NUREG/CR-5114

Findings of a Workshop on Developing a Methodology for Evaluating Effectiveness of Nuclear Power Plant Training

Prepared by
J. A. Cantor, R. E. Fuller, C. L. Walker/DDL OMNI
T. J. Mazour/MAI

DDL OMNI Engineering Corporation

Mazour Associates, Inc.

Science Applications International Corporation

Prepared for
U.S. Nuclear Regulatory Commission

AVAILABILITY NOTICE

Availability of Reference Materials Cited in NRC Publications

Most documents cited in NRC publications will be available from one of the following sources:

1. The NRC Public Document Room, 2120 L Street, NW., Lower Level, Washington, DC 20555
2. The Superintendent of Documents, U.S. Government Printing Office, P.O. Box 37082, Washington, DC 20013-7082
3. The National Technical Information Service, Springfield, VA 22161

Although the listing that follows represents the majority of documents cited in NRC publications, it is not intended to be exhaustive.

Referenced documents available for inspection and copying for a fee from the NRC Public Document Room include NRC correspondence and internal NRC memoranda; NRC bulletins, circulars, information notices, inspection and investigation notices; licensee event reports; vendor reports and correspondence; Commission papers; and applicant and licensee documents and correspondence.

The following documents in the NUREG series are available for purchase from the GPO Sales Program: formal NRC staff and contractor reports, NRC-sponsored conference proceedings, international agreement reports, grant publications, and NRC booklets and brochures. Also available are regulatory guides, NRC regulations in the *Code of Federal Regulations*, and *Nuclear Regulatory Commission Issuances*.

Documents available from the National Technical Information Service include NUREG-series reports and technical reports prepared by other Federal agencies and reports prepared by the Atomic Energy Commission, forerunner agency to the Nuclear Regulatory Commission.

Documents available from public and special technical libraries include all open literature items, such as books, journal articles, and transactions. *Federal Register* notices, Federal and State legislation, and congressional reports can usually be obtained from these libraries.

Documents such as theses, dissertations, foreign reports and translations, and non-NRC conference proceedings are available for purchase from the organization sponsoring the publication cited.

Single copies of NRC draft reports are available free, to the extent of supply, upon written request to the Office of Administration, Distribution and Mail Services Section, U.S. Nuclear Regulatory Commission, Washington, DC 20555.

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at the NRC Library, 7920 Norfolk Avenue, Bethesda, Maryland, for use by the public. Codes and standards are usually copyrighted and may be purchased from the originating organization or, if they are American National Standards, from the American National Standards Institute, 1430 Broadway, New York, NY 10018.

NUREG/CR-5114
RX

Findings of a Workshop on Developing a Methodology for Evaluating Effectiveness of Nuclear Power Plant Training

Manuscript Completed: February 1992
Date Published: April 1992

Prepared by
J. A. Cantor, R. E. Fuller, C. L. Walker, DDL OMNI Engineering Corporation
T. J. Mazour, Mazour Associates, Inc.

I. Schoenfeld, NRC Project Manager

DDL OMNI Engineering Corporation
14303D Sullyfield Circle
Chantilly, VA 22021

Mazour Associates, Inc.
2 Broadway
Mystic, CT 06355

Mazour Associates, Inc. Under Contract to:
Science Applications International Corporation
1710 Goodridge Drive
McLean, VA 22102

Prepared for
Division of Systems Research
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555
NRC FIN L1232

ABSTRACT

In October 1990 the Nuclear Regulatory Commission (NRC) sponsored a workshop to develop a proposed methodology for use by the NRC in determining the effectiveness of nuclear utility training. The workshop developed a framework on which to base a methodology which draws together current NRC and nuclear industry processes and initiatives in training evaluation and plant performance monitoring. The framework recognizes that utilities, under current NRC and industry guidance, operate closed-loop training systems that incorporate methods for self-correction. The model proposes that by monitoring/sampling indicator data at various points in the utility's closed-loop system the NRC can determine whether the loop is operating properly to maintain training program effectiveness. This training loop includes the training process, the performance of trained workers, and plant operations. Monitoring/sampling of indicators is planned such that each indicator provides data which complements data derived from the other indicators. Agreement between indicators is used to confirm either effective training or to detect training problems. Inconsistency between indicators triggers further investigation.

TABLE OF CONTENTS

SECTION **PAGE**

TABLE OF CONTENTS (CONT.)

1.0 INTRODUCTION

This report describes the results of the U.S. Nuclear Regulatory Commission (NRC) training effectiveness workshop conducted from October 29 to November 1, 1990. This workshop was conducted: (1) to review existing training evaluation methodologies, (2) to develop recommendations on the selection of an evaluation methodology to be used by NRC, and (3) if necessary, to provide recommendations on how to develop such a methodology. It is expected that this training effectiveness evaluation methodology would be applied by the NRC staff to evaluate training programs at nuclear power plants by measuring a training program's effectiveness in providing personnel the knowledge and skills necessary to fulfill their job responsibilities.

1.1 BACKGROUND

To fulfill its responsibility to evaluate training in the nuclear power industry, the NRC developed Training Review Criteria and Procedures (NUREG-1220) as incorporated in NRC Inspection Procedure 41500. These criteria and procedures are used by the NRC staff to review training programs at nuclear utilities. In addition, the NRC observes and monitors the training accreditation program managed by the Institute of Nuclear Power Operations (INPO). Both the NRC program and the INPO program are primarily oriented toward evaluating the training process rather than its outcomes. It is important that the NRC develop a methodology which extends evaluation beyond the training process to the outcomes as observed in the performance of both personnel and the plant. This training effectiveness evaluation will in turn help ensure safe nuclear reactor operation and hence protection of public health and safety.

2.0 WORKSHOP ACTIVITIES

Planning for the workshop included defining specific goals of the workshop, obtaining the participation of subject matter experts (SMEs), providing background information to the SMEs, and organizing the day-to-day activities of the workshop. This planning, as well as facilitation of the workshop was the responsibility of DEL OMNI Engineering Corp., under contract to the NRC.

2.1 GOALS AND OBJECTIVES OF THE WORKSHOP

The general goal of the workshop was to define an evaluation methodology that would provide information to the NRC on the effectiveness of licensee training programs, and, in the aggregate, the state of training in the industry. The general goal was supported by five specific objectives:

(1) Propose a training effectiveness evaluation methodology (or methodologies) for use by the NRC and describe the methodology/methodologies in terms of input, process, and output elements.

(2) Specify the internal standards that are required of a robust training effectiveness methodology.

(3) Propose a method or methods for establishing external criteria that determine whether training is "effective."

(4) Propose necessary considerations for pilot-testing the proposed methodology.

(5) Describe research projects that should be undertaken to provide information needed to develop the final methodology.

2.2 PARTICIPANTS

There were four categories of participants at the workshop:

(1) Subject-Matter Experts (SMEs) - the principal persons whose deliberations produced the recommended framework.

(2) Facilitators - those persons who focused the deliberations of the SMEs on the goals of the workshop; in addition, the facilitators all had backgrounds relevant to training effectiveness evaluation.

(3) Resource People - NRC staff and industry representatives who had specific knowledge of current NRC and industry practices and procedures and provided background information.

(4) Observers - individuals with an interest in the discussions leading to the formulation of a methodology.

The list of participants by category is contained in Appendix A.

The success of the workshop in achieving its goal and objectives depended, in large part, on selecting workshop participants representing those specific disciplines that needed to be considered in the development of a training effectiveness evaluation methodology. A literature search, conducted as part of the preparation for the workshop identified evaluation disciplines, and the names of persons prominent in various disciplines relevant to training evaluation. The specific disciplines which were identified in the literature search were:

- Needs Assessment
- Testing and Job Performance Measures
- Analysis of Operational and Maintenance Data
- Statistics
- Training Evaluation
- Program Evaluation

The following criteria were applied to judge the relative merits of the qualifications of individuals considered as workshop SMEs:

a. Direct relationship between the specific discipline in which the individual had experience and the disciplines identified as necessary to support attainment of the workshop goals.

b. Review of specific publications written by the individual and/or the individual's record of published works.

c. The expertise of the individual as derived from peer recommendations and/or reference citations of the individual's work by peers.

d. Specific projects in which the individual had participated that appeared to have direct application to the NRC training effectiveness evaluation project.

2.3 WORKSHOP ORGANIZATION

Prior to the workshop, each participant was provided with an information package that included:

- a description of the workshop purpose, goals, and objectives
- information on the training evaluation method being applied by the NRC staff
- a concept paper describing a review of the literature on training evaluation
- copies of relevant literature describing evaluation processes, strategies, and programs that could potentially provide a basis for the development of the training effectiveness methodology
- representative samples of NRC training inspection reports

Selected participants were asked to prepare brief workshop presentations on specific topics.

In general, the first step in accomplishing each workshop objective was an introduction by the facilitators of the objective and a description of the workshop process for achieving the objective. Following the introduction of each objective, presentations related to the objective were made by designated workshop participants. The presentations were followed by discussions of the objective by all participants. Small-group discussions followed these discussions. The first set of small-group discussions considered an appropriate evaluation methodology for designated job positions. After an evaluation approach had been discussed for each job position, a second set of small groups was formed to consider the application of specific evaluation/measurement techniques.

3.0 SUMMARY OF RELEVANT LITERATURE

In preparation for the workshop, a thorough search of the literature on training effectiveness evaluation was conducted. This search included a review of computerized databases in the physical sciences, energy, government, and social sciences. The review of the literature did not identify any methodology which directly met the needs of the NRC for a training effectiveness evaluation methodology. The review did, however, identify many components of a conceptual framework for developing a methodology. The following sections are a synopsis of these components; the complete report on the literature review is found in a concept paper, Appendix B, which was provided to the SMEs prior to the workshop.

3.1 INTERNAL STANDARDS

Internal standards specify characteristics of a methodology to ensure that the methodology consistently produces intended results. A set of potential internal standards identified by Stufflebeam & Webster (1980) and Bushnell (1990) formed the basis for discussions by the workshop SMEs. These were: (1) internal validity; (2) external validity; (3) reliability; (4) objectivity; (5) relevance; (6) importance; (7) scope; (8) credibility; (9) timeliness; (10) pervasiveness; and (11) efficiency. Appendix B, Section 3.1 provides a definition for each of these standards. Table I in Appendix B presents these standards in a matrix linked to the authors reporting on them.

3.2 EVALUATION STRUCTURES

The literature search produced information concerning existing training evaluation programs, models, methodologies, and techniques. An evaluation program will typically employ a model, implemented by a methodology and techniques, to meet its objectives. An evaluation program provides the plan or system under which action may be taken toward achieving management's goals and objectives for training program evaluation. An evaluation model provides additional details of an evaluation program, including the program structure, elements of the program and their interrelationships. An evaluation methodology describes the operational processes to be used for implementing an evaluation model. Monitoring training program effectiveness indicators is an example of an evaluation methodology. Techniques are means of gathering, processing and reporting the information associated with an evaluation methodology. Surveys are examples of evaluation techniques. Methodologies and techniques are not unique to a specific conceptual model nor limited to any particular program. The models and methodologies identified through the literature search, although not necessarily developed or used in the nuclear power industry, were considered by the reviewers as useful for workshop consideration because their structures have broad application.

3.2.1 Training Evaluation Programs

Several existing training evaluation programs were identified by the reviewers as being potentially applicable to the NRC's objectives. These programs are:

a. The U.S. Navy Strategic Weapon System Training Program (SWSTP) and its training effectiveness component, the Personnel and Training Evaluation Program

-was judged to best apply to the NRC's needs. The SWSTP recognizes a closed loop system of training and job performance. Its evaluation methodology provides information to monitor and evaluate training effectiveness from sample points throughout the loop.

b. The Discrepancy-Based Process for Nuclear Training Program Evaluation

-based on a methodology developed for the U.S. Navy Strategic Weapon System Training Program. This program is designed to identify discrepancy information in order to produce a change in either the training program or in the training and development specification under which the training program operates.

c. The U.S. Navy Job Performance Measurement Program

-links enlistment standards to actual job performance data and includes a description of future prospects for service performance. The Job Performance Measurements Program provides the Navy with the capability to measure and predict job performance.

d. IBM's Corporate Evaluation Strategy.

-uses Input-Process-Output approaches in assessing trainee and workforce effectiveness.

These programs are described in Appendix B.

3.2.2 Models

The appropriateness of "decision-based" evaluation models as means of evaluating training effectiveness in the context of commercial nuclear power training, was addressed at the workshop. Different "decision-based" evaluation models are structured to collect and analyze different kinds of data. The types of data to be collected and the specific analysis procedures vary according to the types of decisions to be made from the evaluation. Several evaluation models, including decision-based models, are described and discussed in Appendix B. The following models were presented for particular consideration by the workshop SMEs because they were judged to be most potentially applicable.

3.2.2.1 Decision-Oriented Evaluation (CIPP) The Decision-Oriented Evaluation Model is conceptualized in terms of the decisions for which the evaluation is to provide information. For purposes of training evaluation these are defined as; (1) Context evaluation; (2) Input evaluation; (3) Process evaluation; and (4) Product evaluation. Hence, this particular decision based process is generally referred to in the literature as the CIPP model. Context evaluation relates to the goals and objectives of the training program. Input evaluation is concerned with the system's capabilities and strategies for overcoming difficulties in meeting program objectives. Process evaluation is concerned with the relationship between actual training program operation and training program design. Product evaluation is concerned with the relationship between program outcomes and program objectives.

3.2.2.2 Discrepancy Evaluation Model (DEM) The Discrepancy Evaluation Model was developed by Provus (1969). In this model, the purpose of program evaluation is to determine whether to improve, maintain, or terminate a program. Evaluation is the process of (a) agreeing upon program standards, (b) determining whether a discrepancy exists between some aspect of the program and its associated standard, and, (c) using discrepancy information to identify the weaknesses of the program. The DEM evaluation always consists of a comparison of performance against standards which yields the discrepancy information. Discrepancy information then serves as the basis on which decisions are made.

3.2.3 Methodologies

The literature search did not identify an existing methodology specifically applicable to the NRC's goal of developing a program for assessing the effectiveness of training programs in the commercial nuclear power industry. However, the search did identify two basic types of evaluation methodologies with possible application to conducting training effectiveness evaluations: (1) the monitor/investigate process; and (2) the review/extract process.

3.2.3.1 Monitor/Investigate The monitor/investigate process forms the basis of most good, in-house training feedback in that it provides continuous monitoring of certain indicators and specific investigations when the indicators fail to meet some criterion/set of criteria. The actual data monitored can vary from program to program and, indeed, multiple data sources may be monitored within a single program. In general, programs of this nature make use of internal records maintained for purposes other than training evaluation. In addition, some evaluation process specific records may be created. The ability to track program change over time is a characteristic of the monitor/investigate programs. Inherent in these programs is an assumption that training is generally acceptable and that only deviations from the norm require investigation.

3.2.3.2 Review/Extract Review/extract is the process generally followed by "outside" investigators (e.g., auditors/inspectors). This process consists of reviewing, over a short period of time, the records maintained by an organization, along with any other indicators considered relevant, followed by a site visit for the purpose of gathering additional information which can either confirm or clarify the impressions developed in the review stage. This review of data in a concentrated format can be helpful in putting into focus trends and problems which otherwise can be missed in a monitoring regime. This is especially true when the changes are gradual and the relationship between events may be lost because of time lapse.

3.2.4 Techniques

Techniques identified in the literature for gathering, processing, and reporting training effectiveness information were:

- Direct Measurement - written tests and performance tests
- Drills/Exercises - simulations of responses to critical events
- Exception Reports - exceptions from normal operations
- Normal Job Records - operating logs and maintenance records
- Performance Measures - samples of tasks performed under controlled conditions
- Surveys - formal systems for collecting on-the-job feedback
- Work Sample Observations - observation of persons on-the-job

Appendix B provides an expanded discussion of these techniques.

4.0 THE WORKSHOP OUTCOMES

The workshop goals and objectives were described in Section 2.1 of this report. The goal of the workshop was to define an evaluation methodology that would provide information to the NRC on the effectiveness of specific licensee training programs and in aggregate the status of training in the industry. While the workshop developed many of the attributes of an evaluation methodology, a complete methodology did not result. Rather, a framework, on which an evaluation methodology can be based, was developed. The framework recognizes and exploits the closed-loop nature of utility training programs which are developed in conformity with a "systems approach to training." A systems approach to training ensures that training is developed and conducted based on job-related needs, and includes a feedback loop to continually monitor the performance of the training program. The framework proposed builds upon process evaluation procedures currently in use by the NRC and the nuclear industry. Eventual implementation of a complete methodology is dependent upon the outcome of research proposed by the workshop and successful pilot-testing.

4.1 THE TRAINING EFFECTIVENESS EVALUATION FRAMEWORK

As previously stated in Section 2.1 of this report, the first objective of the workshop was to propose a training effectiveness evaluation methodology for use by the NRC. This section reports on the steps taken during the workshop to achieve this objective.

4.1.1 Framework

The workshop participants developed an evaluation framework that focuses on safety and integrates current processes and initiatives with respect to training evaluation and plant performance monitoring. The framework is discrepancy-based in that evaluators would check for deviations from expected outcomes and then attempt to confirm and/or clarify such deviations through correlation with other indicators. According to the framework, utility training (measured by training process evaluation) is part of a closed loop that also includes on-the-job performance of trained workers (measured by personnel performance indicators) and plant performance (measured by plant indicators) as shown in Figure 1. The system operates within a multi-layered environment, encompassed by the NRC and society at large. The framework indicates that the needs of society stimulate the development of the NRC's safety requirements for the operation of nuclear power plants. The performance of the plant is compared by the utility to the standards derived from NRC safety requirements and other standards. The root causes for discrepancies are identified. Some of these discrepancies are caused by training program weaknesses. When these weaknesses are identified improvements in the system are implemented. By monitoring/sampling data at various points in the utility's closed-loop system, the NRC can evaluate system performance and determine whether training programs are being effective relative to system performance. The monitoring/sampling points are depicted in Figure 1 by the words MONITOR and/or SAMPLE. The training-personnel-plant loop has three monitoring/sampling points that provide access to potential indicators of training effectiveness:

a. Plant-Derived Performance Indicators

b. Personnel Performance Indicators

c. Training Process Evaluation.

Strategies for implementing the framework are presented in Section 4.1.2. In general, the framework would work through a combination of monitoring and/or sampling at each of the points shown in Figure 1. Each monitoring/sampling point would have an effectiveness standard. Indications at a monitoring/sampling point that did not meet the effectiveness standard(s) for that point would trigger investigation at the other points for data that would either confirm or dismiss a suspected training-related problem. Workshop SMEs agreed that emphasis in implementing the framework should be on monitoring indicators as the trigger for additional review. For example, if a plant-derived performance indicator showed a trend in safety system failures, additional monitoring/ sampling points would be reviewed. If review of the data from these points indicated that the trend was the result of a non-training-related problem (e.g., equipment failure due to the equipment not qualified for the environment in which it was required to function), then the framework would return to steady state until another parameter exceeded its effectiveness standard. If, however, the additional monitoring/sampling indicated that the trend was due to equipment improperly installed by plant personnel, the indicators would be reviewed to determine whether the substandard performance was due to inadequate skills or knowledge. If this was the case then monitoring/sampling would focus on those aspects of training programs specifically related to these skills and knowledge to determine whether the need for corrective action was indicated.

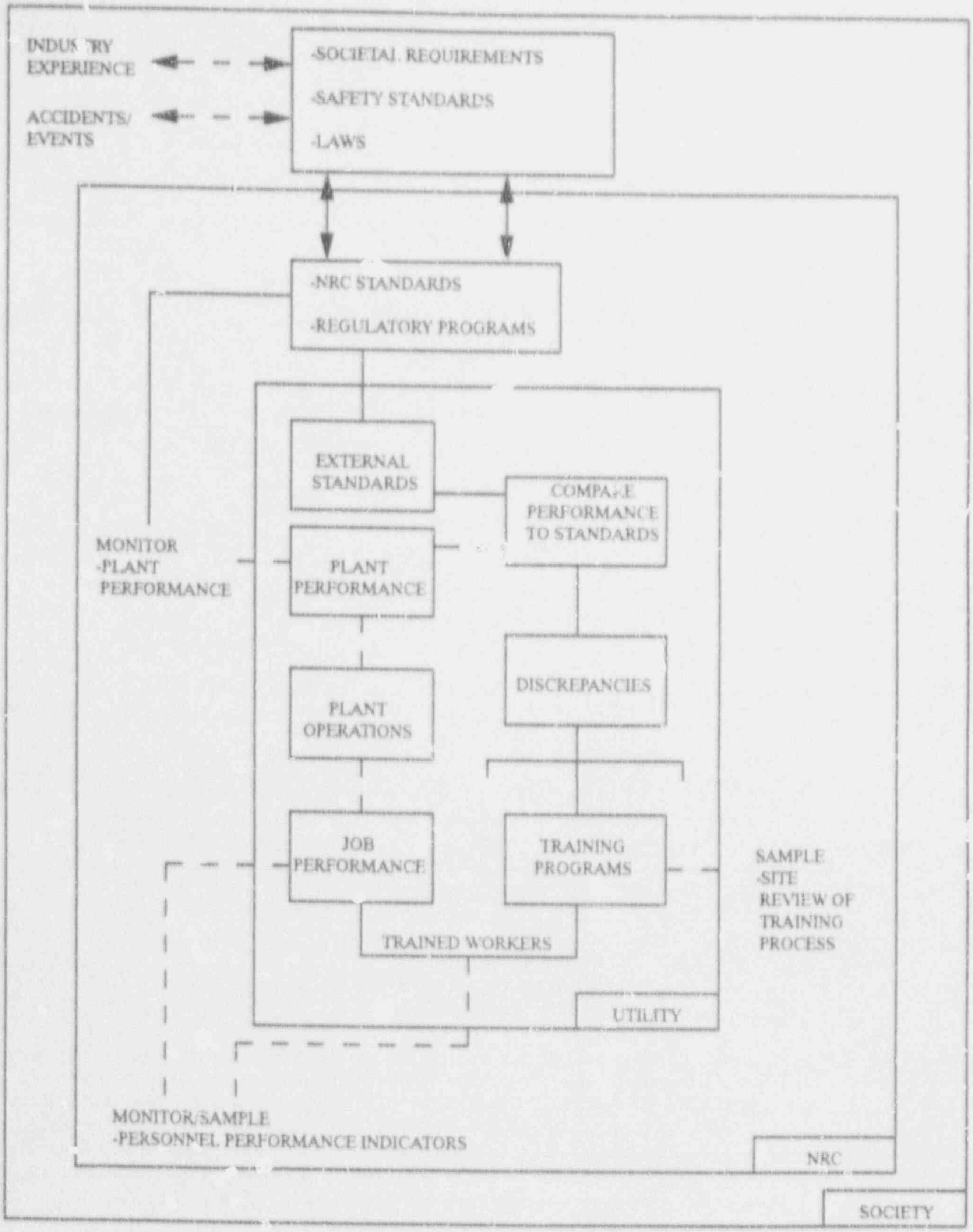

Figure 1. Training Effectiveness Evaluation Framework

Workshop SMEs concluded that there were three important characteristics of this framework that made it particularly well-suited for NRC (and licensee) needs:

1. much of the monitoring and sampling related to each of the three monitoring/sampling points of the framework is currently being performed by either the [illegible] licensees; thus this framework builds upon existing methods (the workshop SMEs concluded, however, that improvements were needed in existing data monitoring and sampling to make them more suitable to evaluating training effectiveness, and that some on-going NRC research efforts had the potential to address these needs).

2. the feedback loop of the systems approach to training, which is being used in some form by all licensees, relies on monitoring and sampling of the same data as does this framework; thus this framework should not impose significant additional monitoring/sampling requirements on licensees.

3. the framework is well-suited to the NRC's performance-based inspection approach in that if a monitored indicator falls below an established effectiveness standard, additional monitoring and sampling is conducted to establish the reason for the substandard performance. The framework encourages a "graded-approach" to allocation of inspection resources based on monitoring of plant-derived and personnel performance indicators.

The following provides additional discussion of the characteristics of the framework's components and their interrelationships.

4.1.1.1 Plant-Derived Performance Indicators The plant-derived performance indicator monitoring/sampling point provides access for monitoring or sampling plant operating and maintenance records and exception reports produced and maintained by licensees and the NRC. Examples of these indicators are: NRC and INPO performance indicators, LERs, SALP reports, inspection reports, QA audit reports, and maintenance backlog, personnel turnover, staffing levels, and maintenance effectiveness statistics. Such information might be expected, through trends or exception reports, to be an indicator of training problems. Although the data reported in the documents might not be directly linked to training, they are closely related to safety. Any discrepancies between actual plant performance and safety standards revealed in the data are cause for checking personnel performance and the training process for corresponding, measurable problems. Alternatively, problems uncovered in either the training process or in personnel performance should stimulate an investigation of related operational data to determine whether training problems have developed to a point where they can be observed in plant performance.

Plant-derived performance indicators provide a view of the plant during the time period covered by the records and reports. These data are normally related to training that took place at some previous period. Because of time lag, training discrepancies or training corrections may not be reflected in plant-derived indicator data until some time after the training has taken place. The amount of time lag between the data at different monitoring/sampling points varies with the nature and severity of the discrepancy observed or corrected.

As indicated above, the NRC is using some plant performance indicators. It is also investigating other indicators of plant performance in operational data. These plant performance indicators can be used in the training effectiveness evaluation methodology.

In addition to the operational data being investigated, other plant-derived performance indicators, such as maintenance records, could provide indications of training effectiveness. The data are available and their use in training effectiveness evaluation is expected to pose little additional data accumulation requirements on utilities or the NRC.

4.1.1.2 Personnel Performance Indicators Monitoring/sampling of personnel performance indicators occurs both immediately upon completion of training and when personnel are on-the-job. Personnel performance indicators can be used to confirm effective training or to isolate or confirm problems found in either the plant-derived performance indicators or in the training process review. However, personnel performance in the work environment is a result of both formal training and the total environment in which people work following training. Therefore, if measurement points to a performance deficiency, it does not necessarily indicate a training problem. In addition to training, some of the other factors involved in personnel performance are experience, policies and procedures, motivation, and health.

SMEs identified three personnel performance measurement techniques as being particularly well-suited to application within the framework; they are: direct measurement (written and performance tests), surveys (interviews/questionnaires) and work sample observations (Appendix C provides a brief description of these and other personnel performance measurement techniques considered). Interactions between SMEs and NRC personnel at the workshop indicated that all of the above techniques are used to a certain extent by the NRC or licensees in evaluating personnel performance, but that the only area where direct measurement was routinely used by the NRC was with respect to licensed operator initial licensing and requalification examinations.

Time-lag effects must also be considered when measuring personnel performance. These effects are related to the time elapsed between training and measurement. When sampling in any population, one must recognize that these time-lag effects will vary because the personnel in the sample are likely to have received their training at different times.

4.1.1.3 Training Process Evaluation Evaluation of the training process is an essential element of the systems approach to training. Once a program is operational, process evaluation is normally performed to seek causes for, or correlation with, performance discrepancies observed at other monitoring/sampling points, and to provide for continual modifications and improvements based on identified needs. In cases where there is no suspected discrepancy to stimulate investigation, NRC evaluators, in applying the training evaluation effectiveness model, might apply selected parts of the process evaluation methodology to related training programs to ensure that the actual training process is continuing to meet established criteria/standards. Because of time lag, changes in the training process may not affect the other indicators to any measurable extent for some time. This can delay correlation from other measurement points of incipient problems and confirmation of corrective action on previous problems.

Training process evaluation is currently the principal element of NRC training inspections. As envisioned within the training effectiveness evaluation model, the existing training process standards contained in NUREG-1220, Training Review Criteria and Procedures (as incorporated in NRC Inspection Procedure 41500), form the basis of the training process evaluation monitoring/sampling point. . A review of NUREG-1220 for the purpose of integrating it into the training effectiveness evaluation model might lead to extension of its coverage or possibly to an extension of its quantitative features.

4.1.1.4 Complementary Characteristics of Sampling Points The use of multiple monitoring/sampling points is important to the effectiveness of the framework because each point makes a different and complementary contribution to the evaluation process.

RELATIONSHIP OF TRAINING EFFECTIVENESS MONITORING/SAMPLING POINTS TO:		
SAFETY		TRAINING
HIGH	Plant-Derived Performance Indicators	LOW
MODERATE	Personnel Performance Indicators	MODERATE TO HIGH
LOW	Training Process Evaluation	HIGH

Figure 2.
Relationship of Monitoring/Sampling Points to Safety and Training

Figure 2 depicts the relationship with respect to safety and training (as assigned by workshop SMEs) of the three monitoring/sampling points. For example, plant-derived performance indicators (first row of data in Figure 2) are highly related to safety but have a low relationship to training. The low relationship to training reflects the fact that plant-derived indicators are influenced by many things in addition to training. The plant-derived indicators are closely related to safety, however, because they directly report the operating status of the plant at any given time.

The internal standards required of the methodology, as discussed in Section 3.1, also underline the complementary nature of the three monitoring/sampling points. Based upon the internal standard descriptors identified in the literature search and provided in the Concept Paper, the SMEs at the workshop identified and defined 10 internal standards that were considered important to the selection of a training effectiveness evaluation methodology, i.e., an acceptable methodology would meet these standards. These 10 standards are listed below and defined in Appendix D.

1. internal validity
2. external validity
3. reliability
4. objectivity
5. relevance
6. importance
7. scope

8. credibility
9. timeliness
10. efficiency

The workshop SMEs concluded that to use these internal standards to evaluate a methodology it was necessary to group them into logical categories. Three categories were identified; these are: (1) technical adequacy (measurement properties), (2) practicality (resources needed to obtain information, and ease of use), and (3) usefulness (application to the decision process). The groupings of the internal standards as assigned by the SMEs to the three categories is shown in Table 1.

Table 1 Categories of Internal Standards

Technical Adequacy	Practicality	Usefulness
internal validity	importance	relevance
reliability	efficiency	credibility
objectivity		timeliness
	scope	scope
external validity		external validity

Figure 3 shows values assigned by the SMEs to the relationship between the internal standard categories and the monitoring/sampling points. Each monitoring/sampling point was judged on a five-point scale with five representing the highest degree attainable. For example, the plant-derived performance indicator monitoring/sampling point (first data column in Figure 3) was considered moderately adequate technically (rating of 3); eminently practical (rating of 5); and of limited usefulness (rating of 2). The "technical adequacy" rating reflects the mix of high objectivity and reliability for indicators derived from plant performance, and the weak linkages of these indicators to training. The high "practicality" rating was given because plant-derived performance indicators are broad in scope, address aspects of plant operation important to safety, and are available without additional collection efforts. The low "usefulness" rating assigned to plant-derived performance indicators reflects the difficulty in isolating a causal relationship between training and plant-derived performance indicator values. The SMEs concluded that this result indicated that the internal standards pointed to a need for a method which integrated the three monitoring/sampling points in a way that builds upon the strengths of each type. This was the case because each monitoring/sampling point complements the other and no one point provides all the information needed to assess overall training effectiveness. As indicated in Figure 3, the uniformly high ratings for training process evaluation were based on an assumption of proper front end analysis. Another complementary aspect of these three monitoring/sampling points is that the plant-derived performance indicators, and personnel performance indicators will aid in identifying weaknesses in the front-end analysis. Where weaknesses in front-end analysis are identified this can be fed back into the system, improving the training process evaluation.

INTERNAL STANDARDS CATEGORY	MONITORING/SAMPLING POINTS		
	PLANT-DERIVED PERFORMANCE INDICATORS	PERSONNEL PERFORMANCE INDICATORS	TRAINING PROCESS EVALUATION
TECHNICAL ADEQUACY	3	4	4
PRACTICALITY	5	2	4
USEFULNESS	2	5	4*

*Assuming proper front-end analysis
SCALE: 1 (lowest degree) - 5 (highest degree)

Figure 3
Relationship of Monitoring/Sampling Points and Internal Standards Categories

4.1.2 Method

The training evaluation framework developed at the workshop lends itself to (1) a monitoring strategy, where full-scale inspections are triggered by discrepancies in indicators, to (2) a strategy of periodic evaluation, or to (3) a combination of both strategies. However, it is critical that those who implement the framework recognize that its driving force is the training-personnel-plant loop operating within the utility. This loop incorporates feedback features that should permit internal correction of deficiencies. By monitoring/sampling at several locations within the loop, the framework accommodates an external check (by the NRC) on the effectiveness of the loop. The workshop SMEs concluded that the preferred use of the framework was a monitoring strategy, augmented by a combination of periodic and discrepancy-driven inspections. Discussions with NRC resource personnel indicated that this approach was consistent with the NRC's graded approach to inspections where SALP reports, the NRC's "troubled plant" list, past inspection results, and other information are used to allocate inspection resources.

Figure 4 provides a graphical presentation of the proposed training effectiveness evaluation method. The fundamental characteristic of the method is its continual feedback and improvement mechanism. Information collected from indicators for individual licensees is used not only to make decisions about subsequent inspections and indicators for that licensee, but for other licensees as well. There are a variety of ways in which this method could be implemented. It could be primarily self-reporting with NRC specifying the information to be reported and licensees providing, or having available, the specified monitoring/sampling information, NRC could collect all or most of the monitoring/sampling information itself, or some combination of the two. The following is an example of an implementation approach which includes a combination of licensee self-reporting of monitoring/sampling information where NRC would specify the monitoring/sampling information to be provided by each licensee. A standard format and content would be specified to allow integration of results among licensees (this is already done for current plant-performance indicators). NRC training process evaluations (e.g., NRC Inspection Procedure 41500) would continue to be conducted to independently verify the quality of licensee sampling. NRC might also conduct independent assessments of the licensee assessment of personnel performance, much as job-performance measures (JPMs) are currently used for the assessment of licensed operator requalification programs. This method could be initiated without a great deal of effort because it builds upon and integrates existing information. Results from implementation of this method would point to needed improvements in the training effectiveness evaluation system. For example, one possible result is that monitoring existing plant performance indicators and personnel performance indicators does not provide sufficient information to ensure the NRC that training programs are effective. This result would indicate a need for regular sampling (training process evaluation and personnel performance) to augment the information provided through these indicators. Because sampling is more labor intensive than monitoring, there is an incentive for both licensees and the NRC to improve monitoring techniques to provide for efficient training effectiveness evaluation.

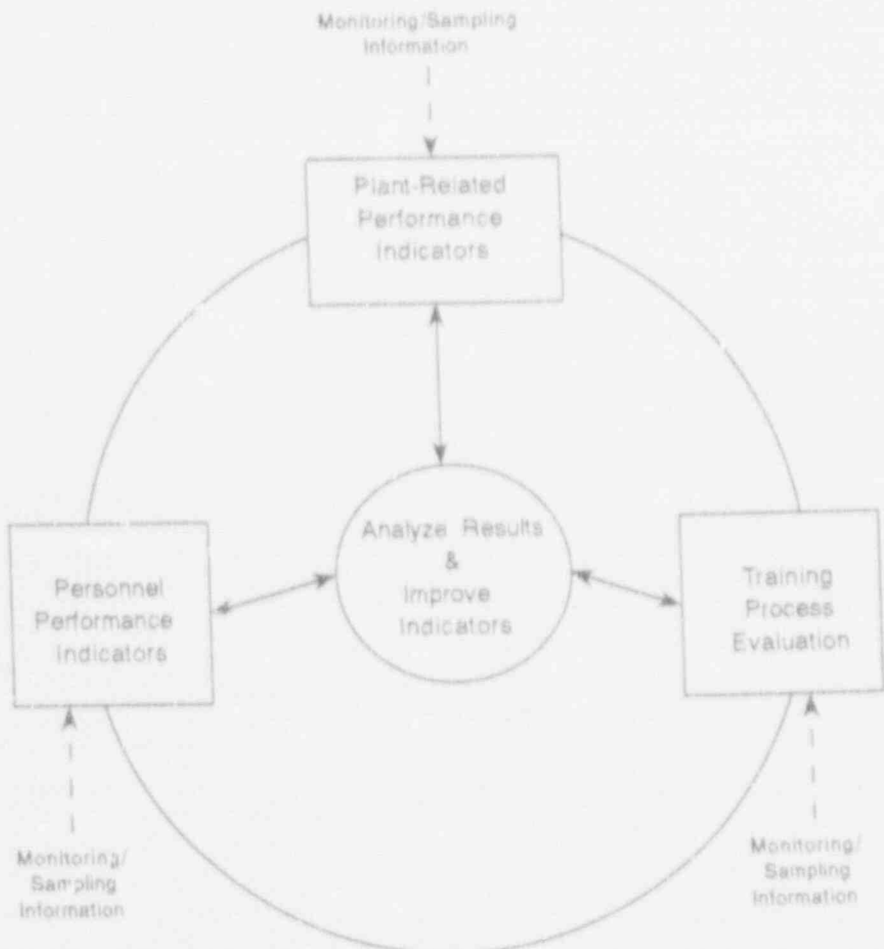

Figure 4 Proposed Training Effectiveness Evaluation Method

4.2 EXTERNAL STANDARDS

The third objective of the workshop was to propose a method for establishing external criteria that determine whether training is "effective." For example, for personnel performance indicators, external criteria could include the jobs and particular tasks to be monitored, and the standards to be used in monitoring/sampling. The workshop SMEs determined that describing approaches for developing external standards that define whether or not training is effective was premature until some of the recommended research had been completed (see Section 4.4). Of particular concern to some of the SMEs was completion of research related to establishing linkages between personnel performance and plant-derived indicators and "safety." These linkages will enable evaluators to determine which indicators must be monitored to determine if training is effectively supporting safe operation. By extension, determination of the indicators to be monitored will permit identification of the best techniques for monitoring/sampling each indicator. Once the techniques for monitoring/sampling an indicator are determined, it will then be possible to set standards for the indicator.

4.3 PILOT-TESTING

The fourth objective of the workshop was to identify necessary considerations for pilot-testing the proposed methodology. The workshop did not proceed to the point of developing a detailed methodology. However, the SMEs recommended that the following be considered in developing a pilot-test after a methodology is developed:

a. The methodology should be pilot-tested on a representative sample of training programs.

b. Pilot-testing should be designed to determine the useability of the methodology by NRC inspectors.

c. During pilot-testing, attention should be given to identifying problems in applying evaluation techniques, including the possible need for specialists to administer some types of evaluation instruments (e.g., personnel performance sampling).

d. As part of pilot-testing, an effort should be made to retroactively validate (or partially validate) the training effectiveness evaluation model and method by: (1) reviewing the results of previous training process evaluations; (2) using the sampling/monitoring points of the model as a guide for collecting additional data concerning conditions as of the time of previous evaluations; (3) evaluating the selected training programs at the time of previous evaluations; and (4) comparing (to the extent possible) the results of the new evaluation with the results of previous evaluations and the results of any relevant evaluations performed by the utility.

4.4 RESEARCH PROJECTS

The fifth and final objective of the workshop was to describe research projects that should be undertaken to collect information needed to define the final methodology. As stated earlier, the workshop did not reach the point of developing a complete training effectiveness evaluation methodology; thus the following developmental efforts are necessary to more fully develop the methodology:

a. Develop a detailed training effectiveness methodology based on the framework and methods recommended by this workshop.

b. Develop external standards.

c. Validate the training effectiveness methodology.

d. Develop a training program for application of the training effectiveness evaluation methodology by NRC personnel.

The workshop participants also identified the following as research questions needing attention in the context of the proposed methodology:

a. Measured against the internal standards proposed by the workshop, what is the quality of existing evaluation techniques and data sources, especially those developed by utilities for evaluation of plant performance, training processes, and personnel performance? To what extent can they be used for NRC training effectiveness evaluation?

b. What effect does the time lag between training and performance have on the evaluation of training effectiveness?

c. How can this methodology be implemented in a way that will not make significant additional resource demands on the NRC or licensees?

The workshop participants identified another research topic that, while not directly related to development of a final methodology, does potentially affect the utility of the methodology; that is the need to develop an operational definition of "safety" in terms of the attributes of safety. These safety attributes can provide an objective basis for identifying safety-related activities, performance indicators, and job positions and tasks.

APPENDIX A

WORKSHOP PARTICIPANTS

SUBJECT-MATTER EXPERTS:

Edward M. Connelly, Systems Analyst, Concord Associates

Bernard Cordier, Assistant, Training Division, Electricité de France

Roger Kaufman, Professor and Director, Center for Needs Assessment and Planning, Florida State University

Arthur L. Korotkin, Senior Study Director, Westat, Inc.

Gerald J. Laabs, Division Head, Performance Assessment Division, U.S. Naval Personnel Research and Development Center

Thomas J. Mazour, Consultant, Professional Engineer

Jo Palchinsky, Nuclear Curriculum Specialist, Florida Power & Light Co.

Charles J. Sener, President, Professional Communication Consultants, LTD

Amiel Sharon, Personnel Research Psychologist, U.S. Office of Personnel Management

Kenneth Stough, Associate Professor, Industrial, Technological & Occupational Education, University of Maryland

John Wreathall, Manager, Technology & Hazard Evaluation Division, Science Applications International Corporation

NRC RESOURCE STAFF:

Christopher Gratton, Operator Licensing Branch, Nuclear Reactor Regulation

Dolores Morisseau, Human Factors Branch, Nuclear Regulatory Research

Richard Pelton, Human Factors Assessment Branch, Nuclear Reactor Regulation

Julius J. Persensky, Human Factors Branch, Nuclear Regulatory Research

Isabelle Schoenfeld, Human Factors Branch, Nuclear Regulatory Research

Susan Shankman, Human Factors Assessment Branch, Nuclear Reactor Regulation

Catherine McCoy Thompson, Human Factors Assessment Branch, Nuclear Reactor Regulation

OBSERVERS

Floyd Croy, EG&G Idaho

Robert Evans, Nuclear Management and Resource Council (NUMARC)

Ronald Fritchley, Institute of Nuclear Power Operations (INPO)

Erik Jende, Swedish Nuclear Power Inspectorate (SKI)

Gerald Scholand, Westinghouse Nuclear Training Center

John Yoder, Department of Energy

APPENDIX B

CONCEPT PAPER

A CONCEPT PAPER
PREPARED FOR DISCUSSION AT
A TRAINING EFFECTIVENESS WORKSHOP

Sponsored by:

The U.S. Nuclear Regulatory Commission

October 29 through November 1, 1990

Conducted and Facilitated by:

DDL OMNI Engineering Corp.

Chantilly, VA

A CONCEPTUAL FRAMEWORK AND LITERATURE REVIEW

This concept paper describes the findings of an analysis of the literature, citing potential internal and external criteria which training effectiveness assessment methods should meet in order to measure training effectiveness and be capable of predicting success of nuclear power industry training programs. This paper also provides a composite abstract of the significant training evaluation methodologies used in various industries, emphasizing performance effectiveness measures. It highlights the methodologies which may apply to the nuclear power industry, as well as possible opportunities for development of other such measures. We hope this information will help you to prepare for the workshop. We look forward to your input and contribution in making the conference a success.

POTENTIAL INTERNAL AND EXTERNAL CRITERIA

For an evaluation methodology to be effective, it must incorporate both internal and external criteria. Internal criteria specify acceptable characteristics for the structure of the methodology, i.e. to ensure that the methodology consistently produces the intended results. External criteria define acceptability for the program characteristics that the methodology is intended to evaluate, for example the characteristics of "effective" training.

Internal Criteria. To support the primary goal of this project [i.e., to propose a training effectiveness evaluation methodology (or methodologies) for use by the NRC, and describe such a methodology in terms of its input, process, and output elements] a sub-goal will be to identify and describe the internal criteria that are required of a training effectiveness methodology. Criteria described by Stufflebeam & Webster (1980) and Bushnell (1990) are the most inclusive of any such criteria suggested. These include, and are enhanced by others, as follows: (1) internal validity -- Does the evaluation design provide the information it is intended to provide? (2) external validity -- To what extent are the results of the evaluation generalizable across time, environment, and human involvement (first cited by Deming, 1975)? (3) reliability -- How accurate and consistent is the information that is collected? (4) objectivity -- The evaluator should strive to collect information and make judgements in such a way that the same interpretations and judgements would be made by any informed and skilled person evaluating the program (Feuer, 1985). (5) relevance -- How closely does the data relate to the objectives of the evaluation study? (6) importance -- Given a set of constraints on the design of an evaluation project, what priorities are placed on the information to be collected or program components to be evaluated? (7) scope -- How comprehensive is the design of the evaluation project? (8) credibility -- Is the evaluator believed by his audiences? Are his audiences predisposed to act on his recommendations? (9) timeliness -- Will evaluation reports be available when they are needed? (10) pervasiveness -- How widely are the results of the evaluation project disseminated? (11) efficiency -- What are the cost/benefits of the project? Have resources been wasted when the waste could have been avoided? Table One presents these criteria in a matrix linked to the authors reporting on them.

TABLE 1 CRITERIA CITED

Authors	Internal Validity	External Validity	Reliability	Objectivity	Relevance	Importance	Scope	Credibility	Timeliness	Pervasiveness	Efficiency
Braun	X	X	X	X	X	X	X	X	X	X	X
Bushnell	X	X	X	X	X	X	X	X	X	X	X
Cantor	X	X	X	X	X	X	X	X	X	X	X
Cantor & Walker	X	X	X	X	X	X	X	X	X	X	
Carter et al.		X	X	X	X	X	X	X	X	X	
Deming	X	X	X	X	X	X	X	X	X	X	X
Erickson & Wentling				X	X	X	X	X	X		
Feiza & Klemm	X	X		X	X	X	X	X	X	X	X
Feuer	X	X	X	X	X	X	X	X		X	
Gammuto	X	X	X		X	X	X	X			
Goldstein	X	X	X	X		X	X				
Hobbs	X	X	X	X	X	X	X	X	X	X	X
Kaufman	X	X	X	X	X	X	X				
Koehorst & Verhoeven	X	X									
Kusy								X	X		X
Laabs & Baker	X	X	X	X	X	X	X	X	X	X	X
Laabs, Berry, Vineberg & Zimmerman	X	X	X	X	X	X	X	X		X	
Lapp	X	X	X		X	X	X		X		X
Matfield	X	X	X	X	X	X	X	X	X	X	X
Morisseau et al.	X	X	X	X	X	X	X	X	X	X	X
Palchinsky & Waylett	X	X	X	X	X	X	X	X	X	X	X
Stufflebeam & Webster	X	X	X	X	X	X	X	X	X	X	X
Swanson & Sleezer	X	X	X						X	X	X
Wentling	X	X	X		X	X	X	X			

External Criteria. The following external criteria are proposed in order to guide us in our workshop efforts: (1) clear, objective standards derived from actual work performance requirements and acceptable to subject-matter experts; (2) a satisfactory design of experiments, tests, and surveys, and examination of existing data. The design of a new project should include samples of the specified material; a description and record of the conditions under which the evaluation is conducted; procedures for carrying out the investigation; and statistical controls to aid supervision of the investigation. (3) methods for presentation and interpretation of the results of the evaluation that will lead to action different from the action that would be taken on the basis of the original data; and (4) an official or group of people authorized to take action.

POTENTIAL EVALUATION MODELS

The appropriateness of "decision-based" evaluation models, as a measure of effectiveness evaluation, and within the context of commercial nuclear power training, is the operational question which we will be addressing at our workshop. Different "decision-based" evaluation models are structured to collect and analyze different kinds of data. The types of data to be collected and the specific analysis procedures vary according to the types of decisions to be made by means of the evaluation. In the context of a training evaluation, typical factors to be considered (i.e, factors that contribute to making decisions) include the input to training, the training process, and the outcomes of training. The framework for considering the use of decision-based evaluation models can be seen in the previously cited works of Stufflebeam & Webster, Kaufman, Cantor, Carter et. al, and also Laabs & Baker, (1989), and Cantor & Walker, (1988). Various evaluation models and processes incorporating decision-based strategies will be described and discussed here as a general introduction or review of the topic.

Decision-Oriented Evaluation (CIPP). Abramson (1979) describes "Decision-Oriented Evaluation," which Stufflebeam and Webster initially proposed as an evaluation model which provides data for decision-makers which are relevant to the specific judgments that are to be made. The model is thus conceptualized in terms of the decisions for which the evaluation is to provide information. Decisions are classified along two dimensions: (1) ends-means, (2) intention-actualities. A 2X2 matrix of these dimensions yields four cells: intended ends, intended means, actual means, actual ends. These same constructs appear in over 85 per cent of all of our reviewed literature (see Table One). These four types of decisions, and their associated evaluations (CIPP) (Kaufman, 1988), are classified as: (1) planning decisions (intended ends) - context evaluation; (2) structuring or programming decisions (intended means) - input evaluation; (3) implementing decisions (actual means) - process evaluation; and (4) recycling decisions (actual ends) - product evaluation.

"Context evaluation" is concerned with the identification of needs, conditions, and problems within a defined and operating educational environment (Bushnell, 1990). This description of the setting gives rise to the goals and specific objectives of the program (input phase). Data from theoretical and empirical studies which compare the intended and actual system performance lead to analysis of possible discrepancies. Cantor (1988) in a review of program evaluation in vocational education found that studies of goals and objectives were usually conducted by asking panels of "experts" to judge the appropriateness of existing objectives.

"Input evaluation" is concerned with the identification and assessment of the system's capabilities and strategies for overcoming difficulties in meeting project objectives. Cost-benefit analyses of human and material resources (in terms of alternative strategies for achieving project objectives) provide the basic data for making comparisons and decisions.

"Process evaluation" is concerned with the relationship between actual training program operation over its life cycle and the original training program design. Process evaluation data help determine whether the program is being operated as it was originally planned. These data also indicate whether the goals and problems identified in the context and input phase and their suggested solutions are occurring as planned. This type of evaluation is basically interested in examining the day-to-day management of the project in terms of those areas which may give rise to project failure, such as personnel, resource, and allocation problems (see Hobbs, 1990; Gammuto, 1987; Lapp, 1987; Kusy, 1986).

"Product evaluation" is concerned with the relationship between program outcomes and program objectives, and the relationship between these outcomes and the context, input, and process evaluation data. Student outcome data are generally collected for this type of evaluation, and are interpreted and related to the context, input and process data.

Although Stufflebeam and Webster (1980) described four phases or types of evaluation in their CIPP model, there is only one strategy for conducting all four types of evaluation, consisting of six major components: (1) focusing the evaluation; (2) data collection; (3) information organization; (4) information analysis; (5) information reporting; and (6) administration of evaluation.

"Focusing the evaluation" requires identification of the decision-making levels, definition of the decision situations within the levels, specification of variables that define the criteria within the situations, and description of the policies to which the evaluator must adhere (as discussed by Feiza & Klemm; Kusy; Morisseau et. al; Palchinksy & Waylett; Swanson & Sleezer; Hobbs; Laabs & Berry, 1987). "Data collection" requires specification of subjects, instruments, sampling, administration procedures, and scheduling. "Information organization" and "information analysis" require an appropriate data processing format, and selection of and means for data analysis (see Hobbs; Lapp). "Reporting of information" requires specification of the audience(s) for the evaluation reports, the means, and schedule of dissemination activities. "Administration of the evaluation" requires planning and conducting the evaluation design by specifying resources, schedules, personnel and budget.

Instructional Systems Model. The ultimate objective of training effectiveness evaluation in the nuclear power industry is to ensure that the program produces competent technicians/operators capable of safe reactor operation. To meet this objective an evaluation process must: provide a means to systematically collect, review, and analyze critical employee (technician/operator) and plant (engineering) performance data; match the data against utility personnel procedures and engineering specifications against industry (INPO) and government (NRC) guidelines and policy; and ultimately match the data against operator performances and the utility's own overall performance record. It is important to understand that nuclear utility technician/operator training is designed and developed using an instructional systems design approach (ISD), according to specific objectives which describe the personnel requirements and responsibilities, and training design requirements and methods. The ISD process (Figure 1) includes a structure for training needs analysis, training program design, development, implementation and evaluation, following the CIPP Model. ISD is a logical systematic process for defining worker competency requirements, developing worker performance objectives, instructional delivery media, and training and trainee effectiveness evaluation strategies. Most all of the training programs in the nuclear power industry are predicated upon this systems model for instructional design. As such, a training effectiveness evaluation methodology must parallel the ISD process (Cantor, 1990; Cantor & Walker, 1988).

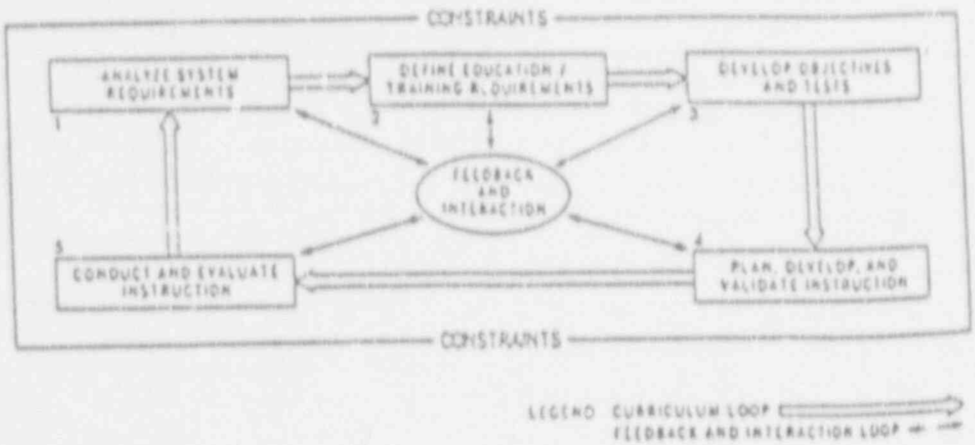

Figure 1.

The first part of such an ISD model is concerned with needs assessment. As the "input phase," it is concerned with resources, budget, and time requirements necessary for achieving the desired goals as identified in the context phase.

Discrepancy Evaluation Model. Another approach to evaluation as a means for providing administrators with information leading to appropriate decisions was developed by Provus (1969). In this model, Provus described the purpose and processes of evaluation as follows:

"The purpose of program evaluation is to determine whether to improve, maintain, or terminate a program. Evaluation is the process of (a) agreeing upon program standards, (b) determining whether a discrepancy exists between some aspect of the program, and, (c) using discrepancy information to identify the weaknesses of the program."

This model which has come to be known as the Discrepancy Evaluation Model (DEM) is conceptualized in terms of four major stages resulting from program development. As described in CIPP, the stages consist of a program definition, program installation, program implementation, and program goal or product attainment. Within each of these stages, the DEM is conceptualized as consisting of continuously repeated sequences of questions through implied criteria, and new information, until ultimately a decision is made. The DEM evaluation always consists of a comparison of Performance (P) against Standard (S) which yields the Discrepancy (D) information. Discrepancy information then serves as the basis on which judgments are made and decisions rendered. Each stage of the program goes through a series of SPD cycles in attempting to provide the necessary information to the program personnel. The content categories -- Input, Process, and Output provide for an in-depth review and analysis of the respective stages in terms of: 1) Design Adequacy, 2) Installation Fidelity, 3) Process Adjustment, 4) Product Assessment, and 5) Cost-Benefit Analysis.

The flow chart (Figure 2) illustrates this process. Here S is Standard, P is Program Performance, C is Compare, D is Discrepancy Information, A is Change in Program Performance or Standards, and T is Terminate. Stage 5 represents the Cost-Benefit option available to the evaluator only after the first four stages have been negotiated. The use of discrepancy information always leads to a decision to either (1) go on to the next stage, (2) recycle the stage after there has been a change in the program's standards or options, (3) recycle to the first stage, or (4) terminate the project.

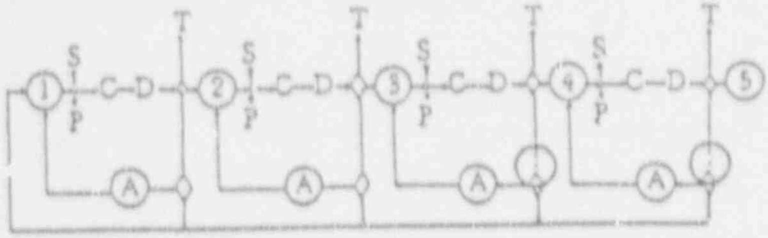

Figure 2.

Cantor (1990) describes the use of DEM in nuclear training environments. Decision-oriented evaluation is appropriate for nuclear power training evaluation because as a product-oriented process, it can be applied within the framework of ISD. Furthermore, it permits analysis of relationships between program outcomes and program objectives; and relationships between these outcomes and the context, input, and process evaluation data. Thus it facilitates: (a) stipulating program standards; (b) determining whether a discrepancy exists between some aspect of the nuclear training program and the standards governing that aspect of the program; and (c) using "discrepancy" information to identify the weaknesses of the training program or its components ("SPD"). Discrepancy analyses provide information leading to appropriate decisions, including immediate or formative decisions, as well as long-range summative decisions (Rog & Bickman, 1984).

An important characteristic of the DEM is its ability to provide information to address several layers of training program design and operation. Data is needed which will permit training program managers to immediately correct individual operator training courses or component parts (simulator, on-the-job training, laboratory, etc.) of the program; to assess overall outcomes of the program (in terms of technician job performance) over time; and/or to redefine aspects of the program's conceptual framework (instructional design procedural standards, performance objective formats, item writing standards, etc.). Therefore, formative and summative evaluation components are necessary.

A Framework for Discussion

In summary, this paper thus far has described the findings of an analysis of the literature citing potential criteria which effectiveness assessment methods must meet to measure training effectiveness, and to predict success of training programs. We have discussed certain fundamental criteria such as the requirements for a process to be both content and construct valid, as well as face valid (internally and externally). It must also be psychometrically reliable, objective, relevant and important, and operationally feasible in terms of credibility, timeliness and efficiency. Design aspects of the process were also discussed including ISD, CIPP and DEM.

A primary purpose of training effectiveness evaluation is to determine whether to maintain, improve, or terminate a program or portion thereof. The training effectiveness evaluation process ultimately identified and described in this workshop will need to incorporate some of the conceptual bases of decision-oriented evaluation models such as the Discrepancy Evaluation Model (DEM) (Provus, 1969; Stufflebeam & Webster, 1980) with the concepts and elements of instructional design.

TRAINING EFFECTIVENESS EVALUATION PROGRAMS

The next part of this paper will provide a composite abstract of the significant training evaluation methodologies emphasizing effectiveness measures. It will highlight the methodologies which best apply to the nuclear power industry, and possible opportunities for development of such measures. While there are numerous training programs operating nationally and internationally, some of which were cited in the literature, four significant training evaluation programs and methodologies emerged as significantly useful to our needs, and will be cited and briefly abstracted. These are: (1) The Discrepancy-Based Process for Nuclear Training Program Evaluation; (2) The Strategic Weapon System Training Program and its training effectiveness component -- Personnel and Training Evaluation Program (PTEP); (3) The Navy Job Performance Measurement Program; and (4) IBM's Corporate Evaluation Strategy.

The Discrepancy-Based Process for Nuclear Training Program Evaluation. The Discrepancy-Based Process for Nuclear Training Program Evaluation is designed in three phases. Phase One of the methodology permits an evaluator to systematically review the overall utility training process including training standards and pinpoint any lack of congruence with accepted industry standards (INPO). Phase Two of the methodology assesses individual training components and their outcomes in order to determine congruence to its respective program development standard. Phase Three permits a synthesis of Phase One and Two discrepancy analysis findings, and includes a description and discussion of the overall indications. For example, in one application of the discrepancy-based process, a commercial nuclear power plant had its training programs evaluated by in order to identify:

a) discrepancies between industry (INPO) training program development standards/guidelines and the utility's own training and development process and standard;

b) discrepancies between the utility's training standard and the utility's actual training program; and,

c) discrepancies between the utility's training program goals and objectives and the nuclear control operator performances.

Each stage of a training program being evaluated goes through a series of "SPD" cycles in attempting to provide the necessary information to address these questions:

1. Is the program defined?
2. Is the program installed?
3. Are the enabling objectives being met?
4. Are the terminal products achieved?

Based on a methodology developed for the U.S. Navy Strategic Weapon System Training Program (Braun, 1981; Cantor, 1986), the Discrepancy-Based Methodology for Nuclear Training Program Evaluation (Figure 3) is designed to identify discrepancy information in order to produce a change in either the operation of the training program or in the training and development specification under which the training program operates. Specifically, discrepancy information can be used to redesign a training standard and process, its relationship to the overall organization, or to better control the process in the training environment (Cantor, 1988; Montague et. al., 1983; Cantor & Hobson, 1981).

The Strategic Weapon System Training Program. The Strategic Weapon System Training Program (SWSTP) was designed and developed in order to serve the constant training needs of the U.S. Navy's Polaris, Poseidon, and Trident submarine force. This training program possesses unique characteristics in that it follows an instructional systems approach for program design from the identification of training needs, through the design and development of instruction, followed by the evaluation of both training effectiveness and trainee readiness. The model constructed for this training program involves the active participation of both Navy activities and a group of contractors providing engineering services support. This program has been in operation for over thirty years, and has become recognized by the Navy as a most effective and desirable training program model (see Cantor, 1986; Braun, 1979).

Figure 3.

The Personnel and Training Evaluation Program (PTEP) is the evaluation and feedback loop in the SWSTP and follows an ISD model. In the SWSTP it is viewed as the quality assurance element. The PTEP measures, evaluates and reports the effectiveness of the program. It is designed to assist overall training measurement by providing the capability for monitoring, evaluation, feedback, and improvements. PTEP accomplishes its objectives by means of personnel testing, collection of test and nontest data, evaluation, and reporting. Both test and nontest data are evaluated to determine trends, identify deficiencies, and formulate recommendations for corrective action. Evaluation results are reported in various types of reports which are distributed to assist commands in increasing personnel proficiency and implementing improvements to SWSTP. Nontest data include personnel history data, group performance data, the status of training facilities/hardware/documentation, and other statistics. An operational data assessment group comprised of fleet personnel, other navy commands and support contractors, assists in this nontest data evaluation effort.

The test instrument basis for PTEP is the criterion-referenced system achievement test (CR SAT) which is built for each technician naval enlisted classification (NEC). The CR SAT is developed through the use of fleet personnel feedback used to identify elements of their jobs that are most crucial. Through a workshop which is periodically conducted, these data are collected and used as a basis for the development of a test design specification for that NEC. For the test design specification, a prototype test is built at two levels per NEC for the technician and technician supervisor. A separate cut-score workshop is conducted, again using fleet technician inputs, to ascertain proper performance standards for these tests.

SATs are administered periodically to all technicians in the SWSTP to measure overall system knowledge and skills levels. The PTEP evaluation program uses all accumulated data to measure the effectiveness of the training program. Analyses are conducted to measure individual and crew status, curriculum appropriateness, and hardware reliability. In addition to formal reports at six-month intervals, quick-look reports and training flags are also provided when test data indicate a problem that cannot wait until a normal reporting cycle.

The Navy's Job Performance Measurement Project. The Navy's Job Performance Measurement Project (JPM) is part of a Congressionally-mandated Joint-Service Project. Specifically, this program links enlistment standards to actual job performance data and includes a description of future prospects for service performance. The JPM provides the Navy with the capability to measure and predict job performance and contributes to major improvements in their personnel system functions and measurement technology (see Laabs & Berry).

The Joint-Service Project had its genesis in congressional concerns that the Armed Services Vocational Aptitude Battery (ASVAB) was not being validated against eventual job performance. The project developed both hands-on and surrogate measures of job performance. This research strategy was endorsed by the Manpower Accession Policy Steering Committee and the National Academy of Sciences (NAS) Committee on the Performance of Military Personnel. Research was conducted in 32 occupational specialty areas across the four services. This constituted about 20 percent of the total enlisted Armed Forces.

The program covered seven ratings that fit well in the Joint-Service effort because together they covered about 25 percent of the total Navy enlisted force, originated from the top 26 critical Navy ratings, represented 5 of the 24 officially recognized occupational fields, and used 5 of the current 10 ASVAB composites that are in operation in the Navy's classification and assignment system. In addition, two original research tasks (i.e., development of a prototype JPM data base and demonstration of new computerized predictor tests) have become separate projects. The JPM data base can be used by the research community to guide criterion development and by the operational community to answer policy questions and guide decisions. The work on new predictors was conducted late in the life of the project and new test development technology progressed at a very fast rate. With the advent of the computerized ASVAB, there has been an increasing amount of interest in evaluating new predictors to supplement this selection battery.

The JPM Program will provide the Navy with the capability to measure and predict job performance. Such a program in the Navy will lead to major improvements in personnel selection and other personnel system functions and significant contributions to measurement technology in particular and applied psychology in general. This program offers certain components to the NRC training effectiveness effort. It also demonstrates a capability to measure a training program's effectiveness from outside that system.

IBM'S Corporate Education Strategy. Bushnell describes IBM's Corporate Education Strategy and its process for evaluating corporate training effectiveness. This program is reported to be tightly linked to all other human resource management programs. Termed an integrated-system approach to training improvement, IBM's top management needs data to anticipate personnel requirements, and to offer retraining opportunities before employee skills become obsolete. Another component of the program is its new-hire skills training.

The IBM process uses Input-Process-Output (IPO) approaches in assessing trainee and workforce effectiveness. IPO is based on a process model in which desired output from a process (i.e. an evaluation process) is identified, the necessary inputs are specified, and a procedure is developed for using the inputs to achieve the desired output. IPO requires evaluation designers to (a) identify their evaluation goals (output); (b) develop an evaluation design strategy; (c) select and construct valid, reliable measurement tools; and (d) make conclusions and recommendations based on analysis (process) of the data (input) collected using the measurement tools.

IBM training directors need to balance the cost and results of training. In the past much of the cost occurred at the delivery stage. Today design and development costs are rising rapidly as technology takes the responsibility for training delivery.

IBM has found that an IPO approach to training evaluation enables decision-makers to select, from several options, the package that will optimize the overall effectiveness of the training program. Those who use the IPO model can readily determine whether the training programs are achieving the right purposes. It also enables them to detect the types of changes they should make to improve cost design, content, and delivery. Most importantly, it tells them whether
students actually acquire the needed knowledge and skills.

EVALUATION METHODOLOGIES AND TOOLS

Methodologies are procedural; i.e., they are ways of doing things. Tools are particular instruments and devices (such as questionnaires and tests) that are used to collect or analyze data. Methodologies and tools are neither unique to a particular conceptual model nor limited to any particular program.

METHODOLOGIES

A training evaluation methodology is a defined procedure for determining the value (for the purpose of this work, effectiveness) of training. Consequently, a training evaluation methodology describes how to evaluate training. Evaluation methodologies vary with the purpose of the evaluation (such as assessing productivity, efficiency, or safety). The NRC's interest in training effectiveness is different from that of the nuclear utilities. The NRC's emphasis is primarily on safety, not specifically on productivity, efficiency, or other economic factors. The literature search did not result in identifying any particular, ready-made methodology specifically applicable to the NRC's goal of developing a program for assessing the effectiveness of training programs in the commercial nuclear power industry. There are, however, two basic types of evaluation methodologies with possible application to conducting training effectiveness evaluations: (1) a process of continuous monitoring of certain indicators with specific investigations when the indicators fail to meet some criterion/set of criteria; and (2) a process whereby data is collected and reviewed in a short period of time followed by a site visit to extract specific information. These two processes will be referred to as "monitor/investigate" and "review/extract."

Monitor/Investigate. The monitor/investigate process forms the basis of most good, in-house training feedback. The actual data monitored can vary from program to program and, indeed, multiple data sources may be monitored within a single program. In general, programs of this nature make use of internal records maintained for purposes other than training evaluation. In addition, some evaluation-process-specific records may be created. The ability to track program change with time is a characteristic of monitor/investigate programs. Inherent in these programs is a generally unexpressed assumption that training is generally acceptable and that only deviations from the norm require investigation. The Personnel and Training Evaluation Program (U.S. Navy, 1989) is an example of a monitor/investigate program.

A positive feature of the monitor/investigate process is that it provides continuous review. In order to be effective, a monitor/investigate process requires constant attention. The appropriate level of attention is sometimes difficult to achieve when evidence of change comes infrequently and when reviews of data are delayed because it will be possible to "catch up tomorrow."

Review/Extract. Review/extract is the general process followed by "outside" investigators (e.g., auditors). The process consists of reviewing, over a short period of time, the records maintained by an organization, along with any other indicators considered relevant, then making a site visit for the purpose of gathering additional information which can either confirm or clarify the impressions developed in the review stage. This review of data in a concentrated format can be helpful in putting into focus trends and problems which otherwise can be missed in a monitoring regime. This is especially true when the changes are gradual and the relationship between events may be lost because of time lapse. The dangers of the review/extract approach lie in the fact that the review is intermittent and that the "snapshot" gained in the review may provide a distorted picture of the overall situation. Abrams et al. (1979) report on processes for conducting a review/extract evaluation. The NRC has developed a review/extract process, Training Review Criteria and Procedures (NUREG-1220) for use by the NRC staff in reviewing nuclear power industry training programs. NUREG-1220 is oriented to evaluating the training process (rather than outcomes).

DATA COLLECTION TOOLS

Data collection methods employed in measuring training effectiveness may apply to either monitor/investigate or review/extract processes. The exact form may either vary with application or remain unchanged. The following items are common tools addressed in the literature (Abrams et al., 1979; Lapp, 1987; Walker, 1987; Worthen & Sanders, 1973):

- Direct Measurement
- Surveys
- Work Sample Observations
- Drills/Exercises
- Normal Job Records
- Exception Reports
- Performance Measures

Experience has shown that training effectiveness can seldom be either validated or challenged on the basis of any single one of the tools cited. It is true, however, that some investigators (e.g., Walker, 1987) report good results using data from either two or more of the tools in combination.

Direct Measurement. Well-managed direct measurement is one of the most useful and best accepted means of determining training effectiveness. Measurement instruments include written tests and performance tests. Performance tests may be administered either at actual work stations or through some form of either simulation or stimulation. Most investigators recommend the use of some form of direct measurement in conducting summative evaluation.

Surveys. Formal systems for collecting on-the-job feedback generally involve some form of questionnaire. The questionnaires can be administered by mail (internal/external) or may be administered by means of structured interview. Survey data is included by most investigators as a product evaluation tool.

Work Sample Observations. This form of data gathering usually involves the observation of persons on the job. Data are recorded on a check sheet, which displays expected actions and has space for recording actual worker actions.

Drills/Exercises. Drills/exercises are simulations of responses to critical events (e.g., large-scale reactor emergency preparedness exercise). Drills/exercises are conducted for the purposes of demonstrating capabilities, maintaining proficiency, and evaluating performance. Jamison et al. (1987) report that drills and exercises provide indications of both personnel and program performance.

Normal Job Records. Normal job records consist of operating logs and maintenance records. They provide the most complete objective record on how a plant operates on a day-to-day basis. The problem with these records lies in discriminating which data clearly reflect upon the adequacy and effectiveness of training (rather than some other variable). These records are important when used in conjunction with other tools (such as tests and surveys) but have proved difficult to use as a primary "training effectiveness" data source. The NRC has an ongoing project on the selection of programmatic performance indicators of training that may include maintenance data as performance indicators.

Exception Reports. Exception reports are made when reporting exceptions from normal operations. For example, nuclear licensees must report certain types of safety-related events to the NRC. Exception reports are of importance both with respect to the nature of the problems and the response to the problems reported. According to Morisseau et al. (1987), the NRC performs evaluations of training programs in response to reportable operating events.

Performance Measures. Performance measures are samples of job tasks performed under controlled conditions. Consequently, they are similar to performance tests and work sample observations. However, they generally have a different purpose. The importance of performance measures (for purposes of performance appraisal, hiring, job selection/classification, compensation) has been described by Ridgev[illegible] & Zucco (1987) and Laabs & Baker (1989). Laabs describes task sampling in terms of Guion's (1979) paradigm for reducing a job to a job sample. The paradigm includes the following steps: (a) define the job-content universe, (b) determine the job-content domain, (c) define the test-content universe, and (d) determine the test-content domain.

Carter, Connelly, and Krois (1989) developed a method of synthesizing performance measures for evaluating overall system (plant and team) performance as a crew responds to an off-normal event in a nuclear power plant simulator. They identified factors important to performance assessment, developed example crew performances and obtained ratings from instructors on the examples, then derived the performance measures using the instructors' assessment rules as a baseline. Consequently, they quantified subjective assessments by instructors, making the assessments more objective.

APPENDIX C

PERSONNEL PERFORMANCE MEASUREMENT TECHNIQUES

Evaluators or investigators develop plans for personnel performance measurement to look for specific kinds of information. This information will be used to attempt correlation with data from other sampling points or to look for changes in response at the Measurable Performance sampling points from that received in previous evaluations. The basic list of measurement techniques or "tools" identified in the literature search and refined by the workshop follows:

a. Direct Measurement

(1) Written Tests
(2) Performance Tests

Well-managed direct measurement is one of the most useful and accepted means of determining training effectiveness. Measurement methodologies include written tests and performance tests. Performance tests can be administered either at actual work stations or through some form of simulation or stimulation. Direct measurement techniques applied in a work environment are generally referred to under the term "Job Performance Measures."

b. Surveys

(1) Interview
(2) Questionnaire

Formal systems for collecting on-the-job feedback generally involve some form of survey consisting of questions for which answers are sought from selected persons. The questions can be asked in an interview or prepared in the form of a questionnaire to be administered by mail (internal or external). Survey data are included by most investigators as a product evaluation tool.

c. Self Reports

These are statements volunteered by individuals outside of the context of a structured response system. The self-report differs from the survey in that it is both randomly supplied and unstructured.

d. Work Sample Observations

(1) Obtrusive
(2) Unobtrusive

This form of data gathering usually involves the observation of persons "on the job." This differs from performance tests, which consist of situations designed for the sole purpose of measurement. Data are recorded on a check sheet that lays out expected actions and reactions. Observations can be made such that they are transparent to the person or persons performing the work (unobtrusive), or can involve interaction between the observer and the worker(s) (obtrusive).

e. Normal Job Records

(1) Raw
(2) Trend

Normal job records consist of operating logs and maintenance records and provide the most complete objective record on how a plant operates on a day-to-day basis.

f. Logged Data

Logged data represent data collection on normal operations that is requested and maintained in addition to normal plant records.

g. Exception Reports, Licensee Event Reports, etc.

These are reports that identify exceptions from normal operations. They are of importance both with respect to the nature of and the response to the problems.

APPENDIX D

INTERNAL STANDARDS
OPERATIONAL DEFINITIONS FOR USE
BY WORKSHOP SMES

1. Internal Validity:

   The methodology must produce the information that it is intended to produce.

2. External Validity:

   The extent to which the results of the evaluation are measurably linked to some external criteria, *e.g.*, safety, organizational results, jobs, and tasks.

3. Reliability:

   The information that is collected must provide consistent indicati

4. Objectivity:

   The evaluation methodology must be constructed so as to control for bias in data collection and interpretation.

5. Relevance:

   The evaluation must serve the information needs of the decision-making audience.

6. Importance:

   The information to be collected and program components to be evaluated are prioritized.

7. Scope:

   The design of the evaluation project must be comprehensive enough to meet, but not exceed, the objectives of the project.

8. Credibility:

   The evaluation process must provide results that are believable to the audiences (i.e., NRC, utilities, public).

9. Timeliness:

   Evaluation reports must be available when needed.

10. Efficiency:

    The costs of the project should be less than the value of the benefits derived.

APPENDIX E

BIBLIOGRAPHY

Abrams, A.J., Peterson, J., Walker, C.L., & Toth, J. (1979). Ship maintenance training & manpower appraisal system. Washington, DC: Naval Sea Systems Command.

Abramson, T. (1979). Issues and models in vocational education evaluation. In Abramson, T., Tittle, C.K., & Cohen, L. (Eds. _ state-tial _, CA: Sage Publications.

Ballentine, R.D. (1986). Air Force job performance measurement research. Proceedings of the Psychology in Dept. of Defense Tenth Symposium, 314-318. (USAFA-TR-86-1)

Bingham, R.D. & Felbinger, C.L. (1988). Evaluation in practice. Longman Braun, F. (1981). The Strategic Weapon System Training Program: Part I -Description. Paper presented at the 23rd Annual Conference of the Military Testing Association, Washington, D.C.

Braun, F. (1979). The strategic weapon system training program. Paper presentation to the annual meeting of the Military Testing Association, Washington D.C.

Braun, F.B. & Walker, C.L. (1974). FBM weapons system training program. In Proceedings of the 16th Annual Conference of the Military Testing Association.

Bryson, J.M. & Cullen, J.W. (1984). A contingent approach to strategy and tactics in formative and summative evaluations. Evaluation and Program Planning, 7, 267-290.

Bushnell, D. (1990). Input, process, output: A model for evaluating training. Journal of Training & Development, March 1990.

Campbell, J. Maintenance procedure upgrade programs. Columbia, MD: Systematic Approach Communication Performance Group Inc. (Compendex Plus database No. 02793468)

Cantor, J.A. (1990). Maximizing the impact of program evaluation: A discrepancy based process for educational program evaluation. Paper presented at the Montclair State University Assessment Conference: Strategies and Prospects for the Decade.

Cantor, J.A. (1988). The training effectiveness algorithm. Journal of Educational Technology Systems, 16(3), 207-229.

Cantor, J.A. (1986). The strategic weapon system training program. Journal of Educational Technology Systems, 14(3), 229-238.

Cantor, J.A. (1984). An effective methodology for evaluation of change in vocational education. Journal of Vocational Education Research, IX(3), 1-13.

Cantor J.A., & Hobson, E.N. (1986). The development of a model for construction of criterion referenced system achievement tests for the strategic weapon system training program. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. April, 1986.

Cantor, J.A. and Hobson, E. (1981). The Strategic Weapon System Training Program: Part II - Executive Steering Group's Role - SWS Personnel and Training Evaluation Program. Paper presented at the 23rd Annual Conference of the Military Testing Association, Washington, D.C.

Cantor, J.A., & Walker, C.L. (1988). A comprehensive method for evaluation of critical military training programs. Paper presentation to the 30th Annual Conference of the Military Testing Association, Washington, D.C.

Carter, R.J., Connelly, E.M., & Krois, P.A. (1989). Feasibility of quantitative performance measures for evaluating nuclear power plant operators. Paper presented at the American Nuclear Society Eighth Symposium on the Training of Nuclear Facility Personnel. (NTIS Document No. DE89007846)

[illegible] & Lipsey, M.W. (Eds.). Evaluation studies reviews annuals. Beverly Hills, [illegible]e Publications.

[illegible]e, S.A., & Korotkin, A.L. (1987). Using FIJA to integrate existing task [illegible]y data for test development. In H.G. Baker, & G. J. Laabs, (Eds.), [illegible]ings of the Department of Defense/Educational Testing Service Conference on [illegible]formance Measurement Technologies. Washington, DC: Office of the Assistant [illegible]ary of Defense (Force Management and Personnel), 65-70..

Davis, L.T. (1983). Performance based training from job and task analysis to training materials. (Department of Energy Database Document No. 140256).

Deming, W.E. (1975). The logic of evaluation. In Struening, E.L., & uttentag, M. (Eds.), Handbook of Evaluation Research (pp. 53-68). Beverly Hills, CA: Sage Publications.

Distefano, M.K., Jr., Preyer, M.W., & Erffmeyer, R.C. Application of content validity methods to the development of a job-related performance rating criterion. Personnel Psychology, 36, 621-631.

Donaldson, L. (1978). Human resource development: The new trainer's guide. Reading, MA: Addison-Wesley.

Draper, J.V. et al. Measuring operator skill and tele-operator performance. Oak Ridge TN: Oak Ridge Labs (NTIS Document No. DE88011393)

Electric Power Research Institute. (1988). Handbook for evaluating the proficiency of maintenance personnel. (INSPEC database Document No. 3295981)

Erickson, R.C. & Wentling, T.L. (1988). Measuring student growth: Techniques and procedures for occupational education (rev.ed.). Urbana: Griffon Press International.

Feiza, A.S. and Klemm, R.W. (1987). Evaluating performance measures to determine training effectiveness. Proceedings of the CSNI Specialist Meeting on Training of Nuclear Reactor Personnel, 390-399.

Feuer, D. (1985). Protecting the public: How much training is enough? Training, 22, 22-24.

Feuer, D. (1985). Training at Three Mile Island: Six years later. Training, 22, 26-27.

Fitzpatrick, R., & Morrison, E.J. (1971).—Performance and product evaluation. In R.L. Thorndike (Ed.), Educational Measurement (pp. 237-270). Washington, D.C.: American Council on Education.

Fitz-Gibbon, C.T. & Morris. L.L. How to design a program evaluation (2nd Ed.). Beverly Hills, CA: Sage Publications.

Gammuto, J. J. (1979). Establishing minimal acceptable performance standards in technical training. Training and Development Journal. (ERIC Document No. EJ210451)

Geroy, G.D. & Wright, P.C. (1988). Evaluation research: A pragmatic, program-focused, research strategy for decision-makers. Performance Improvement Quarterly, 1(3), 17-26. (ERIC Document No. EJ389267)

Goldstein, I.L. (1978). The pursuit of validity in the evaluation of training programs. Human Factors, 20(2), 131-144

Goldstein, I. (1974). Training: Program development and evaluation. Monterey, CA: Brooks/Cole.

Goodstein, L.P. Training diagnostic skills for nuclear power plants - final report of the NKA Project - Lit 4 (NTIS Document No. DE87752878)

Gould, R.B. (1983). Air Force job performance criterion development. In C.R. Curran, Chair, Development of military job performance measures. Symposium presented at the meeting of the American Psychological Association.

Guion, R.M. (1979). Principles of work sample testing: Part III: Construction and evaluation of work sample tests. (ARI Report No. TR-79-A-10). Alexandria, VA: U.S. Army Research Institute.

Gustafson, C.B. (1987). Evaluating learner progress. Journal of Emergency Medical Services.

Hagen, E.W. (1985). Human Factors Activities at EPRI. (INSPEC Database Document No. 1712287)

Harris, D.A. (1987). Job performance measurement and the joint-service project: An overview. In H.G. Baker, & G.J. Laabs (Eds.), Proceedings of the Department of Defense/Educational Testing Service conference on job performance measurement technologies (pp. 33-36). Washington, D.C.: Office of the Assistant Secretary of Defense (Force Management and Personnel).

Henning, P.W. & Christian, K.E. (1985). Supervise and manage trainers at nuclear sites. Security Management, 29, 55-57.

Herman, J.L. et al. (1987). How to communicate evaluation findings. (2nd Ed.). Beverly Hills, CA: Sage Publications.

Hicky, A.E. (Ed.). (1981). Simulation and training technology for nuclear power plant safety. Air Systems.

Hobbs, D.L. (1990). A training-appropriations process. Training & Development Journal May, 109-115.

Jamison, J.D., Moeller, M.P. and Martin, G.F. (1987). Closing the loop: Using the emergency preparedness exercise for self-evaluation and performance enhancement. Proceedings of the CSNI Specialist Meeting on Training of Nuclear Reactor Personnel, 357-368.

Kaufman, R. (1988). Preparing useful performance indicators. Training & Development Journal, 42(9), 80-83.

Kaufman, R. (1986). Obtaining functional results: Relating needs assessment, needs analysis and objectives. Educational Technology, 26(1), 24-27.

Kaufman, R. (1985). Linking training to organizational impact. Journal of Instructional Development, 8(2), 23-29.

Kaufman, R. (1983). A holistic planning model: A system approach for mproving organizational effectiveness and impact. Performance & Instruction Journal, 22(8), 3-12.

King, C.T. Performance standards in job training and job placement programs. Employment and training issues. Washington, DC: National Association of Counties. (ERIC Document No. ED310259)

Kirkpatrick, D.L. (1976). Evaluation of training. In R.L. Craig, (Ed.), Training and development handbook. New York, NY: McGraw-Hill Book Co.

Koehorst, P. & Verhoeven, W. (1986). Effectiveness and efficiency in industrial training, part 1: Sources of ineffectiveness and inefficiency. Journal of European Industrial Training, 10 3, 20-22, (Business Periodical Index No. EJ333932)

Koehorst, P. & Verhoeven, W. (1986). Effectiveness and efficiency in industrial training, part 2: Popular measures for improving effectiveness refuted. Journal of European Industrial Training, 10 (4), 7-10. (Business Periodical Index No. EJ342149)

Koehorst, P. & Verhoeven, W. (1986). Effectiveness and efficiency in industrial training, part 3: How to improve effectiveness. Journal of European Industrial Training, 10 5, 11-16. (Business Periodical Index No. EJ342153)

Kusy, M.E., Jr. (1986). The effects of types of training evaluation on support of training among corporate managers. Paper presentation to the American Vocational Association.

Laabs, G.J. (1983). Performance measurement in support of Navy personnel classification and assignment. In C.R. Curran, Chair, Development of Military Job Performance Measures. Symposium presented at the meeting of the American Psychological Association.

Laabs, G.J. & Baker, H. (1989). Selection of critical tasks for Navy job performance measures. Military Psychology, 1, 3-16.

Laabs, G.J., Baker, H.G., Kroeker, L., & Kidder, P. (1986). Navy Job Performance Measurement Program. Proceedings of the Psychology in Dept. of Defense Tenth Symposium, 309-313. (USAFA-TR-86-1)

Laabs, G.J., & Berry, V.M. (1987). Navy job performance measurement program: Background and inception. (NPRDC TR-87-34). San Diego: Navy Personnel Research and Development Center.

Laabs, G.J., Berry, V.M., Vineberg, R., & Zimmerman, R. (1987). Comparing different procedures of task selection. In H.G. Baker, & G. J. Laabs, (Eds.), Proceedings of the Department of Defense/Educational Testing Service Conference on Job Performance Measurement Technologies. Washington, DC: Office of the Assistant Secretary of Defense (Force Management and Personnel), 79-93.

Lapp, H.J., Jr. (1987). Training evaluation as an integral component of training for performance. (Business Periodical Index No. ED299446)

Lapp, H.J. (1983). A training program evaluation system: An important contribution to plant evaluation. (Department of Energy Database Document No. 275115)

Loeffer, W.R. (1983). A general theory of disambiguation of written texts: An evaluation format for nuclear industry manual design. (Department of Energy Database Document No. 275112)

Love, L. T. et al. (1988). The organizational politics of evaluation: A supplement to the profile section. Hansville, WA: Employment Security Department. (ERIC Document No. EJ378435)

Maier, M.H. (1983). Using job performance tests as criteria for validating qualification standards. In C.R. Curran, Chair, Development of military job performance measures. Symposium presented at the meeting of the American Psychological Association.

Maier, M.H., & Mayberry, P.W. (1986). Improving the design of the Marine Corps job performance measurement project. Proceedings of the Psychology in Dept. of Defense Tenth Symposium, 309-313. (Agency Report No. USAFA-TR-86-1)

Maryland State Council on Vocational Technical Education. (1989). Evaluation of vocational education delivery systems assisted under the Job Training Partnership Act and Carl D. Perkins Vocational Education Act, 2nd Biennial Report. (ERIC Document No. ED305513)

Matfield, R. Quality assurance in nuclear power plants. Harwood Academic,

Mazour, T.J. (1985). Instructional skills evaluation in nuclear industry training. (Department of Energy Database Document No. 1067048)

Mazour, T.J. (1986). Performance based training: Industry commitments. (INSPEC Database Document No. 1850071)

McAllister, J.E. (1984). Industry wide job task analysis: Getting to what's necessary and sufficient. (Department of Energy Database Document No. 368220)

Montague, W.E., Ellis, J.A., & Wulfeck, W.H. (1983). The instructional quality inventory (IQI): A formative evaluation tool for instructional systems development. Monograph: Navy Personnel Research and Development Center, San Diego, CA.

Morisseau, D.S., Koontz, J.L. and Persensky, J. J. (1987). NRC Methods for Evaluation of Industry Training. Proceedings of the CSNI Specialist Meeting on Training of Nuclear Reactor Personnel, 28-32.

Morris, L.J. et al. (1987). How to communicate evaluation findings (2nd Edition). Beverly Hills, CA: Sage Publications.

Morris, L.L., & Fitzgibbon, C.T. (1978). Evaluator's handbook. Beverly Hills, CA: Sage Publications.

Palchinsky, J. & Waylett, W.J., Jr. (1987). Development of a training assurance program. Proceedings of the CSNI Specialist Meeting on Training of Nuclear Reactor Personnel, 306-322.

Palumbo, D. J. (Ed.) (1987). The politics of program evaluation. Beverly Hills, CA: Sage Publications.

Peterson, L.W. Development and implementation of quality based team and diagnostic skills training at the advanced test reactor. EG&G Idaho, Inc. (NTIS Document No. DE89011C52)

Pritchard, R.D. et al. (1987). Feedback, goal setting and incentives effecting on organizational productivity. (ERIC Document No. ED283963)

Provus, M.M. (1969). Evaluation of ongoing programs in the public school system. In R.W. Taylor (Ed.), Educational evaluation: New roles, new means. The 68th Yearbook of the National Society for the Study of Education, Part II. Chicago, IL: National Society for the Study of Education.

Ray, L. (1986). How to Measure Training Effectiveness. New York: Nichols Publishing Company.

Recker, M. Qualification and training of the trainers, verifiers, licensors, and upper management. (NTIS Document No. DE87751504)

Reynolds, M. (1986). Great strides in nuclear training. Electric Perspectives, 42(7).

Ridgeway, R.L. & Zucco, K.K. (1987). Assessing personnel performance, Part I: Nature and Overview of performance assessment. Fire Chief Magazine, 44-47.

Rog, D. & Bickman, L. (1984). The feedback research approach to evaluation. Evaluation and Program Planning, 7, 169-175.

Rose, A.M. et al. (1985). Forecasting device effectiveness: Volume I: Issues (Final Report). (ERIC Document No. ED265186)

Rose, A.M. et al. (1985). Forecasting device effectiveness: Volume II: Procedures (Final Report). (ERIC Document No. ED265187)

Rose, A.M. et al. (1985). Forecasting device effectiveness: Volume III: Analytic assessment of device effectiveness forecasting technique (Final Report). (ERIC Document No. ED265188)

Rumsey, M.G. (1983). Job performance and assessment: A systematic model. In R. Curran, Chair, Development of military job performance measures. Symposium presented at the meeting of the American Psychological Association.

Russell, J.D., & Blake, B.L. (1988). Formative and summative evaluation of instructional products and learners. Educational Technology, 22-28.

Ruzic, D.H. et al. (1987). Computerized training program usage at the Fort Calhoun Nuclear Power Station. (INSPEC Database Document No. 2017437)

Shanaberger, C.J. (1989). Liability in training. Journal of Emergency Medical Services.

Shanaberger, C.J. (1989). Making the grade. Journal of Emergency Medical Services.

Spilberg, S.W. (1983). NRC reactor operator and senior reactor operator licensing examination development and validation project. (Department of Energy Database Document No. 140260)

Stufflebeam D.J., & Webster, W.J. (1980). An analysis of alternative approaches to evaluation. Educational Evaluation and Policy Analysis, 2(3), 5-20.

Stufflebeam D.J. et al. (1971). Educational evaluation and decision-making in education. Itasca, IL: Peacock.

Swanson, R.A., & Sleezer, C.M. (1987). Training effectiveness evaluation. Training Development and Research Center Project Number Sixteen. (ERIC Document No. ED279795)

U.S. Navy. (1989). NAVSEA OD 45953: Strategic weapons system training program personnel and training evaluation program. Washington, DC: Director of Strategic Systems Programs, Department of the Navy.

U.S. Nuclear Regulatory Commission. (1986). NUREG-1220: Training review criteria and procedures. Washington, DC: Office of Nuclear Reactor Regulation, U.S. Nuclear Regulatory Commission.

Wainer, H. (1987). Some problems in statistically adjusting for survey nonresponse. In H.G. Baker, & G. J. Laabs, (Eds.), Proceedings of the Department of Defense/Educational Testing Service Conference on Job Performance Measurement Technologies. Washington, DC: Office of the Assistant Secretary of Defense (Force Management and Personnel), 155-171.

Walker, C.L. (1987). Data collection and data reduction in support of AEGIS combat system training assessment. Chantilly, Va.: DDL Omni Engineering Corp.

Wentling, T.L. (1980). Evaluating occupational education and training programs. Urbana, Ill.: Griffon Press.

Wigdor, A.K., & Green, B.F. (Eds). (1986). Assessing the performance of enlisted personnel: Evaluation of a joint-service research project. Washington, DC: National Academy Press.

Worthen, B.R., & Sanders, J.R. (1973). Considerations in planning evaluation studies. In Educational evaluation: Theory and practice, 218-224. Worthington, OH: Charles A. Jones Publishing Co.

Wright, J.V. (   ). A training program evaluation that works. Carolina Power & Light Company (Nuclear Reactors) 1984778

NRC FORM 335 (2-89) NRCM 1102, 3201, 3202

U.S. NUCLEAR REGULATORY COMMISSION

BIBLIOGRAPHIC DATA SHEET

(See instructions on the reverse)

1. REPORT NUMBER (Assigned by NRC, Add Vol., Supp., Rev., and Addendum Numbers, if any.)

NUREG/CR-5114

2. TITLE AND SUBTITLE

Findings of a Workshop on Developing a Methodology for Evaluating Effectiveness of Nuclear Power Plant Training

3. DATE REPORT PUBLISHED

MONTH	YEAR
April	1992

4. FIN OR GRANT NUMBER

L1232

5. AUTHOR(S)

J.A. Cantor, R.E. Fuller, C.L. Walker/DDL OMNI
T.J. Mazour/MAI

6. TYPE OF REPORT

Technical

7. PERIOD COVERED (Inclusive Dates)

8. PERFORMING ORGANIZATION – NAME AND ADDRESS (If NRC, provide Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)

DDL OMNI Engineering Corporation
14303D Sullyfield Circle
Chantilly, VA 22021

Mazour Associates, Inc.
2 Broadway
Mystic, CT 06355

MAI Under Contract to:
Science Applications International Corporation
1710 Goodridge Drive
McLean, VA 22102

9. SPONSORING ORGANIZATION – NAME AND ADDRESS (If NRC, type "Same as above"; if contractor, provide NRC Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address.)

Division of Systems Research
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555

10. SUPPLEMENTARY NOTES

11. ABSTRACT (200 words or less)

In October 1990, the Nuclear Regulatory Commission (NRC) sponsored a workshop to develop a proposed methodology for use by the NRC in determining the effectiveness of nuclear utility training. The workshop developed a framework on which to base a methodology which draws together current NRC and nuclear industry processes and initiatives in training evaluation and plant performance monitoring. The framework recognizes that utilities, under current NRC and industry guidance, operate closed-loop training systems that incorporate methods for self-correction. The model proposes that by monitoring/sampling indicator data at various points in the utility's closed-loop system, the NRC can determine whether the loop is operating properly to maintain training program effectiveness. This training loop includes the training process, the performance of trained workers, and plant operators. Monitoring/sampling of indicators is planned such that each indicator provides data which complements data derived from the other indicators. Agreement between indicators is used to confirm either effective training or to detect training problems. Inconsistency between indicators triggers further investigation.

12. KEY WORDS/DESCRIPTORS (List words or phrases that will assist researchers in locating the report.)

training effectiveness
performance-based training
systems approach to training

13. AVAILABILITY STATEMENT

Unlimited

14. SECURITY CLASSIFICATION

(This Page) Unclassified

(This Report) Unclassified

15. NUMBER OF PAGES

16. PRICE

NUREG/CR-3114

FINDINGS OF A WORKSHOP ON DEVELOPING A METHODOLOGY FOR EVALUATING EFFECTIVENESS OF NUCLEAR POWER PLANT TRAINING

APRIL 1992

UNITED STATES
NUCLEAR REGULATORY COMMISSION
WASHINGTON, D.C. 20555

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

SPECIAL FOURTH-CLASS RATE
POSTAGE AND FEES PAID
USNRC
PERMIT NO. G-67

120555139531 1 1AN1RX
US NRC-OADM
DIV FOIA & PUBLICATIONS SVCS
TPS-PDR-NUREG
P-211
WASHINGTON DC 20555