A SURVEY AND CRITIQUE OF METHODS AND MODELS OF FACULTY AND PROGRAM EVALUATIONS

JERRY ABERNATHY MERIWETHER

A SURVEY AND CRITIQUE OF METHODS AND MODELS OF FACULTY AND PROGRAM EVALUATIONS

A Research Paper

Presented to the

Graduate and Research Council of

Austin Peay State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in Psychology

---

by

Jerry Abernathy Meriwether

May 1984

To the Graduate Council:

I am submitting herewith a Research Paper written by Jerry Abernathy Meriwether entitled "A Survey and Critique of Methods and Models of Faculty and Program Evaluations." I have examined the final copy of this paper in form and content, and I recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Arts, with a major in Psychology.

Linda Rudolph
Major Professor

Accepted for the Graduate and Research Council:

William H. Ellis
Dean of the Graduate School

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

In order to make improvements, educational institutions have been required to demonstrate accountability while providing a basis for the development of understanding the relationship between available resources and quality in education. Accountability provides the institution with the ability to verify the performance of faculty and staff, and at the same time, to evaluate the extent to which a given program meets its objectives and goals.

Accountability in education seeks the answers to three questions:

1. What is happening in the program?
2. How much does the program cost?
3. Is the program effective?

Evaluation is a primary component of accountability. The main purposes of evaluation according to Tolbert (1978) are (a) to judge the effectiveness of a program, (b) to strengthen weak functions, (c) to revise weak elements, (d) to obtain financial support, (e) to justify continuation of a program or approach, and (f) to assemble information for public relations. Evaluation also enables the evaluator to compare the costs and results of a program, and can suggest methods or procedures to achieve the same goals in a more cost-effective manner. Evaluations completed over an extended period of time serve to give cumulative results which summarize existing programs and make suggestions for improvements or revisions of programs being evaluated.

Although one of the most important forms of accountability in an educational program is the evaluation of that program, evaluation of faculty also serves a multitude of purposes. Faculty may be reviewed

and evaluated as a part of the personnel procedures for promotion, tenure, salary increases, or as part of planning for program improvement. In this paper different forms of evaluation are explored. Some methods of evaluation are extremely accurate and progressive while other forms cling to outmoded methods such as type of clothing worn by the faculty member, attitude toward political parties, and even what type and how much food a faculty member contributed to the faculty parties.

Objective evaluation can be a useful tool in the upward movement of education, but if misused it may become a destructive weapon. Suggestions will be offered for correct evaluative procedures which should be used in both faculty and program evaluations. Suggested references for extended research into the evaluative procedure are included.

CHAPTER II

HISTORY OF THE EVALUATIVE PROCESS

The use of evaluative procedures to assess educational systems and programs has increased rapidly in the last 50 years. The Department of Supervisors and Directors of Instruction of the National Education Association (NEA) devised a self-evaluation checklist for supervisors which was presented in their fourth yearbook (Woody, 1931). The checklist was one of the first attempts to evaluate education. Its main purpose was to assess the supervisors of pupils, teachers, the community, and other supervisors. The self-evaluation, like so many later ones, expected the supervisors to have a good opinion of themselves as well as their work. The scale asked the supervisors (all of whom were men) to tell how they felt about their own work, but the scale did not actually survey the people being supervised or representatives of the community.

Rating scales were improved in 1949 when, for the first time, teachers' personal and educational backgrounds and concepts were explored (Beecher, 1949) and when the Association for Supervision and Curriculum Development of the NEA once more presented guidelines for program evaluation. The Association's guidelines stressed "how to use evaluation as a positive force toward better teaching, better learning, and better balanced curriculum" (Wilhelms, 1967, p. vii).

A major change in program evaluation occurred with the implementation of the Elementary and Secondary Act of 1965, Title I program, designed by Senator Robert Kennedy. This act was one of the first major efforts directed toward social legislation to mandate project reporting. More than $1 billion annually was allocated to the special education needs of disadvantaged children. Kennedy felt that a thorough program evaluation

would provide parents with a method of insuring that the funds in the Title I program would be spent in the most effective interest of disadvantaged children (McLaughlin, 1975).

William Gorham, who was brought to the Department of Health, Education and Welfare to install management principles, contended that through Title I there could be reform of Federal management of education programs which could then be shown in cost-benefit terms. More than $52 million was spent in reform and evaluation, but the program was termed a failure due to lack of cooperation. Both Kennedy and Gorham assumed that the policies would not only be self-executing, but that the reporting requirements would generate useful information. The primary cause of the program's failure was seen to be the lack of incentives for the school districts to collect or report data on their own progress (or lack of progress), and the fact that Federal force was not strong enough to enforce the guidelines or to encourage cooperation with other Federal forms of evaluation. Indications were that the reformers underestimated the resistance of the educational community to the evaluation and changes that are an inherent part of our educational system. The more than 30,000 Title I Federal projects under contract at the time reflected multiple goals and treatments, but these were not easily transformed into measurable objectives. McLaughlin (1975) concluded that no matter how good the intentions, conflict and probable failure in any form of evaluation and reform will result if the evaluation starts at the Federal rather than at the local level. McLaughlin indicated that not only the evaluation but also the interest and desire for evaluation must begin at the local level.

In their evaluation of both state and federally funded public

programs, Wholey, Scanlon, Duffy, Fukumato, and Vogt (1970) attempt to delineate the function of evaluation in both public and private programs:

> Evaluation (1) assesses the effectiveness of an ongoing program in achieving its objective, (2) relies on the principles of research design to distinguish a program's effects from those other forces working in a situation, and (3) aims at program improvement through modification of current operations. (Wholey, et. al. 1970, p. 25)

A symposium covering five areas of evaluation was conducted on the UCLA campus December 13-15, 1967, and was sponsored by the UCLA Research and Development Center for the Study of Evaluation and by the Ford Foundation's Fund for the Advancement of Education. Martin Trow, an expert in evaluative research, was one of the more prominent speakers at the 1967 symposium. The topics discussed at the symposium were: (a) theory of evaluation, (b) instructional variables, (c) contextual variables, (d) criteria of instruction, and (e) methodology of evaluation. Trow stated that "a large part of evaluation in education is best understood as a form of persuasion directed at powerful people who make decisions and control resources." He further contended that "the context and function of such studies affect the way we conduct studies and how much confidence we can place on the findings, and thus are deeply indicated in methodology" (cited in Wittrock, 1967, p. 293).

Trow believes that some colleges as well as some professors see social research or any type of evaluation as a threat to the intelligence of the instructors. These instructors, according to Trow, rely on committee deliberations rather than rational procedures and data. Other more progressive liberal institutions whose goals can easily be measured sponsor and apply all available social research methodology and

procedures to ensure objective, unbiased decision-making in personnel procedures and allocation of resources.

Trow further indicated that even liberal education programs are not immune to biased or inaccurate research information. Trow found the reliability of evaluative procedures and the cooperation of the faculty and staff may be affected by institutional policies and the attitudes of some faculty members.

No matter which evaluation instrument is being used, according to Trow, the farther the research is from those who make decisions the more receptive the decision-makers are to the process of research or evaluation; therefore, research on student dorm life becomes very popular while faculty or administrative evaluations become very unpopular. Program evaluations seem to lie somewhere in between (Wittrock, 1967).

Trow suggests that with the steady increase in the percentage of the population attending universities has come an unforeseen problem. The academic quality of the entrants to the more select universities has risen, while the less selective universities have a large percentage of students attending because they have nothing else to do. Large numbers of these students, in Trow's opinion, are unmotivated and have no strong academic traditions. This places the burden of generating interest in the subject on the classroom teacher. Trow believes that now curricula must be related to the lives and interests of the students rather than to the specifics of academic discipline. He contends that this factor has forced a change in the form and content of academic instruction at the introductory and undergraduate levels. The curriculum changes, as well as other forces, have begun a fermenting process for evaluation of existing techniques, procedures, and the traditional syllabi.

Popham (1973) presented a guide for use by the general public which specified to whom teachers should be accountable and how these teachers were to be evaluated. Popham believes that learner growth cannot be associated with any single teaching behavior or method and contends that identical results can be achieved by using completely different techniques. Popham also presented guidelines for developing useful evaluations. He concluded that observational techniques and procedures can be used for evaluations, but cautioned that these outdated procedures focused on instructional progress rather than instructional outcome. Additionally, Popham believes that the judgments of most educational raters of teacher performance are often wrong because of the outmoded methods being used and the lack of substantial evidence to support the opinions of what these raters consider to be a "good teacher."

Grobman (1968) encouraged discussion between teachers, parents and students concerning teaching techniques and curricula at the lower levels; however, she concluded this is not always possible at the higher levels. Still, Grobman contends that no avenue or source should be eliminated if a valid evaluation is desired.

Smock, as quoted in Perlberg (1979), as well as Grobman, considers students one of the most valuable sources of information concerning actual performance in the classroom, and urges that this evaluative source be tapped as soon as possible. Smock suggests several questionnaires for assessing student opinion (see Appendix I).

One of the newest forms of educational evaluation is direct student evaluation of both teachers and programs. Since this method is often used for promotion or salary merit raises, an increase in reliance of students as raters in universities has accompanied an overall increase in

evaluation. These instruments, as well as all instruments of evaluation, should not be used to the exclusion of others, but as a part of the total data triangulation according to most writers.

CHAPTER III

THE EVALUATOR

Often the evaluator is as important as the person or program being evaluated. Johnson (1979) thinks that the role of an evaluator is defined by the decision that is anticipated as a consequence of the evaluation. Additionally, Johnson stated that "since this consequential decision makes use of the results of an evaluation, it is not part of the evaluation itself. If an evaluator makes the consequential decision, he or she does so not as an evaluator but in some other capacity. An evaluator may make recommendations regarding action, especially if asked to do so, but the action decision itself is the responsibility of someone else" (Johnson 1979, p. 122).

Johnson describes the role of an evaluator and the evaluation process as follows:

1. a judgment of the inherent or instrumental worth
2. of some educational entity or process (evalund)
3. for the purpose of enlightening an anticipated decision-making process
4. arrived at by establishing explicit absolute or relative standards
5. pertaining to relevant criteria or attributes of the evalund
6. that have been weighed in accordance with their perceived contribution to the evalund's overall worth
7. and applying the standards, according to appropriate rules
8. to a full and accurate description of the evalund
9. based on reliable observation pertinent to the criteria (Johnson, 1979, p. 122).

In contrast, Rippy (1979) states that "the idea of the evaluator, in the name of objectivity sitting and doing nothing while a project goes down the chute does not appeal to me" (p. 133). Rippy also refers to TNR (The New Rhetoric, Perelman and Olbrechts-Tyteca, 1969) in which he found 102 new terms with which "we can all write evaluation reports which absolutely no one will understand." Rippy hints that most evaluators try to be inprecise so that they cannot be held responsible for the outcome of the evaluation. In the same article, Rippy praises TNR and its authors for "delineation of the differences among impartiality, objectivity, partisanship, fanaticism, and skepticism" (Rippy, 1979, p. 134).

Reichardt and Cook (1980) emphasize the flexibility that evaluators must have in choosing methods for evaluation. They argue that not only should the quantitative methods of the past be used, but also many of the more modern qualitative methods available should be included. Multiple-method use is encouraged whenever resources allow in order to receive the benefits of the strengths of each method.

Ostrander, Goldstein, and Hull (1978) emphasize the importance of evaluator independence. This includes job security as well as political independence. Power access appears to be the key to evaluator independence; the less an evaluator must depend on others, the more power that the evaluator will have at his/her command. It appears that along with power comes respectability for project results.

Editors of the Journal of Educational Evaluation and Policy Analysis (1981) invited Cooley and Kean to submit separate "counterpoint" articles on the objectivity and subjectivity of evaluators respectively. Cooley contends that the most important part of evaluation is the

dialogue between evaluator and client. Cooley argues that the better a client understands the kinds of information that can be gathered and how it may be used, the better the evaluator will understand what type of information the client needs. Cooley suggests that the evaluator and client plan the studies and consider the results together.

Kean argues that objectivity must remain stringent, but possibly through neutrality. Kean states "neutrality refers to the state of being not aligned with a particular political or idealogical group. . . . Objectivity, on the other hand, refers to the state of independence of mind. An objective individual is able to use facts without distorting them or allowing personal feelings or prejudices to interfere with her or his power of reason" (p. 87). Kean believes that although some compromises may be necessary, it is the wise evaluator who provides the client with an opportunity to save face. Kean suggests that should a client ask the evaluator to alter the information gathered, it might be explained that if an evaluator will lie <u>for</u> the client, then the client cannot be certain that the evaluator is not lying <u>to</u> her or him.

Fathy (1980), an Assistant Professor of Extension Education at the University of Alexandria, Egypt, offers criteria that evaluators should consider in making their evaluations. Fathy states that "adults evaluate according to their own criteria, which depend on their desires, interests, and surroundings. . . . Besides being skillful in collecting information, the evaluator needs to be effective in human relations -- possess insight into self. Some of the skills related to this area of competency are those pertaining to recognizing one's own motivations, interests and needs; realizing one's own biases; avoiding wishful thinking; and being critical of one's own interpretations. . . .

However . . . if the evaluator does permit his values and experience to affect conclusions, he should make it clear to all potential users of the evaluation that this has occurred" (pp. 16-17).

CHAPTER IV

FACULTY EVALUATIONS

The primary purpose of faculty evaluation usually is for the assessment of a faculty member's performance, improvement, and/or completion of instructional goals. However, Dressel (1960), as quoted in Perlberg (1979), argues that " . . . all too frequently the attempt to evaluate teaching in connection with promotion and financial increments ends up being nothing more than an attempt to find out whether an individual's teaching is so bad that promotion must be refused" (p. 143).

The most pressing question seems to be "who is qualified to evaluate faculty?" Menges (1973) is quoted by Perlberg (1979) as arguing that students should be considered as reporters rather than evaluators. Perlberg considers them interested observers rather than judges. Further, he states that "even though students and teachers seem to be the only direct source of information on what occurs in the college classroom, teachers seem to be less accurate reporters than their students" (Perlberg, 1979, p. 145).

Perlberg (1979) argues that "many professors believe that intrusion into their classroom by others, except when invited, violates their academic freedom. However, very often, the concept of academic freedom becomes a license which permits the professor to teach as badly as he wishes" (p. 146). Educational consultants and video taping are two of the alternatives suggested by Perlberg as means of instructional critique and improvement.

Perlberg thinks that "evaluation of instruction by students . . .

should be voluntary and geared mainly to the purpose of training, development, and improvement. It should be the professor's prerogative to submit the results of this evaluation to the administration for the purpose of reward, promotion, tenure and accountability" (p. 147).

Perlberg also contends that if teaching is honored on our campuses, it will be cultivated there and will finally be done well there. If it does not find honor, expressed in the respect and prestige granted the teacher by his/her colleagues and by the dollars paid him or her by the comptroller, it is not likely to be cultivated nor, he argues, to improve. He describes the role that student-organized evaluation teams, which have appeared on some campuses, have served in the evaluation process. Not only did these organizations publish a "customer report" concerning faculty and their quality of instruction, but according to Perlberg the reports were considered by the administrations for salary increases, tenure, and promotions. This so infuriated the faculties that some of the members "demanded" formal institutional evaluation.

The problems that arise with negative feedback from student reports are a dilemma to the faculty member. Poor teaching methods may be criticized, but innovative or creative instructors do not always receive good student ratings (Perlberg, 1979). Perlberg quotes Smock and Crooks (1973) who state that "any evaluation plan which does not include an adjunct service which can effectively assist faculty members in the improvement of their performance is ethically questionable" (Perlberg, 1979, p. 151).

According to Perlberg, faculty evaluation is only the first step toward development and change in education. He argues that in the process "we might fall into the trap of being happy with the large

numbers of faculty members that are being evaluated by students without realizing that this is only the first step in a faculty development program of a long and tedious process towards improvement" (Perlberg, 1979, p. 154).

In their contention that tenure and promotion methods seem to be in trouble and in need of an overhaul, Hoyle and Klewer (1980) have devised the HEAR Model which assumes that different responsibilities and expectations are required of faculty in all disciplines. Hoyle and Klewer agree that "faculty within each professional or liberal arts discipline should be charged with the role of determining the process and criteria for faculty evaluations" (Hoyle and Klewer, 1980, p. 67).

The acronym HEAR stands for four different evaluation components used to determine the difference between the expected performance and the actual performance of the faculty. The H stands for "how" and upon what criteria are faculty evaluated at this time. Problems must be identified and information gathered to establish priorities and satisfy the unmet needs of the present program. Sources of information include faculty, evaluators, senior faculty, administrators, current literature and experts.

The E of HEAR stands for "essential ingredients for effective faculty evaluation." This is determined by groups of administrators and faculty. "Each respective rank . . . instructors, assistant and associate professors, must accept the task to carefully list and prioritize, according to importance, the items and administrative processes upon which the evaluation will be made" (p. 67). Hoyle and Klewer contend that the E step is a critical component to the plan.

The step designated as A "stands for assessment, which determines

the status of the progress being made toward the predetermined criteria established in component E . . . This quality control "check" is vital in order to help the professor and the evaluator analyze and overcome differences while time remains to alter or improve in a deficient area. These precautions will help prevent the professor from creating too wide a discrepancy between predetermined standards or criteria and his actual performance" (p. 69).

The R component stands for review. During this phase, the evaluator determines how closely the mutually determined criteria parallel the accomplishments performed by the professor. "Using the comparison made between expectations and actual results of faculty performance, the evaluator can decide whether to suggest alternative criteria, alter existing criteria and evaluation methods, or recommend the professor for promotion or termination" (p. 69).

Hoyle and Klewer state that the "plan requires a commitment from those in charge of the faculty evaluation program in order to accept the collective effort of faculty for self-determination. . . . The HEAR plan is designed to help faculty and administrators 'improve' their performance rather than to 'prove' their performance" (Hoyle and Klewer, 1980, p. 70).

Eckard (1980) describes some of the processes used in present day evaluations of faculty. The major source of evidence about teaching performance is still student opinionnaires. This can become a problem if faculty members are allowed to choose the most innocuous of questionnaires in hopes that the collected data will be unusable. Since opinionnaires may be constructed poorly or in a manner so as to yield biased results, there are frequently problems with the validity and

reliability of these instruments.

The second most frequently used method, according to Eckard, is an informal commentary by the department head concerning overall teacher-student contributions. This method carries varying amounts of weight in the final decision-making process, and can include personal dress, lifestyle, collegial relationships and political viewpoints. Eckard contends that the faculty should be instrumental in deciding the topics and types of information used for review as well as the amount of weight each topic should carry. Eckard states:

> The review of material collected haphazardly is an unfortunate practice which often occurs in faculty evaluation. Personnel files of faculty members used for the purpose of administrative decision-making, constitute descriptively a collection of miscellaneous materials semi-relevant to established criteria and unorganized to support adequate promotion, tenure, or merit. Information collected in department files can be both nonrepresentative of on-going behavior and, by fragmentary evidence promote prejudicial communication. Following a designated presentation format, the faculty member should organize his file utilizing a data base for decisions regarding tenure, promotion and merit. . . . Secondly, administration and faculty should agree upon consideration for the weighing of general personnel materials prior to the decision-making process, and this weighing should remain constant from one faculty member to another." (p. 98)

Since many individuals may be involved in the decision-making process, and some are tenured faculty who are not themselves to be evaluated, accountability must be assigned to one ultimate source.

> Irresponsible recommendations intentionally forwarded and substantiated only by subjective and noncriteria related comments, which result in a barrier to academic rewards actually deserved, should be identified as the contribution of specific individuals who are in turn held accountable for their unprofessional actions. Conversely, the executive administration should be held to maximum accountability where occurs (a) intentional unacknowledgement concerning viciousness that consistently determines academic rewards, or (b) unwillingness to structure faculty evaluation sufficiently to prohibit irrational outcomes." (pp. 98-99)

Emphasis is placed by Eckard on the need for knowledgeable faculty members and administrators who communicate their wishes for clear, concise evaluation methods. He urges that more weight should be placed on the ability of the instructor rather than on rumors and the personal bias of those attempting to evaluate.

Berk (1979) addresses the confusing issue of rating scales which are used in faculty evaluation. A standard instrument which applies to all faculty members must be developed. He believes that faculty should be evaluated on "(1) presentations, (2) knowledge of the subject, (3) responsiveness to student needs and interests, (4) tolerance of opposing viewpoints, and (5) enthusiasm in the subject" (p. 652).

According to Berk, the instrument which has proved the most popular is the Likert scale, which has both qualitative and quantitative responses and can be scored on separate answer sheets and computerized for additional data gathering. Berk emphasizes that no more than five points should be used on the scale, and that inappropriate or

inapplicable statements should not be included. Berk suggests, also, that at least two or three statements be included about each element, and that one instrument include no more than fifty statements.

The other important rating instrument Berk refers to is a computerized bank devised by Doyle and Wattawa in 1977. This is a computer program with 350 items from which instructors can select a customized evaluation scale to fit their own needs. Items are selected, the instrument administered, and the raw data then fed into a second program which scores the responses and gives an output of the results. The only drawback to this procedure is that much of the required data must be generalized in order to cover all instructors and courses. Berk also includes a set of guidelines for instrument construction, item selection, and key factors important in analyzing the data received from the questionnaire.

White and Means (1978) reviewed 308 student attitude questionnaires with regard to instructors, and conversely, instructors' attitudes toward the students who rated them. They found that students rate instructors by how they, the students, are performing in class, and that " . . . instructors assigned higher grades to classes in which students (a) reported that they had learned more rather than less and (b) had stated that they had been more highly motivated" (p. 1083). In conclusion, White and Means state:

> In short, students do not rate instruction (and their instructor) in an isolated, aloof manner, unrelated to their performances. As one interprets the ratings of a college instructor by students, one should be aware of the possible influence that could be impinging on the ratings -- namely, the level at which the student is

performing in the classroom." (p. 1083)

Ames and Lau (1979) found that instructors may give higher grades in order to make themselves more attractive to students, but that the students who are able to internally attribute the cause of their performance may still rate the instructor positively whether or not the student performs well in class. Therefore, " . . . students may rate courses on the basis of rational cognitive information process rather than on the basis of positive or negative assessment with a good or poor performance and grade. . . . Thus, instructors who want better ratings may need to be more concerned with how students are attributing the causes of their performance rather than the grade they give for the performance" (p. 27). Ames and Lau conclude that if a student believes that luck or ease may play a part in grades or performance, he/she will have little upon which to base a positive rating.

Kaplan, Orr, and Bartell (1978), in their experiment with 111 medical students using a Likert-type scale, found that faculty attitudes and comprehensive notes were considered by the students to be compensation for difficult courses. Emphasis and clarity were considered the most important aspects of examinations, while lecture emphasis and handouts seemed to be the most important part of instruction. They concluded that the difficulty of the course has less to do with the course evaluation than the instructional method used.

Kaplan et. al. recommend the multiple-measurement approach of evaluation which "places no a priori restrictions on the acceptability of various kinds of data. The nature of the problem is the basis for the selection of methods, not vice versa" (p. 155). Kaplan et. al. also found that some of the extremely negative responses to some of the rating

items were due to the presence of an item on the questionnaire and not a genuine concern of the students. The argument is presented that while the collection and subsequent analysis of student-rated data are important first steps in any multiple approach to surveying student opinion, selection of appropriate survey questions may be the most important step of all.

Schumacher (1978) suggests that faculty evaluation has no longer become a question of why, but how. Schumacher cites Miller (1974) as pointing out five critical issues that are linked with faculty evaluation. These five issues include finance, governance, accountability, flexibility, and goals. Since academic units vary widely with regard to their goals, Schumacher states that "recent trends tend to indicate an emphasis on the quality of teaching as equal or more important than the mere examination of a faculty member's publications or other scholarly pursuits" (p. 29). Schumacher argues that the establishment of a clear purpose of the system is the most critical point of developing a total system, and therefore, "data needed for promotion/tenure decisions are quite different from data needed for the improvement of instruction alone" (p. 132).

In the October, 1979, issue of Independent School, Kemerer compares the costs of an evaluation to its benefits. Kemerer states, "In some places, evaluation has come to mean that faculty members' salaries, promotion, and even continued employment are directly related to student scores on national achievement and aptitude tests. The underlying assumption is that teachers have the power to defuse genetic factors, family background, television, peer groups, and a host of other influences on student behavior" (p. 25). Efficiency can only be achieved,

according to Kemerer, by analyzing gathered data on the effectiveness of personnel; and as resources shift, this will be the primary source of decision making. Kemerer also believes that "the evaluation program must not be viewed as an end in itself but rather as an integral part of an effort to link institutional and departmental goals with individual performance" (p. 27).

CHAPTER V

PROGRAM EVALUATION: CRITICISMS AND SUGGESTED MODELS

The primary purpose of program evaluation is to accumulate information about a program in order that a decision may be reached as to whether or not a change in procedure is needed. However, evaluation methods continue to be plagued with problems concerning the validity and reliability of methods used, and consequently, the value of the results obtained. In addition, Ostrander, Goldstein, and Hull (1978) found that power and politics have the most effect on evaluation success or failure. Negative results were found to be suppressed by agencies if it was to their own immediate interest, no matter what the potential worth was to research teams or long range goals. Both professional bias and personal opinions were found to affect funding, as well as administrations who tried to complete the evaluations themselves in order to put the available money back into the project.

These continuing problems have stimulated some attempts by researchers to develop better methods for evaluation. For example, the Cincinnati Public Schools have devised a successful program based on Stufflebeam's CIPP (Context, Input, Process, and Product) Model (Stufflebeam and Webster, 1971). In his 1979 article Felix presented three models used for evaluation in the Cincinnati Public Schools. The high-trust model takes advantage of previous working relationships in which a high trust has been built between faculty and administration. The moderate trust model is an indirect result of voluntary integration of public schools, and has been achieved with a moderate amount of cooperation among staff due to limited financial resources. The

low-trust model involves the use of funds for disadvantaged students, and local autonomy demands as well as funding cutbacks seem to reduce the level of trust between local schools and central administration. In the low-trust model external evaluation teams become a necessity.

Smith (1981) suggests that in order to increase our store of knowledge concerning evaluation methods, we need to know more about (a) the contexts within which evaluation is practiced, (b) the nature of evaluation utility, and (c) the effectiveness of specific evaluation methods. Also reviewed in Smith's article are the criteria for deciding on an evaluation method:

Need: Is there a need for this method of evaluation? Does it provide a unique approach or are there stronger critical competitors?

Utility: Does the method work, providing knowledge and impact it was designed to produce?

Quality: Are the results of the use of this method of high quality? Is the quality of information and human interaction of the highest caliber?

Acceptability: Do evaluation practitioners seek out this approach to evaluation? Do evaluation clients value the products resulting from the use of this method?

Compatibility: Can the method be adapted to fit the context in which it was designed to be used? Is the method flexible with respect to situational constraints?

Cost: Are the resources, expertise, and time required to use this method reasonable and generally available? Is the method marginally cost efficient?

Side Effects: Are the side effects of the use of this method generally well known and acceptable to most audiences?

Smith suggests that it would be best to compare methods and look to conceptual as well as empirical test results as a guideline for method selection.

In discussing career education evaluation, Ryan, Sutton, and Drummond (1979) argue that "if conceptual models for career education in higher education are to be developed and evaluated, attention must be given to performance objectives, pre-post-test designs, instructional activities and implementation strategies" (p. 21). The argument is also presented that "although it is difficult for many university administrators to accept, unrestricted growth in program development and enrollment is no longer receiving enthusiastic public support" (p. 21). Many universities are now implementing career education training for both faculty and students; therefore, new evaluation methods must be designed to accommodate career training as well as academic classes.

Stufflebeam and Webster (1980) have identified and assessed 13 distinct types of educational evaluations and have defined educational evaluation as one that is designed and conducted to assist some audience to judge and improve the worth of some educational concept or object. They conclude that both strengths and weaknesses of any design should be considered, and that more testing and development in the evaluation field is needed.

Page (1979) argues that evaluation is now the most desirable job market for an aspiring researcher, and that evaluation satisfies a human desire to serve and yet play a social role. Page also contends that

"subjectivism makes evaluation, as a discipline, at times unable to discriminate between the effective and the trashy, between the facts of the educational world and the self-interest of the evaluator" (p. 45).

Four main weaknesses in the field of evaluation are listed by Page. These include the ethical dilemma confronting those evaluators who work for administrators who desire certain outcomes. For, if the evaluator tells the truth, he/she may be the one to be hurt first. By telling the truth, the evaluator may find that he/she is being replaced by someone more "flexible."

The second weakness, according to Page, is measurement. The measuring instrument, argues Page, is not at fault, but the fault lies in the attempt to hide measurement results in the appendix of the textbook where it will not disturb the verbal rapport between the author and the students. Page also feels that some areas of evaluation are even hostile to measurement.

The lack of common training and lack of understanding of decision sciences is Page's third listed evaluation weakness. Page contends that the tools that have been developed over the last 30 years are outstanding. The fault lies in the lack of training of evaluators in their use.

The fourth weakness addresses the question of which direction evaluation will take in the next few years. Page tells us:

> On one hand it may slide onto an increasingly politicized, anti-measurement, compliant set of practices, with increasing authoritarianism and decreasing reputation or reliance on objective evidence. . . . On the other hand, evaluation may increasingly bolster itself with good theory, and greater concensus on methods. It may develop computer-aided techniques for establishing values,

> calculating probable outcomes, and optimizing decisions within well-designed trees of alternatives. (p. 46)

In their search for more complete reference materials to assist evaluators, Backer, Attkisson, Barry, Brock, Davis, Kiresuk, Kirkhart, Perloff, and Windle (1980) have compiled an excellent listing of books, journals, newsletters, and professional societies which is the most complete resource reference to date available to evaluators and "evaluators-to-be." They emphasize that professionals in the field should demand more resource materials that are relevant to their needs and to the needs of the real world. They also emphasize that consumers of evaluations must become more active as consultants and advisors. These two types of involvement will proliferate the improvement and value of evaluation materials as well as evaluations themselves.

CHAPTER VI

CONCLUSIONS AND DISCUSSION

Evaluation in education has progressed at a remarkable rate from the first self-evaluations of supervisors to the new computerized evaluation sources and computerized faculty evaluation scorecards. The only area where progress seems to be lagging appears to be in the attitudes of both those who are to be evaluated and the general public who receive evaluation results.

As evaluators become more important and funds become more scarce, educational institutions will want to use the best resources available within the range of fund availability. In the past, it appears that much time and effort have been wasted by those who try make-shift evaluations rather than hire an expert at the beginning of the project. With the research materials and evaluation training programs that are now available it is unlikely that anyone will be forced to start from scratch to do a large scale evaluation of a program.

Cooperation between the evaluators and those being evaluated is as important as the results given to the final decision-makers. Sometimes a neutral, non-interested evaluator may be necessary to start the evaluation process and to recommend the method or evaluation which should be used. These neutral advisors are trained in their field and can be much more objective than someone whose livelihood may depend on the outcome of the evaluation results.

Evaluation of programs and faculty can be a valuable tool toward progress in education if used properly, but it can be a destructive weapon if handled improperly or with prejudice or bias at hand. If evaluators learn to accept what is valuable about a program or a faculty

member, and rate these values according to the needs of the community, then the credibility that evaluators have long desired will be theirs, and the needs of the public will have been met.

The most important part of an evaluation, in this writer's view, is a willingness to accept the results as they are, and a commitment before the evaluation starts to act upon the results of the evaluative process. Unless action is taken following an evaluation, all effort and funds have been wasted, and evaluation once again loses its credibility in the eyes of the public as well as those who funded the study.

Although advances in methods of evaluation have been made during the past 50 years, it is apparent that the area continues to be in need of further research. Materials, methods, and procedures must be refined and validated in order to lend credibility to the evaluative process and to yield better results for decision-making. Attitudes must change to accept the evaluative process as a sign of the future.

BIBLIOGRAPHY

Ames, Russell & Lau, Sing, An Attributional Approach to the Validity of Student Ratings of Instruction, Contemporary Educational Psychology, Academic Press, New York, Vol. 4, No. 1, pp. 23-39, Jan. 1979.

Backer, T. E., Attkisson, C. C., Barry, J. R., Brock, T. C., Davis, H. R., Kiresuk, T. J., Kirkhart, K. E., Perloff, R., Windle, C., Information Resources for Program Evaluators, Evaluation and Program Planning, Vol. 3, pp. 25-33, 1980.

Beecher, D. E., The Evaluation of Teaching Backgrounds and Concepts, Syracuse University Press, 1949.

Berk, Ronald A., The Construction of Rating Instruments for Faculty Evaluation, Journal of Higher Education, Ohio State University Press, Vol. 50, No. 5, pp. 659-667, 1979.

Cooley, Wm. W., The Inevitable Subjectivity of Evaluators, Educational Evaluation and Policy Analysis, Vol. 3, No. 4, pp. 89-90, 1981.

Damms, Jeffery C., An Assessment of the Austin Peay State University Music Department by Its 1957-1969 Graduates and Its 1969-1970 Faculty, Thesis, Austin Peay State University, 1971.

Eckard, Pamela J., Faculty Evaluation: The Basis for Rewards in Higher Education, Peabody Journal of Education, pp. 94-100, Jan. 1980.

Ellis, Agnes B., A Follow-up Study of the Graduates of the School of Business and Economics (Formerly the Department of Business and Economics) of Austin Peay State University, Clarksville, Tennessee, for the years 1963-1968 Inclusive, Thesis, Austin Peay State University, 1969.

Fathy, Shadia H., The Role an Evaluator's Values Should Play in Program Evaluation, Adult Education, Vol. 30, No. 3, pp. 166-172, 1980.

Felix, Joseph L., Research and Evaluation to Improve Instruction: The Cincinnati Strategy, Educational Evaluation and Policy Analysis, Mar-Apr. Vol. 1, No. 2, pp. 57-62, 1979.

Frazier, Pauline, A Follow-up of Teacher Graduates from Austin Peay State College, Thesis, 1961.

Freeman, Howard E. & Solomon, Maria A., The Next Decade in Evaluation Research, Evaluation and Program Planning, Vol. 2, pp. 255-262, 1979.

Grobman, Hulda, Evaluation Activities of Curriculum Projects: A Starting Point, American Educational Research Association Monograph Series on Curriculum Evaluation, Rand McNally & Co., 1968.

Hoyle, John & Klewer, Ed, Faculty Evaluation in Higher Education: Can You Hear?, Planning and Changing, Vol. 11, No. 2, pp. 63-71, Summer 1980.

Johnson, Mauritz, The Locus of Value Judgments in Educational Program Evaluation, Studies in Educational Evaluation, Vol. 5, No. 2, pp. 109-122, 1979.

Kaplan, Judith S., Orr, T., & Bartell, P. F., Using Multiple Measures in the Interpretation of Student Ratings of Instruction, Evaluation and the Health Professions, Vol. 1, No. 4, pp. 139-156, Dec. 1978.

Kean, Michael H., Compromising Positions: The Objectivity of Evaluators, Educational Evaluation and Policy Analysis, Vol. 3, No. 4, pp. 87-88, Jul-Aug. 1981.

Kemerer, Frank R., Evaluating Evaluation -- Are the Costs Worth the Benefits?, Independent School, Vol. 39, No. 1, pp. 25-29, Oct. 1979.

Landiss, Fred L., A Survey to Collect Information for Use in the Adequacy of the Present APSU Graduate Business Education Program, Thesis, 1953.

McLaughlin, Milbrey W., Evaluation and Secondary Act of 1965, Title I, Cambridge, Mass., Ballinger Pub. Co. (Originally published as the author's Thesis at Harvard) 1975.

Ostrander, Susan A., Goldstein, Paul, & Hull, Donald, Toward Overcoming Problems in Evaluation Research: A Beginning Perspective on Power, Evaluation and Program Planning, Vol. 1, pp. 187-93, 1978.

Page, Ellis B. & Stake, Robert E., Should Educational Evaluation Be More Objective or Subjective? Educational Evaluation and Policy Analysis, Vol. 1, No. 1, pp. 45-47, Jan.-Feb. 1979.

Perlberg, Arye, Evaluation of Instruction in Higher Education: Some Critical Issues, Higher Education, Vol. 8, pp. 141-57, Elsevier Scientific Publishing Co., Amsterdam-Printed in the Netherlands, Mar. 1979. (Technion-Israel Institute of Technology, Haifa, Israel.)

Popham, W. James, Evaluating Instruction, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1973.

Reichardt, Charles S. & Cook, Thomas D., "Paradigms Lost": Some Thoughts on Choosing Methods in Evaluation Research, Evaluation and Program Planning, Vol. 3, No. 4, pp. 229-236, 1980.

Rippy, Robert M., Effects of the New Rhetoric, Studies in Educational Evaluation, Vol. 5, No. 2, pp. 133-136, 1979.

Ryan, Charles W., Sutton, John M., Jr. & Drummond, Robert J., Evaluation in Career Education: A Model for Higher Education, Journal of Research and Development in Education, Vol. 12, No. 3, pp. 21-28, Spr. 1979.

Schumacher, Linda C., The Administrator and Faculty Evaluation, Evaluation and the Health Professions, Vol. 1, No. 4, pp. 127-134, 1978.

Smith, Nick L., Evaluation Reflections -- Requirements for a Discipline of Evaluation, Studies in Educational Evaluation, Vol. 5, No. 1, pp. 5-7, 1979.

Smith, Nick L., Evaluating Evaluation Methods, Studies in Educational Evaluation, Vol. 7, pp. 173-181, 1981.

Stufflebeam, Daniel L. & Webster, Wm. J., An Analysis of Alternative Approaches to Evaluation, Educational Evaluation and Policy Analysis, Vol. 2, No. 3, pp. 5-20, May-Jun. 1980.

Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., & Provus, M. M., Educational Evaluation and Decision Making, Itasca, IL., Peacock Pub., 1971.

Tofte, Terry, Factors Affecting Evaluation Utilization, The North Central Association Quarterly, Vol. 55, No. 4, pp. 400-407, Spr. 1981.

Tolbert, E. L., An Introduction to Guidance, Little, Brown, Co., Inc. Boston, 1978.

White, Wm. F., Hsu, Yi-ming & Means, Robert S., Prediction of Student Ratings of College Instructors From Multiple Achievement Test Variables, Educational and Psychological Measurements, Vol. 38, No. 4, pp. 1077-1083, Win. 1978.

Wholey, Joseph S., Scanlon, John W., Duffy, Hugh G., Fukumato, James S., Vogt, Leona M., Federal Evaluation Policy Analyzing the Effects of Public Programs, The Urban Institute, Washington, D. C., 1970.

Wilhelms, Fred T., Chairman & Ed., Evaluation as Feedback and Guide. Prepared by the 1967 Association for Supervision and Curriculum Development, NEA, 1202 Sixteenth Street, N. W., Washington, D. C., 20036, 1967.

Wittrock, M. C. & Wiley, David E., (Eds.), The Evaluation of Instruction Issues and Problems, Holt, Rinehart & Wilson, Inc., N. Y. 1967.

Woody, Clifford -- Chairman, Evaluation of Supervision 4th Yearbook, Bureau of Publications, Teachers College, Columbia University, N. Y., N. Y., 1931.

APPENDIX I

Questionnaires Suggested by Smock for Assessing Student Opinion

1. College Student Questionnaire (CSQ) published by the Educational Testing Service
2. Learning Climate Questionnaire developed by Bowan and Kilmann
3. College Experience Inventory developed by Field and Schoenfeld
4. College and University Environment Scales developed by Educational Testing Service
5. An 80-item instrument used to measure some of the characteristics of doctoral programs developed by Educational Testing Programs