GUIDEBOOK

For Evaluating Teaching

Larry A. Braskamp
Dale C. Brandenburg
Eileen Kohen
John C. Ory
Paul W. Mayberry

Abstract

Part I of the **Guidebook** *contains four sections: a general description of intended use, some thoughts on defining good teaching, considerations for evaluating teaching, and the general approach (or model) followed in the* **Guidebook.** *Thus, Part I is a synthesis of "where we are coming from," and it provides a framework from which we critique and recommend specific strategies for the evaluation of instruction in Parts II, III, and IV which will be published in the March, June and December, 1984 issues of the* **NACTA Journal.**

PART I:

Rationale and Principles

Section 1: Use of this Guidebook

This guidebook is about the evaluation of faculty teaching performance and competence. Teaching as used in this book includes a faculty member's performance and competence in the classroom, in structuring and organizing a course, in curriculum development, and in advising students. A distinction between two purposes of evaluation — personnel decision and improvement — is emphasized. These two purposes are to be viewed as complementary. Conflicts that emerge by simultaneously evaluating faculty for both purposes need to be recognized and dealth with; but if evaluation is properly designed and handled, both purposes can be served with a minimal amount of conflict and with increased efficiency and effectiveness.

The guidebook is organized into four parts for purpose of this publication. The ways of collecting information about teaching are organized around the five common sources of information — students, colleagues, self, alumni, and records. For each source, a discussion of the technical quality of the evaluative information that can be collected from each source, examples of techniques and instruments, and a list of suggestions for using information from these sources for both personnel decision making and improvement are included.

While developing this guidebook, the authors were members of the Measurement and Research Division, Office of Instructional Resources, University of Illinois, Urbana, IL 61801.

This guidebook was written for three major audiences: (a) departmental and college administrators who have the responsibility to evaluate faculty for annual salary increases and for promotion and tenure, (b) departmental advisory and executive committes, and (c) faculty who desire to collect more and better information about their competence both for personnel decisions and for improving their own teaching.

If you are a college or departmental administrator or member of a committee with the responsibility for evaluating teaching for salary and/or promotion and tenure, the sections headed by "Suggestions for Personnel Decision" and the two latter sections in Part IV are the most relevant sections. If you wish to learn of possible ways to evaluate teaching for improvement, sections headed by "Suggestions for Improvement" are the most relevant.

Section 2: Defining Good Teaching

Based on a review of the research on teaching at the collegiate level, no universal definition of effective or good teaching exists. The history of research on teaching is rich and complex, and the net result of the research reflects a diversity of conclusions as much as a consensus. Thus no one definition is advocated as the standard against which all teaching is to be compared. In general, the evaluation of instruction can be divided by its emphasis on product (i.e., what do students learn or accomplish in the course?) or process (i.e., what do teachers and students do in a course?) A closer look at each emphasis should reveal that effective teaching is defined differently by the emphasis placed on product or process.

Product:

If product is emphasized, the basis for judging effective teaching is amount of student learning. Although this definition has great appeal, there are two major problems in linking student learning to conclusions about effective teaching. First, the measurement of student performance must be done in such a manner that confidence can be given to the test results. Do the tests adequately tap what students learn in a course? Second, student ability, motivation, and

prior knowledge influence what students learn in a class. Thus these and other factors need to be taken into account in judging the effectiveness of an instructor. Because of these problems, student learning is seldom used as the sole basis for judging teaching competence.

Process:

If process is emphasized, the focus is on what the instructor does both in the classroom and in organizing and managing the course. The questions relevant to process include: What does the instructor require of the students (e.g., assignments, workload); how does the instructor behave in the classroom (e.g., lecture, discussion); and how does the instructor relate to the students both in and out of the classroom? The basis for judging effective instruction centers around teacher rather than student behaviors. However, the linkage between how an instructor behaves and amount learned by students is not always clear and thus sole reliance on process factors is also not recommended.

Despite the problems inherent in defining good teaching, certain generalizations about teaching can be made:

> Teaching is related to student learning and deals with establishing conditions for facilitating learning. Thus evaluating teaching is best accomplished by including both product and process factors.
>
> No single instructional strategy is always superior to any other. Faculty who lecture are not necessarily better teachers than faculty members who use discussion techniques.
>
> Good teaching means more than good performance in front of the class.
>
> Instructors have different skills, abilities, and preferences, and they should be aware of them and be encouraged to use them.

Section 3: Evaluating Teaching: Some Considerations

Evaluation serves many purposes and functions. The three primary purposes are: (a) assist individual instructors to improve as teachers, (b) provide information to colleagues and administrators for decisions about promotion, tenure, and annual salary increases, and (c) provide information to students for course selection. The purpose of an evaluation is the cornerstone of an evaluation program. It may influence the type of information collected, the analysis and portrayal of the information, and the dissemination and use to be made of the information. Because of the importance of purpose, considerations for a formal evaluation program are presented separately for personnel decisions and for improvement.

For Personnel Decisions

1. **An inherent paradox in evaluation can not be avoided.** The paradox centers around an individual's quest for excellence, which is central in the life of a faculty member. On the one hand, faculty interested in improving their instruction specify goals and receive feedback about their progress towards achieving them. On the other hand, faculty value and need freedom to explore and to fail, while not continuously being judged by others. Both approaches for striving for excellence are valid but in conflict. While faculty have an obligation to demonstrate their accountability to those who support them, they also need autonomy and freedom. The major question to ask is: How can evaluation be designed so that the institution can fulfill its responsibility and faculty still have sufficient autonomy?

2. **The linkage among performance, the evaluation of performance, and reward for quality of performance is necessary for an evaluation program to have any utility.** If no contingencies exist among performance, evaluation, and rewards, then evaluation loses its potential and becomes an unnecessary expense in time and effort. Furthermore if only negative consequences result from an evaluation, then evaluation is doomed to fail. Those evaluated will probably resent the process and seek ways to subvert and discredit the evaluation. The consequences in the linkage between performance, evaluation, and rewards do not need to be solely tied to external rewards, such as salary increases or promotion although these extrinsic rewards are often more important than is admitted publicly. Knowledge that one is doing good work is a condition for high internal motivation. The major question to ask is: Does the evaluation lead to any positive consequences?

3. **Evaluation with criteria, standards, and types of evaluative information to be collected is a powerful means by which the faculty learn of departmental or institutional expectations.** Evaluation cannot be ignored, because policies, values, expectations about goals, workload, and excellence are dealt with in evaluation. The determination of criteria, standards, and types of evaluative information to be collected and used is a departmental faculty matter, and one which is often not without controversy and disagreement. In designing an evaluation program, it should not be so explicit that faculty feel the need to behave in certain ways in order to "look good" on an evaluation. If so, evaluations may be usurping a general working principle of academe — faculty are basically interested in their work, and they receive considerable satisfaction from doing their work well. A reward system that replaces reliance on internal motivation with dependence on tangible external rewards may result in ultimately reducing rather than increasing faculty productivity. Thus a balance between communicating

expectations through evaluation and allowing faculty freedom to pursue their own interests must be made. The major question to ask is: Does the evaluation tend to encourage the faculty member to "look good" as opposed to "being good?"

4. **The evaluation procedures need to be incorporated into the departmental and institutional policies for awarding promotion, tenure, and salary adjustments.** The measures and types of information used by the departments as indicators of instructional quality need to be consistent with the policies and communicated to the faculty in advance. The major question to ask is: Do faculty know what information is accepted as legitimate and appropriate evaluative information?

5. **Evaluation must have credibility to both the faculty and to the administration.** Credibility is largely a political matter. Gaining credibility requires the support of both the administrators and faculty, especially the senior faculty. Furthermore those implementing the program must remain impartial and respect the prerogatives of the individual instructor, and establish guidelines regarding the confidentiality of evaluations. The major question to ask is: Does the evaluation have sufficient credibility?

6. **Information used in evaluation must be fair.** Fairness refers to the extent to which the information adequately represents both the criteria used to evaluate instruction and the complexity of the teaching activities. Criteria are often difficult to specify, but instructors being evaluated need some understanding of the basis on which they are judged. Furthermore, if the information to be collected does not accurately mirror the activities of the instructor or student learning, the information is incomplete. The major question to ask is: Does the information used in evaluation adequately represent the teaching efforts of each faculty member?

7. **Information used in an evaluation must be of sufficient technical quality.** Technical quality refers to the extent to which the information is comprehensive, reliable, and valid. At a minimum, the administrative procedures, the instruments, and methods used in the data collection need to be consistent for all faculty. Student, course, and instructor characteristics (e.g., class size, type of course, elective/required status) also often need to be taken into account when the information is interpreted for assessing competence. The major question to ask is: Is the information accurate, trustworthy, and properly used for the purpose for which it was intended?

8. **Evaluation must be based on acceptable legal principles and practices in personnel appraisal.** Evaluation procedures need to be based on how well faculty members fulfill their responsibilities in teaching, justifiable methods of data analysis and interpretation, and needs to incorporate "due process" into the evaluation. In general, the courts have not dictated the contents (i.e., selection of criteria or standards of quality) but have focused on procedural due process (i.e., how the evaluation was carried out). The key question to ask is: Does the evaluation process — the specification of criteria, collection and interpretation of the information, and dissemination — follow legal principles?

9. **Levels of review built into the program make the evaluation more comprehensive, fair, and credible.** Multiple interpretations of the information are superior to a single person's judgment of teaching quality. Furthermore, factual errors are more apt to be detected and corrected if opportunities for review are built into the process. On the positive side, a consensus achieved through multiple reviews helps make the evaluation more credible and fair. The major question to ask is: Can errors and misinterpretation be detected and corrected before a final assessment is communicated to the faculty member being evaluated?

10. **Evaluation is as much a social and human activity as it is a technical undertaking.** Evaluation often is sensitive and deeply personal, especially to faculty who are not yet tenured. Thus the manner in which evaluative information is communicated is a key factor in an evaluation. Personal communication of feedback by a departmental administrator in an annual review has been rated as especially effective by faculty because it provides opportunities for a faculty member to respond to an evaluation and to discuss their career. The question to ask is: How are evaluations communicated to the individual faculty member?

11. **Alternative evaluation procedures can be examined for their benefits to the institution.** A comprehensive set of procedures, while meeting most of the previous considerations, may not be feasible due to lack of time and financial resources. The major question to ask is: How realistic is the evaluation; i.e., which procedures must be included and which can be altered or eliminated?

For Improvement

1. **Information collected for improvement is collected for the instructor only.** Instructors may wish and generally can benefit by sharing information with a colleague, but the instructors should be able to do it at their discretion. This restriction is necessary so that instructors have the freedom to ask questions about problems and that the asking of them cannot be used against them in making personnel decisions. The major question to ask is: Do faculty have the freedom to collect evaluative information for their private use?

2. **Information can be frequently and informally collected.** Since the information is not intended for personnel decisions, any type of evaluative information can be immediately examined to help the instructor in assessing a course. The trustworthiness of the information does, of course, depend on the reliability

and validity of the data. The major question to ask is: Are faculty collecting enough information to monitor their progress?

3. **Evaluation tied to self development maximizes its long range utility.** Minimally, a faculty member needs to think of evaluative information as a starting point for further analysis and problem solving. Evaluation, professional development, and improvement in instruction are inseparable. The major question to ask is: Does the faculty member accept the principle that self evaluation of teaching is a necessary condition for change and growth?

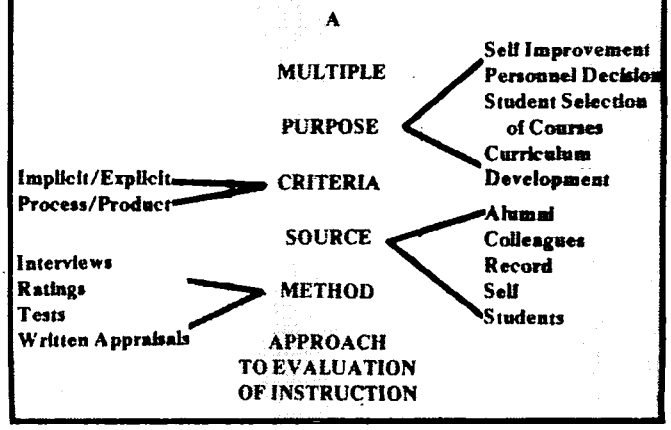
4. **Information collected that is highly detailed, diagnostic, and focused on specific teaching behaviors and course characteristics (e.g., tests, text) increases the usefulness of the information.** Information about specific teacher behavior and course features that need improvement is very helpful before specific changes can be considered. Specific information does not result from asking students or colleagues general questions like "Did you like this course." Instead written comments to specific questions or responses to highly diagnostic scaled items are needed. The major question to ask is: From the information collected, does the instructor know specific strengths and weaknesses?

5. **Information shared with another often increases the usefulness of the information.** A consultative relationship between an instructor and another faculty member or a professional staff member trained in faculty development and evaluation is beneficial for many reasons. The relationship allows the instructor to work through some of the personal reactions to evaluations, especially the negative ones. The consultative relationship also provides an opportunity to both learn about and explore alternative teaching strategies. To improve, a teacher often needs to know more than they are "fair," "average," or "bad" in their teaching. Based on research on the utility of student feedback during the semester, faculty who received feedback within a consultative arrangement improved their teaching (as measured by student ratings at the end of the semester) more than did professors who only received responses from ratings or written suggestions. The major question to ask is: Do faculty who desire to discuss their teaching have opportunities to receive consultative assistance?

Section 4: An Approach For Evaluating Teaching

Evaluation is ultimately a subjective undertaking. Evaluation is more than description; it requires judgments and interpretation. In this guidebook, the key principle is that competence in teaching can best be evaluated if it is assessed from a variety of perspectives. To incorporate this principle a "multiple purpose, criteria, source, method approach" is advocated. This approach, as displayed in Exhibit 1, serves as the conceptual framework for this guidebook. Each element in this approach is briefly described below.

Exhibit 1. An Approach to the Evaluation of Instruction



Multiple Purpose: As noted in Section 3, evaluation is undertaken for a variety of reasons. Four major purposes for evaluating instruction are to provide information to:

1. The instructor for his/her improvement as a teacher.
2. Colleagues for any decisions about the future of a faculty member; e.g., promotion, tenure, termination, special salary adjustments, and annual salary increases.
3. Students to guide their course selection.
4. Colleagues involved in course and curriculum development.

The first two purposes are the primary functions of instructor evaluation. Giving information to students for course selection has not worked very well since the information is often too terse and incomplete and not all faculty or courses are included in a published list. The fourth purpose of evaluation for curriculum development focuses primarily on course rather than instructor evaluation. However, this focus is too often unheeded, especially for courses which are prerequisite or a part of a series of courses in a field of study. Much of the information collected for improvement is appropriate for curriculum evaluation. Instructors who are part of a curriculum committee or team very often can collect information to satisfy these two purposes.

The remainder of this guidebook will focus on the first two purposes. Exhibit 2 presents major features of these two types. Two features are worth noting: the purpose affects data to be collected and how the information is to be disseminated. If the purpose is for personnel decisions, then information that measures overall competence is preferred. If the purpose is for improvement, then highly detailed diagnostic information in which strengths and weaknesses of the instructor are described and assessed is preferred.

The purpose of the evaluation also influences the dissemination of the evaluative information. If the

Exhibit 2. Major Features of the Two Major Functions of Evaluation

Feature	Evaluation for Personnel Decisions	Evaluation for Improvement
Primary Use	Institutional Accountability	Personal Development
Primary Audience	Decision Makers	Faculty Member
Primary Types of Information	Judgments of Quality Global Integrative Assessments High Inference	Descriptions of Behavior Diagnostic, Detailed Specific Low Inference
Primary Evaluation Strategy	Formal, Standardized, Legal	Informal, Frequent
Primary Other Person in the Evaluation	Administrator of Unit	Trusted Colleague, Consultant
Primary Type of Information Communicated	Judgments of Worth and Value to Institution	Suggestions for Alternative Ways

purpose of the evaluation is for **personnel decisions,** then those responsible for the decisions will have access to the data; e.g., department head, promotion committee. If the purpose is for **improvement,** then only the instructor and colleagues working with the instructor on a consultative basis should have access to the information.

Multiple Criteria: Judgments about the value or worth of an instructor and/or course are based on certain criteria. Criteria are dimensions or characteristics that are used for assessing the effectiveness of instructor and/or course. A number of criteria can be used in judging instructor effectiveness. For example, one is student knowledge of common concepts in a subject matter. Others are the instructor's ability to communicate effectively, rapport with students, appropriateness of sequence of the course topics, and clarity of the course objectives. In thinking about criteria, it is important to differentiate between the selection of criteria and selection of indicators or measures of a criterion. For example, instructor ability to communicate effectively can be measured by several indicators — student ratings, colleague assessments through observation, and review of lecture notes.

In selecting criteria, it is helpful to distinguish between **process** or **product.** Is the focus on how well instructors are performing or how much students have achieved? The selection of process and/or product criteria should reflect the importance given to each in defining effective instruction. If teaching effectiveness is defined as the amount of progress students make, then student learning and accomplishments are the primary criteria. If teacher behavior is considered the most relevant factor, then the instructor's teaching skills and ability to design a course should be used as criteria. The emphasis on process and product depends on the values of the discipline and a department's view of good teaching.

Criteria also vary in the extent to which they are specified, described, and measurable. Explicit criteria, like student test scores and attendance, are directly observable. If these are used, everyone knows the basis for an evaluation. As such they are often regarded as the most logical and rational approach to assessing value. On the other hand, implicit criteria, such as colleague judgments based on classroom observation, are more qualitative in nature. They are often regarded as subjective because no tangible observable pieces of data are collected. Instead, judgments based on experience are used to assess value or worth. Both types of criteria are often needed in an evaluation since they complement each other and thus expand the diversity of information collected in an evaluation. If a multiple perspectives approach is adopted, then a variety of criteria are recommended.

Multiple Sources: Information about an instructor can be collected from a number of different sources, since not everyone judges an instructor in exactly the same way. Sources include **self, alumni, students, records,** and **colleagues** (which include other faculty members, departmental administrators, and professional staff trained in faculty development and evaluation).

Multiple Methods: The final step in the multiple perspectives approach is selecting methods or techniques for collecting the information that best serves as indicators of the selected criteria. A number of ways can be used to collect information about teacher competence and course quality — **achievement tests, ratings and surveys, written appraisals** (comments and critiques in response to open-ended questions), **interviews,** and **observations.** The selection of a method is essentially a measurement task; i.e., what procedure or technique should be chosen to obtain the most reliable and valid information?

CONCLUSION: PART I

In sum, the approach outlined in this guidebook emphasizes the importance of multiple perspectives. Information collected from a number of sources and by a variety of methods, each reflecting a diversity of criteria, is the ideal for obtaining a fair and credible assessment of the teaching competence of a faculty member. However, adopting this approach in its entirety is seldom feasible. Selections must be made. In the next section, some of the more common source/method combinations are described, including the importance and appropriate use, technical quality of the information collected by the combination, and suggestions for collecting and using the information. This material will be in the March issue of **NACTA Journal.**