Purposes of Evaluation of University Instructors: Definitions, Delineations and Dimensions

NAFTALY S. GLASMAN AND WALTER H. GMELCH*

ABSTRACT

Research and writings on evaluation of college and university instructors are expanding. Not unrelated to this phenomenon are controversies which focus on justification and use of ratings of instructors[1] as well as attempts to meet demands and ward off pressures for accountability. Also not unrelated to it are uncertainties and confusion surrounding purposes of evaluation of instructors and a widespread feeling that no single evaluation instrument can best suit more than one purpose. This paper describes considerations for establishing purposes of evaluation. We discuss definitions, delineations, and dimensions of purposes, and propose general models which can serve as guidelines for further development and specification of purposes.

In Search for Purposes

The pessimist would argue that such a search is pointless. The argument is based on the link which connects teaching effectiveness, evaluation and purposes of evaluation. Because there are two unbridgeable views of effective teaching, there cannot be a common definition of evaluation. Thus, a search for a model which would help delineate evaluation purposes is pointless.

More specifically the pessimist's argument would run as follows. A model which guides the development and specification of purposes has to be based on a clear-cut definition of evaluation which is acceptable to all. We do not have such a definition. Some define evaluation as a system of measurement and testing while others view it as the formulation of statements of congruence between performance and objectives. Still others accept a multiplicity of definitions, each being a function of the role of different individuals associated with evaluation – the instructor, the administrator, the student, the economist, the politician, the taxpayer. These differing views cannot be easily bridged.

The pessimist will also say that a common definition of evaluation would have to rest on the assumption that there is a clear-cut definition of effective teaching which is accept-

*Field Training and Service Bureau, College of Education, University of Oregon.

able to all. However, we do not have such a definition. Some define effective teaching as a unique personal activity which cannot be studied and explained in full, and about which no meaningful generalizations can be made. Others define certain aspects of teaching as describable in ways which may lead to a better appreciation of current practices if one develops first adequate theoretical models and techniques of assessment.[2]

The optimist believes that our knowledge of teaching effectiveness and evaluation is more promising for a search for purposes. Faculty and students tend to hold, with varying degrees of consistency and depth, several conceptions of what the term "effective teaching" actually means. It would be naive to believe that professors themselves do not have some sense of their effectiveness or ineffectiveness.[3] The profession, too, has substantial information about faculty job components at the university level.[4] Research and service can be included under the domain of teaching if they have direct impact on teaching.[5] Scholarship, delivery and advising are three specific areas which are directly related to the definition and execution of teaching.[6]

Scholarship is an integrative type of activity which could be labeled research, but is more related to the instructional function of teaching. Specifically, scholarship refers to the instructor's breadth of knowledge, analytic ability, and conceptual understanding of the literature and research in his field as these relate to the courses he is teaching.[7] Questions which can be asked about an instructor's scholarship ability in teaching include: Does he discuss views other than his own? Does he present facts and concepts from related fields? Does he contrast the implications of various theories, and discuss recent developments in his field?

The second area, delivery, concerns the instructor's skill at presentation. It is subject related as well as student related and is not merely a matter of his theatrical or rhetorical skills. Questions which may be asked of his delivery skill in teaching are: Does he state course objectives? Does he summarize major points? Is he well prepared? Does he invite criticism of his own ideas? Does he encourage discussion? Does he know when students are bored? Delivery, therefore, includes not only teaching behavior in the classroom, but also the planning of class activities, preparation, evaluation, and continuing improvement of the instructional process.

Advising, the third area of teaching, has received little attention. It encompasses the instructor's interaction with students in and out of the classroom setting. Is he interested in his students? Is he friendly toward them? Does he assist students in academic and personal problems out of class? Is he accessible and approachable to students? Does he relate to and respect students as persons? Such questions need to be asked about the instructor-individual student interaction, which through mutual respect and rapport creates an atmosphere where advising is available, natural, and effective.

The optimist would start his task of clarifying what evaluation of instructors is by admitting that the need for such evaluation has been consistently argued as being that of improvement of the teaching learning process. This goal has been supported by the intimate relationship between evaluation and such a process.[8] He would extend his task, however, by recognizing that evaluation has been defined not only in terms of how it functions to improve the quality of teaching, but also in terms of how it functions to safeguard the teaching profession.[9]

The first function, to improve the quality of teaching, has been argued continuously.

Complaints about the quality of undergraduate teaching are both current and chronic.[10] They are being voiced at all levels of education, from educational policy makers, university and college administrators, faculty members, and students alike.[11] Properly conducted evaluations with corrective feedback mechanisms can serve as a tool to improve the quality of teaching.[12]

The importance of the second function, to safeguard the quality of teaching received by students, has arisen from political pressure both from outside of the university and from within it. Valid and reliable evaluation could well serve to protect universities and instructors from political interference with their professional autonomy.[13]

Already existing literature suggests to the optimist several purposes for evaluation of instructors in higher education. Among them are improving teaching, rewarding teaching,[14] supplying information for administrative decision making,[15] supplying information for students,[16] protecting individuals and organization, aiding in selection, setting public policies, forcing communication between teacher and others in the institution,[17] and improving research on teaching.[18]

If evaluation results are to be used as inputs for the process of decision making, choices among such purposes must be made. The evaluator must know specifically who wants to know what and with what end in view. Otherwise evaluation is likely "to be mired in a morass of conflicting expectations,"[19] especially when upon closer inspection of purposes some are in conflict with each other,[20] some overlap,[21] and some vary according to their institutional environment.[22] The multiple purposes also require different kinds of data,[23] from different sources,[24] at different times,[25] and using different designs.[26] In sum, an all-purpose evaluation is a myth.

The optimist, then, would attempt to pull together in a systematic way information and ideas on the purposes of teaching evaluation. Over a sixty-six year period more than 3,000 studies have attempted to isolate the variables related to effective teaching.[27] Problems which focus on *purposes* of teaching evaluation have received comparatively little attention.[28] The rest of this paper is devoted to extend the latter effort and to deal with problems of specification, organization, and systemization of purposes of evaluation of teaching. The task would be complete if four essential components of purposes of the evaluation of teaching are taken into account: their definition, delineation, dimensions, and instrumentation. This paper will deal with the first three. The fourth was dealt with elsewhere.[29]

Detailed definitions of specific purposes will serve as a departure point, and necessary precursor, for subsequent discussions of the delineations and dimensions. An attempt will be made here to delineate the "simple verbal" definitions by specifying their logical parts. Since it is apparent that for any given purpose some dimensions are better suited than others, the way chosen here to promote a better fit between a given purpose and the selection of instruments and techniques is to isolate a reasonable number of dimensions, and then categorize the measures according to the dimensions they fulfill. This could establish a defensible and valid selection among the array of instruments and techniques. The dimensions which are included in this paper are the nature of the data (from descriptive to judgmental); the level of specificity (from detailed to summary); the method of reporting (from comparative to noncomparative); the timing of the evaluation (from continuous to end of term); and the audience (from private to public). Other dimensions which are specific to one specific purpose are reported in a separate category.

The optimist's search for purposes may start with the most relevant audience for the results of the evaluation. There are at least four clearly distinguishable audiences — the professor himself, his colleagues and administrators at his institution, his current or potential students, and the public at large including any of its segments. Four corresponding purposes emerge and to each a separate section is devoted. These are self improvement, administrative decision making, information to students, and research. The last section of this paper will compare the four purposes in terms of their definitions, delineations and the various dimensions.

Evaluation for Self-Improvement

Definition: Evaluation for self-improvement requires conditions under which instructors can *acquire and diagnose feedback information on their teaching as a means of developing their own teaching competencies.* This evaluation does not include a pronouncement of judgment on the quality of an instructor's teaching; rather the results are used as one uses requested and respected criticism: to provide both the drive and direction for self-improvement.[30]

Delineation: The roots of teaching improvement are found in commonly accepted perceptions of the nature of effective teaching. Some perceive that teaching, by nature, is a great art, rare, and protected by a system of traditions and myths such as the Ph.D. is a license to teach; teaching cannot be taught; good scholarship assures good teaching; and all teachers can please some of the students some of the time, some teachers can please all of the students some of the time, but not all the teachers can please all the students all of the time. This belief system suggests that teaching effectiveness is predetermined by one's genetic ability or by some gracious act of God.

An alternative perception of the nature of teaching is that the teacher is a developing individual who is motivated to realize his potential to become a more effective teacher. Few attempts have been made to conceptualize faculty development.[31] Sanford argued[32] that college professors develop as individuals, in much the same way that others develop through their professions. Their development is distinguished by progressive stages which are only loosely related to chronological age. The first is the achievement of a "sense of competence" in one's discipline, prior to which the professor is unprepared, as a general rule, to move on to the stage of "self-discovery," in which he attends to other interests, aspirations, and abilities. The third stage, the "discovery of others," is the final stage. Ideally the three stages follow in much the same order as Erickson's stages of identity, intimacy, and generativity.[33]

The faculty member must be motivated and stimulated as he moves through these developmental stages if improvement in his teaching is to occur. Earlier concepts of motivation would have held that faculty members inherit most of their capability to perform and that the capabilities of becoming effective teachers can only be maximized by reward and punishment. Additional and alternative theories suggest different premises on which to base improvement of teaching. One postulate is that individuals are constantly striving to satisfy one of a number of hierarchical needs, the apex of which is self-actualization. Achievement is one of man's basic needs.[34] Given this context, the professor would not only be concerned with achieving effectiveness in his teaching but would derive considerable satisfaction from striving for it. McClelland's theories X and Y are also relevant.[35] Theory X of human nature assumes

that man is inherently lazy, unwilling to assume responsibility, and resistant to change. Theory Y is based on the assumption that man does wish to grow and maximize the worth of himself and other people. If it is assumed that professionals will inherently act in accordance with Theory Y, then within the realm of the teaching profession, Theory Y may be the most appropriate modus operandi. McClelland's Concept E management, the belief in individual self-induced development, would be the proper basis on which to begin improvement in teaching.

An all inclusive premise, then, is that motivation is not primarily a fuel to be injected into a system. Rather, it is more an attribute of individuals, linked to their physical vitality. In higher education it may be stimulated by social forces and related to the tone of the educational system and to the presence or absence of opportunity.[36] A faculty member will be self-induced to improve his teaching if the campus provides worthwhile evaluative feedback to him about his teaching. Such a system should be able to serve the need for the improvement in teaching.[37] It should also increase the professor's willingness to expend energy and imagination on his teaching and to enhance the teaching profession.[38]

Dimensions: The nature of evaluation for self-improvement suggests that within each of the first five dimensional ranges, the information for improvement should be respectively diagnostic, detailed, continuous, non-comparative, and private. Elaboration of each of these dimensions follows:

1. Nature of Data: Diagnostic. Evaluation data on individual style, course objectives, and teaching needs of the instructor becomes important in giving the teacher desired feedback on his own personal teaching skills. Data should be specific and intensive enough to permit the instructor to diagnose his teaching strengths and weaknesses. Data is not to be seen here as *against* the faculty, *solely* for the student, nor *by* the administration; it is meant *for* the individual instructor.

2. Level of Specificity: Detailed. Global summary information is not much help to an instructor in search of self-improvement. The information must be specific and detailed enough to provide diagnostic data about instructional problems.

3. Method of Reporting: Noncomparative. Comparisons here become meaningless. Descriptive characteristics and styles of the instructor demand that the information be reported as individual statements of the faculty's teaching profile.

4. Timing: Continuous. Assessment must be available when a person thinks he needs it, continuously if desired. End of term evaluations are essentially of little help if an instructor is to improve his performance while the course is still in progress. It is also unlikely that development and self-improvement occur at certain times during the course of teaching. Development is continuous and needs immediate but continuous feedback and assessment.

5. Audience: The Instructor. The diagnostic information should be accessible only to the instructor himself. He should also have complete discretion as to whether he shares the information with students, colleagues, or administrators. When personnel decisions are pending, he should have the option of using the diagnostic information as testimonial to substantiate any changes he has made toward improvement.

6. Additional Characteristic: Collaborative. Evaluation for improvement is not a zero-sum game. The institution should encourage collaboration among all participants in the educational community. Students have certain perceptions that are helpful in diagnosing a professor's skill in delivery and advising. Self and peer evaluation are essential feedback sources on the instructor's quality of scholarship.

Evaluation for Administrative Decision-Making

Definition: Evaluation for administrative decision-making requires conditions under which administrators can improve their decisions on teaching-related issues. This evaluation *provides a basis on which administrative decisions can be made concerning personnel, modification of assignments, and allocation of learning resources.*

Delineation:

A. Personnel Decisions

Personnel policies and decisions affect faculty members in the areas of selection, advancement, promotion, tenure, and salary. Exact practices and procedures used in assessing fitness for tenure are apparently seldom clearly defined and stated.[39] Because of the significant subjective nature of personnel decisions, faculty generally oppose decisions based on evaluation of teaching.[40] At least for reasons of equitability, every effort must be made to introduce a larger measure of precision into the procedures of personnel decision-making by means of valid and reliable evaluation sources and data.

B. Modification of Assignments

Modification of faculty assignments include decisions on course assignment, timing and frequency. With a broader information base on teaching and teaching effectiveness, one should be better equipped to make decisions as to increase or decrease faculty loads, or assign a professor to an under-graduate or graduate, introduction or advanced course, offer a course in the morning or evening, and offer a course once in two years or twice each quarter. Further, it could produce other mechanical variations in the teaching environments such as residence hall classes, cluster colleges, and ethnic programs.

C. Allocation of Learning Resources

Allocation and adjustment of learning resources is a third area of decision-making. Such resources include personnel, equipment, and personal activity resources. More specifically they may consist of instructional media packages, teaching assistants, or even time and travel expenses for visiting other campuses where similar courses are taught.[41] The problem here is to assess the instructor's effectiveness, then administratively allocate additional resources to him which may be at the least a necessary minimum and at the most have potential for assisting him improve his teaching.

It is assumed that personnel decisions should be based on the assessment of those teaching variables which are under the direct control of the instructor.[42] Information collected for the other two decision areas encompasses the total learning environment, including variables not directly under the control of the instructor such as the hour the class meets, the environment in which it meets (class size), etc. It is generally not fully recognized that other decisions regarding an instructor require more information than those facts which directly relate to his effectiveness.[43]

Dimensions: Dimensions of evaluation for administrative decision-making will vary with respect to which of the three decision areas are under consideration. The following are only some dimensions which may apply to all three areas. In general the following suggests that evaluation for the purpose of administrative decision-making can be facilitated by emphasizing that the nature of data be judgmental, the timing be end of term, the specificity of data be summary, and the users of the data be both faculty and administrators.

1. Nature of Data: Judgmental. Making judgments about merit is unique to this area of evaluation. Administrators must examine and weigh the evidence about a particular teacher's behavior against some explicit or implicit criteria of effective teaching.

2. Level of Specificity: Summary. Only general, overall evaluations are needed for administrative purposes.[44] The administrator does not necessarily need to know the diagnostic details of the merits or shortcomings of an instructor's confidential attributes. For decisions on modification of assignments and allocation of resources, more specific data is necessary.

3. Method of Reporting: Semi-Comparative. Since there is no ideal type of teacher, one can only come to the conclusion that personnel evaluation reports must be flexible and must allow for the reflection of individual differences in courses, subject matter, teaching styles, and external influences. At the same time it is probable that administrators and review committees deal with data from many hundreds of instructors each year. Both faculty members and administrators should have some sense of what constitutes effective teaching. Therefore, the report might be characterized as semi-comparative, giving some feeling for the institutional norm on teaching, but also leaving room for individualistic qualities to be reflected.

4. Timing: End of Term. Little can be gained from instantaneous feed-back on faculty performance. If an instructor is hired for a certain course of time, he should have the benefit of using that time to develop and display his total teaching package.

5. Audience: Faculty and Administration. Both faculty and administrators need to know; students do not. It would be unethical to make personnel decisions without allowing the instructor to inspect and respond to the data on which the decision is to be made. It would also be fruitless to make assignments and modifications without the faculty member's presence and consent.

6. Additional Characteristic: Corrective. Although a corrective feed-back mechanism is probably more unique to evaluation for the purpose of self-improvement, it is apparent that any evaluation system aimed at judging faculty performance should also provide adjunct services. There are three reasons that adjunct services are needed: (1) most faculty members have had little if any teacher training; (2) without providing the means for improvement, evaluation systems for judgmental purposes are unethical and not very valuable to the individual or institution; and (3) unnecessary resistance may be fostered if adjunct services are not provided with judgmental evaluations.[44] Therefore, designers should ensure that this dimension is included in the evaluation program.

Evaluation for Student Information

Definition: Provision of evaluative information for students *enables students to become enlightened consumers and educated participants in the educational process.* Through the consumption of this information, students will be able to shape their own experience, that is, to coordinate the educational offerings with their interests, needs, and objectives. The information will also encourage broader student interest and participation in the educational process.[46] Self-gratifying learning experiences are thus fostered through the intelligent selection of courses and instructors and culminate in drawing students further into responsible and positive action in the academic community.

Delineation: What is included in the term "student information"? First and foremost it

would include information on instructors and courses which would replace the "rumor system" prevalent in many universities and colleges today and serve as counseling assistance in the selection of courses and instructors. Course and instructor information is further refined into two major types of information: summary-rating data and descriptive data.[47] In many instances, summary-rating data is quite easily obtained. Student groups on numerous campuses collect ratings and report the results in critiques and "counter-catalogs," eventually to be published and sold to fellow students.[48] This, however, is not a perfect information system because data are usually collected from a few professors who wish to release such information, organized in such a manner that only a few comparisons can be made, and distributed through sale in the campus bookstore. Descriptive data on instructors and courses are less readily available. The main sources of such information are course catalogs, but these are usually scanty descriptions of course contents. Many are out of date, give misleading information, provide little indication of the flavor of the course, and present no information on the instructor's style, methods, or characteristics. To aid in the selection of courses and teachers, students do not want merely adjuncts or updatings of the present course catalogs. They need to know about the teacher's style of presentation, his emphases on academic activities, and any idiosyncratic characteristics which may have some effect on their learning. In other words, in order for students to be intelligent consumers of education, they must be presented with information about the instructor's qualities in scholarship, delivery, and advising as well as the mechanics, content, and context of the course being taught. Such information will enable the student to better match course and teacher characteristics with the needs and objectives of his educational endeavor.

Dimensions: Information should be produced so that students become better advised on their selection of instructors and courses. Such an evaluation system must be primarily descriptive; it must summarize the major teacher and course characteristics in a comparative form; it must collect information systematically and at the end of the instructional period; and it must disseminate the results to all students free of charge.

1. Nature of Data: Descriptive. Evaluation data for student consumption should be descriptive. It should enable students to select courses and instructors according to their interests, needs, and objectives.

2. Level of Specificity: Summary. The level of specificity should be general and summative. Students normally do not ask for information on an instructor's belief system, opinions, or prejudices, but rather summary descriptions of the course's content, objectives, and physical setting in addition to information on the instructor's delivery and advising skills.

3. Method of Reporting: Comparative. The general, summative information should be published in such a form that it allows students to make general comparisons across departments, courses, and instructors.

4. Timing: End of Term. Information should be collected and disseminated at the end of each term after the end of the normal instructional period and before conclusions and feedback are given.

5. Audience: Students. The information should be public and made accessible to all students without charge.

6. Additional Characteristics: Systematic. While "systematic" may not seem to be a quality unique to this purpose of evaluation, it must be stressed that the information here be systematically collected and disseminated. To make this operational, it should not be

left to the faculty member's discretion whether information on his course and teaching will be collected or not. Disclosure of the information should be out of the hands of the instructors. The information must include all instructors and courses, for if it is incomplete, enlightened decisions by students are not possible.

Evaluation for Research

Definition: Our intention in this section is to promote the union between evaluation and research, such that evaluation can capitalize on the results of research, and research can benefit from the unique perspective of evaluation. Since both have their own individual characteristics, we should concentrate on bridging between them in pursuit of achieving the goal of effective teaching. The first grounds for the union are defined as *providing research with a criterion of teacher effectiveness aimed at predicting, understanding, and controlling the teaching process.*[49] What is presently known about teaching is relatively little compared with what ought to be known. This retardation has been due to the lack of reasonably valid and reliable measures of teaching outcomes which, if specified, will allow research to move ahead and eventually facilitate the increase of teachers' competency.[50]

The complementary grounds for this fourth purpose are based on an interest on the part of psychological and educational researchers in the nature of teaching and the facilitation of learning.[51] The evaluator's purpose for research inquiry on teaching is defined as *describing accurately what teachers do, searching for correlations and linkages between theoretical variables and learning, and demonstrating the predictive power of teaching variables in "making a difference" in learning.*[52] While some are impressed at the amount of research already done,[53] others are appalled at the amount of investigation still to be conducted.[54]

Delineation: Except for passing comments, the purpose of evaluation research has received little attention and specification in the past. We wish to delineate this purpose by describing and elaborating the distinction between evaluation and research; the teaching areas in need of more investigation; and the styles of research utilized in the evaluation of teaching.

A. The Distinction Between Research and Evaluation

The broad area of "disciplined inquiry" encompasses several common elements in education, two of which are research and evaluation. Conceptually, evaluation and research can be differentiated as follows. Educational research draws upon both historical and philosophical inquiry while educational evaluation relies heavily on philosophical inquiry and only slightly upon historical inquiry. Both, as one might suspect, are solidly rooted in empirical inquiry.[55] More specifically, research and evaluation can be differentiated by activities,[56] intent,[57] methodology,[58] and extended generalizability.[59] These differentiations take into account such issues as replication of results, control of variables, problem selection, value judgments, data collection, motivation of inquirer, decision and conclusion orientations, salience of value questions, investigation techniques, criteria for judging, and generalizability of institutional evaluation, summative evaluation, formative evaluation, and instructional research.

Although there are discernible differences in the specifics of evaluation and research,

there are common ingredients as well. First, each evaluation or research activity produces knowledge that was not previously available.[60] Second, both promulgate activities designed to collect evidence systematically, to translate the evidence into quantitative and qualitative terms, to compare it with established criteria of success, and to draw conclusions about the phenomenon under study.[61] Third, although methodologically the pure experimental design may not be wholly applicable to evaluation, there are legitimate and useful designs available and common to both activities.[62] Fourth, the mission of both is to attempt to describe and understand the relationship between variables and disseminate the results to researchers and educators alike.

B. Teaching Evaluation Areas in Need of Research

Four teaching evaluation areas have been identified as being in need of additional research and investigation.[63] The first area is research on faculty attitudes towards evaluation. Teachers' attitudes vary according to educational philosophies, teaching skills, subject matters, teaching environments, and class compositions. The dilemma is that the extent of the influence of these factors on faculty attitudes is virtually unknown. At the present time only a limited number of colleges and universities have been cited in the literature as conducting systematic surveys of faculty attitudes. The results have been far from conclusive and limited in scope. If faculty views with regard to evaluation and how these views are shaped were known, evaluation means could be constructed to have more relevance to individual faculty members, thereby hopefully also reducing faculty resistance to evaluation.[64]

The measurement of changes in student behavior constitutes the second area in which research is needed. In the last decade an increasing amount of attention has been devoted to student growth as a major criterion for teacher effectiveness.[65] A summary of seventy-five doctoral studies conducted at the University of Wisconsin testifies that student change should be the primary criterion against which all other criteria should be validated.[66] However, the problem of the absence of objective and reliable measures of teacher effectiveness based on student gain still persists.

The third area for important research consideration is the measurement of effectiveness in terms of the instructor's personal attributes. Although gains in student learning are recognized as the best criterion to judge teacher competency, many researchers and educators have resorted to the more readily available measures of teacher attributes.[67] The use of teacher attributes assumes that these specifications are related to student growth. Since the linkages have not been positively identified, teacher attribute measures are considered as second best, a priori measures to student learning. Even with second-best measures the problem still remains with defining teacher competency. While the literature on teacher competency is overwhelming, few if any facts are firmly established about teacher effectiveness, with no approved method of measuring competency.

The fourth area is the classroom environment. In their quest for effective teaching correlates, researchers have investigated the impact of the classroom environment itself, believing that this is the most salient and important trend to emerge.[68] Definitions of the classroom environment have also been suggested,[69] but at present a taxonomic effort toward describing the classroom environment and its interaction with students and instructors is needed for further progress in this direction.

The elusiveness of the evaluation of effective teaching beckons for additional research studies in all four of the areas previously outlined. The inadequacies in the present state of knowledge and subsequent evaluation devices, as well as the damage being produced by using inappropriate criteria of effectiveness, accentuate the need for enlightened research.

C. Styles of Research on Teaching

Three styles of research on teaching have been distinguished: experimental, correlational, and process-descriptive.[70] The classic design for evaluation, in thought if not in practice, is the experimental model. This design calls for the use of experimental and control groups and the manipulation of an independent variable. A teaching method may represent the independent variable while changes in knowledge, understanding, or attitudes of students, the dependent. The second style, correlational, does not manipulate the independent variable and typically uses some measure of the teacher's behavior and characteristics. It is usually recorded by observation, testing, or rating. As in the experimental model, the dependent variable is some measure of student change with the results reported in correlational coefficient form. The third style emphasizes description, with the purpose not necessarily to establish relationships but rather to elaborate on the elements of the teaching-learning process itself.

It has been suggested that correlational and experimental studies have lacked precision in defining experimental variables.[71] The proposed solution is to call for the assistance of process-descriptive studies to specify variables in more definitive terms. Gage[72] suggested 16 years ago that . . . "only after we have raised the homely art of description to a much higher level will we be able to carry out experimental and correlational studies that will yield results not only statistically significant but psychologically meaningful and systematically coherent." This observation probably still holds today.

Dimensions: We believe that evaluation for research can be facilitated by designing an evaluation system which is nomothetic in nature, experimental in design, comparative in method, variable in time, and public in distribution. The following is an elaboration of each of these suggested dimensions.

1. Nature of Data: Nomothetic. Evaluation research is involved in the quest for laws, that is, statements of relationships among two or more variables.[73] Nomothetic inquiry seeks to establish logical linkages between the conceptual framework of the evaluation research problem and the operational definitions of the concepts.

2. Level of Specificity: Detailed. It has been proposed that the application of theoretically derived variables be in small, sharpened, definitive units.[74] Presently, most researchers are utilizing broad variables such as lecture vs. discussion method and seminar vs. lab classroom. If progress is to be made toward understanding the teaching-learning process, these variables must be broken down into smaller components. A taxonomic description of the physical setting of the classroom is one example.

3. Method of Reporting: Comparative. Perhaps the highest correlate between the four purposes of evaluation is the generalizability of the results of the phenomenon being studied. Evaluation research must go beyond merely a comparison of instructors on one campus, at one point in time. An evaluation researcher investigating teacher characteristics should attempt to design, collect, and report his study such that the results are not specific to the

term or year it is conducted; the conclusions drawn should strive toward generalizability over time, geography, and population.

4. Timing: Variable. Results can be collected at regular intervals or intermittantly. It depends entirely upon the design of the study and the nature of the phenomenon studied. If usable indicators of at least intermediate success are possible to come by early in the study, the information should be collected.

5. Audience: Public. The need for dissemination of evaluation research results to researchers and educators is essential and unquestionable. Whereas students are interested in only information within their institution, and faculty are not interested in having their linen washed in public, researchers must make their generalizable conclusions public knowledge if progress is to be made. Not all evaluation research reports possess publishable worth. However, at present a substantial number of studies go unpublished because evaluators are too pressed for time or discouraged with compromises made in research design.

6. Additional Characteristic: Criteria for Judging. Important criteria for judging the adequacy of evaluation research are internal and external validity. One might choose "credibility" as an additional criterion to test the worth of evaluations for administrative decision-making or student information.

Comparison of Purposes

Table 1 brings the four purposes back into a common perspective in terms of definitions, delimitations, and dimensions. With specific regard to the dimensions in the table, it should be borne in mind that each entry is, in a way, an answer to specific questions asked about each dimension for each purpose. The following six questions produced six entries for each purpose:

1. Is the evaluation to be primarily descriptive, judgmental, diagnostic, or nomothetic?
2. Is the evaluation to emphasize summary or detailed information, that is, what is the desired level of specificity?
3. Should the evaluation data be comparative, noncomparative, or both?
4. When should the evaluation be performed?
5. Who are the primary beneficiaries of the results?
6. What additional characteristics distinguish one purpose from another?

An intra-dimensional comparison of entries across purposes and an inter-dimensional holistic view of a given purpose can be useful to further study, organize and design evaluation of teaching. The following are a selected number of intra-dimensional entries across purposes.

There is a basic similarity between dimensions of evaluation for decision-making and for student information. Summary data gathered at the end of the course (or any other unit of study) are suitable for both purposes. Some of the efforts expended on each of the two purposes can thus be combined. In suggesting this similarity we do not overlook divergent dimensions across these purposes such as those of audience and the nature of data, which are dissimilar. Since some dimensions of purposes overlap and others do not, a fundamental implication for designing instruments is that they should be uniquely suitable for each separate purpose.[75]

Table 1

Purposes of Evaluation of Teaching and Their Respective Components

Components of Purposes	Purposes			
	Self-Improvement	Administrative Decision-Making	Student Information	Research on Teaching
Definition	To acquire and diagnose information to improve teaching competency	To provide an information base for administrative decisions on personnel, assisgnments, and allocation	To provide evaluative information for enlightened consumption and participation in education	Evaluation for purpose of research Research for purpose of evaluation
Delineation	Perceptions of the nature Stages of development Motivation	Personnel decisions Assignment decisions Allocation decisions	Course characteristics Instructor characteristics	Research and evaluation Areas in need of research Styles of research
Dimensions:				
1. Nature of Data	Diagnostic	Judgmental	Descriptive	Nomothetic
2. Level of Spec.	Detailed	Summary	Summary	Detailed
3. Method of Reporting	Noncomparative	Semi-Comparative	Comparative	Comparative
4. Timing	Continuous	End of Term	End of Term	Variable
5. Audience	Instructor	Instructor and Administrator	Students	Public
6. Additional Characteristic	Collaborative	Corrective	Systematic	Criteria/Judging

Other observations are pertinent to all four purposes. First, while dimensions do differ across purposes, it is not imperative that each one of the dimensions be operationalized only according to the unique specifications suggested in Table 1. A comparative forum of reporting may be used, for instance, for or during the improvement stages of faculty development, as noncomparative reporting may be used for administrative decision-making or for student

information. Second, degrees of difference exist within each dimension category. Even though the method of reporting for student information and for research are both labeled as comparative, there are great differences in the generalizability potential between the two. Third, while some dimensions are interdependent, other dimensions may be able to stand alone. Fourth, the choices one makes as to the degree and specificity of the dimension to be utilized depends upon the kinds of information needed and the instrumentation avilable.

The rest of this section is an inter-dimensional holistic view of each purpose. Such a view may suggest some general and generic types of models that could be applicable for further organizing and developing each purpose.

Specific evaluation models have already been developed and discussed in the past.[76] Sometimes specific models may be used intact for an evaluative purpose, but more often only particular portions may be utilized.[77] We do not wish to elaborate on the specifics of each one of the models which have emerged or measure their applicability to each of the purposes. We only wish to identify and draw, with a general stroke, four such models and to suggest that each of them may be appropriate to one of our four purposes.

The value of suggesting models at this point or at any time is certainly open to question. However, a most unsatisfactory feature of the evaluation being done at the moment is the lack of a common basic framework.[78] Systematic models are needed which allow purposes and their dimensions to be integrated. It is through this conceptualization process that models frequently help to formalize the complex process of evaluation. We contend that much can be gained from such an activity, for, in the past, models have proven to be useful. They have provided a starting point, a precursor for further discussion and organization by placing dimensions into action and into a system and a sensible whole. They have assisted in examining relationships between dimensions. They have simplified communication with colleagues and sponsors by working within established guidelines. They have aided in the planning, implementation, and evaluation of the total evaluation program, and they have given new directions toward applications, identifying problems, and future study in the field.

Given the benefits of providing models for evaluation purposes, we have identified commonalities between four general models and the four particular frameworks of the purposes. We suggest that models and purposes similar in intent and characteristics are:
1. The formative evaluation model for the purpose of improvement of teachers.
2. The summative evaluation model for the purpose of administrative decision-making.
3. The information choice models for the purpose of student advising.
4. The research model for the purpose of evaluation research. Some details follow.

Several common elements exist between the *purpose of teacher improvement* and the *formative evaluation models*. The formative model provides feedback and correctives; it is developmental in nature; it emphasizes noncomparative data; it is conducted during the unit of instruction.[79] Each of these components coincides with the dimensions of the purpose: the nature of data, level of specificity, method of reporting, and timing. Formative evaluation has already been suggested as having great positive effects on motivation, self-concept, and all other internal needs congruent with the development of faculty's teaching ability.

The particular framework of *evaluation for administrative decision-making* can be supported by including it in the general *summative evaluation models*. Guidelines for these

models suggest that they be comparative, conducted at the end of the unit, and judgmental by providing information for continuation or termination of a practice or an individual.[80] These components seem to be congruent with dimensions of the purpose of administrative decision-making.

The basic premise behind *providing information to students* is that the information will be more than a mere collection of facts and data; that it will be organized to serve the purpose of course and instructor selection. In the same light, *information-choice models* serve to differentiate the alternatives involved in the decision situation.[81] An information-choice model is a means for reducing the uncertainty of the decision, in much the same way that information for students on classes and instructors should serve to aid the student in selection. Essentially, all information systems have sources ("systematic" dimensions in the purpose) and formats (dimensions related to a method of reporting, timing, and audience). The approach to the design of an information model includes delineation of information, obtaining the information, and providing information. The first component defines the system, provides statements of evaluation policies, and articulates the evaluation assumptions. The second component specifies the collection, organization, and analysis of data. The third component specifies the preparation and dissemination of reports. These components are keys to the success of a usable information system for students.

Evaluation for the *purpose of improving research on teaching* may be facilitated by the use of *general research models*. Researchers have forwarded schemes for classifying types of research activities.[82] Evaluators have proposed types of evaluation.[83] Although, as we have already seen, there are differences among the specific types of research and evaluation models, common ingredients can be easily identified between the evaluation-research purpose and the basic research model. Both seek to produce previously unavailable knowledge. Perhaps the highest correlate of the research-evaluation combination is the generalizability of the phenomenon over time, geography, and population.

A Concluding Statement

We began the search for purposes of evaluation by linking a purpose to components of effective teaching (scholarship, delivery and advising) and by arguing that original functions of evaluation have been to improve and safeguard instruction. We suggested that the search for purposes of evaluation of instructors should follow the identification of audiences for whom the results of the evaluation would be relevant. We then defined four purposes (instructional improvement, administrative decision-making, information for students, and research), delineated them and offered dimensional features which described them (nature of data, level of specificity, method of reporting, timing, audience and other characteristics). We argued that these features would be helpful in the actual design of evaluation instruments and procedures. In an attempt to show additional implications of this effort, we first compared some interdimensional entries across purposes and then offered holistic views of each purpose which could be applicable to a more detailed operationalization of purposes as guides for both instrument designs and use of results.

Footnotes

1. A most recent overview of some aspects of these controversies is N.L. Gage, "Student Ratings of College Teaching: Their Justification and Proper Use," and "Reactions" and "Discussions" in Naftaly S. Glasman and Berthold R. Killait (eds.) *Second UCSB Conference on Effective Teachers* (Santa Barbara, California: Regents of the University of California, 1974) Chapter 4.

2. A. Morrison and D. McIntyre, *Teachers and Teaching* (Middlesex, England: Penguin Books, Inc., 1969), p. 13.

3. Kenneth E. Eble, *The Recognition and Evaluation of Teaching* (Washington, D.C.: American Association of University Professors, 1971).

4. See, A.S. Barr, "The Appraisal of Teaching in Large Universities: A Summary of Remarks," *The Appraisal of Teaching in Large Universities* (Ann Arbor: University of Michigan, 1959); Paul L. Dressel, "Evaluation of the Environment, the Process, and Results of Higher Education," in Asa S. Knowles, (Ed.) *Handbook of College and University Administration-Academic* (New York: McGraw-Hill, 1970) Section 2, Chapter 4, pp. 253-280; and Richard I. Miller, *Developing Programs for Faculty Evaluation* (San Francisco: Jossey-Bass Pub., 1974).

5. Paul L. Dressel, "The Current Status of Research on College and University Teaching," *The Appraisal of Teaching in Large Universites* (Ann Arbor: The University of Michigan, 1959).

6. Dressel, 1970, *op. cit.*

7. Milton Hildebrand, Robert Wilson and E. Dienst, *Evaluating University Teaching* (Berkeley, California: University of California, Center for Research and Development in Higher Education, 1971).

8. Dressel, 1970, *op. cit.*

9. Dale L. Bolton, *Selection and Evaluation of Teachers* (Berkeley, California: McCutchan Pub. Co., 1973).

10. Eble, *op. cit.*

11. N. Sanford, "Academic Culture and the Teacher's Development," *Soundings,* Vol. LIV, No. 4, Winter 1971.

12. See, W. J. McKeachie, "Student Ratings of Faculty," *American Association of University Professors Bulletin,* Vol. 55, (1969), pp. 439-444; Eble, *op. cit.;* F. Costin, W. T. Greenough and R. J. Menges, "Student Ratings of College Teaching: Reliability, Validity, and Usefulness," *Review of Educational Research,* Vol. 41 (1971), pp. 511-535; R. Smock and T. Crooks, "A Plan for the Comprehensive Evaluation of College Teaching," paper presented at the annual meetings of the America Educational Research Association, New Orleans, Louisiana, 1973; and Gage in Glasman and Killait, *op. cit.*

13. Naftaly S. Glasman and others, *Evaluation of Instructors in Higher Education: Concepts, Research and Development* (Santa Barbara, California: Regents of the University of California, 1974).

14. T. G. Gaff and R. C. Wilson, "Faculty Values and Improving Teaching," *New Teaching New Learning* (San Francisco: Jossey-Bass, 1971).

15. Gage, *op. cit.*

16. Smock and Crooks, *op. cit.*

17. Glasman and others, 1974, *op. cit.*

18. Gage, *op. cit.*

19. C. H. Weiss, *Evaluation Research: Methods for Assessing Program Effectivess* (Englewood Cliffs, New Jersey: Prentice-Hall, 1969).

20. Bolton, *op. cit.*

21. Naftaly S. Glasman, "Personnel Evaluation Research & Implications for Instructional Improvement," *Canadian Administrator,* Vol. 13, No. 6, (March 1974), pp. 29-33.

22. Miller, *op. cit.*

23. Smock and Crooks, *op. cit.*

24. A. Astin and C. Lee, "Current Practices in the Evaluation and Training of College Teachers," in Calvin Lee (ed.) *Improving College Teaching* (Washington, D.C.: American Council on Education, 1967).

25. Benjamin S. Bloom, J. T. Hastings and G. F. Madaus, *Handbook on Formative and Summative Evaluation of Student Learning* (New York: McGraw-Hill, 1971).

26. Weiss, *op. cit.*

27. D. Musella, *Teaching Effectiveness Research: Another Approach* (Albany, New York: Council for Administrative Leadership, March 1966).

28. N. L. Gage, "The Appraisal of College Teaching: An Analysis of Ends and Means," *Journal of Higher Education,* Vol. 32, (January 1961), pp. 17-22; Gage, 1974, *op. cit.;* Eble, *op. cit.;* Smock and Crooks, *op. cit.;* Richard I. Miller, *Evaluating Faculty Performance* (San Francisco-Jossey-Bass, 1972); and Miller, 1974, *op. cit.*

29. Glasman and others, 1974, *op. cit.,* chapters 4 and 14.

30. Eble, *op. cit.*

31. Paul J. Munson, "So You Want to Try Faculty Development? ", A paper delivered at the annual convention of the American Educational Research Association, Washington, D.C., April, 1975.

32. Sanford, *op. cit.*

33. G. Lippitt, *Organizational Renewal.* (New York: Appleton-Century Crofts, 1969).

34. See Abraham H. Maslow, *Motivation and Personality* (New York: Harper and Row, Pub., 1954), and Chris Argyris, *Integrating the Individual and the Organization* (New York: John Wiley & Sons, Inc., 1964).

35. David C. McClelland, T. W. Atkinson, R. A. Clark, and E. L. Lowell, *The Achievement Motive* (New York: Appleton-Century-Crofts, 1953).

36. J. W. Gardner, *Self-Renewal* (New York: Harper and Row, 1963).

37. Michael Scriven, "Evaluating Higher Education in California" in *Master Plan for Higher Education* (Sacramento, California: California Legislature, 1973).

38. Eble, *op. cit.*

39. See, N. Sanford (ed.), *The American College* (New York: John Wiley and Sons, Inc., 1962); T. Parsons and G. M. Platt, *The American Academic Professions: A Pilot Study* (Cambridge: Harvard University, 1968); Hildebrant, Wilson and Dienst, *op. cit;* and Lewis B. Mayhew, *The Literature of Higher Education* (San Francisco: Jossey Bass, 1971).

40. B. M. Anthony, "A New Approach to Merit Rating of Teachers," *Administrator's Notebook,* 1968; Naftaly S. Glasman, "Subject Matter, Approach and the Reward System," in Naftaly S. Glasman and Stanley J. Nicholson (eds.), *First UCSB Conference on Effective Teaching* (Santa Barbara, California: Regents of the University of California, 1973, pp. 86-91; Naftaly S. Glasman, "Merit Pay: A Case Study in a California School District," *Instructional Science,* Vol. 3 (April, 1974), pp. 88-110.

41. An elaboration of such resources can be found in Glasman and others, 1974, *op. cit.,* chapters 8, 9, and 10.

42. Gage, in Glasman and Killait, *op. cit.,* and Glasman and others, 1974 *op. cit.,* chapter 14.

43. James D. McNeil and James W. Popham, "The Assessment of Teacher Competency," *Second Handbook on Research of Teaching* (Washington, D.C.: American Educational Research Association, 1973).

44. Smock and Crooks, *op. cit.,* and Gage, in Glasman and Killait, *op. cit.*

45. Smock and Crooks, *op. cit.*

46. J. A. Johnson, "Instruction: From the Consumer's View," *Improving College Teaching* (Washington, D.C.: American Council on Education, 1967).

47. Smock and Crooks, *op. cit.*

48. John A. Centra, "The Student as Godfather? The Impact of Student Ratings on Academia," *Educational Research,* Vol. 2, No. 10 (October 1973).

49. Gage, 1961, *op. cit.*

50. John W. Gustad, "Evaluation of Teaching Performance: Issues and Possibilities," *Improving College Teaching* (Washington, D.C.: The American Council on Education, 1967).

51. Dressel, 1970, *op. cit.*

52. McNeil and Popham, *op. cit.*

53. W. J. McKeachie, "Research on Teaching: The Gap Between Practice and Theory," *Improving College Teaching* (Washington, D.C.: American Council on Education, 1967).

54. D. D. O'Dowd, "Closing the Gap," *Improving College Teaching* (Washington, D.C.: American Council on Education, 1967).

55. B. R. Wortham and J. R. Sanders, *Educational Evaluation: Theory and Practice* (Worthington, Ohio: Charles Jones, Pub., 1973).

56. J. K. Hemphill, "The Relationships Between Research and Evaluation," *Educational Evaluation: New Roles, New Means* (Chicago: National Society for the Study of Education, 1969), pp. 189-220.

57. Weiss, *op. cit.*

58. Worthen and Sanders, *op. cit.*

59. Robert Stake and D. Gooler, "Measuring Educational Priorities," *Educational Technology,* Vol. 11, No. 9 (September, 1971), pp. 44-48.

60. Worthen and Sanders, *op. cit.*

61. Weiss, *op. cit.*

62. Campbell and Stanley, *Experimental and Quasi Experimental Designs for Research* (Chicago: Rand-McNally, 1963).

63. See, D. B. Stuit, "Needed Research on the Evaluation of Instructional Effectiveness," *The Appraisal of Teaching in Large Universities* (Ann Arbor: University of Michigan, 1959); J. W. Trent and A. M. Cohen, "Research on Teaching in Higher Education," *Second Handbook on Research of Teaching* (Washington, D.C.: American Educational Research Association, 1973); and McNeil and Popham, *op. cit.*

64. Naftaly S. Glasman, "Faculty Perceptions on Teaching, Teaching Needs, Evaluation and Evaluation Needs," in Glasman and Killait, *op. cit.,* pp. 26-51.

65. See, W. J. Popham, "The Performance Test: A New Approach to the Assessment of Teaching Proficiency," *Journal of Teacher Education,* Vol. 19 (1968), pp. 216-222; A. M. Cohen and W. F. Shawl "Coordinating Instruction Through Objectives," *Junior College Journal,* Vol. 41, No. 2 (October, 1970), pp. 17-19; and C. A. Rose, "The Development of Precision Instruments for Assessing Teacher Effectiveness," Unpublished Master's Thesis, University of California, Los Angeles, 1971.

66. Trent and Cohen, *op. cit.*

67. McNeil and Popham, *op. cit.*

68. See, *Ibid.;* Trent and Cohen, *op. cit.;* R. Sommer, *Personal Space* (New York: Prentice-Hall, 1969).

69. Anthony, *op. cit.*

70. Gage, 1961, *op. cit.*

71. See, *Ibid.,* and Trent and Cohen, *op. cit.*

72. Gage, 1961, *op. cit.,* p. 62.

73. Worthen and Sanders, *op. cit.*

74. See, Gage, 1961, *op. cit.,* and Trent and Cohen, *op, cit.*

75. Glasman and others, 1974, chapters 14 and 15.

76. Examples include Campbell and Stanley, *op. cit.;* Lee J. Cronbach, "Course Improvement Through Evaluation," *Teachers College Record* (1963-64), pp. 673-683; Scriven, *op. cit.,* R. C. Anderson, *Current Research on Instruction* (Englewood Cliffs, New Jersey: Prentice-Hall, 1969). G. Caro, *Readings in Evaluation Research* (New York: Russel Sage Foundation, 1971); Stake and Gooler, *op. cit.;* and Worthen and Sanders, *op. cit.*

77. D. A. Payne (ed.), *Curriculum Evaluation* (Lexington, Mass: D. C. Heath & Co., 1974).

78. Scriven, *op. cit.*

79. See, *Ibid.;* Cronbach, *op. cit.;* Benjamin S. Bloom, "Some Theoretical Issues Relating to Education Evaluation," *Educational Evaluation: New Roles, New Means* (Chicago: National Society for the Study

of Education, 1969); Bloom, Hastings and Madaus, *op. cit.;* and Michael G. Saslow, "Establishing the Purpose for Evaluation," *A Strategy for Evaluation Design* (Oregon State University: Teaching Research, 1970).

80. See, Saslow, *op. cit.;* Scriben, *op. cit.;* Bloom, *op. cit.;* Stake and Gooler, *op. cit.;* and Payne, *op. cit.*

81. Daniel L. Stufflebeam and others, *Educational Evaluation and Decision Making* (Itasca, Illinois: F. S. Peacock, Inc., 1971).

82. Examples include, T. Hillway, *Introduction to Research* (Boston: Houghton Mifflin, 1964; A. J. Golfo and E. Miller, *Interpreting Educational Research* (Dubuque, Iowa: Wm. C. Brown, 1970); and Fred W. Kerlinger, *Foundations of Behavioral Research* (New York: Holt, Rinehart and Winston, Inc., 1974).

83. Worthen and Sanders, *op. cit.*