DOCUMENT RESUME

ED 209 546 CE 030 678

AUTHOR Brandenburg, Dale C.
TITLE Quality Assurance of Human Resources Development Programs: A Socio-Technical Continuum Approach.
PUB DATE Apr 81
NOTE 25p.; Paper presented at the Annual Meeting of the Adult Education Research Association (Los Angeles, CA, April 1981). Best copy available.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Accountability; *Business; Data Collection; *Industrial Training; *Industry; *Labor Force Development; Needs Assessment; Program Design; Program Development; Program Effectiveness; *Program Evaluation; Program Improvement; *Quality Control; Statistical Surveys; Training

ABSTRACT

A useful, values-oriented human resource development (HRD) program evaluation approach is one that is empirical with its orientation derived from a social/technical perspective. The optimum mix between the technical skills of evaluation and the required social skills of evaluation delivery is crucial for the quality assurance process. The core to this approach is grounded in the interrelationship of three continua: (1) the general context of training, education, and development; (2) time dimensions; and (3) action components such as source, types, and collection methods for data. The most critical dimension of the socio-technical continuum approach is time. A general framework can be provided to indicate the initial considerations and the immediate, intermediate, and long-term positions. The direction for evaluation design and understanding of evaluation results should be considered under the initial considerations area by means of evaluation issues, participant and management expectations, and the projected results on future decision-making. The day-to-day development and management of quality assurance tasks focus on the levels of information, sources of data, and methods of collecting data. Evaluating the evaluation can be achieved through the following technical and social components: reliability, validity, utility, credibility, and feasibility. This type of analysis links evaluation to planning and thus to recycling and feedback while also blending judgment with objective evidence. (A 16-item reference section is included.) (BPB)

Quality Assurance of Human Resources Development Programs: A Socio-Technical Continuum Approach

Dale C. Brandenburg
University of Illinois at Urbana-Champaign

ABSTRACT

The intent of this paper is to conceptualize an approach to training program evaluation (specifically, HRD quality assurance) that would be useful in business and industry settings. This approach is a blend of evaluation theory, experience gained from evaluation practice and empirically-derived concepts. Significant stress is placed on the relationship between the technical skills of evaluation research and the required social skills of evaluation delivery. An optimum mix of these skills is viewed as crucial for the quality assurance process to gain credibility with decision makers, especially management, as well as to achieve greater utility for the entire evaluation effort.

The key topics discussed include time as a controlling dimension, the role of issue vs. purpose in the quality assurance mission, the use of levels of information and utility of evaluation results. To show the interrelationships among the variables of time, type of program (training, education or development), issues, sources, methods and levels of information, a matrix schema is outlined to illustrate that the nature of thinking involved in designing quality assurance processes is configural and not linear.

Quality Assurance of Human Resources Development Programs: A Socio-Technical Continuum Approach

Introduction

Human Resources Development (HRD) program evaluation, wherever it is seriously employed, is a difficult and time-consuming process. Compared to an experimental design investigation, a typical HRD program evaluation project generally appears to contain more variables, have less control and offer fewer definitive results and implications. The inherent complexity and political interactions that are an integral part of the program evaluation process, deserve a great deal of credit for this disparity.

One of the most serious of these interactions can be illustrated when the evaluator is encountered with the question, "Why evaluate?" (Thompson, 1978). The asking of this question alone indicates that the credibility of the evaluator and the evaluation process is virtually nonexistent. If this question is considered a legitimate inquiry, the proposed evaluation would have to be outside the mainstream of planning, developing, delivering, and managing HRD activities; this is a situation which should not be accepted nor tolerated. In order for evaluation activities to achieve and maintain credibility, evaluation must be viewed as an integral segment of all HRD developmental and implementation efforts -- its role being supportive as well as judgmental. Thus, program evaluation should never be viewed as a solution in search of a problem.

To assist evaluation practitioners through the commonly encountered maze of problems, evaluation theorists have supplied numerous models (see a review by Stufflebeam and Webster, 1980) for consideration. Although the model presented in this paper is admittedly not unique, it does offer a different view. It is also probably labeled better as an approach than

Paper presented at AERA Annual Meeting, SIG: Training in Business and Industry, Los Angeles, CA, April 13, 1981.

a model, and its application is directed specifically at quality assurance of HRD programs (generically, program evaluation). The basis for the approach is empirical, and its orientation is derived from a socio-technical systems perspective. For its core, this approach posits three interlocking and interactive matrices to demonstrate the interrelationship of three continua: 1) the general context of training, education, and development; 2) time dimensions; and 3) action components: types of data, sources of data, and collection methods for data. Technical, political and social components are interwoven throughout the design.

General Context

The general context of HRD programs in business and industry has been subdivided by Nadler (1971) into training, education, and development. These subdivisions can be analyzed for the purpose of evaluation application as shown in Figure 1. If one views training, education, and development as a continuum of a people development/learning process, the corresponding types of objectives go from concrete to abstract and the length of the HRD programs go from definable to undefinable. Desired outcomes move from direct performance

---

Insert Figure 1 About Here

---

related results under training to career advancement to organizational growth for Development. Relationships between the individual and the organization change from those primarily concerned with the individual to those concerned primarily with the organization. Points for decision-making are clearly defined for Training, projected within a certain time period for Education, but can only be classified as future-oriented for Development.

The nature of these three sets of activities (Training, Education, and Development) therefore, has some direct relevance to the design of quality

assurance programs. The continuous nature also implies a close, working relationship between the planning of HRD programs and their subsequent evaluations. A further implication is that particular evaluation techniques used for Training programs would likely have minimal impact or credibility when used for Development programs. For example, a reaction questionnaire may have direct and important impact in the evaluation of Training, but substantially more depth-oriented techniques, such as interviewing, are likely to yield more impact for evaluation of Development activities. Time and amount of program specificity, therefore, have a relevant bearing on the levels of decision-making that will occur and the complexity of evaluation issues that must be taken into account.

Time as a Controlling Dimension

Probably the most critical dimension of this socio-technical continuum approach to designing quality assurance programs is time. While evaluation as an activity looks back, i.e., records what happened, its function is really forward looking. Evaluation has a requirement to reach beyond the present data and present choices among alternatives not directly studied. Purely objective-based evaluation tends not to give full consideration to unintended outcomes that may have important positive or negative effects on organizational growth and vitality. When viewed longitudinally, all evaluation appears formative (Cronbach, 1978); very few decisions as a result of evaluation are truly ultimate ones.

Time is also important in determining when to begin and end the evaluation effort; oftentimes evaluation seems not to start soon enough nor finish long enough after the actual training experience. Figure 2 provides a general framework from which emanates a general context for evaluation design. This framework was intended to be generic, thus it may be imposed upon any HRD program encountered.

Insert Figure 2 About Here

The rows of boxes at the top and the bottom of the time dimension continuum are probably least written about, but for different reasons. Immediate and Intermediate components are most often described so further elaboration is not presented here. "Long-term" objectives are generally seen as difficult to define and measure. For example, improvement in organizational productivity is probably too abstract of an objective to measure reliably especially if we want to attribute a certain proportion of productivity gains directly to training activities within an HRD program. The "Initial Considerations" area, on the other hand, can be easily utilized to guide the direction for evaluation design and the understanding of evaluation results.

Three primary areas of concern need to be considered under the "Initial Consideration" area: 1) evaluation issues, 2) participant and management expectations, and 3) the projected results on future decision-making. Of primary importance among these concerns are the evaluation issues surrounding the quality assurance process. Evaluation issues are meant as a replacement for Tylerian evaluation purposes or objectives. An issue is a central concern or question such as: Were program participants readily able to comprehend materials? or Did training instructors use effective teaching techniques? An important problem with a purely objectives-based evaluation is the narrow preoccupation with specific objectives without regard to unintended or unforeseen behaviors and results. An issue can more easily guide efforts and may be broken down into specific components.

It would be difficult, if not impossible, to conduct a useful evaluation if the designer was not knowledgeable about the particular organizational structure and have some general background in the politics of interpersonal behavior. Knowledge of antecedent conditions for perspective is a mandatory

requirement to identifying the central issues in an evaluation. By utilizing the "issue approach" to design, through contributions of political informants, program observers, relevant theory and personal experience, the designer is better able to guide the effort around the common pitfalls and tradeoffs usually encountered. Some of these pitfalls include not considering all relevant decision-making audiences, focusing on only minor aspects of the program, or more importantly, limited consideration of the utility and usefulness of evaluation results. It does not matter whether HRD program objectives are tightly or loosely defined, the evaluation issues, as opposed to objectives, should form the basis of the quality assurance mission.

The second area of concern revolves around "audience" expectations prior to program activities. "Audiences" are usually those who have some stake in the evaluation and program such as the participants in the program, the teaching staff, and the organizational management. From the participants we need to know what they hope to learn, gain or achieve as a result of the program. For the teaching staff, historical background information on the participants is helpful in establishing what the participants bring to the program (abilities, attitudes, past performance). Asking the participants what they expect to learn or gain can be compared with opinions or actual results after the program. Similar methods may be used with instructional staff and management. Obviously, these groups are looking at the HRD activity from different viewpoints. For example, the staff may be interested in acceptance and usability of developed materials; management may be looking for "star" performers or changes in identification with the organization.

The third area of concern under "Initial Considerations" is with the impact of projected results on future activities and decision-making. Once evaluation issues have been identified and expectations made explicit, it is time to formulate the evaluation design and do some hypothesizing with respect

to potential courses of action. This hypothetical "what-if" stage is crucial for evaluation management. After all, an evaluation design is meant to control potential and actual tradeoffs which minimize post-evaluation uncertainty within a fixed cost. Concurrently, the evaluator must be aware of the community of decision makers who await the results of the evaluation. Taking these points into account will assist in giving direction to the effort and minimize the number of alternative explanations required to interpret the results.

Action Components

Thus far this socio-technical continuum approach to evaluation has dealt with the establishment of the general context within which quality assurance takes place (Figure 1) and with the variable of time for past and future functional directions (Figure 2). It is now appropriate to illuminate on those aspects of evaluation that have direct relevance to the day-to-day development and management of quality assurance tasks. The following description focuses on three major interacting action components of evaluation design: levels of information, sources of data, and methods for collecting data (See Figure 3).

The first of these action components, Levels of information, may be unfamiliar to most readers, so its definition and utility is given initial consideration. Levels of information are meant to provide guidelines for evaluation design and management. There exist a number of definitions for these "levels," all of them implying similar interpretations. Information may be viewed in two continua--1) in terms of focus: from a general judgment to descriptive behavioral information, or 2) in terms of range: from global information to very specific information. In the first continuum, a general opinion implies a judgment of overall worth -- summative evaluation if you will; description implies observations, reporting of facts or specific behavioral reaction -- thus basic formative evaluation. Consider for example,

the evaluation of teaching staff performance in a given HRD program. If the evaluation issue is how can we improve delivery skills of instructors, one would not be very interested in the participants' answer to the question: How would you rate this instructor's overall teaching ability? This question is really a global question, one requiring a statement of worth (a general judgment). It may be an important question, but it is not directly relevant to the issue of delivery skills. Instead, questions such as: Was the pace of presentation too fast or too slow?, Did the instructor explain new ideas by relating them to familiar concepts?, or Did the instructor have annoying mannerisms which detracted from delivery?, would yield information more directly relevant to the issue of delivery skills.

For the sake of simplicity and utility, we may arbitrarily divide the continua (which are essentially different terms with the same meaning) of global to specific (Smock and Crooks, 1973) or judgment to description (Feldman, 1977), into three categories -- global (Level I), general concept (Level II), and specific (Level III). Global judgmental information (Level I) is primarily directed at forming a general opinion of HRD activity success or failure, i.e., was the workshop successful, did the participants learn anything, did the teaching staff perform well, or how did it go. Global information has primary utility in making comparisons across individuals, groups, or settings. These comparisons may be absolute, relative or both, but experience suggests that relative comparisons would probably yield the most fruitful results. If global questions are appropriately applied, they should have some significant relationship with the general concept and specifc information.

General concept information (Level II) by nature of its in-between status has some characteristics in common with both global and specific information. The primary purpose, however, of general concept information is to identify

general strengths and weaknesses within the evaluation issue being investigated. Returning to the example of the teaching staff member, questions such as -- Was the instructor a good speaker?, Did the instructor appear prepared?, Were the discussions well led?, Did the instructor stimulate participant interest in the subject? -- attempt to diagnose lightly the potential concerns with speaking ability, preparation, leadership, and stimulation, respectively. If one closely examines information collected in most evaluation reports, a majority of it can be classified as Level II. An advantage of Level II information over Level I is that both absolute and relative comparisons may be made with equal effectiveness under most conditions. Comparative uses of such data should be performed cautiously paying particular attention to the relevance and appropriateness across the groups of interest. Level II information lends itself well to making some absolute interpretation of results (better in this respect than Level I usually) because it has a definable base of generalization; however, it still lacks direct knowledge of behavioral manifestations necessary for specific diagnosis. The latter role is fulfilled by Specific or Level III information.

The hierarchical steps of identifying a general issue (Global or Level I information), followed by identification of components of the issue (General Concept or Level II information) is further illuminated and pin-pointed by the most diagnostic information, Specific or Level III. In the example of the teaching staff member, questions about presentation pace, clarity of new ideas and annoying mannerisms (given above) were illustrations of Level III information. Oftentimes the nature of Level III information does not lend itself well to standard or structured questions; answers may be better collected with "softer" techniques of evaluation such as specific open-ended questions, interviewing or videotape analysis. This nature also does not permit useful comparative information. Since the questions are specific (here for a given

staff person in a given teaching setting), they are not meant for generalization, and any comparisons may actually be misleading as well as inappropriate. A potential disadvantage of specific Level III information is that it may bear little or no cause and effect relationship with Level I information. For example, there is no guarantee that once an instructor eliminates annoying mannerisms, overall teaching ability would show marked improvement.

It is worthwhile to note that this hierarchical system, while conceptual, has been applied in the development of evaluative designs (Smock and Crooks, 1973; Brandenburg, 1977). Recently, our investigation of college student ratings items (Brandenburg, Derry & Hengstler, 1978) has yielded an empirical demonstration of the existence of this continua of information (although we showed some justification for four levels, this shows the arbitrary division into three levels). In summary, the concept of levels of information should be viewed as a heuristic tool for guiding the total evaluation process. This concept has direct implications for choice of evaluative methodology, for deciding to whom to report information, for interpreting results, and for generally organizing thoughts from a practical perspective.

The other two action components of evaluation design are the sources of information and the methods for collecting information. While the two are separate, they are often confused in the trade literature (for example, Alden, p. 49, 1978; Kirkpatrick, 1978). Specifically, Kirkpatrick's popular four component model of Reaction, Learning, Behavior, and Results tends to over simplify component interactions. In analyzing Kirkpatrick's (1978) explanation, Reaction implies one source and maybe one or two methods; Learning implies one source and one or two methods; Behavior measures include potentially four sources but undefined methods; Results measures appear to imply numerous unidentified sources, specifies one method and inaugurates a hunt for criteria. What Kirkpatrick advocates in his model deserves important consideration,

however, by mixing sources and methods under one label along a single continuum, a user is likely to be caught in some of the pitfalls discussed earlier. One also gets the impression that an evaluator carries out Reaction, Learning, Behavior, and Results regardless of the HRD program situation or activities. The life of an evaluator is not that simple. Unless the methods for collecting information and the levels of information needed are defined and integrated with sources of such data, evaluation credibility and utility are very likely to suffer. Each method and source should be justified in a congruent sense against the evaluation issue -- evaluators have always been known to have the capability to produce an abundance of useless information. Efficiency, cost-effectiveness and mission should be emphasized; Kirkpatrick's model is not systematized enough nor comprehensive enough to satisfy these criteria.

Figure 3 contains an analytical framework (a set of interlocking matrices) for the differentiation between sources and methods and their joint interaction with evaluation issues and levels of information. Each of the twelve boxes in Figure 2 potentially could be filled with the subsystem of Figure 3. The listing of sources in Figure 3 is not meant to be all inclusive, nor would all sources have to be included in investigating every evaluation issue. The number of data collection methods should be

---

Insert Figure 3 About Here

---

allowed to vary depending on numerous conditions, notably the issue involved and the nature of the information source. The data collection method column may include some of the following:

- expectation statements
- objective questionnaire or survey
- open-ended questions or reactions
- cognitive paper and pencil measures
- performance measures

- observation and anecdotal records
- videotape of instruction
- structured interviews
- previous evaluation results
- demographic and historical data collection
- unobtrusive follow-up techniques
- cost-benefit analysis
- cost-effectiveness analysis
- longitudinal personnel data file examination

Due to the nature of behavioral science data, it is seldom wise to depend on one source and one method to yield valid conclusions. It is important therefore, to use triangulation (multiple methods, multiple sources of cross-checking and verification) among methods and sources to achieve validity and credibility. Thus, one is better off to use three methods on one source or three sources with one method. Without this type of validity cross-checking, the evaluation report is more likely to suffer from a lack of clear direction when it comes to laying out alternative courses of future action for decision makers.

Evaluating the Evaluation

This approach, which is meant to be dynamic and open-ended, would not be complete if attention is not drawn to the quality of information collected. The five major categories of information quality and evaluation design quality include the following technical and social components: reliability, validity, utility, credibility, and feasibility. They are defined below.

Reliability. Reliability is generally defined as the consistency of acquired information. Internal consistency measures (like KR#20 or Coefficient Alpha) are appropriate to use in evaluating the homogeneity of a group of questionnaire or test items. Discriminate reliability, like the Horst coefficient, is important in evaluating Level I data for making reliable comparative distinctions. Here such coefficients must have high values. A third type of reliability, and one that should

be used more often, is that given by generalizability coefficients. Using generalizability theory (Cronbach, et al., 1972) one estimates certain variance components for a data collection scheme and can then generalize over certain dimensions (or facets) by modifying the sizes of particular design components. A fourth type of reliability often associated with softer evaluation techniques is replicability (Guba, 1978). Guba suggests three techniques for replicability: an audit (independent judgments of the same data), overlapping methods (nested use of complementary techniques to collect information on the same issue), and stepwise replication (division of information sources or data collection teams into random halves to undertake independent studies).

Validity. Validity refers to a number of aspects of design and data quality exemplified by internal validity, external validity, content validity, and construct validity. Because these definitions are commonly known and applied, further discussion of the potential biasing effects on the lack of valid information is not included here. For softer evaluation techniques, the terms intrinsic adequacy replaces internal validity and extrinsic adequacy replaces external validity (Guba, 1978). Intrinsic adequacy refers to such aspects as erecting safeguards against possible invalidating facts, establishing structural corroboration (use of triangulation and cross examination) and establishing the credibility of findings. Extrinsic adequacy is related to generalizability through areas such as representative situations, timeliness, special subject samples, and recurring encounters (see Guba, 1978).

Utility. The results of the evaluation endeavor must be viewed as useful to the organization; the information leading to the results needs to be accessible to decision makers, and the information must be understandable to them. Utility has both a social and technical side which is explained in the next section. Attention to data reporting formats should be given early on in the evaluation process to influence readability and interpretability of the findings. If a decision maker has to ask -- what's the point, or now what -- the utility of the information is lost.

Credibility. Credibility of information and design concerns both the sources of data and the decision makers. To a certain extent it reflects the notion of face validity -- it must be seen as important or relevant or it is not worthwhile. Credibility for the evaluation practitioner must generally be earned. It can be gained if those implementing a quality assurance program remain impartial, accurate and respect the perogatives of the individuals involved as well as advocate and protect the confidentiality of the data base.

Feasibility. A comprehensive quality assurance program, while meeting most of the previous considerations, may not be possible due to lack of time, financial or physical resources. Return on investment must be paramount when discussing alternative processes.

Summary and Discussion

The purpose of this section is to summarize some of the key elements of the framework presented and to integrate them with the social/technical perspective advocated.

A first general concern is how the approach presented here may be classified among those that have been or are currently being used in the general area of evaluation. A recent article by Stufflebeam and Webster (1980) provides a good deal of assistance on this point. They define an evaluation study as "one that is designed and conducted to assist some audience to judge and improve the worth of some educational object" (emphasis added, p. 6). They then classified studies as "Politically-Oriented," "Questions-Oriented," or "Values-Oriented." The present approach may be best classified as "Decision-Oriented," a subheading of the Values-Oriented studies. Stufflebeam and Webster characterize decision-oriented studies as those emphasizing "that evaluation should be used proactively to help improve the program as well as retroactively to judge its worth" (p. 12). In comparing a number of evaluation strategies, these same authors characterize each strategy according to certain components. The components and descriptions for decision-oriented studies are presented below (p. 16, 17).

Component	Description
Advance Organizers	Decision situations
Purpose	To provide a knowledge and value base for making and defending decisions
Source of Questions	Decision makers (management, clients, participants, staff), their consituents and evaluators
Main Questions	How should a given enterprise be planned, executed, and recycled in order to foster human growth and development at a reasonable cost?
Typical Methods	Surveys, needs assessment, case studies, advocate teams, observations, and quasi-experimental and experimental design
Pioneers	Cronbach, Stufflebeam

A couple of key points deserve emphasis. First, the evaluator is seen as taking an active role in the total evaluation process (or "proactive" after Patton, 1978). Thus, the evaluator does not function as a disinterested observer. Second, in order to maximize utility of the evaluation, the evaluator must specify alternative decisions and must have gathered sufficient information to defend these alternatives. Third, the decision-oriented approach, like the approach described in this paper, provides information on the continuum from judgment of overall worth to program improvement generally useful for "foster(ing) human growth and development at a reasonable cost" (a primary concern to those of us interested in expanding the concept of HRD).

A second general concern about the approach described here is how does it derive its name as a socio-technical continuum (or system). A best answer is based more on personal experience than either theory or the present vogue of quality assurance in business and industry. Technical knowledge and skills about how to conduct an evaluation can be obtained by studying numerous excellent texts or by classroom study. An evaluator's knowledge and skill application, however, does not necessarily yield a useful, successful evaluation. An additional essential quality of advanced social skills, is of significant importance. Included among these skills are being sensitive to various audiences (the willingness to listen and interact), being intuitive, being responsible in various ways to various groups of people, being tactful, the ability to anticipate questions especially with regard to management, a combination of instinct and experience, and being able to think on one's feet while talking. These skills can not generally be learned through books or classroom experience. An optimal combination of social and technical skills for evaluators should yield greater likelihood for a successful quality assurance program, but would not necessarily guarantee it.[1]

[1] One of my colleagues pointed out, appropriately so, that an evaluator's functional authority also has a great deal to do with evaluation success or impact.

A third general concern that readers may have acquired through this manuscript is the attention given to the utility of evaluation results. (See Brandenburg, 1980). Utility or usefulness also has its social and technical sides which are hopefully not incompatible. A number of references were made to some of the social aspects such as "volumes of data," "lack of organizational knowledge" and lack of social sophistication in the application of evaluation methodology in this manuscript. Why produce evaluation reports that people can not read nor even have any hope of interpreting the results? Why persist in obtaining all the information available on a given objective if nobody is interested in the final outcome? One suggestion given to preventing this dilemma was to center the investigation around issues rather than objectives, i.e., focus on the major points of contention to alleviate future confrontations. A second suggestion was to make use of evaluators that have advanced social as well as technical skills. To sum up, Alkin states "the orientation of the evaluator is a decisive factor - perhaps the most influential - in determining whether utilization will occur" (Alkin and Law, 1980, p. 79).

On the technical side of the utility issue, this writer would like to see more attention given to Bayesian applications to decision-making in the quality assurance process. Bayesian analysis would fit comfortably into the decision-oriented studies described by Stufflebeam and Webster. This type of analysis links evaluation to planning and thus to recycling and feedback while also blending judgment with objective evidence. Furthermore, it can answer the extent to which each decision alternative maximizes the realization of issues related to outcomes by providing utility information on each alternative (Saar, 1980). Although the statistical sophistication required to implement Bayesian analysis does create some problems, these are not insurmountable and may be very well worth the effort.

References

Alden, Jay. Evaluation in focus. Training and Development Journal, October, 1978, 46-50.

Alkin, Marvin, and Alex Law. A conversation on evaluation utilization. Educational Evaluation and Policy Analysis, 1980, 2, 73-79.

Brandenburg, Dale C. Student ratings and the CAFETERIA model: Some dilemmas in design (A rationale for the Instructor and Course Evaluation System, ICES, developed at the University of Illinois). Paper presented at the Northern Illinois Association for Education, Research and Development Meeting. May 13, 1977, Bloomingdale, Illinois.

Brandenburg, Dale C., Sharon Derry, and Dennis D. Hengstler. Validation of an item classification scheme for a student rating item catalog. Paper presented at the National Council on Measurement in Education Annual Meeting, March, 1978.

Brandenburg, Dale C., Reporting quantitative evaluation data: Improving utilization. Paper presented at the National Council on Measurement in Education Annual Meeting, April, 1980.

Cronbach, Lee J., Goldine C. Gleser, Harinder Nanda, and Nageswari Rajaratnam. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: John Wiley & Sons, Inc. 1972.

Cronbach, Lee J. Toward a reformation in evaluation: fourteen points. Presentation in honor of J. Thomas Hastings, University of Illinois. January 30, 1978.

Feldman, Kenneth A. Consistency and variability among college students in rating their teachers and courses: A review and analysis. Research in Higher Education, 1977, 6, 223-274.

Guba, Egon G. Toward a methodology of naturalistic inquiry in educational evaluation. CSE Monograph Series in Evaluation, #8. Center for the Study of Evaluation, UCLA, Los Angeles, 1978.

Kirkpatrick, Donald L. Evaluating in-house training programs. Training and Development Journal, September, 1978, 6-9.

Nadler, Leonard. Developing Human Resources. Houston. Gulf Publishing Co., 1971.

Patton, Michael Q. Utilization - focused evaluation. Beverly Hills, CA: Sage Publications, 1978.

Saar, Shalom S. The role of decision-analysis and Bayesian approach in planning and evaluating education programs. Educational Evaluation and Policy Analysis, 1980, 2, 63-72.

Smock, H. Richard and Terrance J. Crooks. A plan for the comprehensive evaluation of college teaching. Journal of Higher Education, 1973, 44, 577-5786.

Stufflebeam, Daniel L. and William J. Webster. An analysis of alternative approaches to evaluation. Educational Evaluation and Policy Analysis, 1980, 2, 5-20.

Thompson, John T. How to develop a more systematic evaluation strategy. Training and Development Journal, July, 1978, 88-93.

Figure 1

General Context of Evaluation in Three Types of HRD Programs

Characteristic	Training	Education	Development
Type of Objective for HRD Program	Competency-based, Behavioral	Stated general goals, Skill development	General Personal Development, Variable to Undefinable
Length of HRD Program	Short, defined term, Immediate	Intermediate to long term, but defined.	Undefined term or long term
Desired Outcome of HRD Program	Performance, Results	Improvement, Career Advancement	Organizational Growth, Improve QWL
Individual/Organizational Relationship	I→O I→I. (Ind. placement in org.) (Ind. affect on him/herself)	I within O (Ind. within Org.)	O→I O→O (Org. affect on ind.) (Org. affect on org.)
Decision-making Points (distribution)	Spiked (defined point)	Cyclical (definable decision points)	Continuous (ongoing)

Figure 2

General Framework:
Evaluation by Time Dimensions

Time Dimensions Continuum	Type of HRD Program Continuum		
	Training	Education	Development
Initial Considerations			
Immediate			
Intermediate			
Long Term			

Figure 3
Action Component Subsystem

Data Sources	Data Collection Methods	Illustrative Evaluation Issues/Types of Data											
		Course Improvement			Instructor Development			Employee Skill Acquisition			Division Productivity		
		I	II	III	I	II	III	I	II	III	I	II	III
Trainee	a. b. c. d.												
Training Instructors	a. b. c.												
Material Developers (including subject matter experts)	a. b. c. d. e.												
Trainee Supervisors	a. b.												
Trainee Subordinates	a. b.												
Middle-Level Managers	a. b. c.												
Clients	a. b.												
Peer Associates	a. b.												