THESIS

A MODEL FOR EVALUATION OF EMPLOYEE TRAINING PROGRAMS

RICHARD DUANE JONES

1978

A MODEL FOR EVALUATION OF EMPLOYEE TRAINING PROGRAMS

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment for the Degree of

Doctor of Philosophy

by

Richard Duane Jones

May 1978

A MODEL FOR EVALUATION OF EMPLOYEE TRAINING PROGRAMS

Richard Duane Jones, Ph.D
Cornell University 1978

The continual growth of industrial training has resulted in training assuming a more important role in determining success of businesses. Yet improvement of the training function has been difficult because of the lack of a comprehensive and effective evaluation design. Existing attempts at training evaluation are fragmented, lack specific criteria, do not follow an experimental design or fail to identify specific courses for improvement.

Purpose - The purpose of this study is to develop a model for evaluation of employee training programs.

Methodology - The basic components for the proposal model originated from an extensive review of existing models and studies in education and training evaluation. A number of training directors were surveyed to identify specific evaluation needs and constraints. The proposed evaluation model and

suggested techniques were field tested in an existing business training program.

Findings - The most important areas of evaluation, as indicated by training directors, were in identifying needs for programs and determining changes in behavior or improvements in performance. The most common types of evaluation used are reaction surveys and achievement tests. The most difficult aspect of evaluation to measure is behavior or performance. Directors further indicated a primary concern in evaluation is the lack of appropriate standards. Finally, the most significant factors that limit their progress in evaluation, is lack of expertise and resources.

The proposed evaluation model in the study emphasizes Summative evaluation of a total program based on measurement of job behavior. Complimenting this Summative evaluation is a Formative evaluation of the Planning and Process of training. Only by collecting information about training components can a decision-maker begin to identify changes for improvement. Finally, the model identifies the sources of evaluation data and standard for making comparisons.

Approximation of job performance through an attitude scale

is proposed as a follow-up evaluation tool when it is too difficult to obtain specific job performance data. Instruments for use in this technique were field tested in a training program.

Techniques and procedures are also suggested for use of the model and conducting evaluation of the Planning and Process of Training.

The study takes an initial step in improving evaluation of employee training programs. Further research and validation should build on this effort. Evaluation can be manageable and contribute significantly to improving programs in employee training if they are planned and follow the logical sequence outlined in the proposed model.

BIOGRAPHICAL SKETCH

Richard Duane Jones was born December 27, 1947. He grew up on a dairy farm near Hamilton, New York. His aspiration to become a veterinarian led him to Cornell University after graduation from high school in 1965.

That ambition faded in the reality of anticipated years of intensive academic study. He turned his interest to the Future Farmers of America organization. Taking two consecutive leaves of absence from Cornell, he served as State President and National Vice President of the FFA.

After returning to Cornell in 1968, he continued his studies in agricultural education. He received his Bachelor of Science degree in 1971. Following graduation, he was employed as a teacher of agriculture at Vernon-Verona-Sherrill Central School, Verona, New York. In addition, he was also working at home on the dairy farm. Teaching became more interesting than farming and he moved to Holland Patent Central School, Holland Patent, New York. There he enjoyed teaching students as well as agriculture from 1972 - 1975.

In 1974 he married Kathleen Sportelli of Syracuse, New York, probably the best decision of his career.

In 1975 he returned to Cornell University in Graduate School, eventually pursuing M.S. and Ph.D. degrees. While at Cornell he

served as a Research Assistant and Teaching Assistant in Agricultural Education.

He is currently employed as Associate in the Bureau of Agricultural Education, New York State Education Department.

ACKNOWLEDGEMENTS

I would like to acknowledge the assistance, cooperation and encouragement of several individuals who helped tremendously in the completion of this thesis.

Dr. William E. Drake, Chairman of my graduate committee, has been a trusted advisor for many years. I thank him for his advice, constructive suggestions and encouragement to develop a better study and personally learn in the process.

Dr. Charles C. Russell, minor committee member, always provided enthusiasm and meticulously and artfully suggested improvements in the thesis.

Dr. Wendell Earle, minor committee member, has continuously encouraged me to think and do beyond my immediate vision.

I thank Mr. Robert Engfer, and Mr. Charles Collins, Agway, Inc., for their cooperation and assistance in field testing evaluation instruments.

Finally, I wish to thank my wife Kathy for her love, patience, and constant encouragement, which made an impossible task, possible and an arduous journey, enjoyable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

NATURE OF THE PROBLEM

Introduction

Public education is one of the largest industries in this country requiring an annual investment of 120 billion dollars,[1] however it does not have a monopoly on education. Expanding technology has fostered the growth of education in all aspects of our society in addition to public schools and colleges.

It is estimated that currently 82 million people are enrolled in non-degree education programs compared to only 67 million in degree programs in colleges and the public and private school systems. This increase in non-degree programs is evidenced by the fact that in 1970 there were only 60 million persons in non-degree programs compared to 64 million in formal education.[2]

A large segment of non-degree education is in the form of training conducted by major businesses and industries. A survey by The Conference Board, of 610 firms that employ over 500 persons, showed they spent a total of $2 billion on training during 1974-75. During that year, a total of 17.8% of the 32 million employees in these companies, participated in either an

---

1/ Statistical Abstracts of the United States, U.S. Department of Commerce, 1976, p. 117.

2/ Glen B. Davis, "Zero Population Growth: Effect on Adult Education," Adult Leadership, Jan. 1974, p. 245.

in-house training program, outside company program or other tuition-aid program.[1/]

Businesses are continually developing educational programs to upgrade, retrain and develop employees. Training has become a significant function in the successful management of most large organizations.

The Conference Board Report went on to state:

> "Education and training is seen by management in more and more instances as an investment in human capital - an instrument for profit, growth, and corporate vitality - rather than as an onerous cost." [2/]

Education programs in the business setting are often referred to as training; probably because the content of original programs was very skill oriented in providing employees practice in developing specific tasks. While the term training has remained, training programs have become more sophisticated and very similar to general education. Richard Marcotte, President of the 18,000 member American Society for Training and Development, states:

> "Training is not only a growing profession in terms of numbers, but also in scope and knowledge span. Today's professional is involved in all aspects of changing human behavior and performance of management." [3/]

1/ "Education in Business," The Conference Board, 1976.

2/ Ibid.

3/ "Training Talks to the Top," Training, Vol. 14, No. 10, October 1977, p. 29.

Training today includes programs in human relations, management development, leadership, and highly technical skills like computer programming and electronic engineering. Increased technology and responsibilities within organizations have increased the demand for more and improved training programs. Currently, formal training programs are an integral part of most all companies and available to all organizations through many different consulting firms.

Management development is becoming an increasingly important segment of the training field. Many large firms are conducting in-house programs and others are using one or more of the many consulting firms, trade association or college programs. (American Management Association, the largest management trade association, grosses well over $10 million annually from over 1,600 seminars.) In all, there are approximately 18,000 trade associations and consultants and more than 2,000 private and public educational institutions conducting seminars in business and management development.[1/] Barton-Dobenin and Hodgetts concluded from a study of 385 business firms on the scope of management training programs that there would be a continuous increase in the number of these programs even though there is no concrete

---

1/ "There's No Business Like the Seminar Business," Dun's Review, September, 1967, p. 36.

evidence as to their effectiveness.[1/]

The types of training programs can be divided into four basic categories. First, is orienting the new employee. This involves acquainting all new employees with the business operations, organization, company philosophy and usually benefits and services. There may be several levels of these programs depending on the job level of new employees.

The second type is skill training. However, no longer are skills taught that are exclusively simple or routine. They may include computer programming, accounting, or equipment modification. Sales training is another of the highly sophisticated skills that employees are learning. The distinction of sales training is that employees develop human relations skills rather than typical manual skills.

The third type of training is management development. Management in every organization is emerging as an important and distinct occupation. Decision-making and effective allocation of resources are critical responsibilities in any organization. The current trend is to encourage the complete development of existing employees as successful managers. Consequently, a large number of firms are now conducting management development programs to achieve these goals.

1/ J. Barton-Dobenin and Richard M. Hodgetts, "Management Training Programs: Who Uses Them and Why?" Training and Development Journal, March 1975, pp. 34-40.

The final type is general education. This includes human relations training, particularly for supervisors, safety programs for production and transportation workers, and other programs that personally improve the individual. These programs often deal with providing specific skills beneficial to the organization. However, these are best classified as general education because the programs often dwell in the area of affective behavior and frequently benefit the individual beyond his or her specific job responsibilities.

Discussion of the Problem

Allocation of funds for training is basically a judgmental decision. It is logical that better trained employees will be greater assets to any organization. However, it is often difficult to assess the monetary return from an investment in a specific training program. The best available means for determining the effectiveness of training dollars is to insure that training programs are based on needs, effectively planned and evaluated.

Administration of a training program can become an exhaustive responsibility. Unfortunately, the planning and evaluation of training programs can easily be overshadowed by the "nuts and bolts" of carrying out the program itself. Currently, evaluation of training programs is largely a judgmental process based on bits and pieces of information and use

of a variety of evaluation techniques.

A survey by Raphael & Wagner (1972) of training directors supported the premise that there was little consistency in the use of training evaluation methods. They found reaction surveys were the most common surveys used. They summarized their findings as follows:

> "It is obvious that a great deal of dissatisfaction exists among those responsible for training. Program effectiveness can only be determined by well designed evaluations, but from the studies reviewed, evaluation receives the least amount of training efforts." 1/

An earlier study by Catalanello and Kirkpatrick (1968) on the "state of the art" in industrial training evaluation, drew the following conclusion:

> "The evaluation "state of the art" is still in its infancy. This survey of 154 selected companies indicates that most organizations are measuring reactions to training programs. As we consider the important and difficult steps in the evaluation process (i.e. learning, behavior, and results) we find less and less being done and many of these efforts are superficial and subjective." 2/

1/ Michael A. Raphael and Edwin E. Wagner, "Training Surveys" Training and Development Journal, December 1972, reprinted in Evaluating Training Programs, D. L. Kirkpatrick, ed., 1975, p. 295.

2/ Ralph F. Catalanello and Donald L. Kirkpatrick, "Evaluating Training Programs - The State of the Art," Training and Development Journal, May 1968, reprinted in Evaluating Training Programs, D. L. Kirkpatrick, ed., 1975, p. 261.

Thorley (1972) identifies the current techniques for evaluating training programs as the use of opinion surveys, subjective measurement of job performance and a general observation of how people grow in an organization. The first problem he identified was the inadequacy of opinion surveys. Other problems were that outside observations were superficial and rarely scientific and most mail surveys often had low response rates. He concluded by stating that evaluation needs an objective measure of performance and results must be compared to control groups.[1/]

The first problem to overcome in evaluation of training programs is to expand the limited scope of existing evaluations and assess programs on a more comprehensive level. The existing superficial approach to training evaluation is most likely a symptom of some real problem that does prevent trainers from making a more comprehensive and objective evaluation of training programs.

Evaluation efforts in training have been fragmented. Numerous attempts have been made to determine the effectiveness of individual aspects of various training activities. Some of these have been successful and others quite unsuccessful. Yet, there have been few attempts to develop a systematic approach to eval-

---

1/ S. Thorley, "Evaluating Management Development," Training in Business and Industry, Feb. 1972, p. 34.

uation which will provide appropriate feedback to all aspects of training. Moreover, this feedback should be in a perspective that will place the greatest emphasis on the most important aspects of the program. Wentling and Lawson (1975) identify the problem of the lack of an integrated approach to evaluation of training:

> "A common problem with traditional evaluation practices results from a failure to approach evaluation in a systematic manner. Even in cases where evaluation has been conducted on a formal basis, activities themselves have not been properly integrated." 1/

Another major problem facing training directors is that of establishing objectives or criteria for measuring training results. These objectives not only include expectations for ultimate organizational performance but many of the immediate objectives of participant learning, reaction and effectiveness of selected strategies.

A summary of interviews with members of the American Society for Training and Development by Blumenfeld and Crane (1974) identified this problem.

> "Typically, not enough attention is paid to the criterion, resulting in meaningless inadequate research. For example, one of the most popular

---

1/ Tim L. Wentling and Tom E. Lawson, Evaluating Occupational Education and Training Programs, 1975, p. 21.

> and least meaningful criteria measures is participant questionnaires for rating such factors as course structure, content and quality of instruction." 1/

The article goes on to say that trainers find this information from common reaction surveys as satisfying to "trainer-ego" because the results are invariably positive.

A model and guideline for evaluation of training must give greater direction to the types of criteria that should be used in training evaluation. Obviously each separate training program will need unique objectives but there are several common characteristics of good instruction which could be the basis for a generalizable evaluation design.

Another critical problem identified by Blumenfeld and Crane is the need for a minimally acceptable experimental design. 2/ In any evaluation research the inclusion of a pre and post test and a control group add to the complexity of evaluation. However, these research techniques offer the benefit of making results more definitive and generalizable.

Dubin, Mezack and Neideg (1974) analyzed reported evaluation studies in training. They specifically examined the apparent influence of using an experimental design in the study. They made

---

1/ Warren S. Blumenfeld and Donald P. Crane, "Opinions of Training Effectiveness: How Good?" reprinted in Evaluating Training Programs, D. L. Kirkpatrick, ed., 1975, p. 300.

2/ Ibid, p. 300.

the following conclusion:[1]

> "The implementation of a more complete experimental design does not appear to reduce the probability of obtaining significant results. On the countrary, with respect to the 16 studies, better designs improved evidence of effective training. It is therefore suggested that future management development evaluation might well benefit from the application of more complete design to allow the experimenter to tap the extent, source and meaningfulness of change."

Existing evaluation techniques also fail to provide information directly applicable to the improvement of a training program. Wentling and Lawson (1975) point out that frequently evaluation may show who or what was at fault in an ineffective program, but they fail to show direction to take in improving the program.[2] An ideal evaluation should somehow give direction to specific aspects of training that need improvement and also what form that improvement might take.

The problems already described are sufficient in the author's opinion to warrant further research into evaluation of training programs. Yet, there is another problem that increases the need for improving evaluation. Evaluation techniques will never realistically be adopted by training personnel if they demand

---

1/ S. S. Dubin, M. Mezack and R. Neideg, "Improving the Evaluation of Management Development Programs," Training and Development Journal, June 1974, reprinted in Evaluating Training Programs, D. L. Kirkpatrick, ed., 1975, p. 308.

2/ Tim L. Wentling and Tom E. Lawson, Evaluating Occupational Education and Training Programs, 1975, p. 22.

too much time, money or personnel. Even if an ideal evaluation system could be perfected to solve all of the needs in training evaluation, it would likely be doomed to failure because of lack of acceptance by training directors. Evaluation must be conducted by individual trainers, and any system that is effective must have the support of individual trainers. To accomplish this training evaluation must be easy to use and understandable by training personnel.

Most training directors realize the potential improvements resulting from evaluation but admit there are many obstacles to effective evaluation. A group discussion on training evaluation by training directors reported some of these concerns in the BACIE Journal. In answer to the question, "What are the obstacles preventing us from making more headway (in evaluation)?," they identified the following items:[1/]

> "Lack of expertise and effort in objective setting and evaluation, arising from a lack of understanding that evaluation is an intrinsic part of total training activity;
>
> "Lack of resources;
>
> "The pressure to satisfy training needs which are considered obvious and not, therefore, demanding a more systematic approach, including validation and evaluation;

1/ R. J. Ayres, "Training Thoughts in a Think Tank," BACIE Journal, Vol. 28, No. 7, July 1974, pp. 83-86.

"The sheer magnitude of the task. This can inhibit some from tackling the problem in a forthright way;

"The problem of setting initial criteria in measurable terms, which calls for awareness of and access to necessary data base;

"Management with a 'school-like' attitude to training does not, therefore, expect it to be validated or evaluated;

"Low expectations of training from management, and therefore, little pressure to check results;

"Low status of training and of trainers can affect the training officer's ability to pressure management into a clear concept of its expectations."

Management of the training function is a demanding task. This discussion reveals some of the problems not only associated with evaluation but the entire management of training. Effective evaluation and the resulting improvements in planning, implementing and conducting training programs can begin to increase the effectiveness of the training role within the organization. This becomes a bit of a "vicious circle" of not being able to have an impact on management until a program is evaluated and not being able to effectively evaluate without the support of management. Yet, the problem can be solved by developing a sound plan for evaluation that management should accept. Evaluation designed as part of the training program and developed in a systematic fashion, rather than a hit or miss observation of parts, should be supported by higher levels of management. Attempts do need to be made to conduct some

assessment in order to make some definitive statements about the results of training efforts.

The training setting in business and industry has unique characteristics that differentiate it from other educational settings. For example, training deals exclusively with adult learners, programs are intensive sessions of a few days or hours in duration, and programs specifically relate to a worker's daily job responsibilities. The most effective evaluation system for training should consider and satisfy these unique characteristics and needs of training in business and industry.

A wealth of research in formal educational evaluation offers many models and techniques. However, few have been specifically related to the training setting. A contribution of education is needed to solve the problems outlined and adapt existing research in educational evaluation to the unique aspects of the extensive education programs in business and industry.

Statement of the Problem

The continual development of industrial training has resulted in training assuming a more important role in determining the success of business. Yet improvement of the training function has been difficult because of a lack of a comprehensive, effective evaluation design. Existing attempts at evaluation;

* are fragmented and lack a comprehensive approach
* do not establish criteria for measurement
* have emphasized reaction surveys without specific purpose
* have not established a minimally acceptable experimental design
* failed to identify specific courses for improvement
* are often awkward, complex, impractical and unused

If training is to improve it must be evaluated in a systematic, specific and practical manner.

Purpose

The purpose of this study is to develop a model for evaluation of employee training programs.

Objectives

The purpose of the study will be accomplished by meeting the following specific objectives:

1. Review and analyze existing evaluation models and techniques.
2. Survey training directors to determine specific evaluation needs and constraints in employee training programs.
3. Propose an evaluation model appropriate for employee training programs.
4. Develop and field test instruments for completing evalu-

ation using the proposed evaluation model.

5. Develop guidelines for using the evaluation model in appropriate training programs.

Discussion of Definitions

The term evaluation means in its simplest form to "determine or fix value."[1/] We evaluate most every action we take from whether we got enough sleep the night before to whether we accomplished anything from our daily efforts. However, the process we go through to make that evaluation, is quite different depending on the activity. For example, a farmer evaluates the "worth" of planting corn by measuring the yield or gross receipts from the crop. On the other hand, a teacher must use less direct and much more complex criteria to measure how much a group of students may have learned.

Worthen and Sanders (1973) offer a simple and broad definition in a discussion of educational evaluation. They define evaluation as;

> ..."the determination of the worth of a thing. It includes obtaining information for use in judging the worth of a program, procedure or objective, or the potential utility of alternative approaches designed to attain specific objectives."[2/]

1/ Webster's Collegiate Dictionary, 1967 edition.

2/ Blaine R. Worthen, and James R. Sanders, Educational Evaluation: Theory and Practice, 1973, p. 19.

Worthen and Sanders point out that their definition does not include the actual collection of evaluative information, the description of programs or monitoring of ongoing programs. They feel that these are important activities that an evaluator might undertake but they are not considered evaluation per se.[1/]

Sufflebeam et. al. (1971), in their noted research on evaluation, defined evaluation as "the process of delineating, collecting and providing information useful for judging decision alternatives."[2/] This definition emphasizes the collection of information for evaluative decisions. It places the evaluator in the position of deciding what information should be collected, actually collecting the information, and providing the information to another party. Therefore, another person is the one that then judges the program. This author feels Stufflebeam's definition is a little too broad and the emphasis in evaluation should be placed on judging and not collecting the information.

Evaluation of training in Great Britian is defined in an interesting fashion which aids in making a distinction between different aspects of evaluation. The British Glossary of Training Terms defines evaluation as follows:

1/ Ibid, p. 38.

2/ D. L. Stufflebeam, et. al. Educational Evaluation and Decision-Making, 1971.

> "Evaluation: the assessment of the total value of a training system, training course or programme in social as well as financial terms. Evaluation differs from validation in that it attempts to measure the overall cost-benefit of the course or programme and not just the achievement of its laid down objectives. The term is also used in a general judgmental sense of the continuous monitoring of a programme or of the training function as a whole."[1]

This definition is very narrow in scope because it only applies to the final results of a training program. Hamblin (1974) points out that it is very idealistic, for it is difficult to quantify "social value" or "cost-benefit".[2] If this were the only type of activity considered under the definition of evaluation, Hamblin states there would be few evaluation studies.

Evaluation, if it is to be useful to a training director, must have some measure of the worth of the specific activities that make up a training program in addition to the total effects of the program.

An analogy of the Apollo Space program can illustrate this point. The objective of this program was to safely place a man on the moon. Evaluation of that program under the narrow definition might have been to simply take periodic measures to see if there was really a man on the moon. Once an observation

---

1/ Department of Employment, British Glossary of Training Terms, London, England, from The Psychology of Training, Stammers and Patrick, 1975, p. 116.

2/ A. C. Hamblin, Evaluation and Control of Training, 1974, p. 12.

verified that a man was safely there, the program could be termed a success. Under this type of evaluation, scientists might have made a million launchings before the objective was achieved. However, the program was successful in a very few well planned attempts because each separate phase of the program was evaluated. This comprehensive and systematic evaluation provided feedback on every step thus allowing scientists to make adjustments during the program with the end result of success on the first full attempt. The same can be said of any training programs. If the parts are evaluated and improved then there is a greater likelihood the end result will be more successful. Evaluation of training is an attempt to measure the worth of training. In order to accomplish that, information is collected on which to judge all parts of the program as well as the ultimate objective.

The British do recognize these additional aspects of evaluation only they are described under validation. The Glossary of British Training Terms defines validation as follows:

> "Validation (of a training program): 1. Internal validation: a series of tests and assessments designed to ascertain whether a training programme has achieved the behavioral objectives specified. 2. External validation: a series of tests and assessments designed to ascertain whether the behavioral objectives of an internally valid training programme were realistically based on accurate initial identification of training needs in relation to the criteria of effectiveness adopted by the organization."

Internal validation parallels what most evaluators would term as student evaluation or learning assessment, measuring achievement in comparision to the stated objectives. External validation is an activity most trainers would perform in the assessment of needs for a training program, evaluating whether there is a definite need.

Another distinction that should be discussed relative to the definition of evaluation is the approach taken by Scriven (1967).[1/] He separated evaluation into two separate categories based on the purpose of the evaluation in relation to the program. This includes aspects like validation of instructional units, and teacher observation, where the intent is to suggest improvements which will increase the impact activities. This Scriven termed as Formative Evaluation. The second type is called Summative Evaluation which is a look back at a completed program. Follow-up studies and external evaluations are examples of Summative Evaluation. While not all evaluation activities fall neatly into one of the definitions, this does point out the different aspects and purposes of evaluation.

Other distinctions on evaluation will be discussed in Chapter IV. In this study, the author will establish a relatively broad definition of evaluation. This will include the judg-

---

1/ Michael Scriven, "The Methodology of Evaluation," in Curriculum Evaluation, R. E. Stake, ed., 1967.

ment of overall effects as well as individual activities and also include the activity of collecting information that will aid in making those judgments. The judgment of worth of a training program cannot be made by simply looking at the ultimate ends. Training programs are not wholistic activities but a collection of activities that contribute toward the ultimate objective. The goal of improvement can only be achieved by fine tuning the machinery and not simply scrapping those that fail and preserving only those that work.

Evaluation as defined by this study will include the judgment of worth, the collection of information as well as the systematic planning of these activities. Pyatte does an excellent job of defining evaluation, combining several of the aspects already discussed. He defines evaluation as:

> "Evaluation is the deliberate act of gathering and processing information according to some rational plan the purpose of which is to render at some point in time a judgment about the worth of that on which the information is gathered.[1]

Primarily to avoid adding one more definition to the literature that is only distinguished from others by a word or two, this study will use Pyatte's definition of evaluation as a basis for discussion.

---

[1] Jeff Pyatte, "Functions of Program Evaluation and Evaluation Models in Education", High School Journal, vol. 53, no. 7, April 1970, p. 387.

There are several definitions of training. Hamblin (1974) defines training as "any activity which deliberately attempts to improve a persons skill in a job". Hamblin differentiates training and education by noting that education "is mainly concerned with personal development" in contrast to direct job relevance.[1/]

Realistically it is difficult to distinguish between training and education for the two do overlap. Glazer (1965) pointed out a difference between training and evaluation based on two criteria: (a) the degree of specificity of objectives, and (b) on minimizing or maximizing individual differences. Training has more specific objectives and attempts to minimize individual differences. He concluded,

> "Training and education are two aspects of the teaching process; the two terms refer to two classes of the teaching process that are not mutually exclusive. Certain dimensions which form the continuum along with the distinctions fall are specificity of behavioral goal and uniformity vs. individual development. Although one may wish to distinguish between training and education in terms of behavioral goals and the methods of attaining them, the technological practices required to carry out either are built upon principles for modifying, developing and grading behavior that are generated from behavioral research."[2/]

1/ A. C. Hamblin, Evaluation and Control of Training, 1974, pp. 6, 7.

2/ R. Glazer, ed. Training Research and Education, 1965, p. 5.

Some authors, e.g., Nadler (1971),[1] have emphasized differences between training in that some are concerned with present jobs and other programs are more applicable to future jobs, e.g. management development.

The British Department of Employment's Glossary of Training Terms offers the following definition of training:[2]

> "Training: The systematic development of the attitude/knowledge/skill behavior pattern required by an individual in order to perform adequately a given task or job."

This definition is the basis of the one developed by this author. Training and education are different but there are enough similarities in techniques as pointed out by Glazer that much of the behavioral research in education can be applied to the training setting.

Training of employees include orientation to the job, formal classroom training and frequently on-the-job training. However, the study will not include on-the-job training as part of the training discussion. Only those formal training activities where the employee is temporarily removed from their job responsibil-

---

1/ L. Nadler, "Using Critical Events to Develop Training Programme," Supplement to Industrial Training International, vol. 6, no. 4, 1971.

2/ Department of Employment, Glossary of Training Terms, 1971, from The Psychology of Training, Stammers and Patrick, 1975, p. 10.

ities will be discussed. On-the-job and classroom training are significantly different and consequently cannot be discussed with a similar evaluation model. However, an effective classroom training program must relate to on-the-job realities.

Therefore, the following definition of training will be used: "the planned development of an individual's attitude/knowledge/skill behaviors through off the job activities."

Definitions

In order to establish some common ground for discussion, the following definitions are used for key terms in this study:

Training: The planned development of an individual's attitude/knowledge/skill behaviors through off the job activities.

Evaluation: The deliberate act of gathering and processing information according to some rational plan the purpose of which is to render at some point in time a judgment about the worth of that on which the information is gathered.

Model: A graphic representation of the concepts of a process.

CHAPTER II

SIGNIFICANCE OF THE STUDY

Identification of the significance of this study will begin with an analysis of existing training evaluation models. A review of reported research in evaluation will also indicate the degree to which current efforts are meeting training needs. Those potential areas for improvements will designate where this study can make a significant contribution.

Existing Training Evaluation Models

One of the major efforts to define a model for evaluating training programs was developed by Kirkpatrick (1960).[1/] This model is still used by many evaluators as a framework for categorizing different aspects of training evaluations. Kirkpatrick summarizes four levels of evaluation: Reaction, Learning, Behavior, and Results.

Reaction is defined as how well the trainers liked the particular program. Kirkpatrick states, "To evaluate effectively, training directors should begin by doing a good job of measuring reactions and feelings of people who participate."[2/]

1/ Donald L. Kirkpatrick, ed., Evaluating Training Programs, 1975, pp. 1-14.

2/ Ibid, p. 4.

Learning evaluation is defined as the principal facts, techniques and attitudes that were understood by the participants. In this level Kirkpatrick points out that measurement should be compared to objectives, quantified if possible, compared to pre-test results and control groups. This type of evaluation is most frequently used in public education; assessing what students have "learned" from the program.

Level 3 in Kirkpatrick's model is Behavior, which examines how the program participants behavior is changed on the job. Since learning is generally defined as a change in behavior, participants who gained something from a training program should show some modification of behavior in their job responsibilities.

The ultimate objective of training in business and industry is to improve organization results. The last level of evaluation, defined by Kirkpatrick, is to determine if the training program had an effect on organization performance. This is obviously the most difficult level to evaluate. Many factors influence organization results and it is very difficult to assess the effects attributable to a specific program.

Kirkpatrick's model is very easy to conceptualize and makes a great deal of sense in categorizing the different aspects of evaluation. It has been used as the framework for describing a great number of the training evaluation activities in this country over the last 15 years.

Numerous evaluation studies have attempted evaluation techniques at either the Reaction, Learning, Behavior or Results level. While Kirkpatrick's contribution has been useful in categorizing evaluation techniques, there are several weaknesses in this framework as a model for evaluation of training program. First of all, it fails to put evaluation in perspective. It is easy to ask questions, "Are all levels of equal importance? Which level should come first? Should all levels be attempted to perform a comprehensive evaluation?"

Kirkpatrick's model also gives little direction to training personnel in determining objectives. By offering the flexibility to include nearly all evaluation efforts, the model does not give enough specific direction as to what criteria should be used to "determine the worth" of programs.

Finally, the Kirkpatrick model does not examine the internal process of training to determine the effectiveness of specific activities. Evaluation in this model is the final assessment of the program which will most likely determine the degree to which the program was successful but it will contribute very little understanding to the why and how.

The review of related models in training evaluation travels across to Great Britain to discover other points of view in evaluation. The efforts of Warr, Bird and Rackham (1970) out-

lined a similar system to Kirkpatrick only terming the four levels as Reactions, Immediate, Intermediate and Ultimate evaluation.[1/]

Reactions are measured during a training program. Immediate results are measured at the conclusion of the program and compared to the program objectives. Intermediate evaluation is a follow-up evaluation of participants to measure job behavior and is compared to the identified needs for the program. Ultimate evaluation is a later measurement determining organizational performance compared to organizational goals.

While this model is also sound, the labeling of levels may be inaccurate for some types of training. Some objectives may not fall easily into a chronological format of immediate, intermediate and ultimate. This model, like Kirkpatrick's, fails to give direction in whether all programs should be evaluated on all four levels, and if they all are equally important.

The one significant contribution of the Warr, Bird, and Rackham model is the emphasis on comparing the observations to objectives. All training programs should have organizational goals, specific assessed needs and training objectives. Results at each level should be compared to the appropriate objective

---

1/ P. B. Warr, M. W. Bird and M. Rackham, Evaluation of Management Training, 1970.

and used to determine progress.

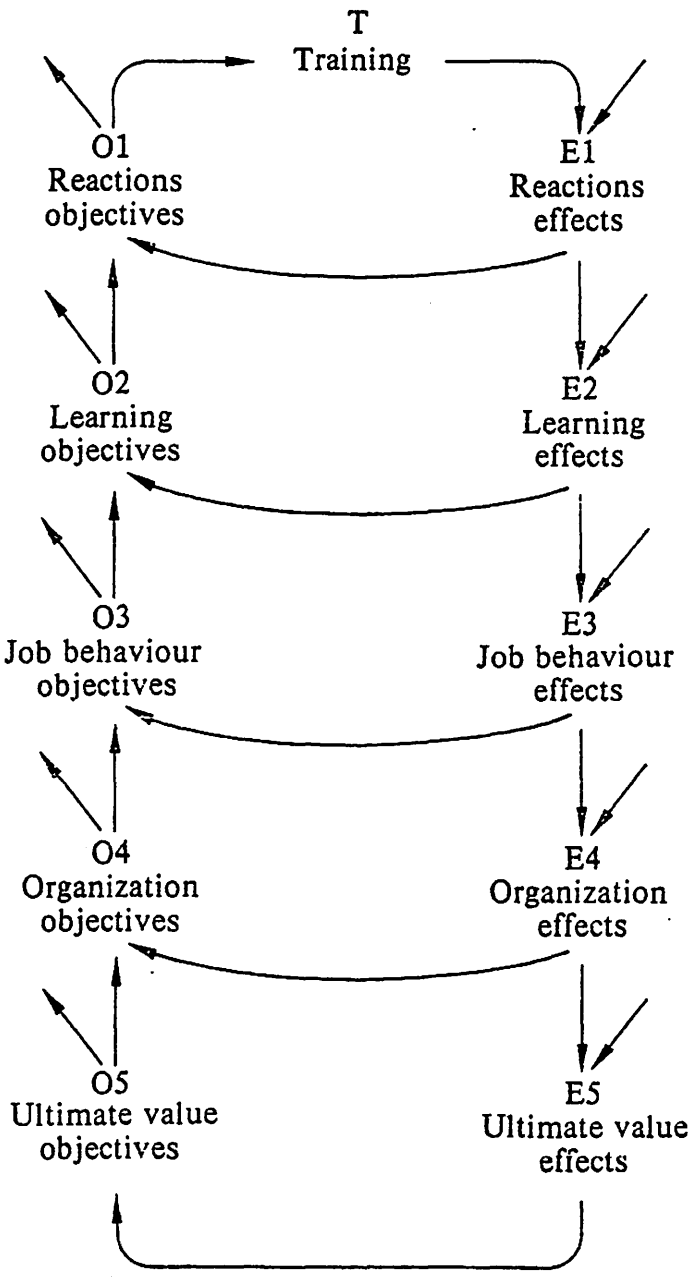
Hamblin (1974) builds upon these two models by using the initial terms of Kirkpatrick and adding a fifth level of Ultimate value.[1/]

Hamblin emphasizes that these five levels of evaluation form a cycle that training personnel can enter or leave at any time depending on the training program. This model is a little more complicated to visualize and therefore is exhibited in Figure 1.

At each of the five levels some information is collected and compared to the specific objectives for that level. Reaction, Learning, Job Behavior and Organization are similar to the levels previously described by Kirkpatrick. Hamblin's fifth level of Ultimate Value is the effect of training that goes beyond the specific expectations of the training program. This includes several areas of special benefits that result from training but were not directly anticipated from the planned training activity.

1/ A. C. Hamblin, Evaluation and Control of Training, 1974.

FIGURE 1. HAMBLIN'S MODEL OF EVALUATION



SOURCE: A.C. Hamblin, Evaluation and Control of Training, 1974, Appendix.

For example, the ultimate effects, in an organization that has attempted to decrease accidents through a training program, might be increased profits because of the resulting reduction in

lost time due to accidents. Ultimate effects are not always expressed in monetary terms. A community-minded organization might be able to provide greater service to the community as an ultimate result of a time management training program that increased the efficient use of time by persons in the organization. Frequently ultimate objectives involve personal achievement where an individual may move to a higher level of the organization beyond the specific objective of the training program.

One important aspect of this model by Hamblin is that it emphasizes a continuous cycle of training which starts by establishing objectives, which leads to the actual training, which leads to measurement of job behavior which returns to establishment of objectives. Training efforts should be a continuous formative process.

A second application of this model is the cycle of evaluation and training activities in which the trainer is involved. Objectives are set at each level prior to the training and in a sequence of the five levels, evaluation takes place after the training.

The chief advantage of this model is the flexibility suggested by Hamblin. He states there is no one entry point and no end point to the evaluation. The model is a map giving

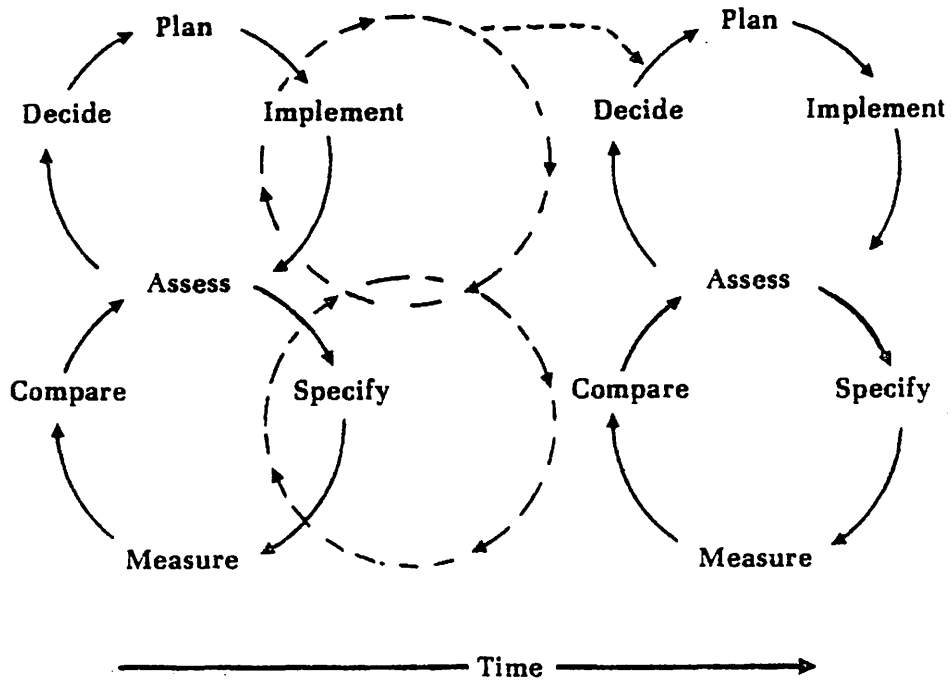
training personnel a direction to follow, starting at the point at which they entered. Hamblin suggests the routes that should be followed and the specific evaluation will depend on the type of training program and the desires of the organization.

This model is superior to the Kirkpatrick model and the Warr, Bird, and Rackham model because it emphasizes the establishment of objectives. It suggests that the trainer establish reaction objectives, learning objectives, etc. Unfortunately, this requires a great deal of additional effort for the trainer. A good training program should have organization and perhaps behavior level objectives in mind but learning and reaction are only intermediate steps in the process of creating some type of effect from the training program. The Hamblin model, like Kirkpatrick and Warr, Bird and Rackham, fails to look at the internal processes of training.

Van Maanen (1973)[1/] proposes a "process of program evaluation" for training evaluation as a continuous cycle of two concentric circles. In order to aid explanation, a copy of this model is shown in Figure 2.

1/ John Van Maanen, The Process of Program Evaluation, 1973, pp. 7-16.

FIGURE 2. VAN MAANEN'S MODEL OF EVALUATION



SOURCE: John Van Maanen, The Process of Program Evaluation, 1973, p. 15.

In the initial circle are the primary activities of evaluation. Regardless of the program, all go through a cycle of Planning, Implementation, Assessing and Deciding. He expands the model by adding the second circle as part of the Assessment stage. This includes the separate activities of Specifying, Measuring, Comparing. This is a relatively simple and broad model which can be applied to such specific activities as selecting instruction media, or to the broad activity of planning an entire training program.

The model does explain the decision-making process quite well. However, it is too broad to give much specific direction to training personnel in planning specific evaluation activities. The one significant contribution of this model is to place the cycle of evaluation in a perspective and proper sequence. Van Maanen suggests his model as a framework for decision-making and will undoubtedly contribute to designing a more specific training evaluation model.

The most popular current education evaluation model is the CIPP model developed by Stufflebeam et. al. (1971).[1/] This model is described in the context of training and occupational education by Wentling and Lawson (1975).[2/] CIPP is an acronym for Context, Input, Process and Product Evaluation, which are the four levels of the CIPP model. Wentling and Lawson define four basic decisions a trainer must make which relate to the four levels of evaluation under the CIPP model. Context evaluation relates to planning decisions. Input evaluation leads to programming decisions. Process evaluation results in implementing decisions and Program evaluation results in decisions about

---

1/ D. L. Stufflebeam, et. al., Educational Evaluation and Decision-Making, 1971.

2/ Tim L. Wentling and Tom E. Lawson, Evaluating Occupational Education and Training Programs, 1975, pp. 24-29.

recycling the program. This is a fundamental concept of the CIPP model it, uses a broad definition of evaluation and is designed to provide information for most all decisions related to a training program.

Context evaluation defines the environment in which a program will take place, it seeks to determine needs, constraints, problems underlying those needs and also existing opportunities for satisying those needs. In training this includes assessment of needs which could lead to training objectives as a means of satisfying those needs.

Input evaluation is the process of identifying and assessing various alternative methods of delivering the training. What methods are available or could be developed? How much will each alternative cost? What is the potential of success for each alternative? Input evaluation seeks to provide information to contribute to making good decisions in response to questions such as these.

Process evaluation is designed to "dectect or predict defects in the procedural design of a program or course during the implementation."[1/] Process evaluation is designed to determine if the actual program is in line with what was planned. Some types of

---

1/ Ibid, p. 27.

information collected here are: Evaluation of instructors, adequacy of facilities and resources, involvement of participants, and timing. Process evaluation can be accomplished by many methods, some of which may include participant feedback and external evaluation.

Product evaluation looks at whether the program objectives were achieved. This may include learner assessment at the conclusion of the program, but more recently training programs have considered the long term objectives related to job and organization performance. All levels of evaluation discussed in the Kirkpatrick; Warr, Bird, and Rackham; and the Hamblin models fall entirely in Product Evaluation as defined by Stufflebeam and Wentling and Lawson.

A strong point of the CIPP model is that is examines the entire process of education and can help to locate areas for improvement. If an evaluation exclusively examined program effects and found that objectives were not achieved, program directors would have difficulty correcting the problem because they would have little information about the internal aspects of the program. This broad definition and system for collection of information is the chief advantage of the CIPP model.

However, the CIPP model is extremely complex and quite general in its components. A typical training director, who

is most probably not a professional evaluator would probably find little practical direction from this model. In addition, the approach of the CIPP model is emphasized in collection of information by outside evaluators, who present findings to program directors for making judgments on the program.

This author feels training personnel will best support and benefit from a training evaluation that is understood and administered by the training department itself. The type of evaluation that can achieve the greatest improvement is self-evaluation. While self-evaluation is always biased and limited in scope, recommendations resulting from self-evaluation are usually accepted and improvements made. An ideal evaluation/improvement system should take full advantage of self-evaluation and develop an evaluation system that would increase the amount of information available to individuals and organizations by which they could perform a useful self-evaluation.

Summary of Existing Models

A review of the principal evaluation models in training evaluation point to a need for additional ideas for more effective evaluation. Specifically, existing evaluation models are limited in the amount of useful information they have outlined. Also, traditional models have failed to take a close examination

of the planning and process of training. The recent CIPP model has filled some of this gap, yet it has not presented specific criteria for training personnel to use in judging programs. Consequently, existing models of training evaluation either fail to provide sufficient direction to training evaluators or they do not emphasize all important aspects of evaluation.

Significant Training Evaluation Studies

There have been a number of training evaluation studies reported through trade publications in the industrial training field and other published research in education. The following discussion will briefly summarize some of the most significant studies and outline the important contributions of each.

First, several studies have made contributions to the evaluation of training programs through assessment of reactions of participants. While this may appear to be the least valid of evaluation alternatives, measurement of reactions is quite popular. In some cases it has provided useful information and may yield some potential contributions for the development of a systematic evaluation design.

Kohn and Parker (1969)[1/] reported on their success in two

1/ Vera Kohn and Tredway C. Parker, "Some Guidelines for Evaluating Management Development Seminars", Training and Development Journal, vol. 23, no. 7, July 1969, pp. 18-23.

separate evaluations of management training programs conducted by the American Management Association. They followed the evaluation model of Kirkpatrick in identifying evaluation areas. In the first study, they selected two different types of programs and assessed participant reaction to the programs. The sample consisted of 2,000 participants divided evenly between the two different types of programs. At the conclusion of the program participants rated their "overall reaction" to the program on a scale of 1 to 20 with 1 equal to poor and 20 corresponding to excellent. This rating scale for the programs was an existing rating scale used by the American Management Association.

In addition participants were asked their reactions to selected aspects of the program. Ratings on these separate variables were on a five point scale ranging from "very satisfied" (effective) to "very dissatisfied" (ineffective). These items on the questionnaire originated from an open ended questionnaire developed in an exploratory phase of the study, which sought to identify "prevailing attitudes towards the program in some depth". Those aspects seemed most important through the open ended questionnaire were used in the reaction survey evaluation.

Examples of the questionnaire items were, "practical value for application" and "comprehensiveness of coverage", under the

broad heading of Program Content. Under the other heading of Registrant Group, typical items were "diversity of group background" and "extent of personal interest".

Participant ratings on a 20 point scale was averaged to determine a mean rating. A correlational analysis was done to determine overlap of variable measures and to draw conclusions about the "adequacy of questionnaire design." Multiple regression analysis was done to determine which combination of variables proved the best predictors of "overall evaluation" and the relative importance of each variable in the combination.

Kohn and Parker reported that the most important contributions to participate satisfaction were: 1) Subject matter had practical value, 2) Balance of background characteristics, (e.g. experience, organizational level, company, type of business) to assure meaningful communication among the learners, and 3) Opportunity for learner participation.

This study, in the opinion of the author, strikes upon a practical format for evaluating training. The techniques used in the study are basically sound, however, they do make the one significant assumption, namely that the 20 point overall rating is a valid measure of participant satisfaction. This, of course, affects the validity of the entire study.

Kohn and Parker were able to determine specific factors in training programs which contribute to favorable participant

reactions. The limitations of this study are that it also assumes that participants can accurately perceive each variable in the questionnaire. A good study should include questions for which participants have adequate information to base an answer. If this exists, a majority of the participants will give consistent responses.

A problem with this type of reaction survey was mentioned in the discussion earlier by a number of training authors which indicates the inadequacy of simply knowing whether participants liked the program. This method is satisfactory if the objective of the program is to have participants enjoy the program, but if the objective is to change job behavior of performance then a reaction survey may not necessarily indicate success in job performance.

The significant contribution of this study is that it may serve as a model for an evaluative instrument which could be improved by including other variables which may give greater indication of job behavior or performance. Another improvement would be to use reaction variables that are proved to be valid indicators of job performance. The result may be an improved design that measures more than just how well participants enjoyed the program.

The second study by Kohn and Parker reported a design of

evaluation of learning based on the Kirkpatrick model. In this study the authors administered a pre-training assessment of "personal background data, the respondent's expectations about the course, attitudes towards management development and the management process," Pre-training responses were also collected on a "Time Distribution Profile" which asked the amount of time they felt a manager should spend on various responsibilities.

During the program, participants were asked to record in a daily journal comments on each topic, usefulness of subject matter and effectiveness of instructional techniques. A post training questionnaire was used to measure the participant attitude and also post responses were measured on the "Time Distribution Profile". A follow-up interview was conducted several years after completion of the program to "determine participant assessment of the program and the ways in which each was able to apply the course material to their job.

Statistical tests were applied to compare pre-test and post-test attitudes and ratings on the "Time Distribution Profile". The authors reported they were satisfied with their findings. They reported the following conclusion on the basis of significant attitude change:[1/]

> "... measuring attitudes on a pre-test and post-test training basis gave us insight into what participants gained from the educational experience as well

---

1/ Ibid, p. 21.

as corroborating evidence, in another realm of measurement, of subjective data."

They found a statistically significant change in the participant responses to the "Time Distribution Profile". This measurement, as was pointed out, is also an attitude assessment rather than a learning assessment because it takes the amount of time that should be spent by a manager rather than the amount of time they actually spend. A learning assessment would have sought to determine how their behavior had changed.

The Kohn and Parker study offer a sound approach to assessing attitude change as a result of training. This approach will be considered as a possible design in the future development of evaluation designs.

Belasco and Trice (1969)[1/] set up a true experimental design for evaluating a supervisory and management training program for 258 supervisors within a single organization. They used the Solomom four-way design to most accurately measure the effects of the training program. The four-way experimental design uses four separate groups, one has a pre-test, the training and a post-test, another has only the training and the post-test, another has a pre-test, no training and the post-test, and the final control group has only the post-test. The four-way design

---

1/ James A. Belasco and Harrison M. Trice, "Unanticipated Returns of Training", Training and Development Journal, vol. 23, no. 7, July 1969, pp. 12-17.

gives a control group comparison and also allows a researcher to determine any effects from a pre-test.

In this evaluation, participants were given a questionnaire on knowledge of the program content, attitudes toward using the concepts, and action statements simulating the program objectives. The results showed very little significant change in the measurement as a result of the training. However, the most significant conclusion made by the authors was a result of the type of experimental design they used. They drew the following conclusions;

> " (1) the changes associated with training alone are small; (2) training serves many unintentional cerimonial functions; (3) the administration of questionnaires before training "opens up" the supervisor and makes him more receptive to the training material; (4) testing is a potent change agent independent of training; (5) one way to improve the probability of change associated with training is through the selection of individuals for training on the basis of the match between their predispositions and the demands of training"[1]

These apparent effects of a pre-test will be considered in the future design of evaluative studies. Also this reported study raises some interesting conclusions that would support many casual observations by industrial trainers, specifically that employees often get more out of a program than the content in program objectives. The fact that employees spend some time away from their job, exchanging points of view with other

---

1/ Ibid, p. 13.

employees and trainers, can have great benefits of providing new ideas, reestablishing enthusiasm or giving encouragement. These unanticipated returns of training should be considered and possibly measured in conducting an effective evaluation of training programs.

Hayes and Williams (1971)[1/] reported on a study of change as a result of a supervisory training program. Changes were measured by comparing responses on a pre-test and post-test of different, but highly correlated, standard tests on supervision. They found a significant change in scores as a result of the program. The authors also correlated changes on several specific personal variables and found that there was a significantly greater change in participants who were younger, and had fewer responsibilities. They found no significant differences on the variables of length of service or extent of academic training.

Mindak and Anderson (1971)[2/] developed a program evaluation

---

1/ William G. Hayes and Eugene I. Williams, "Supervisory Training-An Index of Change", Training and Development Journal, April 1971, reprinted in Evaluating Training Programs, D. L. Kirkpatrick, ed., 1975, pp. 83-86.

2/ William A. Mindak and Robert E. Anderson, "Can we Quantify an Act of Faith", Training and Development Journal, May 1971, reprinted in Evaluating Training Programs, D. L. Kirkpatrick, ed.

for a management development program. Their plan focused on the change in job behavior and attitudes as a result of the program. They selected the semantic differential scale developed by Osgood as the instrument for assessing attitudes. They also used a social perception analysis technique which asked participants to rank various job titles on several criteria.

The sample used in the study was a group of 25 middle-to-top management representatives of large national companies, in an intensive five week training program. Participants were given a pre-test, a post-test at the conclusion of the program and a follow-up 2 months after the program.

The study drew the following conclusions:

> " (1) Management training can be quantified and measured by the use of such tests as the semantic diffential and social perception analysis. Specific shifts in attitudes were registered as a direct result of the course. (2) Changes in attitudes result not only from instruction but also from other experiences in the program. (3) Shifts in attitudes were short lived. External training courses makes sense only when the participants have enjoyed some degree of follow-up. (4) The next step would seem to be to relate quantitative shifts in attitude ... to specific behavioral performance on the job."1/

Again there is the observation of effects of training out-

1/ Ibid, p. 95.

side the specific training objectives. This study also gives support to using attitude measures as a means of measuring program effects. The most significant conclusion of the study is the fact that results are short lived. Training evaluation should measure the results at some point after the conclusion of the program to accurately assess on-the-job application of training objectives. What's more, if training is to be effective it must seek to create carryover of objectives and instructional techniques. Evaluation should assess how effective training has been in creating long term behavioral changes .

The studies examined thus far have used participant reaction and learning to varing degrees to measure the changes as a result of training. The next step as indicated by Mindak and Anderson is to examine changes in behavior. Measuring these changes are more difficult, however, several studies have attempted to examine the changes in job and organizational performance resulting from training.

Thorley (1969)[1/] attempted to evaluate training on the basis of behavioral change. He reported an evaluation of an in-house

---

1/ S. Thorley, "Evaluating An In-Company Management Training Program," Training and Development Journal, September 1969, reprinted in Evaluating Training Programs, D. L. Kirkpatrick ed. 1975, pp. 139-141.

management training program in London. A questionnaire was administered at the conclusion of the program which asked participants to assess how much they had learned and how they expected the course to improve performance. Job behavior changes were assessed by a follow-up interview of immediate supervisors of each of the participants.

Results indicated that a majority of the supervisors felt the performance of trainees had improved, however, most were unsure if the improvements were the direct result of the training program. This comment points to the principal weakness in the study. It is difficult to prove that any progress is the result of the training program unless there is a control group.

An evaluation study of supervisory training was reported by Holder (1972).[1/] In this study trainers and subordinates were sent a follow-up questionnaire to assess improvements in the areas of supervision covered by training program objectives. Participants rated their personal improvement and their subordinates, through an anonymous questionnaire, also rated their supervisors who had participated in the program.

1/ Jack L. Holder Jr., "Evaluation of an In-Company Management Training Program," Training and Development Journal, April 1972, reprinted in Evaluating Training Programs, D. L. Kirkpatrick, ed., 1975, pp. 160-163.

These results showed participants and subordinates reported significant changes in behavior. This study as well as the one by Thorley use observation by other employees as a means of assessing behavior and avoiding personal bias. It does introduce some error, however, in that reliability may be a problem when using different subordinate or supervisor for each observation. These other employees also have biases that might influence accurate observation of job behavior.

In a similar study, Kirkpatrick (1969)[1/] sought to assess behavior changes resulting from a supervisory training program at a management institute. Interviews were conducted 3 - 4 months after the program with both the participants and their supervisors. Participant reactions were also measured at the conclusion of the program and compared to the same responses three months after the program.

Results showed significant changes as a result of the program. In all phases of program content, the participants showed positive changes. This was substantiated by supervisor responses. Reactions to the program were most favorable at the conclusion of the program and declined slightly in the measure-

---

1/ Donald L. Kirkpatrick, "Evaluating a Training Program for Supervisors and Foremen," The Personnel Administrator, vol. 14, no. 5, Sept. - Oct. 1969.

ment at three months following the program.

This study emphasizes importance of follow-up by assessing results after the training program and it utilizes other employee observation of behavior.

An evaluation study by Walker (1972)[1/] had participants in an Air Force supervisor training program identify specific changes they planned to make as a result of the training. A follow-up questionnaire was sent to each participant after six months and asked how many of these specific changes had been accomplished and what barriers there were to accomplishing more. This approach to behavior change evaluation seems to yield very specific information. While it would require a great deal of time and effort to accomplish the evaluation, it does yield practical and specific information that could be greatly beneficial in specific types of training programs. The weaknesses of this study are that it relies on self reporting which could be subject to bias.

A recent study by Kelly (1976)[2/] examined evaluation

1/ Pascal M. Walker, "Evaluation of Air Force Employee Development Specialist Training," Training and Development Journal, reprinted in Evaluating Training Programs, D. L. Kirkpatrick, ed., 1975, pp. 169-172.

2/ Francis J. Kelly, "Methods of Evaluating Public Sector Management Development Programs" unpublished Ph.D. dissertation, State University of New York at Albany, 1976.

techniques appropriate for public management development programs. He divided existing management evaluations into three categories: (1) experimental, (2) cost-benefit, and (3) eclectic. He further divided evaluation into four criteria: (1) Participant Reactions, (2) Learning and Attitude, (3) Behavior and Results, and (4) Process. Combining the categories and the criteria results in a 12 cell matrix into which Kelly proposes to classify all training evaluation studies. Evaluation theory, as Kelly points out, indicates that a majority of efforts should be focused on experimental evaluations using behavior and result criteria. In practice, however, most public sector management development programs are evaluated through eclectic designs using participant reaction criteria. Few experimental evaluations have been attempted and those that have been tried have generally failed to yield conclusive results. Kelly further states, that cost-benefit evaluation methods are still under development and none have been applied successfully to management development programs.

Kelly concludes that;

> ... "further attention should be focused on developing participant reactions and process qualitative evaluations. These methods show promise of providing useful information to training managers and decision-makers." 1/

1/ Ibid, p. 128.

It is interesting that Kelly added process evaluation to the other three areas of evaluation that are generally grouped as product evaluation. Kelly's recommendations will be taken into consideration in the development of a model in this study.

Summary

The review of existing models and reported research has shown there is a need for further research and a fresh approach in the realm of training evaluation. This study will seek to develop a model for training evaluation and suggested techniques for implementing the model. In order to build upon successes of previous research and solve existing problems, the author will keep the following criteria in mind in order to solve some of the current problems in training evaluation and weaknesses in existing training evaluation models.

1. The evaluation should describe the evaluation process in sequence and relative importance of various components.

2. The evaluation model should emphasize self-evaluation and utilize training personnel input in developing an evaluation system.

3. Evaluation techniques should provide specific feedback to trainers on the strengths and weaknesses of the process of training programs.

4. Evaluation techniques should be easy to administer, require a minimum of expertise and consume a minimum of time.

5. The evaluation model should carefully consider the assessment of behavior change and job performance.

6. The evaluation model should use a research design to eliminate contaminating factors.

7. The evaluation model should consider the importance of participant reactions and attitudes.

CHAPTER III

METHODOLOGY

Educational Evaluation Models

Objective one, examining related educational models was accomplished by a review of literature, related studies and research on evaluation. A search of literature was initiated by searching all libraries at Cornell University for monographs on evaluation models in education or training. A computer search was conducted of the ERIC system on the topics of evaluation models and evaluation of training. This was instrumental in locating a number of research publications and journal articles. The computer search was updated by referencing the current ERIC and CIJE indexes. A review of Dissertation Abstracts also located several research studies which contributed to the review of evaluation models and the development of a training evaluation model.

Works of several authors are worth noting for their excellent review of evaluation models. These are Popham (1975),[1/] Worthen and Sanders (1973)[2/] and Steele (1973).[3/] Their analysis

---

1/ James W. Popham, Educational Evaluation, 1975.

2/ B. R. Worthen and J. R. Sanders, Educational Evaluation Theory and Practice, 1973.

3/ Sara M. Steele, Contemporary Approaches to Program Evaluation, 1973.

describes and categorizes most of the significant efforts in educational evaluation.

No attempt was made in this study to make an exhaustive discussion of all evaluation models and designs. Only those studies that contributed unique approaches or had been recognized by several other authors were included in this review.

Training Director Survey

A sample of training directors was identified through the membership list of the Central New York Chapter of the American Society for Training and Development. From this membership a total of seventeen training directors were identified that were heads of training in seventeen separate business organizations. All organizations were businesses operating primarily in the Central New York area.

This group was selected as a sample because these individuals should represent leaders in training programs, because of their involvement in professional organizations. They should be knowledgeable of current developments in training evaluation and could best relate evaluation problems in light of current practices. Each of the organizations they represent should have established training departments which would give them experience on which to base opinions.

A questionnaire was developed to gather information about training programs at these organizations. A copy of the cover letter and questionnaire are included in Appendix A. The first part of the questionnaire sought to determine the size of the training program and the types of training offered. Directors were then asked to describe some of their current training evaluation methods and any evaluation concerns.

The second part of the questionnaire sought to examine the perceived importance of various aspects of evaluation. In order to make this information more relative to the real world of training the various aspects were identified by training decisions. Since evaluation is the collection of information and contributes to decision making, it is logical that the training decisions that are the most important would identify the aspects of training that are the most important.

Eight training decisions were identified by the author based on the evaluation works of Wentling and Lawson (1975),[1/] Kirkpatrick (1967)[2/] and Hamblin (1974).[3/] An attempt was made

---

1/ Tim Wentling and Tom Lawson, Evaluating Occupational Education and Training Programs, 1975, pp. 1-59.

2/ Donald L. Kirkpatrick, "Evaluation of Training," Training and Development Handbook, R. L. Craig and L. R. Bittel eds., 1967.

3/ A. C. Hamblin, Evaluation and Control of Training, 1974.

to state these decisions in an understandable fashion that related more to training rather than educational jargon.

The eight decisions identified were: (a) Determine the need for specific training programs, (b) Determine previous knowledge and skills of trainees, (c) Select appropriate instructional strategies, (d) Determine effectiveness of trainers, (e) Determine if trainees enjoyed the program, (f) Determine the extent of learning at the end of the program (g) Determine the extent of trainee application of skills on the job, and (h) Determine if new skills improve job performance.

The following figure shows how these decisions relate to the levels of evaluation identified by various authors.

FIGURE 3. TRAINING DECISIONS DETERMINED BY EXISTING EVALUATION MODELS

Decision	Hamblin	Kirkpatrick	Wentling & Lawson (Stufflebeam)
1. Determine need for specific training programs	____	____	Context
2. Determine previous knowledge and skills of trainees	____	____	
3. Select appropriate instructional strategies	____	____	Input
4. Determine effectiveness of trainers	____	____	Process
5. Determine if trainees enjoyed the program	Reaction	Reaction	
6. Determine extent of learning at the end of the program	Learning	Learning	Product
7. Determine extent of trainee application of skills on the job	Job Behavior	Behavior	
8. Determine if new skills improve job performance	Organization	Results	

Information was collected in three separate areas for each of these decisions. First, directors were asked their perception of the degree of importance of each of the decisions to their training program. Then, each was asked whether these

decisions were made in their organization and finally, they were asked to identify factors that hinder attempts at collecting information for these decisions.

Importance of each decision was rated on a five-point Likert scale ranging from high to low. Assessment of whether decisions were made was determined by three separate questions. These were: if the decision was made, if they collected information and if the information was adequate.

In the final segment, five factors were identified by the author which could hinder the collection of information through evaluation for these decisions. These were: a) lack of resources, b) lack of expertise, c) lack of cooperation, d) lack of time or, e) decision unimportant. Respondents were also given the opportunity to add additional factors. The directors were asked to check any factor that hindered evaluation for each of the eight decisions.

Training directors were asked to complete the questionnaire frankly and all responses would be kept anonymous. Of the 17 questionnaires mailed, ten were received. Two other directors did respond and indicated that they did not wish to complete the questionnaire. Results of the survey are discussed in Chapter V.

Development of a Model

A model was developed based on information collected in

the review of other evaluation research and the information gathered about training program evaluation. The suggestions made by Kirby (Ph.D. Illinois 1971) were kept in mind during the development of the evaluation model.[1/]

1. The evaluation model should assist evaluators in anticipating all information needed for the decision process.

2. The model should be internally logic and complete.

3. The model should be of sufficient clarity so to allow implementation by a trained evaluator without external interpretation.

4. The model should relate elements in such a way they have not previously been presented.

5. The model should be heuristic.

6. The model should be capable of being extended by empirical study.

7. The model should be efficient.

In order to better explain the evaluation model a description will be made of training and the evaluation model superimposed on the representation of training. Just as training and evaluation are interlocking processes, so too should the models of those processes be interlocked.

1/ I. T. Kirby, "An Approach to Decision-Making" Unpublished Ph.D. dissertation, Department of Education, University of Illinois, 1965 in A Taxonomy of Evaluation Models, W. E. Carter, 1975, p. 5.

Field Testing of Instruments

The suggested instruments and guidelines in the proposed evaluation model were field tested through use in actual training programs. A large business organization in Central New York cooperated in testing these evaluation instruments in their training program.

Follow-up evaluation was conducted on a management training program. This instrument developed by the author utilized the semantic diffential scale developed by Osgood et. al.[1/] This instrument was tested to determine what type of responses it generated as an evaluation instrument. The semantic differential scale was selected to assess attitudes towards training program objectives. This is based on the assumption that attitudes influence behavior more than knowledge or specific skills. Further discussion of techniques and designs used are discussed in Chapter VI and VII.

Process evaluation was tested in five different training programs to determine how participants responded to the instrument. This is also discussed in Chapter VI and VII.

1/ C. E. Osgood et. al., The Measurement of Meaning, 1957.

Guidelines

Overall recommendations resulting from this study will be published in trade publications of the training industry. General guidelines for the use of the proposed model are discussed in Chapter VII.

Assumptions

1. Characteristics of training programs identified in this study were based on several large organizations in Central New York. Conclusions made in this study are only specifically applicable to these organizations. However, it is assumed in this study that these business organizations, and training directors, are representative of all other business and training programs.

2. Field testing of instruments occurred at only one business. It is also assumed that the characteristics of training programs and employees is typical of most other training programs.

CHAPTER IV

EDUCATIONAL EVALUATION MODELS

> "The current literature in education is replete with models. There are in it, discussions of models for decision-making, models for curriculum design, models for learning, models for instruction, models for administering educational programs; the list could be made to go on and on". 1/

Discussion of Educational Evaluation Models

The vast amount of reported research in educational evaluation could easily lead an educator to the naive assumption that we know a great deal about the process of education, its effect, its benefits and the best approaches and techniques. However, the truth is quite the opposite, education is too complex to describe in a thousand research efforts let alone a single project. There is no general agreement about the important effects or behavior changes resulting from education. We can make only general statements about the social benefits or cost effectiveness of education. And, I am sure you could not get even two educators to agree on the best method for teaching.

Education deals with people, individual people and it is a personal process. All people react and perform differently.

1/ Jeff Pyatte, "Functions of Program Evaluation and Evaluation Models in Education", The High School Journal, vol. 53, no. 7, April 1970, p. 389.

Developing education programs and evaluating those efforts is a compromise of trying to hit a majority of the typical population. Measuring results are difficult to quantify and often difficult to predict.

For these reasons a great number of evaluation studies have been attempted, trying to strike upon that most effective methods of education. Each individual study can only hope to make a contribution to that understanding and not uncover the whole thing.

The educational models discussed here are those studies that this author feels have made the most significant contributions to creating that understanding.

Steele (1973)[1/] in her analysis of educational evaluation, grouped evaluation models into six categories; a) Evaluation as an input into decision-making, b) Evaluation of program parts, c) Evaluation - kinds of data, types of activities, d) Evaluation process, e) Results - attainment of objectives and f) Results - evaluation of outcomes and effects. These groups are not entirely clear and seem to overlap to a considerable degree. For example, all evaluations should lead to decision-making. Also, process evaluation should look at program parts and types

---

1/ Sara M. Steele, Contemporary Approaches to Program Evaluation, 1973.

of activities. In spite of this overlap, Steele's discussion does allow a framework for comparing and distinquishing several different evaluation models.

Popham (1975)[1/] has described evaluation models in four distinct groups which are much easier to perceive than the categories outlined by Steele. Popham's categorizes models as; a) Goal Attainment, b) Judgmental Models Using Intrinsic Criteria, c) Judgmental Models Using Extrinsic Criteria and d) Decision-Facilitation Models. These groups appear quite distinct and it is quite easy to place various models within them before comparing parts. It is necessary to make some degree of grouping before comparison because models differ so in there objectives and criteria that in some cases it is similar to comparing apples and oranges.

Goal Attainment Models

The most important aspect of evaluation is measuring results. Simply, what happened as a result of the educational activity. The earliest development of evaluation methodology contered on the measurement of results. Tyler[2/] in his early work empha-

---

1/ James W. Popham, Educational Evaluation, 1975.

2/ R. W. Tyler, "General Statement on Evaluation", Journal of Educational Research, 1942, pp. 492-501.

sized the importance of stating objectives for education and then measuring results in terms of those objectives. Prior to this much of the discussion of evaluation centered around disagreements as to what should be measured.

The Tyler models, and others similar, are categorized by Popham in the first group of Goal Attainment. These designs are the foundations of sound evaluations. If an educational program does meet the objectives established for it, then changes should be made.

Under the Tyler model, it is relatively easy to evaluate educational programs. If the objectives are properly designed then the measurement of results is quite straight forward. Generally the more difficult phase is the agreement on objectives rather than the evaluation of those objectives. This type of evaluation assumes that the objectives that are set are valid and are such that the educational activity will lead to attaining those objectives. The major contribution of Tyler was to develop a means of setting a standard for evaluation. Any teacher or program director should be able to evaluate a program under this Goal Attainment model by simply stating objectives and comparing them to measured results.

Weaknesses in the Tyler models were identified by Taba and

Sawain[1] in their model of evaluation developed by the Association for Supervision and Curriculum development of the National Educational Association. They identified the following weaknesses;

"(1) Objectives that form the basis of evaluation are usually too narrow.

(2) The range of instruments and devices that are being used are often too limited.

(3) The focus of attention has been on the end product rather than on the processes by which the end products are attained.

(4) Results have often been interpreted without adequate information about factors which affect learning and achievement.

(5) Results of evaluation haven't been adequately translated into curriculum decisions."[2]

In the process of education there are numerous decisions that must be made and very distinct activities and influences that ffect final outcomes. Evaluation, as in the Tyler model, should assess accomplishment of objectives. However, an improved evaluation design would also look at the process of education. As a program director confronts the alternatives for modification of a program based on the evaluation of results, it would be beneficial to have additional specific information

---

1/ Sara M. Steele, Contemporary Approaches to Program Evaluation, 1973.

2/ Ibid.

on the process of education in addition to data on the degree to which objectives were met.

Hammond (1967)[1/] expanded Tyler's model with the consideration of the process of education. He described a three dimensional design for evaluation. These three dimensions were: terminal behavior, instructional and institutional. The instructional dimension included organization, content, method, facilities and cost. The institutional dimension included students, teachers, administrators, educational specialists, family and community. All of these sub-items represent variables that need to be considered in evaluation. This model makes the needed step to include more of the process of education and the outside factors that influence education. It, like Tyler's model, can be designed and used by local personnel to evaluate educational activities. It stresses self-evaluation of programs and also utilizes behavioral objectives. However, the model is extremely complex and requires a great deal of time and effort to quantify variables in three dimensions. In some cases it may be too difficult to quantify date and because of its complexity it may not be well adopted by local education

---

1/ R. L. Hammond, "Evaluation at the Local Level" cited in Contempory Approaches to Program Evaluation by Sara Steele, 1973, p. 180.

personnel.

These two models (Hammond & Tyler) fall into the Goal Attainment category outlined by Popham. Later models contribute to understanding evaluation by further defining objectives and examining other aspects of education.

Judgmental Models Using Intrinsic Criteria

Popham defines the second and third group of models as judgmental, where the program administrator makes some degree of judgment on the effectiveness of the program. This group is differentiated into two separate categories based on the types of criteria used for judging programs.

The first category is "Using Intrinsic Criteria". The best example of this type of evaluation is the school accreditation program. Professional educators make a judgment on the effectiveness of a program by examining internal factors such as the training of staff, size of classes, number of textbooks, etc.

This has been a popular form of evaluation in previous years. Designated "experts" selected on criteria such as age or years of experience would observe parts of educational programs, review teacher credentials, types of facilities and

resources. From this information they would make suggestions for improvement and judge the effectiveness of the program, giving or denying accreditation.

The weakness in this judgment based on intrinsic factors is the same reason there has been a decline in the use of "experts" for school accreditation. Almost total reliance on intrinsic criteria is an extremely limited base on which to form a realistic evaluation. Many of these intrinsic factors do effect educational outcomes but they cannot be used as the sole predictors of educational effectiveness. This is true in training programs also, that you cannot totally evaluate a program based on class activities.

Judgmental Models Using Extrinsic Criteria

Judging programs on the basis of extrinsic criteria is generally a more sound approach to evaluation. Scriven's model (1967)[1/] of evaluation contributes to understanding in the area of judgmental models using extrinsic criteria. While Scriven's model is not a full-blown complex model, he did make an important distinction in evaluation. He divided evaluation into two aspects, formative evaluation and summative evaluation.

1/ Michael Scriven, "The Methodology of Evaluation", In Curriculum Evaluation, R.E. Stake, ed., 1967.

Formative evaluation are those observations that can change a program while it is still on-going, while summative evaluation judges the effectiveness of an activity that has already happened. Scriven emphasizes summative evaluation using outside observations and factors; he calls this payoff evaluation. These outside factors include such things as performance in a job, or behavior changes. Scriven does state that attention should be paid to intrinsic factors through formative evaluation. He recommends that most evaluators will find a hybrid of intrinsic and extrinsic factors extremely effective in assessing a program.

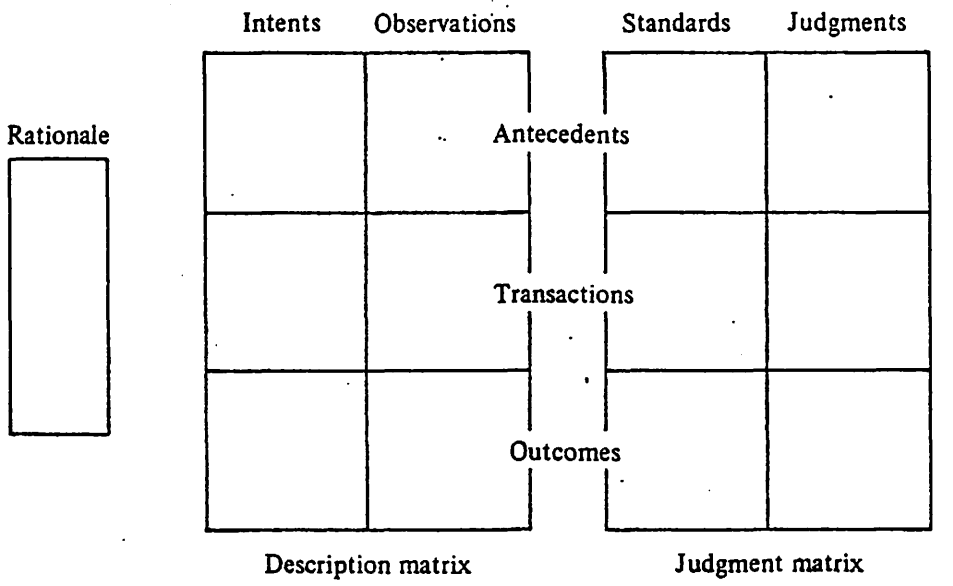
Stake (1967)[1/] proposed an evaluation model which is the best example of a judgmental evaluation using extrinsic criteria. In this model he first states that all programs should have some overall stated rationale. Stake then describes evaluation in three distinct phases. One, of course, is results or as Stake terms it, Outcomes. The process of education is termed Transactions and the conditions prior to learning are called Antecedents. Stake proposed an evaluator go through two different processes. In the first process the evaluator describes the antecedents, transactions and outcomes of the program. This

---

1/ R. E. Stake, "The Countenance of Educational Evaluation" Teacher's College Record, 1967, pp. 523 - 540.

first process is subdivided into describing both the intent and the observations for each of the three phases. In the second process, of judgment, the evaluator establishes a standard and a judgment for each phase. Figure 4 helps to illustrate Stake's model.

FIGURE 4. STAKE'S DEPICTION OF DATA NEEDED IN EVALUATION



SOURCE: James Popham, Educational Evaluation, 1975, p. 31.

Stake's model makes the evaluation process more extensive. It is a systematic approach which gives equal importance to the, now identified, three areas of evaluation, planning, process, and results. For the first time, this model gives consideration to planning as part of evaluation and the influence of outside factors. Both of these can affect the success or failure of a program.

It is obvious that not all students are affected in the same manner by a single education program. These individual differences are an important consideration in assessing the worth of a program. Although Stake's model makes significant contributions, it is limited because of its complexity. It may be difficult to quantify data in all areas and it may be difficult for many educators to understand and adopt it.

Decision-Facilitation Models

The next step in the evolution of evaluation was to remove, or at least decrease, the role of the judgmental part of evaluation. Many educators have proposed that there should be more objective forms of evaluation and the judgments and potential bias of individuals should be minimized. The role of the evaluator should be one of information collector rather than judge. The information collected through the evaluation should contribute to making decisions on alternatives made by those directly responsible for the program. The group of evaluation models that contribute to decision making are grouped by Popham as Decision-Facilitation Models. These models generally go beyond those previously discussed and examine all aspects of an education program.

Marvin Alkin (1969)[1/] proposed a systematic evaluation program which included five parts. These were: systems assessment, program planning, program implementation, program improvement and program certification.

The first phase of systems assessment consisted of examining the present state of affairs of the community, students or organizations, depending on what type of education program is being designed. This assessment should describe a current educational need resulting in a statement of objectives for the output of the educational program.

Program planning, the second phase, provides a decision-maker with information which will assist him/her in selecting between alternative strategies for achieving the established goals. This is a difficult area in which to measure because you are forced to predict how effective a particular strategy might be. Efforts that can be made are to estimate costs of various alternatives and look for results from related research and similar situations.

An evaluation of program implementation determines the

1/ Marvin Alkin, "Evaluation Theory Development", Evaluation Comment, January 1970, pp. 2 - 7.

degree to which the implemented program meets the program described in the planning phase. This might appear to be an unnecessary step, but Alkin emphasizes:

> "There have been numerous examples in educational literature of conflicting results relative to the impact of a specific instructional treatment. We would maintain that in large part this is attributable to the lack of specificity of the precise nature of the instructional treatment that was employed."[1/]

If a program director is to find weaknesses in a program that is not meeting its objectives, there must be come certainty that the program actually used in the classroom was the one outlined in the planning.

Program improvement is an examination of the process of education. An evaluator collects information relative to making improvements in the actual program. Alkin encourages evaluators to, "present data immediately to the decision-makers so that change may be executed within the system to improve the operation of the program."[2/]

Finally, what Alkin terms Program Certification, is measuring the impact of the program. In order for a program to become certified, or worth continuing, it must show results in line with the intended program objectives.

1/ Ibid, p. 3.

2/ Ibid.

The Alkin model is quite similar to the CIPP model developed by Stufflebeam et. al. (1971)[1/] which identified four types of evaluation: Context, Input, Process and Product.

The CIPP model was briefly discussed in Chapter II. In addition to determining the four types of evaluation, Stufflebeam, also distinguishes different decision settings for evaluation. The close relationship of evaluation and decision-making is a key aspect of Stufflebeam's definition of evaluation. As mentioned before, he elaborates the existing definitions of evaluation to include the "deliniation, obtaining and providing of information for decision-making."[2/]

The distinction of decision settings are important because they influence the types of decisions made and consequently the types of information that are necessary from evaluation. The first setting is homeostatic which involves maintaining a normal balance in the educational process. It could include minor decisions of class enrollment, or faculty

---

1/ D. Stufflebeam, Educational Evaluation and Decision-Making, 1971.

2/ D. L. Stufflebeam, In Worthen & Sanders, Educational Evaluation: Theory and Practice, 1973, p. 178.

assignment. Next is the incremental setting which could include activities aimed at continually improving the program. These are minor changes or adjustments in an on-going program in an attempt to improve an existing program. The third setting is neomobilistic which involves major changes in educational activities. This might include testing of innovative teaching methods, a new curriculum or new instructional materials. Finally, the highest setting is called metamorphism, a concept rarely encountered in the real world of education. This setting accounts for a complete change in educational activities.

The CIPP model is appropriate for contributing to making decisions in all four of these decision settings. However, the primary uses for educators will be in the incremental level of seeking to make minor improvements in programs.

Malcolm Provus (1971)[1/] developed a Discepancy Model of educational evaluation. He based his model on a unique definition of education. He defined program evaluation as "the process of (1) defining program standards; (2) determining whether a discrepancy exists between some aspect of program

---

1/ Malcolm Provus, Discrepancy Evaluation, 1971.

performance and the standards governing that aspect of the program; and (3) using discrepancy information either to change performance or to change program standards."

This discrepancy model consisted of five steps. First is Design, which involves documenting the program by identifying program objectives, students, staff and other resources and the instructional strategies to be used. This roughly approximates the first two steps of the Stufflebeam CIPP model and Alkin model.

The next step is Installation, to see if the actual program compares to the planned program. This step is comparable to Alkin's program Implementation evaluation stage.

The next step is process in which the evaluator examines "enabling objectives achievement". This is similar to the Formative evaluation defined by Scriven, Process evaluation in the CIPP model and Program Improvement defined by Alkin. The Discrepancy model does emphasize process observation in comparison to established standards and then basing recommendations on those discrepancies.

The fourth stage of the model focuses on the question "How well does the program achieve its terminal objectives?" Standards established in step one are compared to the resulting product. This again is similar to the final stage of the CIPP

model and Alkin's model.

Provus proposes an additional fifth step, which is unlike the four previous steps which are developmental in nature. This is Program Comparison on a cost analysis basis, comparing the selected alternative to other competing alternatives. It asks the question, "Could the same result be achieved at a cheaper cost?"

The models of Stufflebeam (CIPP), Alkin, Provus, all follow the same mode of providing information to those responsible for the educational program. This information should be as structured, objective and appropriate as possible to assist the program directors in making good decisions and improvements.

The fourth stage of the model focuses on the question "How well does the program achieve its terminal objectives?" Standards established in step one are compared to the resulting product. This again is similar to the final stage of the CIPP model and Alkin's model.

Provus proposes an additional fifth step, which is unlike the four previous steps which are developmental in nature. This is Program Comparison on a cost analysis basis, comparing the selected alternative to other competing alternatives. It asks

the question, "Could the same result be achieved at a cheaper cost?"

The models of Stufflebeam (CIPP), Alkin, Provus, all follow the same mode of providing information to those responsible for the educational program. This information should be as structured, objective and appropriate as possible to assist the program directors in making good decisions and improvements.

Summary

These models do provide some excellent suggestions and observations that could be adapted to a specific model for training programs. This author will make an attempt to weigh advantages of various models and build on significant contributions of each research effort. The plan is to borrow those aspects which seem to be most appropriate to the identified characteristics of training programs. Popham provides an excellent suggestion on building upon the successful efforts of other educators.

> "It should be apparent, of course, that a builder of evaluation models has a difficult time subscribing to Polonius' admonition, "Neither a borrower nor a lender be." But although the effective evaluator will hopefully avoid the perils of becoming preoccupied with model minutiae, these diverse approaches to the task of

> educational evaluation are obviously instructive. To proceed without modest conversance with their major elements, would be foolhardy."[1]

The evaluation models discussed in this review do an excellent job of describing evaluation and have made significant contributions to education. This research effort will primarily describe the evaluation process in a manner appropriate to employee training. The designs suggested by various researchers can be combined and adjusted to yield a more appropriate concept of evaluation.

1/ James Popham, Educational Evaluation, 1975, p. 42.

CHAPTER V

CHARACTERISTICS OF TRAINING FOR EVALUATION

If a theoretical concept for evaluation of training is to be effective, it must be based on the reality of existing training as well as the theoretical concept of education. Information collected from real training programs is the source for establishing common characteristics for evaluation. Selected Training Directors were surveyed to determine this information, since this was the easiest method to obtain data under the limited scope of this study. Responses from this sample were used to determine current evaluation problems and expectations. Their comments were structured through a questionnaire in order to combine responses and make overall conclusions. These conclusions are molded into a set of criteria for the evaluation model.

Results of Training Director Survey

The survey of training directors yielded a relatively low return in actual numbers of directors. A total of ten responses were received from the initial seventeen training directors identified. Even though this number is small, these individuals should represent the most progressive opinions in train-

ing, since they are responsible for large training departments and are active members of the professional organizations.

The number of training programs and participating employees was significant. The average number of training programs offered annually by each organization was 20, with a range of two to fifty. An estimated total of over 15,000 employees participate annually in the training programs offered by these ten organizations.

Directors were asked to categorize the types of training programs offered on a percentage basis between management training, supervisory training and technical training. Sales training was added by some respondents as an additional type of training. These responses were grouped in the technical training category, because this was defined by the author as one type of technical training. Table 1 shows the mean responses and range of responses of each training category.

TABLE 1. TYPES OF TRAINING PROGRAMS

TYPE	MEAN	RANGE PERCENT	S.D.*
Management	18.5	0 - 60	18.8
Supervisory	28.0	0 - 100	27.5
Technical	53.5	0 - 95	29.0

*Standard Deviation

There was a considerable range in the type of training programs offered by these organizations. Most all of the firms had a large percentage of technical training and the mean response indicated over fifty percent of the training emphasis. Seven of the ten firms had between ten and forty percent supervisory training. All but one of the firms had management training programs, however, this was the smallest percentage, with a mean of 18.5 percent. Two firms had half of their program in management. While technical training (including sales training) was a significant part of these training efforts, supervisory and management training comprise almost half of the training programs. Consequently, these latter types of programs should receive attention in addition to traditional technical training.

A question was asked to determine the percentage of training participants who select training programs on a voluntary basis. The fact that employees are in a program because they want to be, can affect the type of training activities and the subsequent evaluation. Responses in the survey showed on the average, 29 percent of the participants select programs voluntarily. For most firms this percentage was very small, usually 10-20 percent. A few firms had 50-70 percent. These

firms with a high percentage of voluntary participants also had a high percentage of supervisory and management training programs. It might be concluded that these types of program were more of a voluntary nature.

The next question asked the current types of evaluation. This yielded a great variety of responses, indicating there was no universally accepted evaluation method. The most common type of evaluation indicated was objective quizzes based on program objectives, which four respondents mentioned. The next most frequently mentioned methods were; reaction questionnaires, skill demonstration and measures of job performance. The importance of reaction questionnaires was in agreement with reported findings discussed in Chapter I. Skill demonstration is apparently an important evaluation method in technical training. Respondents indicated that job performance was primarily evaluated on an informal and often subjective basis by the participating employee's supervisor.

One respondent mentioned informal interviews and narrative, subjective reports by trainers as a means of evaluating programs. This may be a more frequently mentioned method than indicated in the survey, for trainers often evaluate activities even if on a very informal and subjective basis.

Subjective judgments may be frequently used in the absence of, or in supplement to, more formal types of evaluation.

One firm used existing performance information to evaluate programs, specifically they used the number of complaint letters and accident reports to measure the success of training programs. Unfortunately, not all training programs are of a nature that existing data can be used in evaluation. However, whenever it is available it can make an excellent data source for evaluation.

The information collected from these training directors indicated that several of the organizations are making progress in evaluation. While most rely on the more traditional methods of participant reaction and post training quizzes, some were apparently attempting techniques that would provide more information about behavioral change and organizational performance.

The next question in the survey sought to obtain directors' perception of the problems in training evaluation. Responses to this question were many and varied, however, there were a few common responses. Four of the ten respondents indicated a need for developing a good measure of job performance. Also, two stated they need help in establishing standards or base-

lines of performance. These directors indicated that the problem was not measurement itself, but what to compare the measured results with. Two other responses related to this area of behavior measurement, which indicated a need for a means of measurement of retention of skill/knowledge content learned in the course. These responses indicated that a number of training personnel were concerned about evaluating training programs in terms of job behavior but they were having difficulty developing standards and measures of performance.

Three other responses indicated a need to improve evaluation in the newest area of training, namely human relations and management development. It was pointed out earlier that training programs had evolved from simple programs offering instruction in manual tasks to sophisticated seminars seeking to develop skills in working with people and managing resources. Training directors apparently were more comfortable evaluating programs that teach specific psychomoter skills or cognitive information, with the use of skill demonstration or criterion referenced tests. However, these methods are not as effective in the area of human relations and management development.

Improvements in evaluation should place greater emphasis on human relations and management development aspect

of training.

Only one director mentioned the need to measure "bottom line" impact. From the information in this survey, cost effectiveness or cost benefit was not a major concern of training directors in measuring the impact of their programs.

Several other comments by training directors indicate some of the evaluation problems they were faced with. One director expressed a concern that he lacked the time or expertise to evaluate programs. He added "most patented approaches are too vague or unapplicable." Another respondent was concerned with measuring effectiveness of outside speakers and also determining the relevancy of course material.

A couple of training directors also used the questionnaire to relate some of the problems that affect training but may not be directly concerned with evaluation. One indicated a need to improve management follow-up and reinforcement of training efforts. This concern is based on the concept that classroom training is not a sole solution to improving indivdual and organizational performance. Management must support and continue to stress the content and objectives of training programs. This could affect evaluation, because frequently behavior observation systems utilize a participant's supervisor,

but more importantly obstacles such as lack of management support can greatly inhibit retention and application of principles learned in a training program.

Another training director was concerned about training "programs and formulas" in general. His reaction was that there are none that properly meet the needs or can be adopted inexpensively.

Responses to this question on evaluation concerns indicated there definitely are training evaluation needs. Many of the problems identified in the literature were supported by the responses in this survey.

In summary of Part I, evaluation, according to training directors, needs to (1) primarily improve measures and standard of job behavior and performance, (2) develop effective of results of human relations training, and (3) create a design that conserves time, expense and allows for individual needs. Further insight into training evaluation needs will be achieved through responses to the second part of the survey.

Part II of the training director survey sought specific quantifiable data on the importance and needs of various aspects of training evaluation. Eight training decisions, discussed in Chapter III, were presented for training directors

to rate as to importance. A five point scale was used, ranging from high importance to low importance. Table 2 shows the average degree of importance for each of these decisions. A value of one corresponds to high importance and a value of five corresponds to low importance.

TABLE 2. IMPORTANCE OF TRAINING DECISIONS

Training Decision	Degree of Importance		
	Mean	Range	S. D.*
1. Determine need for specific training programs	1.1	1-2	.30
2. Determine previous knowledge and skills of trainees	1.9	1-4	.80
3. Select appropriate instructional strategies	1.6	1-3	.67
4. Determine effectiveness of trainers	1.8	1-4	.97
5. Determine if trainees enjoyed the program	3.1	1-5	1.13
6. Determine extent of learning at the end of the program	2.0	1-3	.78
7. Determine extent of trainee application of skills on the job	1.4	1-3	.66
8. Determine if new skills improve job performance	1.5	1-3	.67

*Standard Deviation

All eight training decisions were rated high in importance with the one exception of, "determining if trainees enjoyed the program." On that particular decision responses ranged from high to low with very little agreement. The average resulted a little on the low importance side.

All other decisions were high in importance with a value of 2.0 or less and responses were pretty consistent as evidenced by the relatively low standard deviations.

The decision that was rated the most important was "Determine need for specific training program". Some may not consider this decision relative to evaluation, but if it is considered, it becomes one of the most important evaluation aspects.

After determining need, next in importance were "Determining extent of trainee application of skills on the job," and "Determining if new skills improve job performance." These responses reinforce the premise that training directors apparently are increasingly concerned with evaluating programs on the basis of job behavior and performance.

The increasing importance of job behavior is further supported by the observation of the lower rating of "Determining

extent of trainee learning at the end of the program." It really is of little importance what a person knows at the end of the program, for the objective should be to achieve job behavior modification. The only importance of measuring post training learning is when it is not possible to make some accurate judgments of behavior back on the job.

Decisions also rated in the important category were "Selecting appropriate strategies" and "Determining the effectiveness of trainers." These two decisions are in the area of input and process evaluation, according to the CIPP model. They were not frequently mentioned in the literature or in existing training evaluation models. It was anticipated by the author that these were important areas. This survey verified that they are important even though they are of slightly less importance than assessing needs or measuring behavior.

"Determining the previous knowledge and skills of trainees," was also rated as important. This may be an important decision training directors make on occasion, but not at the beginning of every training program.

Table 3 shows information on the extent of evaluation for each of the training decisions within the surveyed organizations. Each director was asked: "If the decisions were made?", "If the

information for the decision was adequate?", and "If information for the decision was formally collected?"

TABLE 3. COLLECTION OF INFORMATION FOR TRAINING DECISIONS

Training Decision	Is the Dec. made	Info. adequate	Info. collected
	Percent Responding Yes		
1. Determine need for specific training programs	100	56	30
2. Determine previous knowledge and skills of trainees	70	70	30
3. Select appropriate instructional stragegies	89	80	60
4. Determine effectiveness of trainers	89	60	56
5. Determine if trainees enjoyed the program	60	70	50
6. Determine extent of learning at the end of the program	70	22	33
7. Determine extent of trainee application of skills on the job	60	30	20
8. Determine if new skills improve job performance	50	20	10

The majority of these training decisions were made in most organizations. The decisions which were not made by at least some organizations were also rated by many as lacking adequate information. This information coupled with the fact that all considered these decisions important leads to the conclusion that there is a need to improve methods of collecting information on decisions especially regarding job behavior and performance.

The areas that directors indicated as having the most adequate information were in "Selecting strategies", "Determining if trainees enjoyed the program" and "Determining previous knowledge." Those slightly adequate were "Determining needs", and "Determining effectiveness of trainers." The areas showing the greatest lack of adequate information were job behavior and performance and the area of measuring learning after the conclusion of the program.

It is strange that trainers seem to have adequate knowledge and information about training needs and abilities of employees before they come into the program and yet they have inadequate information about them after they leave. It leaves this author wondering if assessment of needs and determining skills is done using job behavior criteria.

In response to the question whether information was formally collected for decisions, the majority of trainers indicated that there was no formal collection made. Those areas where the most collection of data occurs were "Selecting strategies", "Determining enjoyment" of the program and evaluating trainers. This supports the literature in training which indicates that happiness scales were the most common form of evaluation, however, the literature does not mention trainer evaluation frequently. The low percentage reporting formal selection of information in the areas of needs assessment and determining skills of trainees supports the premise that trainers were making these decisions on limited information. Finally trainers were making very little progress in collecting information of application of skills on the job or performance.

The final part of the questionnaire sought to assess training director perception of the factors that hinder evaluation and the collection of information for training decisions. The objective of this section was to provide direction in the development of an evaluation model and system which might be able to help overcome some of these problems.

Table 4 shows the percentage of respondents who indicated the factor was a hinderance to evaluation.

TABLE 4. FACTORS THAT HINDER EVALUATION

Training Decision	Lack of Resources	Lack of Expertise	Lack of Cooperation	Lack of Time	Unimportant
		PERCENTAGE RESPONDING YES			
1. Determine need for specific training programs	20	40	30	0	0
2. Determine previous knowledge and skills of trainees	20	20	30	20	0
3. Select appropriate instructional strategies	20	30	0	10	0
4. Determine effectiveness of trainers	10	10	10	10	0
5. Determine if trainees enjoyed the program	0	0	0	10	30
6. Determine extent of learning at the end of the program	50	30	30	30	0
7. Determine extent of trainee application of skills on the job	50	60	40	40	0
8. Determine if new skills improve job performance	40	50	60	30	0
Average	26.25	30.00	25.00	18.75	3.75

This data indicated training personnel have difficulty making decisions about job behavior and performance. Half of the respondents indicated the lack of resources hindered evaluation through measuring application of skills on the job. In addition, sixty percent indicated a lack of expertise and forty percent indicate lack of cooperation and lack of time.

Decision number 8 on determining improvement of performance as a result of training also showed a number of factors hindering evaluation. Sixty percent checked a lack of evaluation requires the greatest involvement of line managers and supervisors. When working with a large number of individuals outside of the training department, there will be a number who will not share the same enthusiasm for evaluation of training.

Fifty percent indicated a lack of expertise in measuring job performance, and forty percent expressed a lack of resources as a hindering factor. Time was also a problem with thirty percent of the organizations.

The decision with the next greatest degree of hindering factors was assessment of learning. Fifty percent indicated lack of resources was a problem. Thirty percent checked lack of expertise, lack of cooperation as well as lack of time.

From the information on this decision, lack of expertise does not seem to be as much of a hinderance to training compared to lack of resources, time and cooperation. The literature in training evaluation has several examples of reported successes in assessing learning. However, it appears that training directors in this sample still perceive problems in measuring trainee learning.

Respondents indicate a few problems in determining the need for a training program. Lack of expertise and lack of cooperation seem to be the most significant factors. Lack of cooperation, again is probably a result of having to work with other departments in the organization. As was previously pointed out, an evaluation model needs to address those decisions of determining the need for a training system.

A few training directors expressed hinderance in determining knowledge and skills of trainees. Lack of cooperation was followed by the lack of resources, lack of expertise and lack of time as significant factors. Generally there were few problems with collecting information for this decision.

Lack of expertise and lack of resources were designated as hinderance factors in selecting appropriate strategies. Only ten percent of the respondents expressed concern in collecting information for decisions in determining effectiveness

of instructors.

Finally, determining if trainees enjoyed the program had no factors hindering evaluation. What's more, thirty percent of those responding indicated that this decision was unimportant. From this information it is easy to see that reaction scales were the most frequently used evaluation form.

Looking at the factors that hinder evaluation in an overall point of view, responses were averaged across all eight decisions. This is an approximation of the most significant factors and it assumes each decision has an equal importance, which of course may not be true.

Lack of expertise appears to be the greatest hinderance, followed by lack of resources and lack of cooperation. An evaluation design should consider these obstacles. It first should be in a form that can be easily communicated to and understood by training personnel. Secondly, it should utilize a minimum of resources, including cost and personnel time. Finally, it should be in a form that will minimally threaten and interfere with line managers and their responsibilities in an organization. By satisfying this latter criteria an evaluation will hopefully secure greater cooperation. Lack of time is of minor importance and should be considered in the criteria of using a minimum of resources.

Criteria for Evaluation Model

This survey of training directors leads to several conclusions as to what should be the criteria for a new evaluation model, more appropriate for employee training programs. The training director responses are summarized into several key points which should be considered in developing the model. These conclusions and criteria are:

1. The survey reported that half of all training programs were technical, the other half were divided between supervisory and management training. Since all of these types of training appear to be important, the proposed evaluation model should be appropriate for all three types of training. This will simplify evaluation if the same general framework can be used for all training programs.

2. Directors indicated the most important training decision relates to planning and determining the need for a program. Consequently, the evaluation model should consider program planning and determining training needs.

3. Directors indicated the second most important area was determining on the job behavior and measurement of improved performance. This, too, should be emphasized in the model.

4. There appears to be little importance in the decision of whether participants enjoyed the program. Therefore, this

aspect of evaluation should be de-emphasized or eliminated.

5. The areas in which training directors have the least adequate information is in determining performance measuring job behavior and assessing learning after the program. Consequently, the model should make a contribution to improving evaluation techniques in their area.

6. Since directors indicate that lack of expertise is the most significant factor hindering evaluation, there is a need to improve the tools that they have for evaluation. The model should be presented in a manner to improve understanding of evaluation.

7. Lack of resources is another significant hinderance. The proposed model should use a minimum of resources yet accomplish the task of evaluation.

8. The other significant hinderance indicated by the survey, is the lack of cooperation. One method to improve cooperation between training and the rest of the organization is to incorporate evaluation activities that involve other departments in a constructive manner. In other words, use their expertise and make them feel a part of the training and evaluation.

9. One of the problems indicated by a number of directors comments was the lack of standards for evaluation. Several

mentioned that measurement was not as much of a problem as was determining what standards with which to compare the measurement. It appears that the proposed model should address this problem of identifying the standards for comparison in evaluation.

CHAPTER VI

MODEL FOR EVALUATION OF TRAINING PROGRAMS

Model for Employee Training

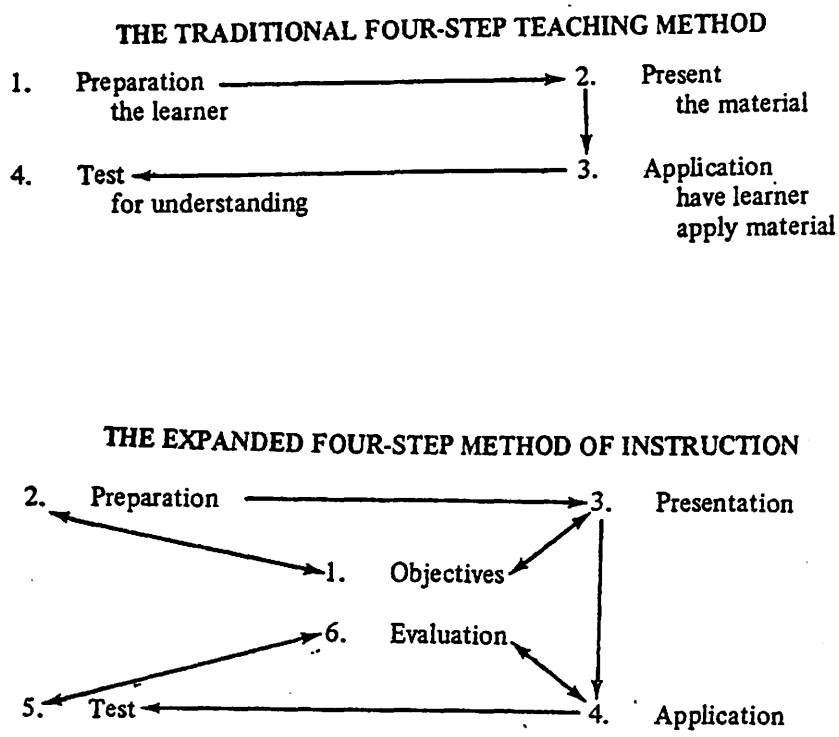
In order to discuss a model for evaluation of training, it is necessary to establish a conceptual framework for the process of training itself. Most evaluation researchers agree that evaluation should be an integral part of the training process, therefore, the evaluation design should fit snuggly around an existing training model.

There are almost as many reported models of training as there are evaluation models, however, most of the differences are minute. Current researchers have described training models as continuous processes rather than separate activities with an established end or beginning.

Ford (1970) proposed an improvement in the traditional view of training which was called the four step approach.[1/] The comparison of the four step method and Ford's expanded model of instruction is shown in Figure 5.

1/ George A. Ford, "Four Steps Are No Longer Enough", Training and Development Journal, vol. 24, no. 7, pp. 24-34.

FIGURE 5. FORD'S MODEL OF TRAINING



SOURCE: George A. Ford, "Four Steps Are No Longer Enough" Training and Development Journal, vol. 24, no. 7, July 1970, pp. 24-34.

In the traditional approach trainers prepare the learners, present material, have students apply course material and finally test for understanding. While this four-step view is basically sound and has undoubtedly resulted in thousands of excellent programs, it lacks the objectives of the expanded Ford model. The addition of objectives improves instruction for it sets boundaries and goals which help to limit extraneous

instruction and encourage achievement of definitive results in a program. Additionally, objectives not only aid in instruction, they are also important to developing evaluation. Mager (1962) states:

> "When clearly defined goals are lacking, it is impossible to evaluate a course or program efficently, and there is no sound basic for selecting appropriate materials, content or instructional methods."[1/]

In addition, the Ford model adds an evaluation step which looks at application of the course content as well as learning at the completion of the program. This expanded concept of evaluation is consistent with the suggestions of other educators as discussed in Chapter IV.

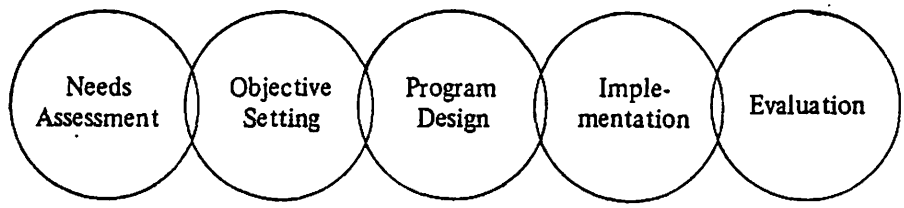
Another view of the changing model of training was discussed by Miller (1969).[2/] He detailed the traditional views of training as a series of activities. Figure 6 depicts his traditional view and the revision.

1/ Robert F. Mager, Preparing Instructional Objectives, 1962, p. 16.

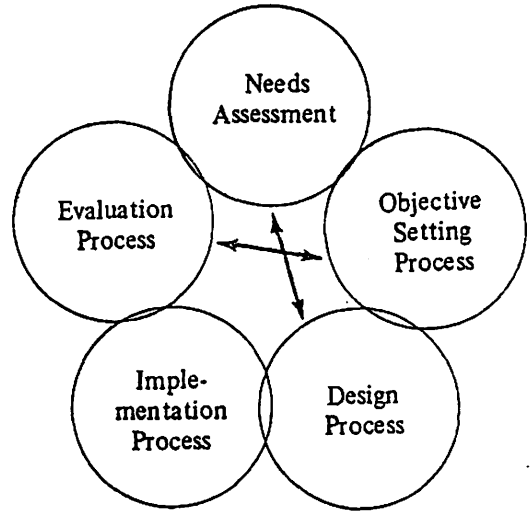
2/ Richard D. Miller, "A System Concept of Training," Training and Development Journal, vol. 23, no. 4, April 1969, pp. 4-6.

FIGURE 6. MILLER'S MODEL OF TRAINING

THE TRADITIONAL CONCEPTS OF THE TRAINING MODEL



CONCEPTION OF THE TRAINING MODEL AS AN INTEGRATED PROCESS



SOURCE: Richard D. Miller, "A System Concept of Training," Training and Development Journal, vol. 23, no. 4, April 1969, pp. 4-6.

In the revised model these same activities become a cycle with interaction between all activities and not simply separate activities that follow chronologically.

The unique feature of training that distinguishes it from

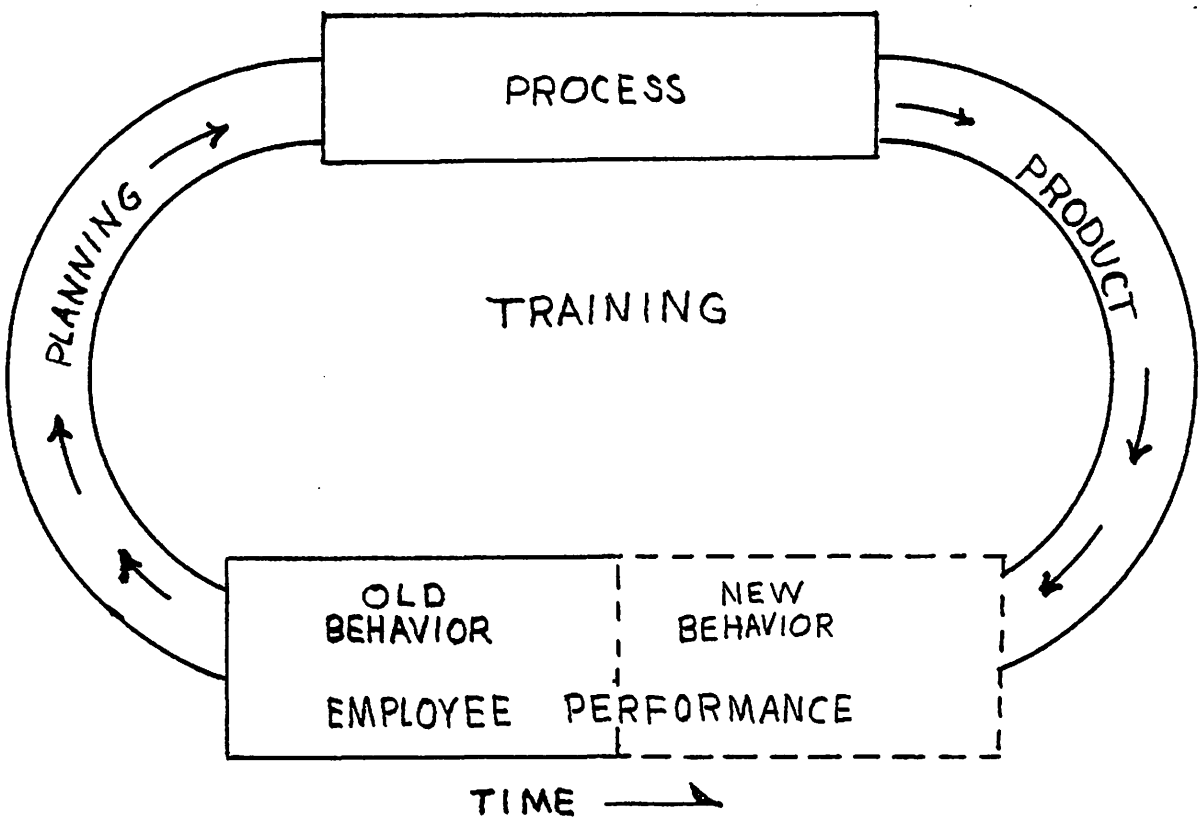
education is the fact that training is based on the job or tasks for which employees are being trained. Even in management development programs or supervisory training, it is possible to identify specific tasks that employees will hopefully be able to perform at the completion of the training program.

Bail and Cushman (1977) identify the unique characteristics of the Employee Training Model.[1/] The purpose of training is to develop technical competencies. The instructional focus is based on job requirements and instructional techniques use the "tell them" approach with the instructor exerting a strong leadership role. This strong leadership role utilizing trainer judgment rather than consensus and, the use of industry standards are key features of the employee training model.

The model of training developed for this study is illustrated in Figure 7. It draws on related models and observation of training programs.

1/ Joe P. Bail and Harold R. Cushman, Teaching Adult Education Courses: The Employee Training Model, 1977, pp. 4-5.

FIGURE 7. TRAINING MODEL



The large rectangle represents organization performance. Relative performance in any organization, compared to other businesses, is based on the collective performance of the employees of that organization. People make the difference in any business organization and their behavior and personal performance is the core of the performance of the organization.

At this point it is helpful to more closely examine the factors which influence individual employee performance. Their performance has as its core the personal knowledge and skills that an individual possesses. The greater a person's knowledge

and skills, the greater their potential to perform. Actual performance is, of course, influenced by personal attitudes and the work environment. Work environment is comprised of working conditions, type of job tasks, wages, benefits, peer influence, etc. Individuals modify their behavior and consequently their performance because of these environmental factors which influence attitudes. Attitudes are influenced by factors other than the environment. A person's self image, values and experience all influence attitudes. Individually employees modify their behavior as a result of attitudes and the work enviornment. Behavior is defined as an individual's actions, generally determine performance. How an individual performs a particular task usually determines the performance level of that task.

Organizational performance is shown in the training model (Figure 7) as changing over time. All people, and consequently the organization, change. Training is only one factor that causes change, for other factors such as, working conditions, personnel changes and reorganization of the business all modify organization performance and change it over time.

Training exists as a function of the organization charged with the responsibility of modifying and improving employee

behavior and ultimately contributing to improved performance. The point should be made that training is not the sole solution to improving performance, for training frequently has little effect on working conditions, peer influence or the make up of the job task. These other factors must be improved by line management and supervision.

Training removes individual employees from the organization, runs them through a training program and returns them to their responsibilities in the organization. Training attempts to influence individual behavior by increasing a person's knowledge or skill level, or improving attitudes. After these individuals return to the organization, it is hoped that their new level of knowledge or skills, or different attitude will modify their behavior and ultimately increase their performance level.

Training, therefore, is a continuous cycle of removing employees from an existing organization, conducting a training program and returning them to the organization where performance has changed. Training programs do not materialize through any magical gesture, nor is it a mystical black box through which employees pass and are transformed into sparkling new enthusiastic employees. Training must be planned, organized, and effectively carried out. Training is described in this model as having three distinct phases: Planning, Process & Product.

The first phase of training is illustrated by the bridge from the organization to training. This phase is Planning; assessment of needs for a specific training program. Training should be determined by needs in the existing organizational and individual performance. Any gap between observed employee behavior and expected behavior or industry standards might dictate a need for a training program.

Training directors should be guided by organization performance. In successfully carrying out the training function, they should take two important examinations of job performance and behavior. The first is in this planning phase and the other is in the product phase after the training program has been completed.

For example, assessment of needs in the area of fork lift operators in a warehouse might be an observation of how safely equipment was operated. The observed behaviors would be compared to industry standards or organization goals or objectives. If there was a significant gap, a training program might be identified as a possible solution. After the training program is set up and conducted, the follow-up of training products would look at the same behaviors to see if the observations more closely matched the industry standards or organization goals.

In between the Planning and Product phases is, of course, the process of training. Process is divided into three distinct activities that take place or at least should take place in all training; Setting Objectives, Selecting Strategies, and Execution. Setting objectives is simply determining what is going to be done in the training. These objectives should be stated in terms of job behaviors and, of course, they should relate to established needs.

Selecting strategies is determining how the training program will be done; what instructional methods, what materials, what types of trainee application activities, etc. Finally, Execution of Techniques is the carrying out of objectives and strategies, doing what was planned for the training. Obviously, some people perform a job better than others. Effective training must have the best personnel executing the training plan, consequently Execution is a distinct activity which should be evaluated.

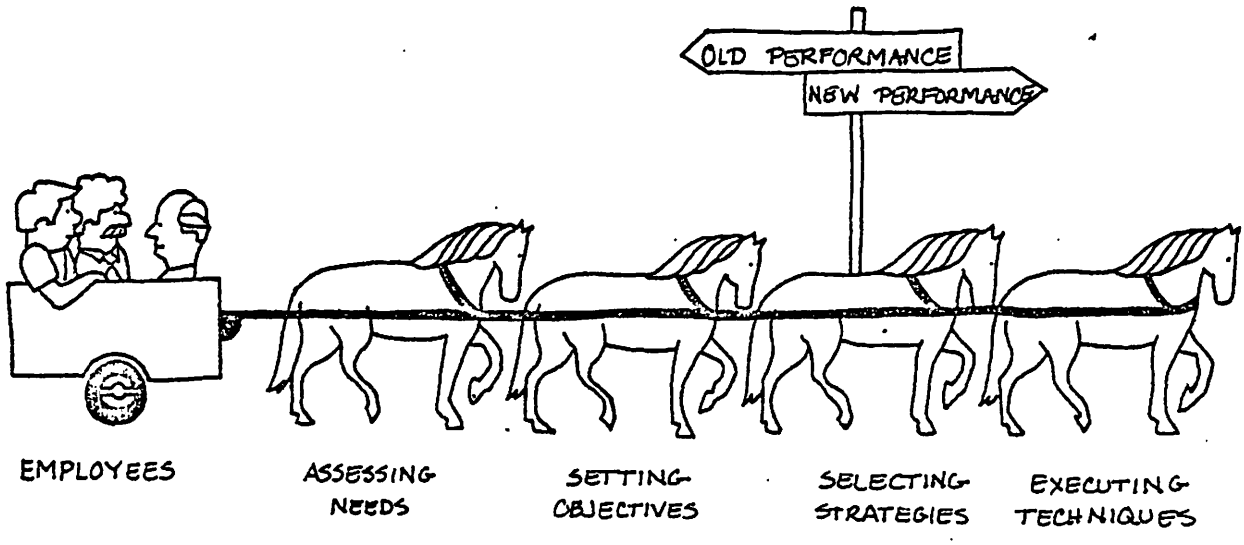
This broad description of training has three separate phases; Planning, Process and Product. Process of training is broken down into three important separate steps of Setting Objectives, Selecting Strategies and Execution of Techniques.

Description of Evaluation Model

Many evaluation models are a conglomeration of circles, squares, rectangles and arrows, which frequently boggle the mind and often create more confustion that the verbal description, they were intended to simplify.

Before describing the proposed evaluation model and attaching this to the training model, it may be helpful to reconstruct the training model in a different format, sort of a Operational Model. This Operational Model is presented in a fashion that emphasizes the aspects of training that need to be evaluated. This operational model of training is shown in Figure 8.

FIGURE 8. OPERATIONAL TRAINING MODEL



This model depicts training as a wagon and team of horses. The passengers aboard the wagon are employees, their point or origin is Old Performance and their destination is New Performance. The power for transportation are four strong horses which are equivalent to the planning and process of training. The horses are labeled, Assessing Needs, Setting Objectives, Selecting Strategies, and Execution of Techniques. The collective efforts of each of these horses, or the parts of training, help employees to reach their destination of New Performance.

Evaluation was previously defined as the collection of information and judging the worth of activities. Training can be viewed collectively as one activity or separately as several sequential activities. Evaluation of the whole of training can be termed Summative Evaluation, borrowing the term from Scriven.[1/] The Follow-Up of Training Products is the same as Summative Evaluation, measuring the worth of the whole of training.

Odiorne points out evaluation of training based on the

1/ Michael Scriven, "The Methodology of Evaluation" in Curriculum Evaluation, R.E. Stake ed., 1967.

follow-up of employees on the job. He states;

> "The systems approach to evaluation of training starts with a definition of behavior change objectives sought through a concious development effort. This definition then remains the yardstick for measurement throughout the course, and achievement against the stated goals is the measurement of success. All other forms of evaluation measure the internal characteristics of the activity itself, not the effectiveness of training.1/

Viewing training as separate activities leads to the conclusion that each of the parts can be evaluated. This type of evaluation could best be termed Formative Evaluation, again borrowing the term from Scriven. The major, identifiable parts of training which should be evaluated are Assessment of Needs, Setting of Objectives, Selecting Strategies and Execution of Techniques.

These four aspects are the parts identified in the Operational Model in Figure 8. If the goal of training is to move a group of employers along the path from Old Performance to New Performance, Summative Evaluation measures how far training moved employees along the path. Training evaluation simply measures progress resulting from the training. However, the

---

1/ George S. Odiorne, Training by Objectives: An Economic Approach to Management Training, 1970, p. 181.

measure of progress is not enough. A Training Director needs to know more than whether progress is made. He needs to examine the individual activities which contributed to that progress. In order to make specific changes or improvements, evaluation information needs to be collected on the program itself. Were there any weak links in the process? Is every segment performing up to its potential? Formative evaluation judges each of the parts of the training program, evaluating the Planning and Process of Training.

FIGURE 9. EVALUATION MODEL

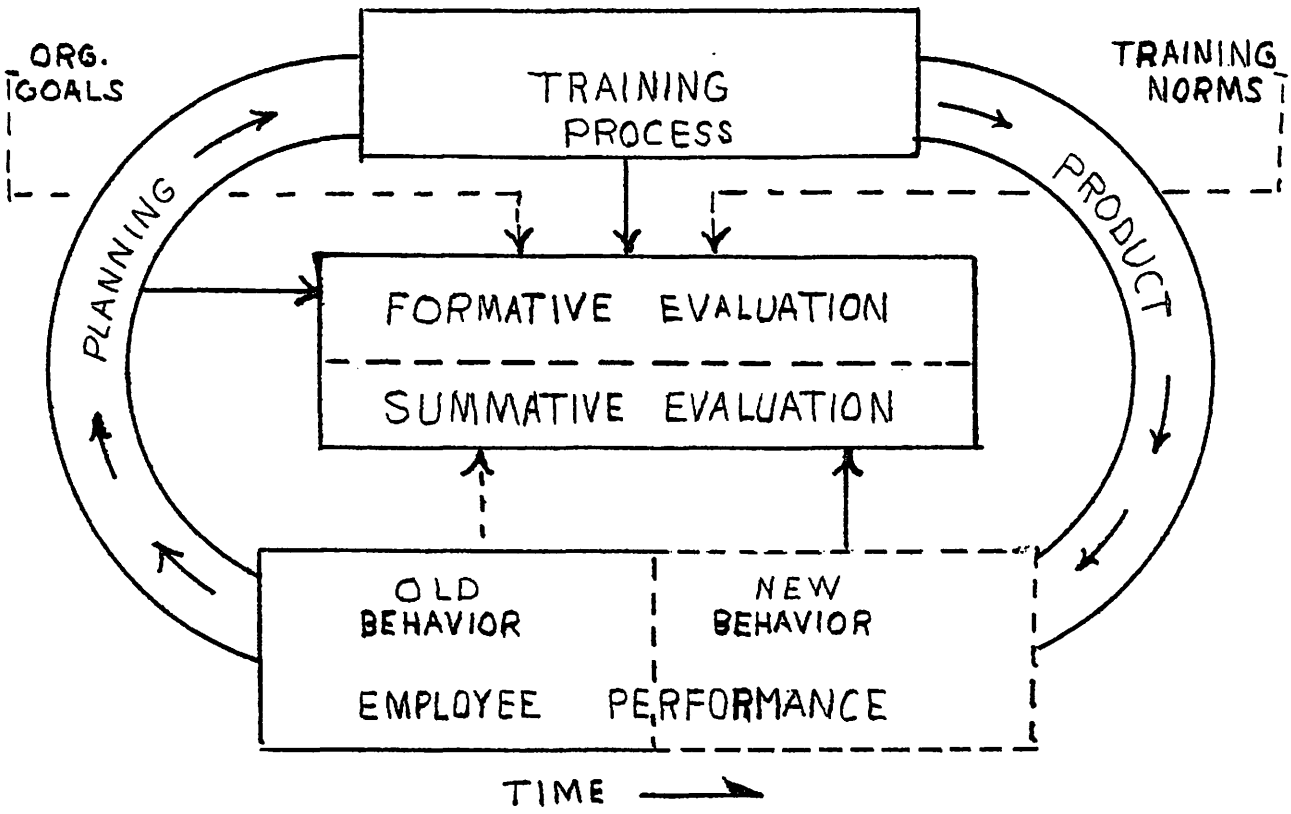


Figure 9 illustrates the proposed evaluation model, which fits with the author's concept of training depicted in Figure 7. Evaluation rests in the center of the continuous training cycle and monitors all activities making judgments on their effectiveness.

The rectangle depicting evaluation is divided into two major segments. The lower part is Summative Evaluation which assesses the total program. The other half compliments the Summative Evaluation and is termed Formative Evaluation. It monitors the planning and process of training and measures their respective effectiveness. Formative evaluation measures four separate segments; Assessing Needs, Setting Objectives, Selecting Strategies, and Execution of Techniques.

Arrows show the various inputs to evaluation. Solid lines are the flow of information or measurement of some aspect. This is the assessment for the evaluation, examples are, observing employee behavior after a training program or observing trainer performance. Broken lines indicate the input of information for standards with which to compare the measurement for evaluation. In each aspect of evaluation some measurement is made and it is compared to a standards.

Evaluation of Planning is necessary to determine if needs were properly assessed and the right objectives set. As

mentioned before, not all organizations needs can be solved through training. Performance needs identified should be sorted to determine which are appropriate for training programs. Joe Harless, President of Harless Performance Guild, Inc., calls this type of evaluation "Front-End Analysis." He states that performance problems must be analyzed to determine if they are skills and knowledge, motivational, environmental or some combination.[1/]

In summary, evaluation is divided into two basic types. Summative evaluation measures the results of the training program. This is the payoff. Did the training accomplish what it set out to accomplish? The second type is Formative evaluation. This is assessment of the parts of training. This information serves as the basis for decision-making regarding improvements in the program. Formative evaluation contributes to improvements in the operation of the program which should lead to greater results in the Summative evaluation.

Measurement and Standards

Now that the types of evaluation have been identified, it is

1/ Harrold Littledale, "Front-End Analysis", Training, vol. 12, no. 3, March 1975, pp. 27-29+.

necessary to focus in on the areas of measurement. Information must be collected for making judgments, therefore, the areas of obtaining this information should be specified. The evaluation model generally identifies these areas, and this discussion will describe them in greater detail. Figure 10 shows a summary of the areas of measurement and source of information on standards with which to compare the measurements.

FIGURE 10. MEASUREMENT AND CRITERIA FOR EVALUATION

Type of Evaluation	Areas of Measurement	Source of Standards
Summative		
Follow-Up	Employee New Behavior	Employee Old Behavior Supervisor Expected Behavior Organization Goals Industry Standards
Formative		
Needs	Employee Old Behavior	Organization Goals Supervisor Expected Behavior Industry Standards
Objectives	Program Content	Organization Goals Supervisor Expected Behavior Industry Standards
Strategies	Program Activities	Training Norms
Techniques	Program Activities	Training Norms

Follow-up evaluation measures employee new behaviors or performance and compares this to one of several standards. One standard might be a control group of existing or old behaviors in the organization. Others might be expected behaviors of supervisors, organizational goals or standards of the particular industry.

Needs evaluation measures the employee old or existing behaviors and compares this to organization goals, expected behaviors or industry standards. Evaluation of objectives examines the content of the program and compares it to organization goals, expected behaviors, or industry standards. Evaluation of strategies and techniques used in training is accomplished by collecting information of training activities and comparing this to training norms. All of the standards in this evaluation could be collected within the organization, with the exception of training norms for strategies and techniques. These can only be established through the evaluation of a large number of training programs. Some large organizations might be able to accomplish this, however, the best method would be for some professional organization to conduct a standard survey and establish criteria for evaluation of training strategies and techniques.

Contribution of the Model

The first significant contribution of this training evaluation model is the recognition of evaluation of the planning and process of training. Since training is a combination of several activities, evaluation is not complete until all aspects have been evaluated. A training director is not fully aware of the worth of a training program until he/she judges all planning and process activities as well as training products.

By evaluating all activities of a training program, a director has the ability to make improvements in the areas that show up as weaknesses. Previous evaluations models in training have exclusively concentrated on the products of training programs. Results from this type of evaluation only tell how effective of ineffective a program may have been. By using the expanded model, proposed in this study, a director can also determine why the program reached its level of effectiveness.

The second contribution of the model is that it specifically identifies target areas for measurement. Training personnel are not left with the problem of determining where to get information, for these areas are designated, for each of the types of evaluation identified. Also the areas of standards are designated. Training directors indicated the problem in obtaining standards and this model identifies the areas in

which to obtain standards. With the collection of this information, the making of decisions will be much easier.

One thing this model does not do is establish specific types of measurement. This must be determined in the specific training programs. As training programs vary, so will the types of measurements made. However, some suggested techniques are made in the following section which should be applicable to a great number of training programs.

Another contribution of the model is that is simplifies product evaluation. The levels of reaction, learning, behavior and results, used by Kirkpatrick, are eliminated. First of all these categories do not fall into a continuum. Reaction in itself is not a valid measurement of program products and is eliminated from consideration as a technique for Summative Evaluation. Learning, measured at the end of the program, is also eliminated for it is of little value to determine the overall effectiveness of a program. Objectives for training programs should be stated in terms of performance or behavior on the job. If the employee does not apply these on the job, it doesn't really matter how much they learned at the conclusion of the program. Training should be directed at creating modified behavior on the job and should be only evaluated in

that context.

The final contribution is that it properly defines the relationship between different levels of evaluation. Summative or Follow-Up evaluation is final measure of the effectiveness of the overall program. Formative evaluation, the evaluation of individual training activities, is the evaluation of parts of the whole. Both Summative and Formative evaluation have different purposes and yield different information, yet both are necessary to fully evaluate training efforts.

Suggested Techniques

There are a number of techniques available for evaluation. The proposed evaluation model presented in this study develops good opportunities for the use of these specific techniques. These suggested techniques are considered by the author most appropriate for use with the model. Figure 11 shows a summary of these suggested techniques.

FIGURE 11. SUGGESTED TECHNIQUES FOR EVALUATION

Type of Evaluation	Suggested Technique
Summative	
Follow-Up	Attitude Scale Performance Review Skill Demonstration
Formative	
Needs	Consultative Team
Objectives	Consultative Team
Strategies	Participant Observation
Execution	Participant Observation

Follow-up evaluation or Summative evaluation examines the entire process of training. As indicated in Figure 10, the measurement area for follow-up evaluation is the new behaviors of employees.

Several techniques available for measurement of behavior include interviews, outside observation, diaries, or supervisor observations. In addition, for some employee tasks, there are performance data, particularly in technical programs. This existing data should be used whenever it is available or can be obtained easily.

A unique approach to behavior measurement is the use of

attitude scales as an approximation of behaviors when it may be too difficult to obtain an accurate observation or measure of the actual behavior. The attitude scale suggested for use in approximating behaviors is the semantic differential scale developed by Osgood et. al. (1957).[1/] This scale was selected because it is difficult for respondents to "out guess" the survey and put down what he/she thinks is the expected response. Also, a great deal of information can be obtained in a little amount of time. This could be a disadvantage in summarizing responses, however, today with the common use of computers, this is a minor consideration.

The semantic differential scale is a measure of attitude toward a particular concept, using rating pairs of adjectives as a measure. Respondents are asked to place the selected concept along a continuum between two opposite adjectives. For example, a respondent may be asked to rate a concept like "woman" between Stong and Weak. This would be repeated for several different rating pairs. The total of these ratings should indicate the individuals attitude toward the selected concept, in the example, "woman."

1/ C. E. Osgood et. al., The Measurement of Meaning, 1957.

The semantic differential scale has been used by several evaluators over the years. Abbatiello (1967)[1/] developed a semantic differential scale to assess attitude change in a group of supervisors participating in a training program. A total of 12 concepts or stimuli were selected based on the program content. The instrument placed each of these with a 10 pairs of rating adjectives. Participants were asked to rate each of the concepts on the seven point scale between the adjectives.

The rating pairs were selected from the initial work done by Osgood, et. al. These pairs measured three separate aspects of attitudes as defined by Osgood; evaluation, potency and activity.

Abbatiello found significant changes between pre and post tests on five of the twelve concepts. He found that rating pairs intended to measure evaluation showed the greatest significant difference and there was little contribution of the pairs selected for potency and activity.

The semantic differential scale has also been used to a limited degree by Hessling (1962)[2/] Warr, Bird and Rackham

---

1/ A. A. Abbatiello, "An Objective Evaluation of Attitude Change in Training", Training and Development Journal, vol. 4, no. 5, May 1967.

2/ Hessling referenced in A. C. Hamblin, Evaluation and Control of Training, 1974, p. 106.

(1970)[1] and Hamblin (1974).[2] All suggested the potential use of the semantic differential scale for assessing attitude changes as a result of training. Hamblin states;[3]

> "Despite their deficiencies, the semantic differential scales appear to be, at the present time, among the best approaches for the quantification of attitudes."

Using participant attitudes as an estimation of behaviors may be a great assumption and it will require additional validation beyond this study. However, attitudes can be a valuable bit of information if properly collected. Elkins found in the evaluation of a program for government supervisors in California that;

> "trainee attitudes were the strongest single determinant of on the job application of new management principles. Stronger by far than the amount of new information learned in the program"[4]

The use of the semantic differential scale will be further explained in Chapter VII as a result of the development of

---

1/ P. B. Warr, M. W. Bird, and M. Rackham, Evaluation of Management Training, 1970, p. 65.

2/ A. C. Hamblin, Evaluation and Control of Training, 1974, pp. 106 - 109.

3/ Ibid, p. 109.

4/ Aaron Elkins, reported by Ron Zemke, "Management Training Development: Measuring the Impact" Training, vol. 14, no. 10, October 1977.

specific instruments and field testing.

A suggested technique for comparing program objectives and content to organization goals and standards is to use a consultative team from within the organization. The use of line managers and supervisors can provide a fresh observation and perspective of training program. This observation can determine if training content and the objectives from which content was derived is in line with the real working situation. Line management is most familiar with job requirements and problems, therefore, they would be in an excellent position to observe the degree to which training programs are related to job situations.

Another advantage of using a consultative team from within the organization is that it will help to build rapport between the training department and the rest of the organization. Other managers will get a first hand view of training practices and programs. Assuming the training program is effectively managed, this will help to build a favorable image of training and develop closer working cooperation.

In using consultative team, care should be taken to give the team members specific directions as to the types of observations they should make. They should be thoroughly oriented so they are familiar with their charge. Initially team members

should be given a description of the program and activities which they will observe.

The consultative team should be used to observe specific aspects of the program. Generally this will be to determine if program content is related to the job situation and if the techniques taught in the program can effectively be used by the employee in their job responsibility. The team may also be used to identify any additional objective content areas or problems not identified by the training staff. The consultative team should have the opportunity to question and clarify observations through contact with the training staff. However, they should reach conclusions and make recommendations on their own and present these to the training staff. This report should include three things; Observations, Suggestions and the Reasons on which the suggestions were made.

The consultative observations and report can be used to evaluate needs, objectives and possibly the strategies used for developing the particular training program.

Evaluation of the process of training has not been emphasized and more often it has not even been mentioned. The model of evaluation in this study emphasizes evaluation of the strategies and techniques as important aspects of

Formative evaluation.

Techniques appropriate for measuring the process of training are centered around observation of training activities. The consultative team observation could provide one approach, but their effectiveness to evaluate the process of training is limited because they only observe parts of the actual program and they also may not be familiar with effective teaching techniques. Consequently, they can only make a limited observation and have nothing in their experience with which to compare their observations.

Training directors are another possibility for an observer, for they should possess the experience of being able to identify effective teaching techniques. This type of observation parallels the situation of a local school administrator observing and evaluating teachers. Observation by training directors is used quite frequently and does give substantial feedback about the effectiveness of training strategies and techniques.

An approach that could be superior to either of these methods is using the training participants as observers. No other group or individual is able to observe every activity

that occurs in a training classroom. Whatsmore, outside observers often conduct their observation in an artificial atmosphere. Any trainer or teacher will naturally modify their behavior and try to be at their best if they are being observed. Participant observation eliminates that problem because the observation is constant.

The major limitation to participant observation is the problem that most employees in training program lack experience to judge effective techniques and they each would probably observe activities a little differently. However, these problems can be overcome by using highly structured observations.

A limitation related to variance in observation is that of different interpretation of the statements on the observation instrument. Individuals may have different interpretations of such terms as enthusiasm, student-oriented or flexible. This problem of interpretation can be minimized by using low inference items. Popham (1975) differentiates between low inference and high inference behavior statements. The examples given by Popham are;

> "A. Engage in aggressive behavior (high inference)
> B. Strike a neighboring child (low inference)"[1/]

1/ James Popham, Educational Evaluation, 1975, p. 99.

Popham further states;

> "Low inference observation categories or rating dimensions require few inferential leaps on the part of the observer or rater.... High inference categories, however, demand that the observer or rater draw inferences regarding what a child's behavior represents."[1/]

The higher the inference, the less reliable the measurement. However, the danger in using too many low inference items is that the resulting list of variables might be of little significance to the objective of the observation. For example, whether a trainer wore a tie would be a very low inference item, that all observers could agree on but this variable probably has little correlation to effective teaching.

The observer form developed should use low inference items and, avoid high inference that might reduce reliability of rating scales. If observer scales are correlated to effective teaching, this will eliminate those low inference variables that are not important to the rating of effective training. If this observer scale is tested for validity against the effectiveness of the course, an effective evaluation instrument can be developed.

1/ Ibid.

It is intended that this technique will observe training strategies and techniques for executing those strategies, to determine the effectiveness of the process of training.

One study which effectively used participant observation of classroom activities is the study of College teaching done by Cushman and Tom (1974).[1/]

Steps in Evaluation

In order to briefly review the proposed evaluation model and explain how it might be used by a training director, the following steps are outlined.

1. Plan the training and evaluation. Determine target employees and select performance areas. Select measures of performance and determine the specific concepts, attitudes or activities that will be measured. Use either existing performance data or approximation of performance through attitude measurement.

2. Determine the need for the training program. Establish a baseline of current job performance using either existing performance data or an assessment instrument. Compare observations

---

1/ Harold Cushman and Fred K. T. Tom, Cornell Diagnostic Observation and Reporting System for Student Description of College Teaching, 1975.

to organization goals or other employee or industry standards.

3. Design and conduct the training program.

4. Collect Formative Evaluation data. Using observer forms and consultative team reviews, identify the types of activities in the training program.

5. Conduct the Summative Evaluation. Using a follow-up measure determine either actual job performance or approximate performance through measurement of attitudes. Use the same method as selected in the planning phase of the training and evaluation process.

6. Compare follow-up results to the pre-training observations to determine the extent of change resulting from the training.

7. If the results concluded from the Summative Evaluation are not satisfactory, seek to identify changes to be made in the program. Compare the Formative Evaluation observations to standards selected from other training programs or profession standards, if they exist. The areas to be changed are the aspects of the actual training that show differences from standards in the Formative Evaluation process.

8. Recycle the training and evaluation steps to determine improvements in training and employee job performance.

Achievement of Criteria for Evaluation Model

The criteria for development of an evaluation model based on the survey of training directors is achieved by the model described in this chapter and the evaluation techniques described for use with the model.

1. The evaluation model can be adapted to all training programs ranging from very specific technical training programs to much more abstract human relations training. Once the behaviors are identified on which to base a training program, they also become the criteria on which to base an evaluation.

2. The evaluation does include a provision to judge the effectiveness of the planning phase of training through the use of the consultative team.

3. The model emphasizes behavior and performance of trainees. Summative evaluation or evaluation of the total program is based on job performance and behavior.

4. The participant reaction scale is removed. However, actual participant reaction is made more sophisticated and used in the evaluation. Participant reaction is a means to an end and not an end in itself. By structuring participant observations rather than just reactions, the evaluation is

able to obtain an objective evaluation of what goes on in the classroom. Also reactions are made more sophisticated in the attitude scale and measured back on the job. It is important what training participants think and if this information can be collected in a valid and reliable manner, it can result in a sound evaluation.

5. The Summative evaluation method of using the semantic differential scale is a unique approach that could be used to greater degrees in approximating behavior and performance. Attitudes influence job satisfaction and performance. Measuring an individuals attitude toward specific job aspects can be used to approximate job performance.

6. The evaluation is simple and requires only three phases for both the Summative evaluation and Formative evaluation. This model should be understood and accepted by trainers.

7. There are only three or four measurements in this evaluation. This should provide training directors with a great deal of information without consuming a large amount of resources.

8. The suggestion of the use of consultative teams to conduct the planning evaluation will not only contribute evaluative information, but will aid in improving cooperation between training and the rest of the organization.

9. The evaluation model is defined in a way to identify the standards with which observations can be compared.

The other criteria established for the evaluation model were the statements made by Kelly (Chapter III) in regards to the development of new evaluation models.

1. This model does identify all of the information evaluators need collect to make a comprehensive evaluation.

2. The model is internally logic and complete.

3. The model is still relatively simple and should be easily understood by training personnel.

4. While the model draws heavily on the work of other educators it does present the model in a form that has not been presented previously.

5. The model is heuristic and provides a framework for future training evaluation research.

6. Future study will contribute greatly to improving the model. The suggested techniques need to be validated as part of the model.

7. The model is efficient and provides a natural sequence of activities.

CHAPTER VII

FIELD TESTING

Within the limited scope of this study it was possible to field test the proposed evaluation instruments. This field test is the first step in the validation process. The instruments were drafted, revised and used in an evaluation. This field test determines how employees react to the instruments and the types of data received. The result is a tested and revised instrument which can be formally tested in major studies in large populations.

Follow-Up Evaluation

The attitude scale suggested as a technique for follow-up evaluation was developed and tested at an in-house training program of a large business organization in Central New York. The training program was a week long management training program for local store managers.

The semantic differential scale developed for the evaluation instruments drew on the observations of Osgood, et. al. (1957).[1/] Twelve pairs of descriptor were selected for testing.

1/ C.A. Osgood, et. al., Measurement of Meaning, 1957.

Those selected were items that showed a high degree of discrimination and contributed to evaluation. The twelve pairs of adjectives selected were ordered as follows in the instrument:

important	_._._._._._._._._._._._._.	unimportant
difficult	_._._._._._._._._._._._._.	easy
weak	_._._._._._._._._._._._._.	strong
worthless	_._._._._._._._._._._._._.	valuable
fast	_._._._._._._._._._._._._.	slow
good	_._._._._._._._._._._._._.	bad
painful	_._._._._._._._._._._._._.	pleasurable
wise	_._._._._._._._._._._._._.	foolish
free	_._._._._._._._._._._._._.	constrained
successful	_._._._._._._._._._._._._.	unsuccessful
soft	_._._._._._._._._._._._._.	hard
active	_._._._._._._._._._._._._.	passive

The specific evaluation instrument was developed for the management training program by examination of program objectives and discussion with the program trainer. Concepts were selected that were fundamental objectives of the program and could easily be described in a few words. Concepts were also selected for which participants should have a favorable attitude,

if they are to be effective managers.

Five concepts were identified for the management training course; Setting Objectives, Developing Action Plans, Delegating Management Responsibilities, Establishing Specific Measures of Performance and Getting Subordinates Involved in Objectives. In addition, one concept was selected for which managers should have a negative attitude, if the training program were effective.

A concept was placed over a scale of the twelve pairs of adjectives. Participants were asked to complete the instrument by placing an X along the seven point scale which best indicated how they felt the particular concept should fall between the two extremes. A copy of the instrument used in the management training program is exhibited in Appendix B.

This evaluation instrument was designed as a follow-up to be sent to participants after they had returned to their job. Four weeks after the conclusion of the program, questionnaires were mailed to selected participants and a control group.

A pretest was not used in this evaluation for two reasons. First, the elimination of a pre-test simplifies the evaluation activities and second, the pre-test has shown to be a contam-

inate in some studies, e.g., Belasco and Trice (1969).[1] Consequently, the design used for this field test included randomization of responses and comparison to a control group.

The ideal experimental design proposed by many researchers is the four way design. This includes four groups and uses a pre-test, post-test and control groups. This type of design can be illustrated by the following diagram:

Group 1	--------	T	-------- M	
				Experimental
Group 2	M--------	T	-------- M	
Group 3	----------		---------- M	
				Control
Group 4	M----------		---------- M	

M - measurement
T - treatment

Through comparison of control groups, experimental groups and groups with a pre-test, differences can be determined. However, this type of study design can be difficult to construct under field conditions. Belasco and Trice (1969)[2]

1/ James A. Belasco and Harrison Trice, "Unanticipated Returns of Training", Training and Development Journal, vol. 23, no. 7, July 1969, pp. 12-17.

2/ Ibid.

used this design to determine specific effects of training programs.

Under field conditions, it is difficult to set up an experiment where only part of the sample receives the treatment and it also may be difficult to give only part of the group a pre-test. Also, it is very difficult to achieve similar groups for comparison.

An alternative design was selected for this follow-up evaluation. This design involves the randomization of participants when applying a post-test to the selected trainees and the comparison control group. The effect of randomization is that it creates equal groups for comparison. If the participants and the control group are randomly selected from the same pool of individuals, it can be assumed that they are equal. Campbell and Stanley state, "the most adequate all purpose assurance of lack of initial biases between groups is randomization."[1/]

Popham comments on the use of the post-test only control group design.

1/ D. T. Campbell and J. T. Stanley, "Experimental and Quasi-Experimental Designs for Research on Teaching", in Handbook of Research on Teaching, N. L. Gage ed., 1963, p. 195.

> "The basic dividend of this post-test only control group design is that by measuring an untreated, randomly assigned control group, we secure an estimate of how the treated and control groups would have responded on a pre-test..."[1/]

In other words, since groups can be assumed equal, the observations of the control group can also be used as the observations of the pre-test of the group that participated in the training program.

In the field test, even though the actual participants in the management program were not selected on a random basis, only a random sample of the participants were surveyed. This was accomplished by the following procedure. In each of the training programs, a pool of employees was identified in which the employees worked. Employees were randomly identified from this pool until the total number selected included at least 10 of the participants of the training program. There were a total of 20 participants in the program. This random selection process identified a participant sample of 10 and a control sample of 42.

Questionnaires were sent to individuals in the sample by

1/ James W. Popham, Educational Evaluation, 1975, p. 210.

the training department of the organization. Responses were fairly good, eight responses were received from the participant group and 31 from the control group. This is a return rate of 80 percent for the participant group and 74 percent for the control group.

A mean value was calculated on each pair of adjectives for all six concepts. This resulted in a total of 72 variables. A value of one was given to the more favorable of each adjective. The more favorable of each of the pairs were; Important, Easy, Strong, Valuable, Fast, Good, Pleasurable, Wise, Free, Hard, Successful, and Active.

The mean value for each variable was compared between the participant and control groups. Theoretically, if the training program has any effect, the participant group should have a lower mean value than the control group. The actual mean values are exhibited in Appendix C.

A t test was performed on each comparison to determine which variables indicated a significant difference. Table 5 shows the summary of t value for the management training program. A negative t value indicates that the control group had a lower value than the participant group. This would be the opposite of the expected value. An asterisk indicates which

of the individual variables are significant at the .10 level. A .05 level is customarily used for significance test but since this test demands a one tailed test rather than looking for significance in either direction, a .10 level was used.

As a result of this data, the author recommends the elimination of four pairs of adjectives from future evaluation. Those eliminated are; Important-Unimportant, Good-Bad, Successful-Unsuccessful, and Soft-Hard. The first two pairs did not show very significant differences because, they may be more obvious than other rating pairs and respondents might be more likely to give the expected response. Successful-Unsuccessful variable may be influenced by other factors having to do with the employees work situation and may not reflect how a person's behavior may have changed as a result of training. Finally, Soft-Hard may be too abstract a pair to provide any meaningful data.

TABLE 5. t TEST VALUES OF DIFFERENCE BETWEEN PARTICIPANT AND CONTROL GROUP MEANS ON MANAGEMENT TRAINING EVALUATION

Variable	Concept					
	1	2	3	4	5	6
A	.489	.650	.213	.041	.627	3.340*
B	.426	1.204	1.283	1.121	1.875*	2.420*
C	-1.164	.814	.421	1.037	1.070	1.605*
D	.979	.680	-.117	1.030	.353	3.424*
E	1.046	.386	.798	.553	1.272	1.230
F	.482	0	.173	.304	-.121	4.923*
G	.063	.992	1.956*	2.335*	.150	1.660*
H	.616	.898	.225	1.204	.076	3.967*
I	-.385	.024	2.015*	2.380	1.185	2.258*
J	-1.342	1.253	.502	.416	-.084	3.976*
K	.990	-.367	-.492	1.526*	1.147	.083
L	1.250	.886	1.144	.201	.731	3.392*

*significant at .10 level

TABLE KEY

Concept

1. Setting Objectives
2. Developing Action Plans
3. Delegating Management Responsibilities
4. Establishing Specific Measures of Performance
5. Getting Subordinates Involved in Objectives
6. Performing Manual Tasks

Variables

A. Important - Unimportant
B. Difficult - Easy
C. Weak - Strong
D. Worthless - Valuable
E. Fast - Slow
F. Good - Bad
G. Painful - Pleasurable
H. Wise - Foolish
I. Free - Constrained
J. Successful - Unsuccessful
K. Soft - Hard
L. Active - Passive

Removing these four pairs results in 48 variables in the six concept instrument. On these measures a total of 45 of the 48 were positive movement for the participant group, and 12 of 48 were significant at the .10 level.

Figure 12 shows the suggested pairs for use in future validation and research on the use of the semantic differential for follow-up evaluation of training programs.

FIGURE 12. SUGGESTED SEMANTIC DIFFERENTIAL SCALE FOR EVALUATION

difficult	_._._._._._._._._._._._._._._.	easy
weak	_._._._._._._._._._._._._._._.	strong
worthless	_._._._._._._._._._._._._._._.	valuable
fast	_._._._._._._._._._._._._._._.	slow
painful	_._._._._._._._._._._._._._._.	pleasurable
wise	_._._._._._._._._._._._._._._.	foolish
free	_._._._._._._._._._._._._._._.	constrained
active	_._._._._._._._._._._._._._._.	passive

Trainer Evaluation

A trainer observation form was developed and used on five training programs. The participant observation instrument was developed using low inference statements. Training personnel were first contacted and asked to identify the aspects of

training on which they would like to be evaluated. This information was converted to low inference items and included in the instrument. The reason for this was to increase the feeling of self evaluation because trainers could feel their input into the process.

The principal research used to identify a list of low inference items for trainer observation was the work done by Cushman and Tom[1/] in the rating scales for college teaching they developed.

The initial scale developed included 43 behavior statements. The scale asked participants to what degree instructors demonstrated the various behaviors. The questionnaire used is shown in Appendix D.

After administration in several training programs, the author decided this instrument was too long. It took most participants between ten and fifteen minutes to complete. Many participants noted that several items did not apply and that there are some items that repeated similar statements. Therefore, the number of items were reduced to 26. Those eliminated were those that showed a very high degree of variability (a standard deviation of over 1.0 in most programs). Others were

---

1/ Harold Cushman and Fred K. T. Tom, Cornell Diagnostic Observation and Reporting System for Student Description of College Teaching, 1975.

eliminated that did repeat or measure the same thing. The revised questionnaire is shown in Appendix E. This instrument should be tested and validated in future studies.

CHAPTER VIII

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

As a result of this research effort there are several conclusions from the background research, the study itself and the attempt to develop an evaluation model based on the findings and observations.

Current designs for evaluation are inadequate for they fail to consider evaluation of the planning and process of training. The activity of evaluation must look beyond the evaluation reports and anticipate what will be done with evaluation results. The conclusion in this case is that little can be done unless the evaluation provides a judgment of the planning and process of training.

Current evaluation of the results of training do not emphasize individual behavior. Behavior is the most important aspect of evaluation. Other measures such as participant reaction give only an indication of how employees enjoyed the program which may or may not be correlated to effectiveness on the job. Likewise, learning measured at the end of

the program does not directly measure effects on the job. Consequently, employee behavior should be the most important measurement in the evaluation of training.

In addition to being the most important area of measurement, behavior is also the most difficult area to measure. Any observation of behavior is not a true observation of behavior because the presence of the observer often causes the employee to modify their behavior because of the observation. Evaluation should try to obtain undistorted measures of behavior. The use of attitude scales and other rating forms may be a superior approach to quantifying behavior. Since all of an employees knowledge, skills, frustrations, values and rewards determine his attitudes, this may be an excellent assessment of performance.

Another problem in the evaluation of training programs is the establishment of standards. It is difficult to assess the effectiveness of any activity without having some standard with which to compare the observation. An evaluation design should include a provision for establishing standards.

Participant observation can be used as a potential evaluation, the the observations are structured and use validated instruments to obtain the best data possible.

Another conclusion is that many people feel threatened by new evaluation programs. A way of reducing this perceived theat is to involve the people who might be evaluated in the development of the evaluation. Any evaluation design should establish an overall framework and allow local input in developing the finished product so individual persons will feel more a part of the evaluation and more likely to follow the recommendations resulting from the evaluation.

Suggestions for Future Studies

Characteristics of training programs for evaluation determined in this study were based on a limited number of organizations from one geographic area. Future studies should be made to determine if the conclusions drawn from these organizations are consistent with organizations in other areas and all training programs. Another study might take a closer analysis of training to characterize training programs and the types of activities most commonly used. Since training director reactions were used as a basis for designing the proposed evaluation model, the next logical step is to submit the proposal to training directors to determine if it actually does meet their needs. If the objectives of the study are met,

then the model should be accepted by the training personnel.

One of the standards for use in judging the training process is established instructional standards on the training process. Currently, this data does not exist. A major study should be undertaken, using the instruments suggested in this study, to develop standards which could be used by all training programs. These observation instruments should be validated against results in training program and average values established for different types of training program. For example, these observation criteria might be established for outside resource people, in-house training programs or college executive programs.

Another area for future studies is the validation of the attitude scales proposed in this study. It is suggested that these measures be used as an approximation of behaviors. It remains for future studies to validate these scales to determine the actual degree they do approximate behaviors. This validation could be done by testing training participants and correlating results to an observed and specific record of behaviors or performance.

Recommendations

The major recommendation of this study is that the model

developed in the study be used for evaluation of employee training programs. It should be used in evaluation of programs and in future studies to learn more about evaluation.

In using the model there are several recommendations which will contribute to more effective evaluations.

1. Evaluation should be planned. Just as the author suggests evaluating the planning of training, the evaluation should be planned. If evaluation data is anticipated, instruments can be designed which will provide data for setting objectives and also assessing the effectiveness of training participants. All evaluation efforts should be planned to most efficiently use staff and resources.

2. Existing performance data should be used wherever possible. No matter how exciting or dynamic a survey instrument may appear, existing data on employees is generally superior. This should always be incorporated whenever possible in evaluation as a measure of follow-up of training participants.

3. Objectives for training should be stated in behavioral terms. This not only benefits a training program by making objectives more explicit and realistic, it is helpful to evaluation in describing the behaviors that should be observed to assess effectiveness.

4. Attitude scales like the semantic differential scale proposed in this study, should be used to approximate behavior and performance when it is frequently difficult to obtain other data on job performance.

5. Participant observation should be used in evaluation, not to assess the effectiveness of the program, but to rate trainer behaviors and strategies. When this data is compared to training standards, the process of training can be evaluated and improved.

6. The consultative team of internal managers and supervisors should be used to evaluate the planning of training programs. This type of evaluation will not only contribute to improved programs, but will improve relations within the organization.

Evaluation can be manageable and contribute significantly to improving programs in employee training if it follows the logical sequence proposed in this training evaluation model. Planning through staff involvement, properly delegating responsibilities and encouraging training staff to utilize results in self evaluation and improvement can increase the effect of training on the performance of any organization.

BIBLIOGRAPHY

SELECTED BIBLIOGRAPHY

Abbatiello, A. A. "An Objective Evaluation of Attitude Change in Training", Training and Development Journal, vol. 4, no. 5, May 1967.

Alkin, Marvin C. "Evaluation Theory Development", Evaluation Comment, Los Angeles, California, Center for the Study of Evaluation, University of California, January 1970.

Anderson, Scarvia; et. al. Encyclopedia of Education Evaluation, San Francisco: Jossey-Bass Publishers, 1975.

Arnstein, George. "Trial By Jury: A New Evaluation Method, II: The Outcome" Phi Delta Kappan, vol. 57, no. 3, November 1975.

Ayres, R. J. "Training Thoughts in A Think Tank", BACIE Journal, vol. 28, no. 7, July 74, pp. 83-86.

Bail, Joe P. and Cushman, Harold R. Teaching Adult Education Courses: The Employee Training Model, Ithaca; College of Agriculture and Life Sciences, Cornell University, 1977.

Barton-Dobenin, J. and Hodgetts, Richard M. "Management Training Programs: Who Uses Them and Why", Training and Development Journal, vol. 29, no. 3, March 1975, pp. 34-35, 37-40.

Bass, B. M. and Vaughan, J. A. Training in Industry: The Management of Learning, Belmont California: Brooks/Cole Publishing Company, 1966.

Belasco, James A. and Trice, Harrison M. The Assessment of Change in Training and Therapy, New York: McGraw-Hill Book Company, 1969.

Belasco, James A. and Trice, Harrison M. "Unanticipated Returns of Training", Training and Development Journal, vol. 23, no. 7, July 1969, pp. 12-17.

Blumenfeld, Warren S. and Crane, Donald P. "Opinions of Training Effectiveness: How Good? in Evaluating Training Programs, Donald L. Kirkpatrick, ed., Madison, Wisconsin; American Society for Training and Development, 1975.

Brandt, R. M. Studying Behavior in Natural Settings, New York: Holt, Richard and Winston, 1972.

Campbell, D. T. and Stanley, J. T. "Experimental and Quasi-experimental Designs for Research on Teaching", Handbook of Research on Teaching, N. L. Gage (Ed.), Chicago: Rand McNally, 1963.

Carter, W. E. "A Taxonomy of Evaluation Models: Use of Evaluation Models in Program Evaluation", A paper presented at the AERA Annual Meeting, Washington, D.C., April 1975.

Catalanello, Ralph F. and Kirkpatrick, Donald L. "Evaluating Training Programs - The State of the Art" in "Evaluating Training Programs", Donald Kirkpatrick ed., Madison, Wisconsin: American Society for Training and Development, 1972.

Chabotar, Kent J. "Management Development: A Training Evaluation Model", Training, vol. 11, no. 7, July 1974, pp. 42-44, 46, 48, 50.

Cote, D. Phillias, "Measuring Results of Supervisory Training", Training and Development Journal, vol. 23, no. 11, November 1969, pp. 38-46.

Cushman, Harold and Tom, F. K. T. Cornell Diagnostic Observation and Reporting System for Student Description of College Teaching, Ithaca: College of Agriculture and Life Sciences, Cornell University, 1975.

Davis, Glen B. "Zero Population Growth: Effect on Adult Education", Adult Leadership, January 1974.

Dean, Robert Arthur. "An Objective Approach to the Evaluation at Management Training", Ph.D. dissertation, The Ohio State University, 1976.

Dubin, S. S., Mezack, M. and Neideg, R. "Improving the Evaluation at Management Development Programs", in Evaluating Training Programs, D. L. Kirkpatrick, ed., Madison: Wisconsin; American Society for Training and Development, 1975.

"Education in Business," New York, The Conference Board, 1976.

Findlay, Donald C., "Application of the CIPP Evaluation Model to a Center with Multiple Programs Areas and Levels, Educational Technology, October 1971, pp. 43-47.

Ford, George A., "Four Steps Are No Longer Enough", Training and Development Journal, vol. 24, no. 7, July 1970, pp. 29-34.

Glazer, R., "Psychology and Instructional Technology", In Training Research and Education, R. Glazer (ed.), University of Pittsburgh, Reprinted, New York: Wiley, 1965.

Goodacre, D. M., "The Experimental Evaluation of Management Training: Principles and Practice", Personnel, 33 (May 1957).

Hamblin, A. C., Evaluation and Control of Training, London, England: McGraw-Hill, 1974.

Hamblin, A. C., "Evaluation and Control of Training", Industrial Training International, vol. 9, no. 5, pp. 154-156, May 1974.

Hayes, William G. and Williams, Eugene I., "Supervisory Training - An Index of Change" in Evaluating Training Programs, D. L. Kirkpatrick, ed., Madison, Wisconsin; American Society for Training and Development, 1975.

Kelly, Francis Joseph, "Methods of Evaluating Public Sector Management Development Program", Ph.D. dissertation, State University of New York at Albany, 1976.

Kirkpatrick, Donald L., "Evaluating a Training Program for Supervisors and Foreman", The Personnel Administrator, vol. 14, no. 5, September-October 1969.

Kirkpatrick, Donald L., ed. Evaluating Training Programs, Madison, Wisconsin; American Society for Training and Development, Inc., 1975.

Kirkpatrick, Donald L., "Evaluation of Training" in Training and Development Handbook, R. L. Craig and L. R. Brittel, eds., New York: McGraw-Hill, American Society for Training Development, 1967.

Kohn, Vera and Parker Treadway C., "Some Guidelines for Evaluating Management Development Seminars", Training and Development Journal, vol. 23, no. 7, July 1969. pp. 18-23.

Littledale, Harold "Front-End Analysis", Training, vol. 12:3, March 1975, pp. 27-29, 50-52, 54.

Lucco, Robert J., "Conceptualizing Evaluation Strategy: An Evaluation Systems Framework", A Paper presented to the Annual Meeting of the AERA, San Francisco, California, April 1976.

Mager, Robert F., Preparing, Instructional Objectives, Palo Alto, California, Fearon Publishers, 1962.

Miller, Richard D., "A System Concept of Training," Training and Development Journal, vol. 23, no. 4, April 1969, pp. 4-6.

Mindak, William A. and Anderson, Robert E. "Can We Quantify An Act of Faith" in Evaluating Training Programs, D. L. Kirkpatrick, ed., Madison, Wisconsin, American Society for Training and Development, 1975.

Mirsberger, Gerald E., "The Four Crucial Phases of Evaluation" Training, vol. 11-8, pp. 34-35, August 1974.

Moursund, J. P., Evaluation: An Introduction to Research Design, Monterey, California: Brooks/Cole, 1973.

Nadler, L., "Using Critical Events to Develop Training Programmes", Supplement to Industrial Training International, vol. 6, no. 4, 1971.

Odiorne, George S., Training by Objectives: An Economic Approach to Management Training, London, England: Macmillan, 1970.

Osgood, C. E., Suci, George J., and Tannenbaum, Percy H., The Measurement of Meaning, University of Illinois Press, 1957.

Popham, James W., Educational Evaluation, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1975.

Provus, Malcolm, Discrepancy Evaluation, Berkeley, CA: McCutchan Publishing Corporation, 1971.

Pyatte, Jeff A., "Functions of Program Evaluation and Evaluation Models in Education", High School Journal, vol. 53, no. 7, April 1970, pp. 385-399.

Raphael, Michael A. and Wagner, Edwin E., "Training Surveys Surveyed", in Evaluating Training Programs, Donald L. Kirkpatrick, ed., Madison, Wisconsin: American Society for Training and Development, 1975.

Scriven, Michael, "The Methodology of Evaluation" in Curriculum Evaluation, edited by R. E. Stake, American Association Monograph Series on Evaluation, no. 1, Chicago, Illinois: Rand McNally, 1967.

Sommers, Paul A., "An Inferential Evaluation Model", Educational Technology, May 1973, pp. 65-67.

Spoor, Jeremy and Schofield, Andrew, "The Assessment of Training Effectiveness", BACIE Journal, vol. 31, no. 5, May 1977.

Stake, R. E., "The Countenance of Educational Evaluation", Teachers College Record, 1967, pp. 523-540.

Stammers, Robert and Patrick, John, The Psychology of Training, London, England: Methuen and Co., Ltd., 1975.

Steele, Sara M., Contemporary Approaches to Program Evaluation: Implications for Evaluating Programs for Disadvantaged Adults, Syracuse, N.Y. - ERIC Clearinghouse on Adult Education, 1973.

Stufflebeam, D. L., et. al., Educational Evaluation and Decision Making, Itasca, Illinois: Peacock, 1971.

Thorley, S., "Evaluating An In Company Management Training Program", in Evaluating Training Programs, D. L. Kirkpatrick, ed., Madison, Wisconsin; American Society for Training and Development, 1975.

Thorley, S., "Evaluating Management Development", Training in Business and Industry, February 1972, p. 34-36.

"Training Folks to the Top", Training, vol. 14, no. 10, October 1977, pp. 24-32.

Tyler, R. W., "General Statement on Evaluation", Journal of Educational Research, 1942, pp. 492-501.

VanMaanen, John, The Process of Program Evaluation: A Guide for Managers, Washington, D.C., National Training and Development Service Press, 1973.

Walberg, Herbert J., ed. Evaluating Educational Performance Berkley, California; McCutchan Publishing Corporation, 1974.

Walker, Pascal M., "Evaluation of Air Force Employee Development Specialist Training", in Evaluating Training Programs, D. L. Kirkpatrick, ed., Madison, Wisconsin; American Society for Training and Development, 1975.

Warr, P. B., Bird, M. W., and Rackham, N., Evaluation of Management Training, London: Gower Press, 1970.

Webb, E. J., et. al., Unobstrusive Measures: Non Reactive Research in the Social Sciences, Chicago: Rand-McNally, 1966.

Weiss, Carol H., "Alternative Models of Program Evaluation", Social Work, vol. 19:6, November 1974, pp. 675-681.

Wentling, Tim L and Lawson, Tom E., Evaluating Occupational Education and Training Programs, Boston, Mass.: Allyn and Bacon, Inc., 1975.

Wolf, Robert L., "Trial by Jury A New Evaluation Method - I: The Process", Phi Belta Kappan, vol. 57, no. 3, November 1975, pp. 187-188.

Worthen, B. R. and Sanders, J. R., Educational Evaluation: Theory and Practice, Worthington, Ohio: Charles A. Jones Publishing Co., 1973.

Worthen, B. R., "Toward A Taxonomy of Evaluation Designs", Educational Technology, vol. VIII, August 1968.

Zemke, Ron, "Management Training Development: Measuring the Impact", Training, vol. 14, no. 10, October 1977.

APPENDIX A

TRAINING DIRECTOR SURVEY

September 16, 1977

Dear

I am conducting a research project that should make a contribution to the evaluation of training programs in business and industry. In order to achieve an effective result, I need your input of ideas and expectations.

I have worked some with Mr. Robert Engfer, Director of Training at Agway, Inc., in applying some of my ideas on evaluation. He suggested the membership in the Central New York Chapter of American Society of Training and Development as an audience that I might use to get some useful information.

I would appreciate your taking a few minutes to complete the enclosed survey form and returning it to me as soon as possible. The information in the survey will be anonymous and will only be published as group responses.

If you are interested in receiving any of the results of this survey, please indicate so on the form. Thank you for your co-operation.

Sincerely,

Richard D. Jones
Graduate Assistant

EVALUATION OF TRAINING SURVEY

Information will be kept anonymous and confidential

What is the total number of different classroom training programs offered annually by your organization? _____

Of the total training program, what percent are:

management training _____%
supervisory training _____%
technical training _____%

What is the approximate number of employees annually served by training programs? _____

Approximately what percent of training participants select training programs voluntarily? (circle one)

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Briefly describe the current methods you use for training evaluation.

What are some of the special concerns or needs you have in regards to training evaluation?

TRAINING DECISIONS	Degree of Importance to an effective training program (High … Low)	Evaluation in your organization: Is this decision made?	Is the information you have adequate?	Is information formally collected for this decision?	Which of these severely hinder evaluation? (Lack of Resources (People and Money); Lack of Expertise; Lack of Cooperation; Decision is Unimportant; Lack of Time; Other (Specify))
	(Circle One)		(Circle One)		(Check any that apply)
1. Determine need for specific training programs	1 2 3 4 5	yes no	yes no	yes no	
2. Determine previous knowledge and skills of trainees	1 2 3 4 5	yes no	yes no	yes no	
3. Select appropriate instructional strategies	1 2 3 4 5	yes no	yes no	yes no	
4. Determine effectiveness of trainers	1 2 3 4 5	yes no	yes no	yes no	
5. Determine if trainees enjoyed the program	1 2 3 4 5	yes no	yes no	yes no	
6. Determine extent of learning at the end of the program	1 2 3 4 5	yes no	yes no	yes no	
7. Determine extent of trainee application of skills on the job	1 2 3 4 5	yes no	yes no	yes no	
8. Determine if new skills improve job performance	1 2 3 4 5	yes no	yes no	yes no	

APPENDIX B

QUESTIONNAIRE USED IN MANAGEMENT TRAINING EVALUATION FIELD TEST

EVALUATION

AGWAY MANAGEMENT ACTION PROGRAM

The following questionnaire was developed to assist the training department in determining the effectiveness of Agway training programs. Please be frank in your responses and complete the questionnaire as quickly as possible.

DIRECTIONS: Place your birth date at the top of the form as an anonymous identification. Read each of the following six concepts listed in bold print. Based on your reaction, place an X in a space between each pair of adjectives. Complete all responses and stick with your first reaction. Judge each of these relative to your job responsibilities.

EXAMPLE

JIMMY CARTER

liberal __.__.__X__.__.__.__ conservative

peanut __X__.__.__.__.__.__ cashew

successful __.__.__.__X__.__.__ unsuccessful

SETTING OBJECTIVES

important __.__.__.__.__.__.__ unimportant

difficult __.__.__.__.__.__.__ easy

weak __.__.__.__.__.__.__ strong

worthless __.__.__.__.__.__.__ valuable

fast __.__.__.__.__.__.__ slow

good __.__.__.__.__.__.__ bad

painful __.__.__.__.__.__.__ pleasurable

wise __.__.__.__.__.__.__ foolish

free __.__.__.__.__.__.__ constrained

successful __.__.__.__.__.__.__ unsuccessful

soft __.__.__.__.__.__.__ hard

active __.__.__.__.__.__.__ passive

DEVELOPING ACTION PLANS

important __.__.__.__.__.__.__ unimportant

difficult __.__.__.__.__.__.__ easy

weak __.__.__.__.__.__.__ strong

worthless __.__.__.__.__.__.__ valuable

fast __.__.__.__.__.__.__ slow

good __.__.__.__.__.__.__ bad

painful __.__.__.__.__.__.__ pleasurable

wise __.__.__.__.__.__.__ foolish

free __.__.__.__.__.__.__ constrained

successful __.__.__.__.__.__.__ unsuccessful

soft __.__.__.__.__.__.__ hard

active __.__.__.__.__.__.__ passive

DELEGATING MANAGEMENT RESPONSIBILITIES

important __.__.__.__.__.__.__ unimportant
difficult __.__.__.__.__.__.__ easy
weak __.__.__.__.__.__.__ strong
worthless __.__.__.__.__.__.__ valuable
fast __.__.__.__.__.__.__ slow
good __.__.__.__.__.__.__ bad
painful __.__.__.__.__.__.__ pleasurable
wise __.__.__.__.__.__.__ foolish
free __.__.__.__.__.__.__ constrained
successful __.__.__.__.__.__.__ unsuccessful
soft __.__.__.__.__.__.__ hard
active __.__.__.__.__.__.__ passive

ESTABLISHING SPECIFIC MEASURES OF PERFORMANCE

important __.__.__.__.__.__.__ unimportant
difficult __.__.__.__.__.__.__ easy
weak __.__.__.__.__.__.__ strong
worthless __.__.__.__.__.__.__ valuable
fast __.__.__.__.__.__.__ slow
good __.__.__.__.__.__.__ bad
painful __.__.__.__.__.__.__ pleasurable
wise __.__.__.__.__.__.__ foolish
free __.__.__.__.__.__.__ constrained
successful __.__.__.__.__.__.__ unsuccessful
soft __.__.__.__.__.__.__ hard
active __.__.__.__.__.__.__ passive

GETTING SUBORDINATES INVOLVED IN OBJECTIVES

important __.__.__.__.__.__.__ unimportant
difficult __.__.__.__.__.__.__ easy
weak __.__.__.__.__.__.__ strong
worthless __.__.__.__.__.__.__ valuable
fast __.__.__.__.__.__.__ slow
good __.__.__.__.__.__.__ bad
painful __.__.__.__.__.__.__ pleasurable
wise __.__.__.__.__.__.__ foolish
free __.__.__.__.__.__.__ constrained
successful __.__.__.__.__.__.__ unsuccessful
soft __.__.__.__.__.__.__ hard
active __.__.__.__.__.__.__ passive

PERFORMING MANUAL TASKS

important __.__.__.__.__.__.__ unimportant
difficult __.__.__.__.__.__.__ easy
weak __.__.__.__.__.__.__ strong
worthless __.__.__.__.__.__.__ valuable
fast __.__.__.__.__.__.__ slow
good __.__.__.__.__.__.__ bad
painful __.__.__.__.__.__.__ pleasurable
wise __.__.__.__.__.__.__ foolish
free __.__.__.__.__.__.__ constrained
successful __.__.__.__.__.__.__ unsuccessful
soft __.__.__.__.__.__.__ hard
active __.__.__.__.__.__.__ passive

APPENDIX C

RESULTS OF FOLLOW-UP SURVEY
MANAGEMENT TRAINING EVAULATION
FIELD TEST

Follow-Up Evaluation
Management Development Program

Concept - SETTING OBJECTIVES

Variable	Participants n = 8 $\bar{X}$	s	Control n = 31 $\bar{X}$	s	t value of difference
important-unimportant	1.37	.52	1.54	.92	.489
difficult-easy	3.87	1.13	4.12	1.51	.426
weak-strong	3.50	1.41	2.88	1.28	1.164
worthless-valuable	1.63	.74	2.06	1.15	.979
fast-slow	3.50	.75	4.00	1.26	1.046
good-bad	2.00	.92	2.19	.98	.482
painful-pleasurable	2.87	1.35	2.84	1.12	.063
wise-foolish	1.50	.53	1.71	.90	.616
free-constrained	3.62	.91	3.41	1.43	.385
successful-unsuccessful	2.62	.92	2.16	.82	1.342
soft-hard	3.12	.99	3.50	.93	.990
active-passive	2.00	.93	2.54	1.09	1.250

Follow-Up Evaluation
Management Development Program

Concept - DEVELOPING ACTION PLANS

Variable	Participants n = 8 $\bar{X}$	s	Control n = 31 $\bar{X}$	s	t value of difference
important-unimportant	1.37	.52	1.54	.67	.650
difficult-easy	3.62	1.40	4.33	1.49	1.204
weak-strong	2.50	1.00	2.88	1.18	.814
worthless-valuable	1.75	.71	1.94	.68	.680
fast-slow	3.37	.91	3.54	1.12	.386
good-bad	2.12	.64	2.12	.99	0
painful-pleasurable	2.75	1.48	3.17	1.00	.992
wise-foolish	1.63	.74	1.96	.94	.898
free-constrained	3.37	.74	3.38	1.08	.024
successful-unsuccessful	2.12	.98	2.58	.88	1.253
soft-hard	3.50	.75	3.36	.98	.367
active-passive	2.12	.99	2.48	1.00	.886

Follow-Up Evaluation
Management Development Program

Concept - DELEGATING MANAGEMENT RESPONSIBILITIES

Variable	Participants n = 8 $\bar{X}$	s	Control n = 31 $\bar{X}$	s	t value of difference
important-unimportant	1.25	.46	1.29	.46	.213
difficult-easy	2.75	1.90	3.71	1.82	1.283
weak-strong	2.63	1.41	2.88	1.47	.421
worthless-valuable	1.63	1.18	1.59	.72	.117
fast-slow	2.75	1.16	3.19	1.40	.798
good-bad	1.88	1.36	1.93	1.00	.113
painful-pleasurable	2.13	1.35	3.13	1.23	1.956
wise-foolish	1.63	1.06	1.71	.82	.225
free-constrained	2.12	1.35	3.32	1.49	2.015
successful-unsuccessful	2.00	.93	2.23	1.17	.502
soft-hard	3.75	1.58	4.52	1.28	.492
active-passive	2.00	1.30	2.54	1.12	1.144

Follow-Up Evaluation
Management Development Program

Concept - ESTABLISHING SPECIFIC MEASURES OF PERFORMANCE

Variables	Participants n = 8		Control n = 31		
	$\bar{X}$	s	$\bar{X}$	s	t value of difference
important-unimportant	1.37	.74	1.38	.55	.041
difficult-easy	3.25	2.13	4.10	1.79	1.121
weak-strong	2.50	1.31	3.07	1.36	1.037
worthless-valuable	1.13	.35	1.81	1.04	1.030
fast-slow	3.37	1.68	3.71	1.46	.553
good-bad	1.50	1.06	1.61	.84	.304
painful-pleasurable	1.63	1.06	2.87	1.36	2.335
wise-foolish	1.50	1.06	1.90	.74	1.204
free-constrained	1.87	1.12	3.12	1.33	2.380
successful-unsuccessful	2.00	1.30	2.19	1.07	.416
soft-hard	3.13	1.24	4.22	.99	1.526
active-passive	2.38	1.18	2.48	1.23	.201

Follow-Up Evaluation
Management Development Program

Concept - GETTING SUBORDINATES INVOLVED IN OBJECTIVES

Variables	Participants n = 8		Control n = 31		t value of difference
	$\bar{X}$	s	$\bar{X}$	s	
important-unimportant	1.50	1.19	1.77	1.02	.627
difficult-easy	3.13	1.72	4.30	1.48	1.875
weak-strong	2.38	1.18	2.94	1.31	1.070
worthless-valuable	1.63	1.18	1.75	.72	.353
fast-slow	3.00	1.06	3.71	1.44	1.272
good-bad	2.12	1.24	2.06	1.21	.121
painful-pleasurable	3.00	.75	3.07	1.23	.150
wise-foolish	2.00	1.06	2.03	.94	.076
free-constrained	2.87	1.12	3.54	1.45	1.185
successful-unsuccessful	2.75	1.28	2.71	1.13	.084
soft-hard	3.37	1.18	3.87	1.04	1.147
active-passive	2.25	1.03	2.61	1.25	.731

Follow-Up Evaluation
Management Development Program

Concept - PERFORMING MANUAL TASKS

Variables	Participants n = 8 $\bar{X}$	s	Control n = 31 $\bar{X}$	s	t value of difference
important-unimportant	4.13	1.88	2.38	1.08	3.340
difficult-easy	4.13	1.72	2.61	1.49	2.420
weak-strong	3.75	.88	2.91	1.37	1.605
worthless-valuable	3.13	1.64	2.55	.96	3.424
fast-slow	3.75	1.38	3.16	1.12	1.230
good-bad	4.63	1.76	2.35	.91	4.923
painful-pleasurable	3.37	1.06	2.61	1.14	1.660
wise-foolish	4.87	1.64	2.61	1.33	3.967
free-constrained	4.38	1.76	3.12	1.25	2.258
successful-unsuccessful	3.50	.92	2.23	.88	3.976
soft-hard	4.13	.64	4.17	1.28	.083
active-passive	4.38	1.50	2.64	1.19	3.392

APPENDIX D

QUESTIONNAIRE FOR EVALUATION OF TEACHING FIELD TEST VERSION

TRAINER EVALUATION SURVEY

THE INSTRUCTOR(S):	HARDLY EVER	OCCASIONALLY	SOMETIMES	FREQUENTLY	ALMOST ALWAYS
1. Described at the beginning of each class what he/she planned to do	1	2	3	4	5
2. Pointed out what was important to learn in each class	1	2	3	4	5
3. Gave step-by-step instructions when needed	1	2	3	4	5
4. Stated the objectives of the program	1	2	3	4	5
5. Brought needed materials to class	1	2	3	4	5
6. Involved participants that appeared uninterested in the program	1	2	3	4	5
7. Presented material in a well-organized fashion	1	2	3	4	5
8. Promoted teacher-student discussion (as opposed to mere response to questions)	1	2	3	4	5
9. Encouraged all individuals to participate	1	2	3	4	5
10. Initiated conversation with participants before and after class	1	2	3	4	5
11. Resolved any conflicts that arose	1	2	3	4	5
12. Praised participants during class	1	2	3	4	5
13. Was harshly critical of individual responses	1	2	3	4	5
14. Provided individual assistance when needed	1	2	3	4	5
15. Provided relevant information in response to questions	1	2	3	4	5
16. Addressed participants by names	1	2	3	4	5
17. Explained information in a manner understood by participants	1	2	3	4	5

TRAINER EVALUATION SURVEY (CONT'D)

THE INSTRUCTOR(S):	1 HARDLY EVER	2 OCCASIONALLY	3 SOMETIMES	4 FREQUENTLY	5 ALMOST ALWAYS
18. Displayed concern that students learn	1	2	3	4	5
19. Made positive statements about the subject matter of the program	1	2	3	4	5
20. Spoke with expressiveness and variety in tone of voice	1	2	3	4	5
21. Was poised during presentations	1	2	3	4	5
22. Used understandable vocabulary	1	2	3	4	5
23. Used distracting mannerisms in speaking	1	2	3	4	5
24. Moved around the classroom	1	2	3	4	5
25. Exhibited a sense of humor	1	2	3	4	5
26. Used gestures while teaching	1	2	3	4	5
27. Related program to real-life situations	1	2	3	4	5
28. Presented course material at too slow a pace	1	2	3	4	5
29. Used a variety of teaching techniques	1	2	3	4	5
30. Indicated when a new topic was introduced	1	2	3	4	5
31. Used examples to make a point	1	2	3	4	5
32. Presented material at too fast a pace	1	2	3	4	5
33. Adjusted program to the needs of the participants	1	2	3	4	5
34. Used visual aids to complement oral presentations	1	2	3	4	5
35. Summarized material presented in each class session	1	2	3	4	5
36. Provided participants with practice in recalling knowledge	1	2	3	4	5

TRAINER EVALUATION SURVEY (CONT'D)

THE INSTRUCTOR(S):	HARDLY EVER	OCCASIONALLY	SOMETIMES	FREQUENTLY	ALMOST ALWAYS
37. Provided participants with practice in recalling principles or theories	1	2	3	4	5
38. Provided participants with practice in problem-solving & decision-making	1	2	3	4	5
39. Provided participants with practice in organizing & presenting ideas	1	2	3	4	5
40. Provided participants with practice in developing manual skills	1	2	3	4	5
41. Discouraged participants from expressing themselves openly & freely	1	2	3	4	5
42. Used a variety of teaching materials	1	2	3	4	5
43. Provided participants with opportunities to be creative	1	2	3	4	5

44. Compared to other education programs you have participated in, please rate this program as to its educational value:

5 - One of the best

4 - Above average

3 - Average

2 - Below average

1 - One of the worst

APPENDIX E

SUGGESTED QUESTIONNAIRE FOR TRAINER EVALUATION

TRAINER EVALUATION SURVEY

THE INSTRUCTOR(S):	HARDLY EVER	OCCASIONALLY	SOMETIMES	FREQUENTLY	ALMOST ALWAYS
1. Pointed out what was important to learn in each class	1	2	3	4	5
2. Gave step-by-step instructions when needed	1	2	3	4	5
3. Stated the objectives of the program	1	2	3	4	5
4. Presented material in a well-organized fashion	1	2	3	4	5
5. Promoted teacher-student discussion (as opposed to mere response to questions)	1	2	3	4	5
6. Encouraged all individuals to participate	1	2	3	4	5
7. Initiated conversation with participants before and after class	1	2	3	4	5
8. Resolved any conflicts that arose	1	2	3	4	5
9. Praised participants during class	1	2	3	4	5
10. Provided individual assistance when needed	1	2	3	4	5
11. Provided relevant information in response to questions	1	2	3	4	5
12. Addressed participants by names	1	2	3	4	5
13. Explained information in a manner understood by participants	1	2	3	4	5
14. Displayed concern that students learn	1	2	3	4	5

TRAINER EVALUATION SURVEY (CON'D)

THE INSTRUCTOR(S):	HARDLY EVER	OCCASIONALLY	SOMETIMES	FREQUENTLY	ALMOST ALWAYS
15. Made positive statements about the subject matter of the program	1	2	3	4	5
16. Spoke with expressiveness and variety in tone of voice	1	2	3	4	5
17. Used understandable vocabulary	1	2	3	4	5
18. Used distracting mannerisms in speaking	1	2	3	4	5
19. Related program to real-life situations	1	2	3	4	5
20. Presented course material at too slow a pace	1	2	3	4	5
21. Used a variety of teaching techniques	1	2	3	4	5
22. Used examples to make a point	1	2	3	4	5
23. Presented material at too fast a pace	1	2	3	4	5
24. Summarized material presented in each class session	1	2	3	4	5
25. Provided participants with practice in recalling knowledge	1	2	3	4	5