting Institutions

Source of Information	Percentage of Respondents Reporting Utilization			
	Overall Use	For Decisions on Salary Increases	For Decisions on Promotion and Tenure	For Faculty Improvement
Academic Dean or V P	90%	70%	81%	81%
Alumni	18	1	4	14
Colleagues	15	19	29	35
Department Chairman	90	70	82	83
Faculty Committee	39	14	33	23
Joint Committee (faculty/student)	10	2	1	7
Peers (other institutions)	9	1	6	4
President or Provost	17	10	11	36
Self-Evaluation	56	25	33	51
Students	88	30	45	79

While these data show the importance of various evidence gathering techniques, they do not show the extent or use of the techniques for different evaluation areas and criteria. After assessing this aspect of evaluation in the case study institutions, we have concluded that evidence, particularly as gathered in the area of instruction, is the element in current practices which perhaps requires greatest attention. Criteria and evidence for some of the other areas seem more easily

developed and more readily accepted. Evidence related to criteria such as number of publications, number of citations, enrollment trends, number of advisees, and number of grants received, as the means for determining performance level is largely self-evident. Evidence of a more qualitative nature is more difficult to specify however.

It seems most appropriate for each institution to decide for itself how best to gather evidence in connection with the criteria and standards that it has established.

A variety of approaches to compiling evidence was found in the institutions studied.

Self-evaluation.

Table 9 shows that 56 percent of the institutions responding to the survey report use of self evaluation as an information gathering technique. Information from self-evaluations is used by 25 percent of the institutions for decisions on salary increases, by 33 percent for decisions on promotion and tenure, and by 51 percent for faculty development and improvement. The data show that while only slightly over half of the responding institutions use self-evaluation as a source of information at all, this technique is more likely to be used for faculty development and improvement purposes than for purposes of personnel decisions.

It was found in the case studies that most self-evaluations usually are simply an annual listing or report of activities by each individual faculty member. In this report the faculty member may list publications, research projects, committee work, courses taught, number of advisees, and other relevant information, usually in a quantitative format. This technique appears to be used primarily, therefore, as a means of providing quantitative data on faculty activity rather than as a technique for judging quality.

On the other hand, in a few institutions studied, faculty members are asked to write a self-evaluation report reflecting on their year's work, sometimes using data from various sources employed by the institution. The value of such reflective self-evaluations has not been established. Centra (1977) found little agreement between faculty self-evaluations of teaching effectiveness and ratings by students, colleagues, and administrators. He found discrepancies most notably in relation to student-teacher interaction, where about one-third of the instructors generally rated themselves considerably higher than did students.

While self-evaluations may not be very reliable as sources of data for making personnel decisions, they may be more useful for development and improvement purposes. If faculty members are given non-threatening opportunities for self-reflection, the process might be more productive. Examples of such self-reflection are found in institutions having faculty development offices at which faculty members may seek assistance without the knowledge of administrative superiors. In sum,

self-evaluations may be a legitimate technique for compiling some quantitative data on faculty activities, but the subjectivity of evaluative judgments must be recognized.

Student ratings.

The survey results indicate widespread use of students as sources of information on faculty performance. Of the institutions responding to the survey questionnaire 88 percent report using information from students and 30 percent of those institutions use that student information for making decisions on salary increases, 45 percent for decisions on promotion and tenure, and 79 percent for faculty improvement and development. The non-doctoral level institutions use student information more than doctoral level institutions, and two-year institutions are more likely than any other institutional type to use students as sources of information.

Unfortunately, while we know a number of institutions collect information from students, we do not know precisely how the information is used. The case study results indicate that a smaller percentage of institutions than that reported above use systematic student evaluation to provide comparative norms. A surprisingly large percentage of institutions evidently use students in such an informal way that the reliability of information gained is questionable.

Among the institutions studied in depth, only one approaches student ratings systematically and on an institution-wide basis. At this institution, comparative norms and data are compiled, made available and used for summative decision making. In the major doctoral level institutions studied, student ratings are left to each department or to each college. The most common practice in such institutions seems to be for each department to design and administer its own form and to use the results at the departmental level, primarily for summative purposes. Although the survey found students most likely to be used as sources of information in two-year institutions, few of these institutions were found to use student ratings in a systematic, institution-wide manner.

While the use of student ratings has become widespread, such ratings also have become controversial. One problem appears to be that many institutions treat student rating results as if they were a comprehensive evaluation of instruction rather than one means of gathering evidence with regard to specific criteria. Student ratings certainly do not provide all possible evidence for all possible criteria in the instructional area. In designing and using student rating forms, institutions may wish to keep in mind that the only evidence one can obtain from students is that which they have. Students should not be asked to provide evidence regarding a criterion about which they are unlikely to have information.

The use of student evaluation has created the need for readily available and reliable forms. A number of organizations, including Educational Testing Service, Kansas State University, and Purdue

University, market forms on a nationwide basis. Use of these forms allows institutions the advantage of not having to design their own forms—and they provide national norms and data for comparative purposes. On the other hand, many institutions feel there may be advantages to designing one's own instrument or at least adapting others to suit one's own purposes.

There has been a great deal of research into the validity of the data gained from student ratings, of which institutions should be aware. Centra (1977) provides an excellent summary of much of this research. He reported for example, that global questions which ask for an assessment of the overall effectiveness of the teacher and the course correlate more highly with course examination results than do questions assessing more specific teaching characteristics. He suggests that these global-type questions may be valid estimates of students' academic achievement because such questions are not tied to a specific instructional style.

With regard to using student ratings for summative purposes, Centra offers this concluding comment:

> If student ratings are used in making personnel decisions, there should be a sufficient and representative number of students responding for each class and the forms should be administered and collected in a prescribed and systematic manner. In addition, global or overall ratings rather than ratings of specific practices or behaviors ought to be used, and ratings for several courses over a period of time would probably provide the best basis for making judgments. (pp. 98-99)

Another approach to using student ratings is to survey students who have graduated. Many faculty seem to feel that judgments of alumni are more reliable than those of enrolled students. The research summarized by Centra shows, however, that ratings by alumni who have been out of college for 5 to 10 years compare favorably to current students' ratings of the same instructors and courses. Some institutions find that alumni may be more appropriately used to provide information on the usefulness of certain programs or courses as preparation for various professions.

Colleague ratings

The survey results show that colleagues, either as individuals or as members of faculty committees, are used in various ways as sources of evidence. Colleagues as individuals are reported used by 45 percent of the responding institutions, and faculty committees by 39 percent. Respondents in 19 percent of the institutions say they use individual colleagues as sources of evidence for decisions on salary increases, 29 percent use them as sources for decisions on promotion and tenure, and 35 percent for faculty development and improvement. Information

from faculty committees is used by only 14 percent of the institutions for decisions on salary increases, by 33 percent for decisions on promotion and tenure, and by 23 percent for faculty development and improvement.

Colleagues are most likely to be sources of information in doctoral level institutions and least likely in two-year institutions. For summative purposes, 54 percent of the doctoral level institutions report using individual colleagues, and 37 percent report using faculty committees as sources of evidence. Only 32 percent of the two-year institutions use information from individuals, and only 17 percent use faculty committees. Colleagues probably are used more by doctoral level institutions to judge the quality of research and publications. The case study results suggest that if one looks at instruction alone, the use of colleagues as sources of evidence would be unusual in all institutions, and perhaps most unusual at the doctoral level.

Perhaps the least used and least effective colleague ratings are those derived from classroom visits to observe teaching. Strong faculty resistance to visitation was found in the institutions included in the case studies. Moreover, research at several institutions shows that colleague ratings based primarily on observations of classroom performance are not sufficiently reliable for making personnel decisions. Centra, for example, found that such ratings are far less reliable than student ratings, probably because one's colleagues tend to have a positive bias.

Colleague ratings of teaching materials, such as syllabi, reading lists and examinations, may be more valuable than those based on classroom visitations. A number of institutions used colleagues in this way with positive results. In one institution it was found that simply requiring colleague review of teaching materials caused them to begin to improve dramatically.

When one moves from summative to formative evaluation, the use of colleagues as sources of information may be more useful. If an institution uses colleagues in an informal situation to provide positive feedback on instruction, the element of threat may largely disappear. If resistance is reduced, colleagues may be able to provide supportive suggestions and criticism of instructional practices.

Student Learning Outcomes.

A number of writers Rose (1976) and Meeth (1976), for example—contend that criterion-referenced measurement of learning is the ultimate standard of good teaching and that data on the extent of learning is the only legitimate evidence for determining instructional effectiveness. However, using evidence of student learning is uncommon in the Southern region. Because this approach requires one to look at student behavior rather than the behavior of the instructor, a great deal of resistance is generated. Faculty opposed to this approach argue that if their effectiveness is to be judged, it is their own behavior that should be observed. The argument for using student learning outcomes

as evidence is that student success should be the main criterion for judging teaching effectiveness. One cannot be judged an effective teacher, it is argued, unless significant student learning can be identified.

This approach to gathering evidence on instructional effectiveness is most likely to be found in colleges or universities which have Management by Objectives (MBO), behavioral or instructional objectives, competency-based curricula, or a systems approach to instruction. Even few of those institutions use this technique, but they are the types to which it most likely would apply. In more traditional institutions this technique can be found in the performing arts and in vocational and technical programs. Two major doctoral level institutions visited use this approach exclusively in the performing arts, with evident success.

Clearly, measuring student learning outcomes emphasizes purpose and impact. One community college in the region is moving toward an institution-wide system of "performance contracting" in which faculty members "contract" with the institution to produce a certain student success rate. Often a third party determines whether students have met learning objectives. Pre- and post-tests of student achievement are used extensively. It is interesting to note that this college is deemphasizing the use of student-evaluation-of-instruction forms as legitimate measures of teaching effectiveness. Officials at this institution are asking students to judge only those things they can judge, such as the teacher's behavior, rather than their own learning growth.

It appears difficult to dispute the point that student learning is the purpose of the instructional process. The problem with using student learning outcomes as evidence of teaching effectiveness is that such outcomes are not readily available when traditional teaching approaches are used. With newer approaches, such as individualized instruction, competency-based instruction, and contract learning, such outcomes should be more visible.

Multiple Approaches.

In summary, it seems clear that there are advantages and disadvantages to all approaches for gathering evidence in the instructional area. Because none of the approaches to gathering evidence appears to be entirely reliable, or most likely to achieve acceptance among faculty, it probably is best to make use of a variety of approaches. Attention will be given in the next section to the issue of strategy to be used in developing an evaluation program, but it may be well to note here that one of the weaknesses of current evaluation practices is failure of faculty to accept the sources of evidence or procedures for gathering it. No matter how valid the technique may be, if it is instituted in a way that causes faculty resentment and resistance, the evidence gathered will be suspect. Student ratings, for example, are easily influenced by what is said during the process of their

administration. If a resistant, unhappy faculty is doing the administering, student responses stand to lose any possible validity.

Putting All The Elements Together

The three elements which are common to evaluation programs—criteria, standards, and evidence—have been described. Figure 2 provides a hypothetical example of how an evaluation elements chart might look as an institution begins the process of developing the

Figure 2

Example of an Evaluation Elements Chart

Evaluation Areas	Criteria	Standards	Evidence
I. Instruction	1. Student learning gains	75% achieve 90% of learning objectives	Pre- and post-tests graded by department
	2. Course materials	Clear, relevant, current; behavioral objectives	Submitted to and examined by departmental committee; student ratings
	3. Student satisfaction	80% rate course in top level	Student ratings
II. Research	1. Publications	Specific number in refereed journals Excellence in quality	Faculty self-report Testimony of six outside readers; citations in other works
	2. Grants	$20,000 minimum in grants	Faculty self-report; records

evaluation elements. This chart shows just two areas—instruction and research—and could be expanded to include other areas. The evaluation elements chart may be developed on an institution-wide basis at smaller institutions, or on a divisional or departmental basis in larger institutions. As discussed previously, it also is possible to have individualized evaluation plans for specific faculty members.

It has been suggested that the most logical steps in developing these

three elements is to begin with criteria for each evaluation area, then to develop standards for each criterion, and finally to assemble evidence for applying each standard to each criterion. One of the tests for validity in this process is to determine whether or not it is possible to complete each succeeding step. If it is not possible to agree upon standards for a given criterion, it may be that the criterion should not be used. The same point applies to the form of evidence and techniques for gathering it.

4. Procedures for Evaluating

After the purpose of the evaluation program has been established, the areas of faculty activity to be evaluated agreed upon, and the elements of evaluation—criteria, standards, and evidence—developed, the next step is to establish a set of procedures to achieve the purpose. These procedures are a key component in the evaluation program because if nothing or little is done with all the material as it is generated, then little can be expected to be accomplished.

A number of exemplary patterns of procedures were found in the case studies. It was pointed out in Section 1 that one of the characteristics of current faculty evaluation practice is that procedures, summative and formative, are dominated by administrators. Table 10 summarizes the survey results on this subject.

Table 10

Assignment of Principal Evaluation Responsibility for Administrative Decisions and for Faculty Development, Number of Institutions

Source of Principal Responsibility	For Decisions on Salary, Promotion and Tenure	For Faculty Development
Academic Dean or Vice President	190	167
Department Chairman	155	167
President or Provost	54	10
Faculty Committee	36	13
Students	10	67
Colleagues	6	9
Self	2	21
Peers (other institutions)	2	2
Joint Student/ Faculty Group	1	1
Alumni	0	4
Other	3	3
No Response	77	72
Total	536	536

While there are a number of exceptions, it seems that many institutions use unsystematic and informal procedures in making evaluative judgments. Few institutions involve the faculty at significant points in the evaluation process. The absence of significant faculty involvement may be one explanation for the lack of evidence that current practices achieve the formative objectives set for them. The case study results suggest that in many non-doctoral level institutions where due process and fair procedure issues are not vocalized, procedures for making personnel decisions are vague and will vary for different faculty. For faculty development and improvement, procedures are even less well-defined and articulated throughout all types of institutions. The dominant attitude seems to be that the formative area is an individual responsibility, and that an institutional set of procedures or services either is unnecessary or that no one knows what procedures would be appropriate. The remainder of this section provides a description of current approaches to developing procedures in the summative and formative areas.

Summative Procedures.

Procedures for making summative judgments clearly are a necessity at all institutions. Ideally, summative procedures should fulfill the expectations and needs of both the institution and the individual faculty member. The institution needs procedures (1) that will separate the superior, the satisfactory, and the unsatisfactory performers; (2) that will build an evaluation record about each faculty member to satisfy judicial and governmental requirements, (3) that will provide due process and fair procedures; (4) that will be acceptable to the total institutional community, including the faculty; and (5) that will allocate fairly and appropriately the scarce rewards the institution has available. The individual faculty member needs summative procedures that will, in addition to the above, (1) ensure that he or she is judged on the basis of his or her particular responsibilities; (2) provide clear, understandable, and communicated criteria, standards, and evidence; (3) provide for due process and an appeal procedure; (4) involve faculty in the design and operation of the program, (5) ensure against arbitrary and capricious decisions; and (6) provide frequent feedback on performance and expectations.

The case studies show that there are four kinds of summative decisions that can be made about faculty. They are decisions involving (1) promotion, (2) tenure, (3) termination, and (4) merit salary increases. Some institutions have a multiple year contract system rather than tenure, but the procedures remain about the same. Of these four possible decisions, only that involving merit salary increases normally occurs annually. Institutions having systematic faculty evaluation programs often have a separate procedure for making promotion and tenure decisions, and another for merit salary increases. The latter decisions can provide faculty with continuous feedback and progress

reports, building a record that can be used in promotion and tenure procedures later on. Termination can be the end result of the denial of promotion or tenure, or it may be handled through a separate procedure. Typically, procedures for making promotion and tenure decisions are more elaborate than those for annual salary decisions.

The first step in any procedure is compiling the appropriate information into a dossier for each faculty member. This process usually takes place within each department and is the responsibility of either the department chairman or a departmental committee. The dossier includes all evidence called for in the evaluation elements. In addition, the candidate probably will have the privilege of adding additional information.

Once the dossier is complete, the evaluation process begins. Figures 3 and 4 give two examples of procedural flow charts currently being used at major doctoral level institutions in the SREB region. Both charts show procedures for making decisions on promotion and tenure.

Two principles are evident in the systems illustrated by these charts. First, it is clear that both systems provide for multiple reviews and independent judgments regarding the same evidence. Second, both provide significant faculty involvement in the judgment process. These principles seem to be effective in ensuring a high level of faculty satisfaction with the evaluation program and with the program results.

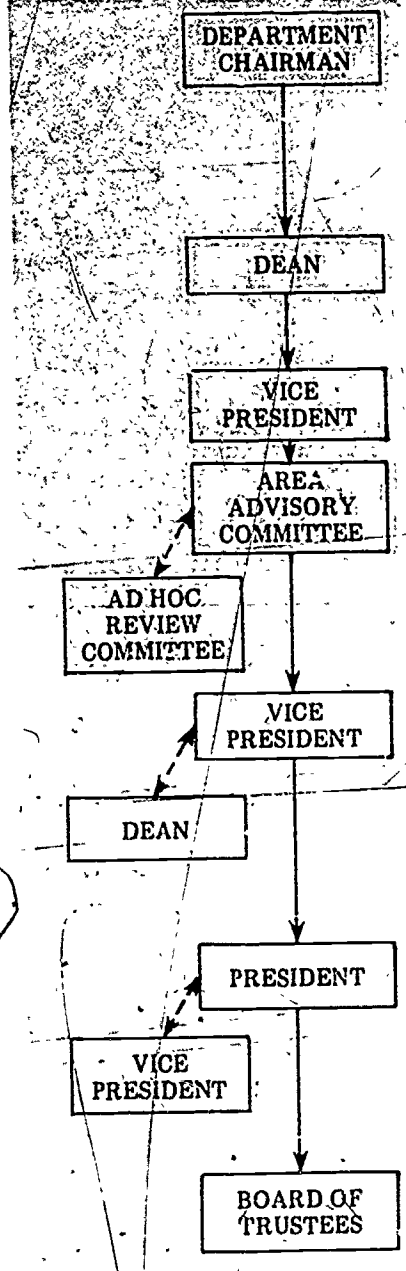
In Figure 3, faculty are involved primarily as members of "area advisory committees." Members of these committees, which exist in each of the major areas of the institution, are appointed by the president from nominations made by the faculty senate. The committees review the same material reviewed by the administrators and make their own independent judgments. Faculty members also are sometimes part of this system at the departmental level, and occasionally at the deanship level, when either the department chairman or the dean wishes outside consultation.

Figure 4 shows that faculty are involved at the departmental level, but primarily as members of the "University Committee on Promotion and Tenure." This is a committee consisting of 21 tenured professors, 15 elected by the faculty and 9 appointed by the president. It should be of special interest to note that this committee sits above all the administrative officials except the president.

In both systems, one copy of the dossier goes through the entire procedure. At each level, judgments are made in written form and added to the dossier. Thus the dossier continues to build as it goes along and, at each step in the procedure, the persons making recommendations see the recommendations that have been made at lower levels. Persons at these institutions feel that these procedures mandate greater objectivity and fairness than would a more informal approach. The institution using the procedures in Figure 4 follows a system whereby each department and school has its own evaluation areas, criteria, standards, and evidence. Consequently, each of the components must be a part of the dossiers and must be carefully examined by the committees and administrative officers.

Figure 3

Procedural Flow Chart for Academic Appointment, Promotion, Granting of Tenure and Termination for the Division of Colleges and the Medical Center



1. Initiates the proposal
2. Provides appropriate vita including a list and sample of publications
3. Forwards recommendation to the appropriate dean with supporting data from tenured members and evidence of external consultation

1. Reviews the proposal and supporting data
2. Adds his endorsement or commentary and forwards the proposal to the appropriate vice president

1. Reviews for completeness and forwards to the appropriate area committee

1. Recommends approval or disapproval
2. May request Ad Hoc committee (appointed by vice president) for further evaluation
3. Forwards recommendation to appropriate vice president

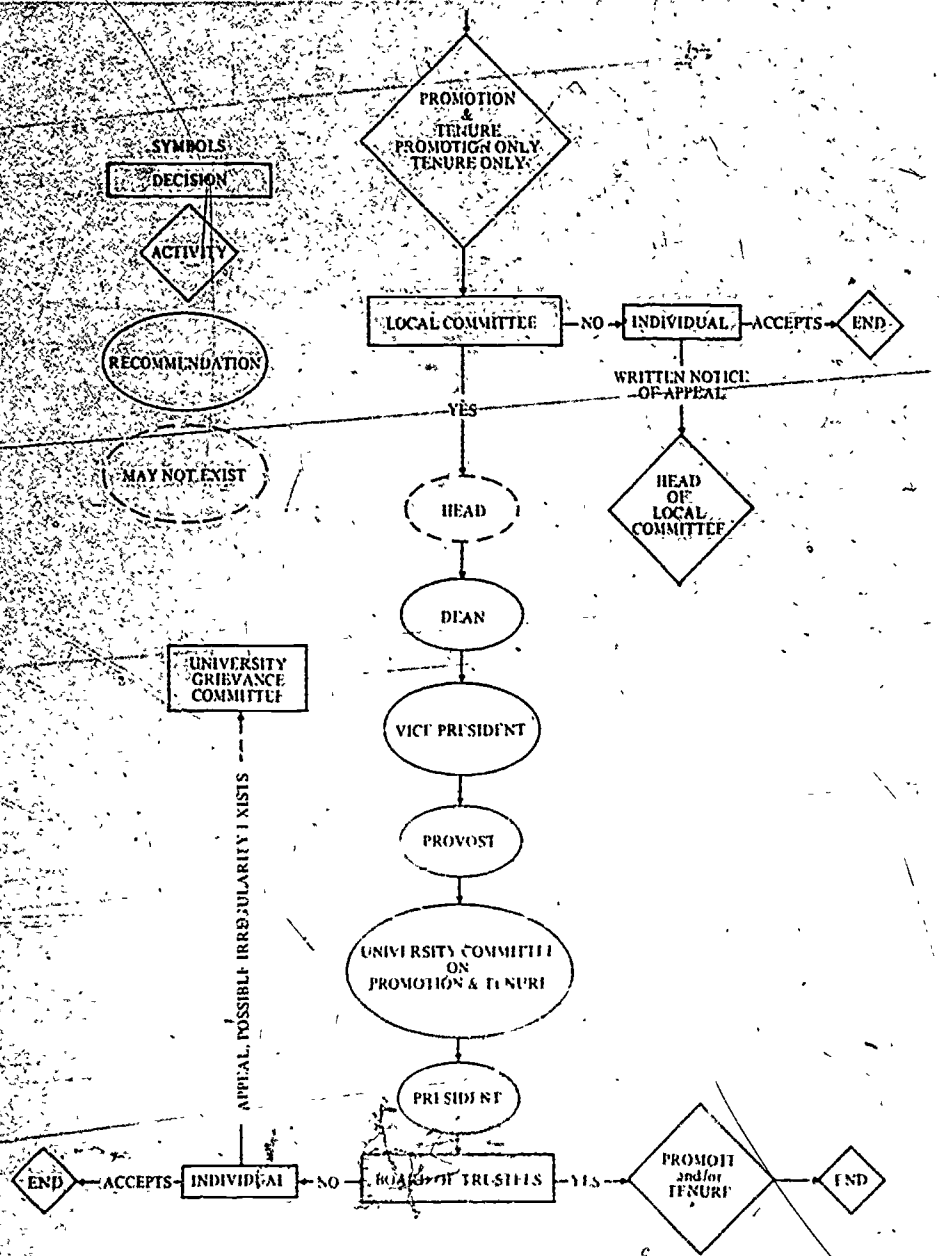
1. Reviews the proposal and all recommendations
2. Approves or disapproves the proposal
3. Advises dean of the action so the dean can respond prior to final action
4. Forwards recommendation to president
5. Maintains documentation file of all recommendations and actions

1. Reviews the proposal and submits recommendation to the board of trustees for final action
2. Consults with appropriate vice president in cases where the vice president recommends contrary to the advice of the area committee

1. Takes final action (approves or disapproves the recommendation)

Figure 4

Promotion and Tenure
All Eligible Faculty



Smaller institutions would not have procedural flow charts as elaborate as the two shown. The major difference in the smaller institutions is that there usually are not as many administrators. Even though this is the case, these insitutions still could follow the two principles of multiple reviews of the same evidence and key involvement of faculty. Selection of faculty review committees could be made in small institutions in a way similar to that in large insitutions.

For annual performance review for merit salary increases, institutions typically use a less elaborate procedure than for promotion and tenure. The institution using the procedure in Figure 3, for example, follows a process that stops at the dean's level in making decisions for annual performance review. Separate copies of each department member's file are provided the department chairman and dean, who review the files independently and make separate judgments. Upon completion of these independent judgments, the department chairman and the dean meet to compare judgments and to resolve differences. If a successful resolution cannot be made, the dean's judgment is final. The vice president is part of the appeals procedure. Also, part of the annual performance review is a mandatory conference between the department chairman and the faculty member, in which the faculty member is apprised of his or her standing and strong and weak points. Faculty members without tenure are apprised of the progress they are making toward tenure.

All the systems studied which have elaborate procedures include appeal processes. These processes usually are limited to questions of proper procedure, rather than to substantive issues. In public institutions, such appeal opportunities usually are required by law. Private institutions often feel pressure to follow due process procedures as well.

A final issue regarding summative procedures is that of the openness of files and the availability of information and evaluation results to the individual faculty member. Table 11 gives responses to a survey question regarding the availability of the results of evaluation to faculty members.

The meaning of the findings in Table 11 remains unclear, even though they may appear straightforward. The reason for this lack of clarity is that respondents could either have understood the question to ask whether or not the final decision was communicated to the faculty member, or whether all information compiled in the decision making process was made available. If the former is the meaning understood, the high percentage of respondents reporting affirmatively comes as no surprise. If the latter was understood as the meaning of the question, however, the high percentage of affirmative responses is surprising indeed in light of further study.

The case studies show that institutions requiring all recommendations and judgments be put in writing normally do not make the file contents completely available to the faculty member concerned. Usually, instead, the faculty member is provided a summary of what

Table 11

Availability of Results of Evaluation to Faculty Members

Type of Institution	Number of Responding Institutions	Percentage Reporting Results Available to Faculty
Doctoral	70	80%
Master's	107	95
Bachelor's	150	95
Two-Year	209	98
Total	536	94

was said, without being given the names of the persons making each response. In some cases, faculty members are given access to the contents of the files with the names of the evaluators deleted. In no case where systematic evaluation practices are used are faculty entirely denied access to the substance of their files. Persons interviewed in institutions seemed to feel that opening files completely, including association of the named evaluators with what they said, would intimidate evaluators and minimize frankness in the evaluations. On the other hand, some persons feel that persons making objective evaluations should be willing to make their judgments known to those whom they are evaluating. This issue, clearly, is a difficult one for institutions and may have to be dealt with on an individual basis.

There is no doubt that the elaborate summative procedures described here constitute an enormous amount of work, require a great amount of time, and sometimes prove burdensome to those involved. Is this type of procedure worth it? Conflicting points of view were found on this question in the case studies. One vice president for academic affairs who opposes elaborate procedures feels that the burden is not worth the trouble and would prefer the more informal, less pressured atmosphere which characterizes institutions with informal evaluation approaches. He thinks faculty and administrators should spend their time on more important duties.

On the other hand, most administrators and faculty at the two institutions used as examples above feel that the advantages of the elaborate summative procedure outweigh the disadvantages. The following are some of the advantages that were cited in the case studies.

a. The criteria, standards and evidence needed for promotion, tenure and salary increases are known throughout the institution; because all know what is expected of them, the procedures stimulate rigor at lower levels.

b. Young, untenured professors seem to feel that the system provides the "potential for equity." They feel protected by the elaborate procedures. They feel that it is nearly impossible for anyone to "get" anyone else.

c. The broad and significant faculty involvement in the system gives faculty members a feeling of control over their own destiny.

d. The procedures provide for "structured conflict" in which people are encouraged to look for weaknesses, to criticize, and to find ways to do things better.

e. The procedures provide top administrators the means to "keep up" with what is going on at all levels of the institution. By reading all those dossiers, administrators find out what departments are doing, what their plans are, what weaknesses exist, and so on. It provides structured opportunities for administrators to meet with each other and communicate.

f. The procedures are seen as a kind of "checks and balances" system in that all institutional perspectives and constituencies are brought to bear on the same issue. Top administrators then can balance the various interests and points of view. Without the procedures, top administrators do not necessarily know what these various and competing perspectives are.

Finally, administrators at institutions using elaborate procedures for summative purposes express the viewpoint that these procedures have the effect of stimulating and assisting faculty development and improvement; in short, a formative effect to a summative process. Not only does the quality of the total faculty improve (which is one approach to formative evaluation), but simply going through the procedures causes faculty to do their best. Knowing that what one does is being observed carefully at several levels seems to have the effect of causing one to perform better. In addition, these institutions find the elaborate procedure brings about a constant examination of the evaluation program itself. Persons at various evaluation levels continually question the appropriateness of the evaluation areas, criteria, standards, and evidence available, and try to improve all the components. The result is a continually changing rather than static evaluation program.

Formative Procedures.

It seems paradoxical that a majority of administrators responding to the survey cited faculty development and improvement as the primary purpose of their faculty evaluation systems, but that almost all existing procedures have been established to facilitate summative personnel decisions. There appear two chief reasons for the existence of so few examples of formative procedures. First, there is widespread feeling among administrators that faculty development and improvement is a personal and individual matter. It will be recalled that 36 percent of the respondents stated that simply providing faculty with information on their own teaching effectiveness was the primary purpose of their evaluation programs. Administrators at these institutions evidently feel this practice brings about improvement, although there is little evidence to support that supposition.

The second reason for the lack of established formative procedures seems to be that there are few examples of what can be done and few new ideas are available. The two most successful approaches found are the establishment of campus faculty development centers or offices and the use of growth contracts.

The work of campus faculty development centers or offices in Southern colleges and universities has been described in detail elsewhere and need not be repeated here (Crow, Milton, Moomaw and O'Connell, 1976). While these centers are directly concerned with providing assistance to faculty for development and improvement, they normally are not related to formal faculty evaluation programs. Instead they are available at the institution for those faculty who wish to receive such assistance. Directors of these centers say they do not want to become part of formal faculty evaluation because their services are best provided in a non-threatening atmosphere. It is hoped that faculty members, upon receiving the results of formal evaluation, go to the centers for the assistance they need. Unfortunately, that does not always seem to be the case. Perhaps there is a failure on some campuses to communicate the services offered by campus centers. A procedure that would encourage individual faculty to use the services of the centers when they find, as a result of the evaluation program, that such services are needed, would be useful. Moreover, it would seem to be desirable to establish a direct link between evaluation and faculty development centers. An option that institutions might consider is a requirement that persons achieving a certain level in the evaluation program have an interview with a staff member of the faculty development center. These interviews could be confidential and not part of future evaluation.

Perhaps the most successful formative procedures currently in use are those that come under the rubric of *growth contracting*. A growth contract focuses on each individual faculty member, assisting each to assess his or her own work and to develop plans and activities for change and improvement. Formative evaluation and the needs both of the individual and the institution are emphasized. The individual

faculty member plays a prominent role in deciding what criteria, standards, evidence, and process are most appropriate and useful. Emphasis is placed on a helping relationship rather than on summative decision making.

In the growth contract model, each individual works with a team of two or more persons who assist the faculty member in assessing strengths and weaknesses. Team members provide supportive observations and helpful criticism. Throughout the process, emphasis is positive and constructive. Institutions using this approach seem to use this formative evaluation system periodically for all faculty members and have a separate summative evaluation system that is used to make personnel decisions. In a growth contract institution the summative evaluation is neither so intimidating nor so surprising as it normally is because the faculty member is continuously involved in evaluation, assessment, growth and development. The individual is aware of strengths and weaknesses and has been striving constantly to improve.

At least one institution which uses the growth contract approach sees it as providing a means of combining both summative and formative evaluation procedures. Smith (1976) proposes that the growth contract should be used in this way. Smith and others involved in faculty development activities believe it is important that summative and formative procedures not be separated. It was found in the case study interviews, however, that the stress factor involved in summative evaluation is very strong and appears to interfere substantially with the relationships that formative procedures require. In addition to Smith, cited above, further detail on the use of growth contracts is available in Mather (1975) and from Gordon College (1977).

Other formative procedures currently in use appear to be quite informal in nature, sometimes no more than appointing a senior faculty member within a department to serve as a counselor to younger faculty. When this is the case there usually is an attempt to separate this practice from the summative evaluation process. Success seems to depend greatly on the willingness of individual faculty to make use of the available resource.

Section 3.

Strategies for Implementing Faculty Evaluation Programs

Section 2 described a framework that might be used for analyzing or developing a systematic faculty evaluation program. Although faculty evaluation programs are best when tailored to suit the individual institution, it is the thesis of this report that there are common basic components which all evaluation programs should contain to be workable. The framework which contains four components—1) purposes, 2) evaluation areas, 3) evaluation elements, 4) procedures—can be used by any type institution to analyze its evaluation goals and design a comprehensive approach that best fits its own circumstances, style and traditions.

We have proposed that the most logical approach to be used in designing a faculty evaluation program is to develop the four components in the sequential order in which they have been presented. While this order may not be essential in all cases, we are certain that the first step should be a clear and straightforward statement of purpose and expected outcomes. This first step greatly influences development of the remaining components and the ultimate test of a successful faculty evaluation program is whether the program achieves its purposes. The best planned faculty evaluation system is a failure if the plan's objectives are not achieved.

In addition to helping identify the essential components and characteristics of current faculty evaluation practices, our investigations also yielded insights into various conditions and strategies which seem to have an impact on the success of an evaluation program.

1. The institution might choose to organize the evaluation program around a central idea or specific program emphasis.

An analysis of the case studies led to the identification of several possible emphases which a program might choose as its organizing focus. One such emphasis, successful in a number of major doctoral level institutions and potentially successful in other settings as well, is to develop complete procedures first of all. The concern for due process

and fair procedure in these institutions seems to be greater than concern for particular detail in criteria, standards and evidence.

Another set of institutions has chosen to focus on a quantitative-mathematical approach, requiring quantification of key program components. This approach is recommended in some of the literature and is practiced in a number of non-doctoral institutions where evaluation emphasizes the quantification of criteria, standards, and evidence, and deemphasizes the importance of elaborate procedures. Elaborate procedures are thought to be unnecessary because judgments are made by following a mathematical formula. This approach is said to offer the advantage of precise communication of expectations for rewards. It may be the best approach for institutions where a high degree of objectivity and specificity is valued.

Placing primary emphasis on student learning outcomes is an approach being considered by a few bachelor's level and two-year colleges in the region. This approach can most likely be used by institutions using instructional methodologies which call for clear measurement of what students have learned. Such measurement appears to be most possible when there is individualized instruction, a systems approach to instruction, a competency-based curriculum, performance-based instruction, and the like. Using these measures for faculty evaluation may not be possible in more traditional institutions where expected student learning outcomes are not precisely stated or examined.

A number of institutions emphasize MBO and use it as an organizing rationale in developing their faculty evaluation programs. In this approach, written behavioral objectives are developed for both criteria and standards in the evaluation framework.

2. Components and elements of a program should be designed to fit the individual institutional context.

Institutions thinking of initiating new programs sometimes try to gather all existing materials on similar institutional programs in order to adopt those that have been successful elsewhere. The case studies suggest, however, that while it is helpful to know how various practices work in other places, success appears most likely when each institution develops program components for itself. Numerous evaluation forms that probably work very well in some circumstances are available for purchase on today's market. But it may be best not to start with the assumption that they will work in all circumstances. The very process of designing activities may be helpful in generating understanding and support among the various institutional constituencies.

3. Program components and practices should be kept open and well communicated throughout the institution.

One of the reasons some faculty evaluation programs fail is the attempt to maintain a degree of secrecy about certain program

components or elements. Keeping secret the names of faculty members who serve on evaluation committees, for example, seems to build distrust among faculty. One procedure might be to have a faculty body choose the members of these committees and to publicize the names of those chosen. It also seems best for the details of the evaluation components and elements to be openly and democratically arrived at, adopted, and publicized throughout the institution.

4. Willingness to be incremental and flexible in developing the program and its components is important.

In developing new programs, institutions sometimes wait until components and elements are totally and completely developed before beginning any implementation. This practice can result in a long and tedious process, with implementation sometimes never taking place because all components and elements of a faculty evaluation program may never be perfectly developed to everyone's satisfaction. The best strategy to follow, therefore, may be to start the program incrementally and maintain a flexible stance toward its further development and modification.

5. It is important that top administrative support and commitment be assured.

The administration should be committed not only to having an evaluation program, but also to doing what is necessary to make it work. There should be a willingness on the part of the administration to expend a large amount of effort, give the faculty a significant voice in program development and operation and provide resources for the necessary evaluation components.

6. The faculty needs to be intimately involved in program design and to play a significant role in program operation.

There is no doubt that faculty evaluation programs and practices can be dictated from the top, but the case studies show clearly that such dictation does not produce workable programs that achieve the purposes set for them. Institutions operating systematic faculty evaluation programs have found that faculty are capable of filling a responsible role in evaluation. Moreover, faculty are more likely to accept an evaluation program and its results if they are significantly involved in the process.

7. Orientation or training for those to be involved in program operations will increase the likelihood of a good beginning.

Serious evaluation roles are new responsibilities for many administrators and faculty. Because so many are unprepared for these new roles, it may be desirable to include role descriptions as part of the program plan and to develop and conduct training sessions, once the proper roles have been determined and before the program itself is implemented.

References

Astin, Alexander, and Calvin B. T. Lee. "Current Practices in the evaluation and training of college teachers," in C.B.T. Lee, ed., *Improving College Teaching.* Washington, D.C.: American Council on Education, 1967.

Boyd, James E., and E. F. Schietinger. *Faculty Evaluation Procedures in Southern Colleges and Universities.* Atlanta: Southern Regional Education Board, 1976.

Centra, John A. "The How and Why of Evaluating Teaching." in J. Centra, ed., *Renewing and Evaluating Teaching.* San Francisco: Jossey-Bass Inc., 1977.

Crow, Mary Lynn, Ohmer Milton, W. Edmund Moomaw, and William R. O'Connell, Jr., eds. *Faculty Development Centers in Southern Universities.* Atlanta: Southern Regional Education Board, 1976.

Gordon College Faculty Development Manual. Wenham, Massachusetts: Gordon College, 1977.

Gordon College Faculty Development—The First Year. Wenham, Massachusetts: Gordon College, 1977.

Hildebrand, Milton, Robert C. Wilson, and Evelyn R. Dienst. *Evaluating University Teaching.* Berkeley: University of California, 1971.

Mather, Anne D. " 'Opening Up' Faculty Evaluation," in *Regional Spotlight,* Vol. X, No. 1 Atlanta: Southern Regional Education Board, 1975.

Meeth, L. Richard. "The Stateless Art of Teaching Evaluation," in *Report on Teaching,* Vol. 2. New Rochelle, N.Y.: Change Magazine, 1976.

Miller, Richard I. *Evaluating Faculty Performance* San Francisco. Jossey-Bass Inc., 1972.

Rose, Clare. "Stalking the Perfect Teacher." *The Chronicle of Higher Education,* October 4, 1976, p. 24.

Seldin, Peter. *How Colleges Evaluate Professors.* New York. Blythe-Pennington, Ltd., 1975.

Smith, Albert B. *Faculty Development and Evaluation in Higher Education.* Washington, D.C.: The American Association for Higher Education, 1976.