

1-1-1977

The development of an evaluation methodology for professional training practitioners to evaluate educational training workshops.

Thomas E. Raynor
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_1

Recommended Citation

Raynor, Thomas E., "The development of an evaluation methodology for professional training practitioners to evaluate educational training workshops." (1977). *Doctoral Dissertations 1896 - February 2014*. 4606.
<https://doi.org/10.7275/13475136> https://scholarworks.umass.edu/dissertations_1/4606

This Open Access Dissertation is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations 1896 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

★ UMASS/AMHERST ★

312066 0296 7127 6

**FIVE COLLEGE
DEPOSITORY**

THE DEVELOPMENT OF AN EVALUATION METHODOLOGY FOR
PROFESSIONAL TRAINING PRACTITIONERS TO
EVALUATE EDUCATIONAL TRAINING
WORKSHOPS

A Dissertation Presented

by

THOMAS E. RAYNOR

Submitted to the Graduate School of the
University of Massachusetts in partial
fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

April

1977

© 1977

THOMAS EDMUND RAYNOR

ALL RIGHTS RESERVED

The Development of an Evaluation Methodology for
Professional Training Practitioners to
Evaluate Educational Training
Workshops

A Dissertation Presented

by

THOMAS E. RAYNOR

Approved as to style and content by:

Kenneth H. Blanchard

Dr. Kenneth Blanchard
Chairperson of Committee

Ronald K. Hambleton

Dr. Ronald Hambleton
Member

Frederic E. Finch

Dr. Frederic Finch
Member

Mario Fantini

Dr. Mario Fantini, Dean
School of Education

ACKNOWLEDGEMENTS

To the many special people in my life who through their individual concern and encouragement inspired me to venture forth and achieve many of my life goals.

To my Mother and my Father for the years of love, support, understanding, faith and personal sacrifice which has enabled me to arrive at this point in my life.

To my Wife whose love and faith over the past years has provided me with the inspiration and desire to take on this project as well as others as I work toward becoming a total person.

To Karine who through those first weeks and later months taught me the new joy of sharing the small but meaningful pleasures of life.

To Nathalie whose most recent arrival has added further joys to my life.

To Hedley Dimock, friend, colleague and teacher who through the years has provided unselfish assistance and advice thus making much of present achievements possible.

To Kenneth Blanchard, friend, advisor, Doctoral Chairman, who through his support, direction and guidance was continually helpful and invaluable during my stay at the University of Massachusetts.

To Ron Hambleton, friend, research advisor and committee member for his endless suggestions, advice and helpfulness while guiding me through the dissertation process.

To Fred Finch, a friend and committee member for his continuing interest and timely support and advice during my stay at the University of Massachusetts.

To Doug Forsythe for his helpful suggestions and continuing interest throughout my comprehensive exams and dissertation.

To Bob Sinclair for his specific assistance, friendship and his guidance through my chosen curriculum.

To Horace Reed who was a pleasure to work with and for whom I will always regard with the highest esteem and greatest admiration.

To all others, and in particular my fellow students John Beck, Susan Schraggle, and Gwen Brown who shared my highs and lows and provided advice and understanding when needed and greatly enhanced my stay at the University of Massachusetts.

The Development of an Evaluation Methodology for
Professional Training Practitioners to
Evaluate Educational Training
Workshops

(April 1977)

Thomas E. Raynor, B.A. Sir George Williams University,
M. Ed. Springfield College, Ed. D., University of Massachusetts
Directed by: Dr. K.H. Blanchard

ABSTRACT

Since the majority of training practitioners lack evaluation skills and knowledge, it was timely that an evaluation methodology was developed to provide them with the expertise necessary to evaluate training workshops.

The purpose of this study was to provide a series of detailed evaluative steps for the inexperienced potential evaluator, thus enabling him/her to evaluate his/her training efforts. The second purpose was to provide evaluation knowledge for non-evaluators enabling them to hire competent external evaluators.

This study describes eight steps to evaluation -- needs assessment and goal clarification, choosing evaluation models, selecting research designs, data collection and instrumentation, analyzing data, reporting

outcomes and giving feedback, evaluating the evaluation, and finally, action deliberation and redesign. Each step is presented as a chapter and contains a theory and theory application section.

Step one, "Needs assessment and goal clarification" reviews the importance of setting goals and objectives and presents eighteen models for practical use. The application section presents four typical situations evaluators encounter when assessing needs and then provides a tentative model for use in each case.

The second step of the evaluation methodology, "Choosing evaluation models", discusses various approaches to evaluation. Six models for application are recommended while highlighting three models for particular use.

Step three, "Selecting research designs" emphasizes the importance of selecting an appropriate research design as well as including the important factors influencing its selection (i.e. control groups, validity, sampling and contamination). The application section provides six practical research designs for consideration for training workshops.

"Data collection and instrumentation", step four, highlights the use of selected data collection techniques such as observation, interviewing and questionnaires.

Suggestions for instrument revision, pretesting and editing are found in the application section. Included also is a detailed discussion on "interviewing" and the constructing and scoring of "Likert scales".

The fifth step, "Analyzing data", is a brief and selected review of statistical procedures presented for the novice. In the application section, examples are given for the use of some of the various statistical procedures mentioned.

Step six, "Reporting outcomes and giving feedback", discusses the importance of the evaluator's report and how information is reported. Report content, presentation methods, data organization, communicating results and ethics are included. In the application section, primary focus is given to principles and consideration for evaluators when organizing and presenting reports.

"Evaluating the evaluation", step seven, presents professional opinions on the check points to use when reviewing evaluations before decision-making. The application section cautions the evaluator in making premature decisions prior to reviewing the evaluation process for omissions and discrepancies. A check list for evaluation models has been presented as a guide.

The final step, eight, "Action deliberation and redesign", outlines various models and methods for

operationalizing action plans. Models of cost analysis and cost projection for decision-making are outlined in the application section of this chapter.

It is intended that this evaluation methodology will provide the stimulus for training practitioners to introduce evaluative procedures within their future training efforts, thus increasing the quality of the training and the training results.

The research of this study has indicated an absence of usable evaluative designs, models, and instrumentation for workshop training. This area should be considered as one for priority development by evaluators. The study recommends the creation of other methodologies to challenge present thinking and interest, as well as suggesting the compiling and documenting of the volumes of evaluation material found in various related fields. Finally, the study suggests that it is timely for the development and planning of a series of evaluation workshops for training practitioners interested in raising the quality and effect of their training endeavours.

This research study has provided the training practitioner with a methodology for evaluating training workshops as well as providing knowledge and guidelines in evaluation for those persons or organizations wishing to hire outside evaluators. As a result, the quality

of training and training outcomes in workshops can be increased while more competent outside evaluators can be selected to evaluate training programs.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
ABSTRACT	vi
LIST OF TABLES	xv
 CHAPTER	
I. INTRODUCTION.	1
1.1 Statement of the Problem	1
What an evaluation can provide	4
Growing support for evaluation	5
The need to evaluate training workshops.	6
Pressures to evaluate.	9
1.2 Statement of the Purpose	10
1.3 Definition of Terms.	12
1.4 An Overview of Evaluation Methodology.	13
II. NEEDS ASSESSMENT AND GOAL CLARIFICATION	16
2.1 Theory	16
Determining needs.	16
Setting goals and objectives	16
Models for assessing needs and clarifying goals	24
2.2 Application.	37
Approaching evaluation	37
Organizing for evaluation.	38
Goal clarification and needs assessment.	40
Specific models for clarification and assessment.	42
III. CHOOSING EVALUATION MODELS.	52
3.1 Theory	52
Introduction	52
Goal free and goal based evaluation.	52
Summative evaluation	53
Formative evaluation	55
Models for evaluation.	56
3.2 Application.	64
Models for evaluation.	64
Program models for evaluation.	66
Pilot program.	66
Model program.	67

CHAPTER

Page

IV.	SELECTING RESEARCH DESIGNS	69
4.1	Theory.	69
	Purpose of the research design.	69
	Variables	70
	Use of control group.	73
	Contaminations.	77
	Experimental validity	78
	Experimental samples.	81
	Selecting research designs.	83
	Soft technique research	84
	Pretesting and posttesting research designs.	86
	One-shot case study and one group, pretest and posttest design.	87
	Expost facto design	88
	Comparative research design	90
	Own control group design.	92
	Time series design.	93
	Nonequivalent control group design.	95
	Pretest, posttest, control group design	97
4.2	Application	105
	Research design	105
	One-shot case study design.	106
	One group, pretest, posttest design	107
	Time series design.	109
	Comparable design	111
	Experimental design	112
V.	DATA COLLECTION AND INSTRUMENTATION.	116
5.1	Theory.	116
	Introduction.	116
	Observation	117
	Questionnaires and interviews	123
	Types of questionnaires and interviews	126
	Sociometry.	131
	Popularity of data gathering techniques	133
	Projective techniques	134
	Attitude and rating scales.	135
5.2	Application	145
	Introduction.	145
	Use of questionnaires	145
	Use of interviews	150

CHAPTER	Page
Use of observation techniques	158
Use of scales	159
VI. ANALYZING DATA	162
6.1 Theory.	162
Introduction.	162
Descriptive statistics.	163
Frequency distributions	164
Measures of central tendency.	164
Graphs.	167
Variability	168
Variance.	169
Correlation and regression coefficients	171
Inferential statistics	172
Significance testing.	173
Reliability and validity.	175
6.2 Application	176
Descriptive statistics.	176
Frequency distributions	176
Measures of central tendency.	177
Variability	179
Correlation coefficients.	179
Inferential statistics	180
Introduction.	180
VII. REPORTING OUTCOMES AND GIVING FEEDBACK . . .	183
7.1 Theory.	183
The report.	183
Communicating results	186
Professional ethics	186
Feedback.	187
7.2 Application	189
Providing feedback.	189
Writing the report.	191
Reporting and feedback in training workshops.	192
VIII. EVALUATING THE EVALUATION.	195
8.1 Theory.	195
Introduction.	195
Further checkpoints in evaluating evaluation designs	196

CHAPTER

Page

8.2	Application	197
	Evaluating the evaluation	197
IX.	ACTION DELIBERATION AND REDESIGN	199
9.1	Theory.	199
	Decision-making	199
	Models for decision-making.	202
	Delphi method	202
	Cost analysis and projection.	202
	Redesign.	206
9.2	Application	206
	Cost projection	206
	Cost analysis	209
	Redesign.	209
X.	SUMMARY, RECOMMENDATIONS, SIGNIFICANCE AND LIMITATIONS OF THE STUDY	212
10.1	Summary of the Study.	212
10.2	Recommendations for additional research	215
10.3	Significance of the study	218
10.4	Limitations of the study	219
	Field test.	219
	Instrumentation	219
	Selective theory review	220

APPENDIX

APPENDIX A:	Comparisons of contemporary evaluation models on selected characteristics	221
APPENDIX B:	A sociometric matrix-friendship chart	229
APPENDIX C:	A self rating scale	230
APPENDIX D:	A self administered, standardized questionnaire entitled "Workshop Evaluation Questionnaire"	231
APPENDIX E:	A semantic differential scale	235
APPENDIX F:	Sample observation questionnaire.	237
APPENDIX G:	An example rating scale	240
APPENDIX H:	Sample table formats for reporting data.	241
APPENDIX I:	An evaluation design checklist.	242
BIBLIOGRAPHY.	245

LIST OF FIGURES

	Page
FIGURE 1. Steps to evaluation	15
FIGURE 2. Independent and dependent variable axes.	72
FIGURE 3. One shot case study	88
FIGURE 4. One group pretest, posttest design. . .	88
FIGURE 5. Expost factor design.	89
FIGURE 6. Comparative design.	92
FIGURE 7. Own control group design.	93
FIGURE 8. Time series design.	95
FIGURE 9. Nonequivalent control group design. . .	97
FIGURE 10. Pretest, posttest, control group design (with random selection "R"). . .	98
FIGURE 11. Some advantages and disadvantages of selected research designs.	101
FIGURE 12. Summary of methods of data collection including comments	138
FIGURE 13. Positive skewed curve	165
FIGURE 14. Negatively skewed curve	166
FIGURE 15. Biomodal histogram graph.	167
FIGURE 16. Biomodal polygram graph	167
FIGURE 17. Scatter plot graph.	168
FIGURE 18. Normal curve graph.	168
FIGURE 19. Positive correlation scattergram. . .	172
FIGURE 20. Negative correlation scattergram. . .	172

C H A P T E R I

INTRODUCTION

1.1 Statement of the Problem

A careful review of training programs in the field of Education has revealed that neither evaluation methods nor information resulting from evaluation studies have been used for evaluating new programs or for providing the basis of information for implementing necessary changes. This practice has been supported by traditional public attitudes which have tended to associate evaluation and experimentation with the natural sciences and not with the social sciences (Lehman & Mehrens, 1971). As a result of the lack of use of evaluation, training practitioners have never fully realized the short and long term effects, or the side effects of their training efforts. Neither have they been able to determine accurately the exact causes of their successes or their failures in training.

Ralph Tyler's study (Tyler & Smith, 1942) applies evaluation technology and philosophy in evaluating curriculum. For education, this study has remained a classic as it represents one of the first attempts to apply an evaluation methodology to an educational program with the intention of determining the degree of success of the program and at the same time assess its strengths and weaknesses. With the exception of this study, there has been little in the

way of evaluation activity in the social sciences until the early 1960's.

Since then, one may find evidence of numerous attempts of evaluating educational training programs by practitioners. Most of the studies recorded demonstrate evaluative procedures which are typically unsystematic, informal and summative in nature while calling upon the subjective responses of the training program's participants to state their degree of pleasure or dissatisfaction with the program they have just completed (Belasco & Trice, 1969; Benne, Bradford & Lippitt, 1964; Schmuck & Miles, 1971). Current literature illustrates that more proficient or scientific methods of evaluation have generally been resisted by training practitioners in the social sciences. Some practitioners have ignored or refused to consider any form of evaluative procedures when implementing change programs or training as it is their belief that evaluation is unnecessary and that "change effort equals change itself" (Belasco & Trice, 1969). Other practitioners have resisted any suggestions to use evaluation methodology as a result of some prior negative experiences with evaluation or they have experienced some disappointment with the outcomes of one or more evaluation efforts (Weiss, 1972).

In 1973, Payne compiled a list of additional reasons why practitioners have resisted opportunities to implement

or support evaluation attempts. Some of these reasons are listed as follows:

1. Traditional thinking has viewed evaluation as interfering with program development.
2. Evaluation has the interests and desires of the program or training at odds with the interests and desires of the program staff.
3. Traditional views of the evaluator's role has the evaluator specifying the program objective (many evaluators would disagree with this view, thus it is a source of conflict for some).
4. Evaluation is seen by many only as a long term proposition, being expensive and not necessarily for short term training or programming.
5. For some, evaluation is only seen acceptable if it is bristling with correlations, tests of significance and random samplings of a thousand persons.

Belasco and Trice (1969) add:

6. Evaluation is seen by many as an invasion of privacy, specifically if it is done from within, by an organization.

Furthermore, it has been the experience of some, after confidentially and willingly sharing opinions and information with evaluators, to have this very same information later used against them by a superior, peer or subordinate. Most often, these cases constitute a minority, but, when a confidentiality is broken, the total organizational system suffers from the negative effects. Persons working in close proximity and wishing not to become victims of such irresponsible sharing of information, avoid all contact with all sources which they see as potentially harmful to their

present status, position and welfare. Unfortunately, it is these isolated instances that receive prominence and become generalized to all evaluation situations. The result is what Payne (1973) refers to as the generally accepted "Old Wives Tales" of evaluation.

Organizations and individuals doing private training have also resisted using evaluation not just because of the cost, time and attention demanded by evaluation but because of the wide range of skills needed by those doing evaluations. Most practitioners or training personnel recognize that they are neither evaluators nor are they researchers. They recognize that today's evaluator must have a multidisciplinary background as well as a well-developed area of specialization related to the field. Even then, an evaluation can be a complicated and difficult activity even for the trained person to perform. Payne, himself an experienced evaluator, expresses that "the new generation evaluator must be part sociologist, economist, social psychologist, anthropologist and philosopher and possess a strong self concept, high tolerance for ambiguity and should have the patience of a United Nations arbitrator".

What evaluation can provide

Despite Payne's previously stated view concerning the skills needed by an evaluator, evaluation is a necessary and required process which can provide a wealth of data, information and insight for the recipient. Matthew B. Miles

(1969), a researcher in the field of education, states "Any evaluation attempt which moves in the direction of helping an organization or individual develop a more systematic and thoughtful assessment of what is happening in a program or training effort is all to the good".

Evaluation can bring the following benefits to a program by providing a continuous flow of useful data:

1. on which program practitioners can base their decisions to modify, expand, cancel, or continue a program.
2. which can be used in judging or deciding if certain activities or treatments being applied are achieving the desired effects and to what extent.
3. which provides the information by which individual segments or phases of a program can be judged as being effective or ineffective.
4. which can provide information to the program participants and practitioners as to whether the program is accomplishing its goals and objectives.
5. which can be fed back to the participants making them aware of their personal progress toward both personal and organizational goals. In addition, this information has been frequently seen to be a positive force which has the effect of increasing individual motivation thus adding to the chances of both the training and the program attaining success.

Growing Support for Evaluation.

Both organizations and individuals are becoming more aware of the benefits and advantages which evaluation can bring to training workshops. Due to budget restraints and insufficient knowledge of the requirements of good evaluation, many organizations and individuals attempt to carry

out an evaluation themselves. Naturally, many of those persons doing evaluations often lack the skill and expertise to make it a success. As a result, the final evaluation product is poorly done and it is only later that it is found to be lacking in reliability, validity and worth. This is not to mention what is lost to the organization in cost, time and commitment by all involved (Buchanan, 1971).

Other organizations which have become newly aware of the advantages of evaluation may respond by inviting an external person into the organization to do the evaluation. Because the organization does not thoroughly understand what has to be done and what is involved, their choice of evaluator may not be suitable for the evaluation at hand. Thus, there are unsatisfactory results for all involved. How can an organization avoid this pitfall? The choices are limited. Firstly, the organization may attempt to increase its knowledge of evaluation and thus be in a better position to hire an external evaluator. Or, secondly, the organization may provide sufficient training opportunities and resources for key staff members to become proficient enough in evaluation methodology to carry on adequately the evaluation requirements internally themselves.

The need to evaluate training workshops

Despite the traditional public attitudes towards evaluation, and the present resistances and problems that arise

when attempting to evaluate programming and training, there is a growing vocal minority who strongly advocate the continued development and influence of evaluation in the planning of change strategies and measuring effectiveness. The following is a list of some of those persons who promote the use of evaluation for determining effectiveness and for planning change: Argyris, 1965; Belasco and Trice, 1969; De Phillips, Berliner and Cribbin, 1960; Finklestein, 1971; Jenks, 1970; Lippitt, This, and Bidwell, 1971; Payne, 1973; Popham, 1974; Schmuck and Runkle, 1972; Scriven, 1967; Stake, 1967; Stufflebeam, 1971; Tracy, 1968; Verner and Booth, 1964; and Vroom, 1964. Each of these persons has been actively involved in educating individuals through training programs, often referred to as "training workshops".

In the field of education "training workshops" are commonly used by those persons doing leadership training, humanistic education, and teacher education. These workshops are presented by practitioners to achieve a variety of outcomes, some of which are:

1. Clarifying communication
2. Establishing and owning goals
3. Uncovering conflict and interdependence
4. Improving group procedures
5. Improving problem-solving procedures
6. Improving decision-making procedures
7. Assessing long and short-term changes

Belasco and Trice (1969) estimate that probably 99% of present day training goes unevaluated. These figures can be readily applicable to training workshops in the field of education. In addition, of the remaining 1% who do evaluate their workshops, it is suspected that a sizable number are performing evaluation incorrectly or at least, under poorly controlled conditions. This is unfortunate, as a well-planned and organized evaluation performance could provide specific information pertaining to the strengths and weaknesses of the program, the degree of improvement of the participants at various points in the training workshop, and the effectiveness of the program moving toward and achieving the desired outcomes. Additional data for planning and improving future programs would also be made available for any needed program adjustments. Data obtained from a well-designed and well-performed evaluation can be invaluable and unlimited in its impact for improving training, motivating participants, and providing high quality information for decision-making. In highlighting the importance of evaluation, Elsie Finklestein (1971, p. 3) states that "evaluation is important as a source of knowledge and direction, it tells which program works, which does not and it gives direction to better formulation of policy and program planning. Without evaluation, educational learning and training processes move at an undetermined speed and in an uncharted direction".

Pressures to evaluate

Times have changed abruptly in recent years for those who have been reluctant to utilize evaluation methods for decision-making and planning. Organizations who for years have been operating on Federal funding and support for a wide variety of projects and training activities are now being faced with an ultimatum -- "evaluate your effectiveness or have your funding terminated". In other areas, training departments are being challenged to justify their expenditures, their time, and their effective utilization of staff. Educators and training practitioners who, for years have been allowed to publish case studies, workshop experiences and other related materials based on speculation and feelings rather than solid evaluative results, now are experiencing severe criticism for the absence of a more scientific approach to their methods of diagnosis and resulting speculations.

As educational training workshops will probably remain one of the more popular vehicles for passing on new information to those persons responsible for the delivery of educational services, it is essential that care be taken to apply good evaluation procedures to workshops during the stages of conceptualizing, planning and implementing of the program. This will ensure that the training will have greater potential for achieving the desired effect, as well as providing the base for higher quality decision-making in training.

In summary, despite the obvious lack of use of evaluative procedures for assessing new programs and changes, despite the resistance to evaluation by some, there is growing pressure from governments, organizations and funding bodies to evaluate programs and training in order to determine the degree of effectiveness in achieving goals or in providing data for improved decision-making for program improvement.

This methodology will reflect the content of Van Mannen's (1973, p. 35) definition of an evaluation methodology. He refers to the ideas of Hirshi and Selvin by stating that "a methodology is not an everlasting truth. It is a living body of ideas that change with time... there is no purely logical method of always linking concepts to indicators".

In this context this study has attempted to provide an evaluation methodology which is flexible and adaptable, yet will provide the new evaluator/practitioner with an instrument which s/he will be able to apply readily and consult as a guide when attempting to evaluate program effectiveness and progress in a training workshop environment.

1.2 Statement of the Purpose

The research and development of a methodology for evaluating educational training workshops was undertaken because of the lack of a methodology in the field which could be easily understood and applied by the training practitioner

in training situations. At the present time, this information is available in only a disorganized and scattered form. In preparing this study, attempts were made to localize and organize this information in a usable format as part of the methodology. In addition, references and sources for this material were provided wherever possible.

It is hoped that the preparation of this study will encourage a new interest and enthusiasm in evaluation by those active in leadership training, humanistic education and teacher education. It is further hoped that the development of this methodology will also serve to reduce the resistance to evaluation as well as to disprove many of the fears and myths surrounding the use of evaluative procedures in training workshops.

The purposes of this research study were as follows:

1. To develop an evaluation methodology for the professional practitioner who is a novice to evaluation, but who is interested in evaluating the effectiveness of his/her training endeavors. This methodology includes a comprehensive review of the current literature related to the evaluation field as well as providing a step-by-step guideline for performing an evaluation of a training workshop.
2. To develop an evaluation methodology which be used as a resource to organizations or individuals who wish to hire an external person to evaluate a training workshop. The evaluation methodology contains instructions, references for referral and discussion relating to goal selection, data gathering procedures, data analysis and interpretation, feedback, reporting and selecting alternative actions. This provides the organization or individual(s) with sufficient information on evaluation and evaluation technology that they could hire the

appropriate evaluator and be informed to ask knowledgeable questions when the occasion presents itself.

It is hoped that the preparation of this study will encourage a new interest and enthusiasm in evaluation by those active in leadership training, humanistic education and teacher education. It is further hoped that the development of this methodology will also serve to reduce the resistance to evaluation as well as disproving many of the fears and myths surrounding the use of evaluative procedures in training workshops.

1.3 Definition of Terms

The following are definitions of key terminology used within the context of this study.

Behavioral objectives - Individuals or group goals which have been stated behaviorally. For the purpose of evaluation, they must be stated clearly, specifically and in measurable terms.

Educational settings - Environments where both informal and formal learning take place.

Evaluation - The process of ascertaining the decision areas of concern, selecting appropriate information, and analyzing information in order to report summary data useful to decision-makers in selecting among alternatives (Alkin, 1969).

Evaluation methodology - A body of methods, rules and postulates employed in evaluation. In addition, it provides an organized, systematic explanation of steps and processes to be employed when carrying through an evaluation.

Formative evaluation - Provides information and data that can be fed back during a training program or workshop for the purpose of improving it and identifying areas where modifications must be made in order to attain the specified goals.

Novice - A beginner. In the context of this study, a person who has little knowledge, skill or sophistication in evaluation methods, but may very well be highly skilled in other professional fields, for example teacher training.

Sponsor - An organization or group financially supporting or requesting a training program of activity.

Summative evaluation - Provides information and data on the results of a training program or workshop after it has been completed. It provides information on the overall effectiveness and success of the training workshop.

Trainer - A name used to identify persons whose responsibility it is to lead or carry on training workshops.

Training practitioner - Another term used for "Trainer".

Training workshop - A brief, systematic and organized educational program of instruction, involving a small group of people in a given field. This usually involves the active participation and interaction of all the participants in developing new skills and increasing their learning.

1.4 An Overview of Evaluation Methodology

After a careful review of evaluation literature and a review of selected studies, it appeared that the various areas or steps necessary for the evaluator to know can be merged into eight areas. They are:

1. Needs Assessment and Goal Classification
2. Choosing Evaluation Models
3. Selecting Research Designs
4. Data Collection and Instrumentation
5. Analyzing Data
6. Reporting Outcomes and Providing Feedback
7. Evaluating the Evaluation
8. Action Deliberation and Redesign

Further explanation of these steps and how they are to be

used is described later in the various chapters of the study. Each chapter is divided into a theory and in application section highlighting one of the eight action steps proposed for evaluating training workshops. The final chapter, Chapter X, is a brief summary and conclusion of the total methodology as well as presenting suggestions for further areas of study and research.

1. NEEDS ASSESSMENT AND GOAL CLARIFICATION	5. ANALYSIS OF DATA
2. CHOOSING EVALUATION MODELS	6. REPORTING OUTCOMES AND PROVIDING FEEDBACK
3. SELECTING RESEARCH DESIGNS	7. EVALUATING THE EVALUATION
4. DATA COLLECTION AND INSTRUMENTATION	8. ACTION DELIBERATION AND REDESIGN

FIGURE 1. Steps to evaluation.*

- * Appreciation is expressed for the guidance and specific suggestions provided by Dr. R.K. Hambleton, School of Education, University of Massachusetts in the development of the various "Steps to evaluation" listed above.

C H A P T E R I I

2.0 NEEDS ASSESSMENT AND GOAL CLARIFICATION

2.1 Theory

Determining needs

The failure or success of training workshops has often been found to be very dependent on how carefully the evaluator or sponsors have identified and isolated the specific needs of those persons for whom the training is being provided. The over-looking of this procedure results in the absence of establishing comparable measures or goals. It is then impossible to determine if the outcomes of a training endeavor have met the desired specifications intended at the outset of the training. For some evaluators, the identification of specific needs is the basis for specifying the goals and objectives required in most popular forms of evaluation. Yet, there are some in the field who see this differently.

Setting goals and objectives

According to Welsh and Hambleton (1976), a recurring question in the evaluation field is "Should goals be used in evaluation?" Scriven (1973) advocates an approach to evaluation where the person(s) performing the evaluation are unaware of the goals of the evaluation effort being performed. Scriven refers to this as "goal free" evaluation. He sees the function of the evaluator as one of attending a

program/training session and developing specific feelings and impressions for what is happening or occurring among the participants of the program. Techniques such as active listening and direct observation are commonly used by the evaluator in forming his/her impressions of the program goals. Once the evaluator feels s/he has grasped the essence of the program being evaluated, s/he shares these formed impressions with those responsible for the program content and goals. If the information provided by the evaluator is compatible with the intentions and the goals of the program, as seen by the program directors, then the program or training is said to be accomplishing the desired outcomes. This approach has been devised by Scriven, as he believes that this "goal free" approach makes the impartiality of the evaluator easier. Thus, this works towards reducing evaluator contamination through preconceived ideas and notions of what the program goals should be.

Other methodologists in the field of evaluation have taken issue with Scriven's approach and have expressed a wide range of opinions and criticism of his thinking. The majority of these people would strongly support the more accepted traditional approach of "goal based" evaluation. In "goal based" evaluation, the goals and objectives are clearly identified prior to program planning and training. The programming or training is designed to provide the opportunity for certain outcomes to occur. In evaluating the effectiveness or success of the program or training,

the evaluator measures the degree to which the participants have achieved the desired goals defined prior to the planned activities.

Weiss (1972) summarizes this approach by saying that the reason for identifying goals prior to evaluation is to enable the evaluator to compare "what is" with "what should be". Some of the many persons supporting the idea of "goal based" evaluation are Alkin (1969), Bloom (1969) Hutchinson (1971), Pace (1968), Popham (1974), Provus (1969), Stake (1970), Stufflebeam (1971), and Tyler (1942).

There exists the possibility of combining both "goal free" and "goal based" methods of evaluation to a situation desiring evaluation. If the investigator/evaluator had the resources of staff, time, sufficient financing and support, the result could be the combining of the best of two worlds. An abundance of rich data would then be available for comparison between the two methods. If the data were similar, the reliability of the resulting information could be assumed.

For the intents and purposes of evaluating programming and training, the bias leans towards a "goal based" evaluation approach as most often the sponsoring client or organization desires to work toward specifically defined and predetermined goals and objectives. A "goal based" approach seems to be more acceptable and logical in situations where the evaluator's role is one of aiding the client system in assessing needs and identifying goals in a spirit of coop-

perative venture.

In 1976, Welsh and Hambleton highlight another issue which has caused considerable division of opinion within the field, that is "Should an evaluative study be designed such as to detect unintended outcomes?" Fortune (1970) and Hutchinson (1971) strongly advocate that the evaluation study's main focus is only to assess the goals of the program and any diversion to detect other outcomes serves only to take valuable time away from the original focus of the evaluation. This position does not exclude the reporting of such unintended outcomes to the sponsors as they occur.

Tyler and Klein (1974) have feelings to the contrary. They feel that provisions should be made for the inclusion of search and measure procedures for such unintended outcomes in that new perceptions and insights may be discovered, thus providing innovative solutions to the problems currently being discussed. Bloom (1969) also supports this position as it is his feeling that one cannot foresee unintended effects; therefore, when they arise they must be both considered and effectively dealt with.

According to Pace (1968), the answer to this question depends entirely on the size and scope of the evaluation effort. He would support the position of Tyler and Klein, and Bloom if the evaluation was large scale where extensive manipulation of variables, extended use of control groups and sophisticated use of experimental designs occurred.

Under these conditions he would feel that pursuit of search and measure activities of unintended outcomes would not interfere with the original intent of measuring the pre-specified goals of the program and training. If this pursuit interfered with the original focus of the evaluation, then he would likely assume the philosophical position of Fortune/Hutchinson and assess only the goals of the program at hand.

Stake (1970) presses this controversy one step further. He not only calls for the inclusion and consideration of unintended outcomes in evaluation, but he argues for inclusion of judgment data. He wants to know not just what motivates people, but what motivates those persons who wish these same people to learn. What are their goals, values, priorities and standards?

"Which individuals should be responsible for stating goals?" This is the third issue that Welsh and Hambleton (1976) highlight as an unresolved problem by evaluators and methodologists. Concerning this question, there are again mixed opinions within the field. Stake and Denny (1969), and Provus (1969) state that it is the evaluator's role to help the program and training personnel state their goals. According to Stufflebeam (1971) this function should be jointly shared between the evaluator and the sponsor or client, with the evaluator noting and recording their basis for selection. At the other end of this continuum is Scriven (1973), who feels that it is the evaluator's role to speci-

fy all variables for evaluation independently of the project personnel.

"Should goals be flexible?" is also a question under consideration. General agreement seems to exist in the field on that issue. Evaluators generally agree that whether goals are flexible or not depends on the purpose of the evaluation. Once the purpose is determined, then the answer is evident. In formative evaluation, goals may be flexible. In goal based summative evaluation there are fixed and non-flexible. Of course, in goal free evaluation there are no problems with goals because they are not used. The next section will deal more specifically with the differences between formative evaluation and summative evaluation.

The most controversial and widely discussed subject in this area according to Welsh and Hambleton (1976) concerns whether goals should be stated in behavioral terms. Bloom (1969), Gagné (1972), and Popham (1972) are supportive of stating goals in behavioral terms. They feel that specifications can be judged for appropriateness for a given group of learners through behavioral objectives. They feel that behavioral objectives can facilitate improved communication between learners, participants, trainers, and sponsors which could result in better programming and training experiences and increased performances of participants. Finally, they believe that stating goals in behavioral objective terms provides a method for finding out if learners

have accomplished the objectives they set out to. Ebel (1973) though, stresses that it is more important to have or achieve agreement in general goals; then to expand the resources to state the general goals more specifically in behavioral terms. Some words of caution on using behavioral objectives have been raised by Stufflebeam (1971). He says that stating behavioral objectives may create tunnel vision for the evaluator. It may bias the evaluation such that only the predetermined behaviors may be observed with the exclusion of all other behaviors even when not appropriate for evaluation. Stufflebeam also feels that stating goals only in behavioral terms has resulted in only a "post facto" approach to evaluation which imposes definite limits (post facto research will be discussed later in the paper). Hogben (1973) agrees with Stufflebeam basically, but goes on to say that at best we are assessing with various degrees of accuracy, how well people can perform predetermined behaviors identified by the evaluator at the outset, as the specific behavior to be measured when using behavioral objectives. When attempting to state course goals in behavioral terms, for example, he adds that it is an enormously time consuming and difficult task for anyone to perform. Attempting to break down each general course goal into measurable behavior can only result in an overwhelming number of behavioral criteria. This also recognizes that it may not always be possible to transfer some goals into behavioral terms. These goals, although they may be of great value and importance, get

lost, omitted, or even classified as being unrealistic.

Hogben acknowledges that describing objectives behaviorally does have the advantage of providing clear end points towards which all persons involved can strive. Because behavioral objectives focus on expected terminal performances, they suggest methods of assessing to what extent the objectives have been realized. This is one reason which explains the popularity for the use of behavioral objectives in evaluation.

The evaluators who accept the behavioral objective approach to formulating goals and objectives usually support the concept of criterion referenced measures and criterion referenced tests in evaluating performance. Criterion referenced measures provide a basis for interpreting student performance relative to the curriculum. This is a constant to norm-referenced measures which facilitate the reliable ranking of one student with another on the ability measured by the test.

Some of those who support a criterion referenced approach rebut the idea that criterion referenced measures are well-suited for assessing skill development. They say criterion referenced measures occupy a greater time proportion in learning than previously thought, but they have been found to be more economical on some perimeters than previously believed. It is also acknowledged that they do not tell us everything, but that they do clearly indicate an

excellence or a deficiency in learner performances. For further reference and discussion in this area see Mager (1962).

Models for assessing needs and clarifying goals

When an investigator or evaluator attempts to assess the success and effectiveness of a specific training program, it is necessary to have a clear, precise and well-defined idea what the goals of the program are ("goal free" evaluation is the exception). Experience has demonstrated that the training programs that have met with the greatest success and satisfaction are ones which have developed as a result of the evaluator or the trainer first establishing the needs and clarifying the goals of all involved.

Typically, an evaluator is faced with one or more of the following situations when assessing needs or clarifying goals:

1. The participants, sponsors and/or organizers have only a general feeling of what the goals are (usually very ambiguous).
2. Those involved or responsible for the program are having difficulty in specifying their goals. The goals exist as "fuzzy concepts".
3. Those involved or responsible have no idea of what the goals are, but feel a change is necessary through one action or another.

The role of the evaluator is to aid these persons in obtaining a clearly defined concept of what their goals are. S/he can help them to develop specifications and measures for these same goals, enabling them in the final outcome

to have a standard by which they can determine the extent to which their goals have been accomplished.

The methods or processes that are presently being used in various fields for helping persons to assess their needs or clarify their goals are many. The method used is generally determined by the investigator's preference or by the description of the situation desirous of needs assessment or goal clarification. The following are names of authors who have presented models for need assessment and goal clarification that an evaluator may refer to when a situation arises where need or goal identification would be required: Dimock (1973), Fordyce and Well (1971), Fortune/Hutchinson (1973), McGill and Horton (1973), Popham (1972), Reed (1975), Schmuck and Miles (1971), Van Maanen (1973) and Weiss (1972). The following is a brief review of the strategies for need assessment and goal clarification of each author.

Dimock, in his monograph (1973), suggests five strategies or methods for determining needs for programming and training workshops. The first method he suggests, is to call upon program members, training participants or staff members to list the problems or need areas they would like to work on. If it is a large group, the evaluator can readily break it down into smaller groups of about sixteen persons each. Each groups' mission would be to combine their ideas listing them in priority. Each group contributes their individual priorities to all groups and selects indi-

vidual representatives to decide and vote on the final priorities. The priorities receiving the most votes become top priority and are turned over to a predetermined planning committee who takes the responsibility for operationalizing them.

The second method Dimock suggests takes the format of a paper and pencil questionnaire sent to members of a program or group who have requested some form of training. Each person is asked to describe their training needs and return the questionnaire. The responses are summarized by a predetermined individual or committee. As two or three areas begin to stand out clearly, they become the target areas for the first program. If none were to stand out, then the ten most frequent responses listed would comprise a second questionnaire which would be sent to the respondents of the first questionnaire. They are instructed to rank the areas listed, according to their interests. These would now be the target areas. Typical questions respondents would be asked to answer would be:

- a. Please list the areas that you think should be covered in a training program for the total staff/group.
- b. Which area interests you the most and therefore is the topic you would like to see dealt with first?
- c. Describe a problem which you are now facing which would be typical of the way this topic affects you?

A third method suggested by Dimock is a variation of

the paper and pencil survey questionnaire. The evaluator lists a variety of training workshop or program interests and asks the staff to check off areas that most interest them in the order of 1, 2, or 3. Number 1 would be their first choice, number 3 their third choice. Any interests omitted could be listed at the bottom of the interest inventory and included in the prioritizing exercise. The listing would include a wide range of training activities as well as including needs for technical skills, new knowledge and development of self-insight and sensitivity. This method has been found especially useful in large groups. This method can also be given to staff or participants after completion of their programs or training in order to evaluate the progress made and identify new areas of need priority.

Method number four Dimock mentions is especially useful when there is a new director of training starting out, or where an organization has had minimal exposure to training. This method consists of interviews and group discussions with a cross-section of potential recipients within a training group or an organization. This method would usually be carried out by the director and can provide valuable insight and understanding of training needs, not to mention the increased possibility of building increasingly good support with potential recipients.

When performing group interviews, Dimock suggests four

to six persons per group would be adequate. He also suggests that where a group or an organization has not had exposure to training, this approach would provide an ideal opportunity to test their readiness for the requested training. Ideal questions for obtaining data from this type of group would be:

- a. What percentage of the staff/participants here do you think are sincerely interested in participating in our inservice training program?
- b. What are some of the reasons for the readiness to participate.?
- c. What reasons exist which may prevent participation?
- d. What do you think would help set a good climate for training and reduce the resistance to participate?

Dimock also feels that this approach is more effective if those being interviewed know each other and occupy similar positions in the organization. He also feels that the information gleaned from this format may not yield as great a depth of information as other methods but it provides a much broader scope of data and helps to establish personal contact.

His fifth and final method for needs and goal assessment is less time-consuming, easier to operationalize, and less demanding and threatening to staff and participants. He simply suggests the use of staff meetings for discussion of staff needs, and requests members to submit their suggestions to a volunteer planning committee or a pre-appointed

program operations committee. The greatest disadvantage of this method is the possible apathy or reluctance of staff to put forward suggestions, the lack of clarity and specificity that may result from some suggestions, and the time pressures the organization feels to cover other agenda items of equal or greater importance.

Fordyce and Weil (1971) offer four models for determining needs and clarifying goals. The first model involves getting participants together in a general session to discuss generally the needs of the gathered group. Sub-groups are formed and the discussion is continued. A recorder in each group records the values and needs that should be included in the program. A spokesman for each sub-group reports the results at a general session. The organizers or chairpersons meet at a later date (as soon as possible), classify the information, consider implications and plan the next step. During this stage, the organizers or chairpersons are closely conferring with a third party consultant. A report is written by the organizers or chairpersons including all classified information and is sent to all participants encouraging responses. A "planning group" is formed from those participants previously attending the general session consisting of a cross-cultural or cross-sectional mixture of persons from the original group. This group develops long and short-term goals as well as general objectives for the tentative program. Through a process of

consensus, a steering committee is formed to set up and implement the program. The final stage is a pilot program offered by the steering committee (a similar model is also forwarded by Delbecq and Van De Ven, 1971).

The second model forwarded by Fordyce and Weil is likely familiar to many who are active in the behavioral science field. This is Lewin's "Forced Field Analysis Model" (Lewin, 1958). This model is usually used as a problem-solving and analysis model, but here it is readily adopted for identifying needs and goals. The "Forced Field Analysis Model" presupposes that there are forces both for and against any situation or condition. Those situations that are seen as being a problem will likely have an over-balance of negative forces or forces preventing the situation from changing. The need or goal would be, in this case, the problem identified by listing both the positive and negative forces blocking any possible change in the situation. To work towards solving the problem, goal or need, one would work towards reducing the negative forces preventing the necessary changes.

The third approach to need assessment presented by Fordyce and Weil is a more direct and shorter approach. In this approach a "boss" or "superior" identifies the goals to be met. He thoroughly discusses the goals with his superiors, peers and subordinates. He ensures that all people involved and approached understand the conditions to be met for goal attainment. Furthermore, he ensures that these

conditions for attainment are subscribed to by all. The final outcome of success is dependent on many variables such as "personal power" (friendly persuasiveness) of the "boss", "position power", (organizational position) of the "boss", organizational climate and effective communication.

The last model proposed by Fordyce and Weil is a model sometimes referred to as a "Career planning model" or a "Life planning model". This model consists of a series of steps as follows:

1. Identifying the problem, the need or the goal.
2. Drawing a horizontal straight line on a piece of paper to represent the problem, need or goal of the person involved.
3. The polar ends of this line represent the past and the future position of the problem, need or goal of the person involved and is usually represented by a symbol such as an "X". Another symbol (most often the letter "Y") represents the person's position in time or the degree of information s/he has on the need, goal or problem indicated.
4. Under "X" "past", all the things are listed that are known to have contributed to the problem or need and under the "Y" all the things are listed which are seen as presently contributing to the problem or need situation at hand (e.g. competence, degree of risk-taking, co-operation).
5. Finally, under "X" "future" all the things are listed which are necessary for fulfillment or satisfaction of the needs or goals identified.

Fortune/Hutchinson (1973) have specifically designed a very extensive, elaborate and multi-step model for assessing needs and identifying goals. Reed (1975) has neatly summarized this elaborate model into a workable and applicable form. Basically, Reed takes the persons meeting in

a general session and divides them into small groups of five to eight persons. He then has them fantasize what the ideal situation would be for them to satisfy their needs. Each member records their individual fantasies as they come to mind. Participants are asked to list what they consider to be the least ideal situation for their need satisfaction. These are also recorded. The least ideal situations are converted to ideal situations and added to the ideal situation list. The members of each group share their lists with other members noting any ideas for addition to their own lists. Each member then prioritizes their lists and the group forms a new list consisting of the top three priorities of each group member. Each group presents its list to the general session of all member groups. Voting and consensus taking occurs after the top priorities and their respective specifications have been reviewed for each of the target areas suggested by the total group.

McGill and Horton (1973) offer three approaches that are very similar to Dimock's models. McGill and Horton suggest the three methods for needs assessment would be the survey questionnaire, the interview and observation. They suggest two questions that would begin to provide a focus for goal clarification and needs assessment. The first question would be: "In relation to goal "x" what would you stop doing today that would contribute to its achievement?" The second is as follows: "What would you start

doing today that would contribute to its achievement?"

Popham (1972) describes a further method of needs assessment. This method has been used to identify educational deficiencies through using measurable objectives. These measurable objectives are selected from "objective depositions" or an "objective bank" previously organized. Those involved in the assessment process rank the objectives from the bank according to their suitability for the situation or program at hand.

An approach with a different variation is presented by Schmuck and Miles (1971). They look at goal setting and clarification in terms of long-term and short-term goals, and suggest the "Delphi method". The "Delphi method" is a way of identifying and predicting long-term goals. Participants involved in the planning of a long-term project or program are asked to conjecture some events that might take place in the next fifteen years which would effect their project or program. Then, they are asked to conjecture things that would effect other things associated with their program or project. What impact or effects would they have? The participants are asked next to assign dates to their conjectures. The information is collected, processed and fed back to the participants. The participants are asked to examine the range of dates provided and assign a value to each event that may occur.

Finally, the participants are asked to write a short

description about the possible consequences of the events in question. The short-term goal setting method suggested by Shmuck and Miles is one that is found to be frequently used in the process of "Management by Objectives". Here a superior and subordinate work together to identify the goals to be accomplished over the next short period of time.

Once the goals have been identified and specifications formulated indicating their degree of accomplishment, and they are mutually agreed upon, the subordinate then works towards accomplishing these goals within the mutually agreed specifications. Periodic sessions for appraisal, review and realignment of the goals and their specifications occur during the pursuit of accomplishing the goals agreed upon.

Van Mannen's (1973) model for need and goal assessment is initially consultant or evaluator-oriented. He says the first step is to read all relevant material available in the beginning to become aware of clues that may give some indication of the unspoken goals or hidden agendas people are working towards. Next, he proposes that the evaluator talk with all those people involved with the situation either formally, informally, individually or by small groups. Then, he proposes the evaluator shall observe the program at length. Once these three actions have been completed, the evaluator specifies the goals of the program as he sees them. The next step is the most crucial. The evaluator must have those people involved accept the ownership and the responsi-

bility of working towards the goals he has specified. This is accomplished by feeding back to the participants the goals which the evaluator has specified from his readings of documents, interviews and observations. The participants are encouraged to modify, expand, eliminate, add, or change any of the goals specified by the evaluator until they are satisfied and consensus is established for the goals. The evaluator may want to assist this process by encouraging the participants to maintain a collective and cooperative procedure.

An alternative to the above procedure would be to have the evaluator ignore the question of goal specification and encourage an open-ended evaluation. This procedure may be adopted when the evaluator realizes that in some cases, goals set very early in a complex, uncharted direction, may be premature and of poor quality.

Weiss (1972) presents a model that closely parallels that of Van Mannen, but adds that goals that are given priority should be given priority on the basis of the following criteria:

1. Potential for use
2. Relevance to decisions which need to be made
3. Time, money, access and commitment
4. Greatest payoffs, most pressing need or problem and urgency

Other variations and designs of needs assessment models for evaluator use in workshops can be found by Beckhard (1969),

Delbecqu and Van De Ven (1971), Gordon (1973), and Kindall and Gatza (1963).

In closing this section on needs and goal assessment, it would be appropriate to mention a word of caution concerning the many designs and variety of models available to the evaluator. Stake (1972) comments that no statement of program or training objectives ever has come close to representing the real world intents of the people involved. The unspoken objectives are usually left to take care of themselves, at least until a crisis arises; then, these objectives may pre-empt all others. He states further, that consensus is one of the great simplifiers, but it may mislead the investigator by indicating that problems, needs and goals are much less important to some people than they really are.

An area of concern that has not been mentioned in this discussion is the shared opinion that all behaviors or needs (specifically affective needs) cannot always be presented or stated in measurable terms. Fortune and Hutchinson (1971) refer to those behaviors and needs that are difficult to place in measurable form as "fuzzy concepts".

Benedict (1971) presents a model discussed in the Fortune/Hutchinson Evaluation Methodology which assists putting some of these "fuzzy concepts" into measurable terms. Briefly, this procedure identifies typical or representative behavior and objectives in the affective domain through the

use of fantasy. Once these numerous behaviors have been identified, friends and colleagues are consulted to add to the list and are asked to refine it using a similar process of fantasy. The final product is a list of affective behaviors that has been transformed into specific behavioral actions. These specific behaviors are now in measurable form and are ready for inclusion within a questionnaire (Instruction for specific application of this method can be found on page 46 of the "Application" section of this chapter).

2.2 Application

Approaching evaluation

Schmuck and Miles (1971) describe a typical sequence of events that lead to an organizational development intervention. On close observation, the same sequence may be adapted and applied in describing the series of events occurring when the decision to evaluate training or programming has taken place. The sequence of steps are as follows:

1. Some authority (an organization, an administrator or a training director, for example) becomes interested in evaluation. S/he feels that there would be a value or a need to evaluate a particular training program. This initial interest may have been stimulated by the person's exposure to the value and benefits of evaluation.

2. The authority decides to investigate the possibilities of an evaluation by either inviting an evaluator in

for further discussion or by consulting himself/herself with informed sources.

3. After the decision as to who will perform the desired evaluation, the potential evaluator/investigator reviews the program areas that would have the potential to be evaluated. When these areas have been identified, a proposal or contract should be constructed (as a record or intent, or as a binding agreement with the sponsoring body) specifying the nature of the evaluation, how it is to be performed, its goals, and the role of the evaluator.

Organizing for evaluation

According to Van Maanen (1973) a sequenced development of a program evaluation model such as suggested by Schmuck and Miles provides the investigator/evaluator with an idea of how things will fit together. It serves to sensitize the evaluator to various possibilities of evaluation strategy and specific consequences that may result from each particular approach. One such model may be suggested in the following:

1. Determine what is to be accomplished at the time of termination of the training program to be evaluated.
2. Determine what shall be measured in each training or program area and what the yardstick of measurement should be.
3. Contract for these (with sponsoring organization, with participants, with others involved).

4. Set check points and points for revision of goal achievement.

5. Specify and document proposed action performance or behavior expectations.

6. Provide for review times and for progress reports at specified, periodic time sequences.

The program evaluation model given as an example informs the investigator where to look for the information and possibly what to measure. At this stage, an evaluator should be interested in how the goals are defined as it is Van Maanen's opinion that a close relationship exists between the goal definition and the resulting selection of appropriate measures of evaluation. Some principles of effective goal clarification and understanding which could be applied in order to ensure that existing goals are well defined would be:

1. Ensure that all goal statements (or statement of needs) are in terminology that all directly involved, understand.

2. Ensure that all necessary information has been included in the context of the statement.

3. Ensure that there are clear definitions and common meanings of all terms and statements used.

4. All statements and terminology used should be unbiased in its intent and attitudes.

5. All goals and needs identified should be directly related to the broader purposes of the organization and persons involved.

6. Restate the goal to those persons defining it, observing the effect and response. If questions, misinterpretations or misstatements have resulted, proceed to redefine or reclarify.

7. Check for missing and relevant information.

Goal clarification and need assessment

The principles and guidelines just described are undoubtedly helpful to the investigator when the sponsoring group or those desiring evaluation are organized to the degree that they can identify what it is they wish to achieve or attain. Fortunate is the investigator/evaluator who has this situation. It is more usual that s/he is called upon to perform an evaluation when the situation or planning processes are in one of the four following common situations:

1. The planning group or body have some generally agreed upon concepts of the things they want done or the training they wish to provide, but there is no specific agreement or specification of the goals and outcomes desired as a result of training or programs needed.

2. The planning body shares an unspecified concern and is in common agreement that something should be done, but is having difficulty in getting started and being able to focus

on what it is to be done. There may be many good but diversified ideas as to what can remedy the situation.

3. The planning group has worked hard at identifying various goals and behaviors desired as outcomes, but they are having difficulty in specifying some in measurable terms; for example, increasing the degree of self-awareness of participants and raising the trust level of participants towards each other or toward subordinates.

4. The planning body is confronted with providing a long-range training effort (about five years in duration) which on termination they expect specific behaviors or outcomes to result. The problem is that they are having difficulty in conceptualizing what problems may occur or what needs may arise over that period of time.

In assisting a group in clarifying their goals or assessing their needs, it is of prime importance that the resulting goals are clearly defined and unmistakably understood by all involved, both in the planning process and in the participation action phase. The following guidelines (Van Maanen, 1973) should be followed as closely as possible.

1. All goals, objectives or needs should be written and recorded in clear terms which effectively communicate the program or training intent by excluding other possible meanings. A recognized and accepted method is to write the goals or objectives in a specific format such as, "to write", "to list", or "to solve". Using ambiguous verbs such as "to

know" or "to understand" should be discouraged as they do not specify a measurable behavior or action and would not adequately satisfy the definitional requirements in writing good objectives. (Mager, 1962).

2. The second guideline to adhere to is to include all important conditions in which the behavior will be expected to occur.

3. Finally, it is advised that the objectives should be written such that a criterion of success can be readily applied. This would involve a description of how well the individual should perform. A base for comparison should be available for use.

Specific models for clarification and assessment

The first model to be outlined in this methodology for clarifying goals and assessing needs is advocated by Burke (1972) and Hambleton (1976). It is best used under the conditions outlined in situation No. 1 "when the planning body have some generally agreed upon concept of the things they want done, or the training they wish to provide, but there exists no agreement or specification of the goals or outcomes desired as a result of the training or programming".

In this situation, the novice investigator/evaluator should attempt to encourage a general discussion by the planning body of what they perceive as the goals of the training or the needs to be satisfied by the program. A sample question would be "What do you see as the goal(s) of

the training?" or "What needs do you see this program satisfying?" Once this discussion is underway, the evaluator or investigator's role should be one of listening and recording the various opinions and answers to the questions presented. When the discussion has terminated (30-40 minutes approximately) and the evaluator feels that s/he has adequate information to work with, s/he formulates a series of goals and objectives from what has been recorded from the general discussion. Then, the evaluator presents to the group these goals or objectives as s/he has interpreted them, allowing the planning group to accept, reject, modify, expand and add wherever they feel necessary. This process allows the planning group to take "ownership" of these goals from the evaluator's interpretation. The evaluator then checks once more for final adjustments to the list. If no adjustments are made, the evaluator presents this list as the generally agreed upon list of goals and objectives for the specified program or training desired by the planning group. On acceptance of this list, the evaluator may want to go through a similar process with the planning group in designing specifications or measures for each goal or objective. A second alternative would be for the evaluator to draw these up himself and present them as the final specifications. The choice depends on the time available, the evaluator's contracted design or the preference of all involved.

The second model has been adapted from Fortune/Hutchinson

(1973) by Reed (1975), and is suggested for use under situation No. 2 where "the planning body has an unspecified concern, but is in common agreement that something has to be done and is having difficulty in starting, or deciding what to do".

Reed outlines the steps as follows:

1. At a predetermined general meeting of the planning body and all those persons directly involved, divide those attending the session into groupings of five to eight people. (If only a small planning group is involved, adapt the model and proceed with step No. 2).
2. Ask each participant to take a sheet of blank paper and imagine what the ideal seminar, program or training situation would be for them. Then, without evaluating their thoughts to the question, list them on the paper as they come to mind.
3. Next, have them list on the reverse of the paper what they could imagine or picture as being the "least ideal" seminar, program or training workshop session.
4. Have them take every "least ideal" item and convert it into a positive, "ideal" situation. Have them add these items to their first list of "ideal" situations.
5. Each of the members of the group in turn reads his/her list of positive "ideal" situations to other group members. As each member reads his/her list, other members jot down items on their own lists which were inspired by the

reading. All participants are instructed to listen carefully while the lists are being read and the readers are instructed to read their lists slowly.

6. At the conclusion of the reading of the lists, each member is instructed to consider any dimensions that come to mind which are not directly related to the training or need situation being planned such as a) cultural, economic, political and social forces, b) short and long-range career plans, c) nature of present or new future organizational situation.

7. Take all items within each personal list and prioritize and rank each (for example, No. 1 first priority, No. 2 second priority). Because all items listed cannot be worked upon due to time and resource limitations, the participants are instructed to select the first three to five priorities on their lists. They are told to share these with other members of their group. The group is then instructed to form a prioritized group list of five items.

Three questions which will help participants carry out this step at this time would be: a) What is most important to you? b) What is most possible in this setting? c) What is most pressing at this time?

8. Each group (if more than one group exists) collects and compiles its lists with all other group members totally prioritizing all their contributed lists together into one final list for the whole group. If only one group exists, then each member's five top priorities are considered.

9. A limited number of priority need items are agreed upon by the participants from the composed lists. Everyone is asked to come to agreement on which needs or goals will be worked upon and when. Action plans are then proposed by the participants and the evaluator.

10. If selected priority needs still appear to be of a general nature, the same process which has just been described may be followed again, but using just the items which seem to be still general in nature. The result should be a list of specific items that can be described as goals or objectives satisfying a specific need or needs.

The third model to be presented in this phase of the evaluation methodology is suggested by Benedict (1971). Its use is intended to assist the planning body or user to place what is seen as a "fuzzy concept", or a goal which is not in measurable form, into a measurable and a definable quantity. This technique has been found considerably useful in developing measurable criteria for behaviors that are in the affective mode such as measuring the amount of love, empathy or degree of rejection, for example. The steps are presented as follows:

1. Fantasize the "ideal" (one hundred percent situation) for fuzzy concept or affective area goal that is to be measured.
2. List and record all that is seen in the fantasy.
3. Fantasize the "least ideal" situation (fuzzy concept).

4. List and record all that is seen in the "least ideal" fantasy.

5. Test for completeness of list by consulting four or five others, requesting they perform the same process for the same situation. Add their fantasy observations to the listing that is being compiled.

6. Take the two lists, the "ideal" and "least ideal" and include the new data and information received by consulting others to the list being compiled.

7. If there are items that were previously overlooked or not recalled include them in the list at this time.

8. Fantasize and record those items which may not appear to have a direct influence on the "fuzzy concept". Once listed, check to see if they still have no relationship or stimulus to the "ideal" or "least ideal" lists.

9. Take each concept on the "ideal" and "least ideal" lists and ask: "Can this dimension be observed directly?" If it still cannot be directly observed (with behavior specifications), then repeat the "fuzzy concept" process for each of those dimensions. (If you can, good!)

10. Develop a priority list of those items that are felt to be most important with the greatest need for measurability.

11. Using these items and their respective behavioral actions seen in the fantasy, develop measures indicating the degree of existence of the action representing the be-

havior in a reality situation.

12. An optional step for the enthusiastic novice evaluator would be to develop a "Likert Scale" to measure the extent that the described item exists in the reality situation being assessed (for a modified Likert Scale, see appendix G).

The fourth and final model presented is an adaptation of the "Delphi Method" of predicting and focussing on long-term goal outcomes in programming and training. Schmuck and Miles' (1971) presentation of this method provides the vehicle for assisting planning bodies which are having difficulties similar to that has been described in situation No. 4 "the planning body is confronted with providing long-range training or programming where specific behavior outcomes are expected at the termination of the training workshop or programming". As the reader will recall, the problem is that the planning body is experiencing difficulty in conceptualizing what problems or needs may occur in the projected time perspective of five years hence. The general procedures of the method for this model are presented as follows:

1. The planning body (additional persons familiar with the situation would be an asset when added to the planning body for this exercise) is asked to conjecture some prospective events which may take place during the next five years that could be perceived as possibly having some influence or impact on what it is that is being attempted to be achieved

over that period of time.

2. In the conjecturing, the planning body (plus others?) are also asked to think about things that may happen which would affect other needs such as continued availability of training budgets, manpower availability for training, urgent reassessments of system wide needs, changes or new fads in training, or claims of cheaper and more highly productive methods of training and so on.

3. The planning body is requested to place specific dates on these conjectures which are then processed either manually or by computer (depending on the number of respondents and the number of conjectures) and are fed back.

4. The planning body is then asked to examine the range of dates fed back, and to assign a value estimate to the events conjectured.

5. The planning body is then asked to write a very short description about the possible consequences of the events in question.

This strategy helps the planning group to set the goals, to assess strategy options for their goals against the beliefs, attitudes and consequences felt and seen by various other planning groups, and to help them determine the value of achieving their individually identified goals.

The four evaluation models suggested by the author are just that -- suggested. The interested reader or informed novice may know of, or wish to use, other models more appro-

priate to his/her individual situations. Such models may be directly taken from sources identified in Chapters II, or created or adapted from those sources and their references. Presently, there exists abundant literature in the area of goal clarification and needs assessment awaiting those interested in its use.

Once the appropriate model for clarification and assessment is chosen and applied to the situation at hand, the wise investigator would use the following list of questions suggested as checkpoints to clarifications by McGill and Horton, Jr. (1973) and Scriven (1974):

1. Have all involved group members participated in the goals selection?
2. Do all members see goals important to themselves and significant to the sponsoring organization?
3. Is there agreement on the goals and are people committed to them?
4. Are the goals consistent with the sponsoring and parent organization (if it exists)?
6. Are the identified goals reasonable?
7. Are they feasible, given the existing time and resource constraints? Are there ways of finding out?
8. Are there provisions for orienting new members to these goals?
9. Is there provision for modifying or changing goals as the result of experience?
10. Are the alleged goals the real goals?

11. Are the goals stated, in conflict? Which ones should have priority, and who decides?
12. What anticipated side effects are there? Speculate.
13. What may be the unanticipated effect? (Goals are a subset of anticipated effects.)
14. Are goals to be evaluated in testable terms?

Assessing needs and determining goals and objectives are a necessary part of every training workshop. If the evaluator is engaged to evaluate a workshop after much of the initial planning has taken place, it is even then a worthwhile endeavor for him/her to review briefly the clarity and the specificity of the goals and objectives desired as a result of the training. This action serves to familiarize the newly-arrived evaluator with the interests and desired outcomes of the sponsor and allows the evaluator the opportunity to suggest any modifications or adjustments s/he may feel necessary in attaining the desired results.

C H A P T E R I I I

3.0 CHOOSING EVALUATION MODELS

3.1 Theory

Introduction

An evaluation model is a formal and logical framework for evaluation which represents the patterned thinking or considerations of an evaluator when s/he proceeds to carry on an evaluation. As the choosing of the evaluation model occurs early in the methodology, it cannot be over-emphasized that the selection of the appropriate model must be done with care and consideration of all the potentially confounding variables and environmental factors involved. When considering evaluation models for selection, all possible known advantages and disadvantages of each model must be weighed in respect to its over-all effect and compatibility with the evaluation project at hand. This action will move to ensure that the evaluation model finally selected will assist the evaluator in obtaining the information desired within the conditions and perimeters outlined by the sponsoring persons. Examples of some of these conditions or perimeters are time, budget, availability and internal resources.

Goal Free and Goal Based Evaluation

Goal based evaluation and goal free evaluation has

been mentioned previously within the context of assessing needs and clarifying goals. As the reader will recall, a goal free evaluation model is where an evaluator has no prior concept of what the goals or intents of a program are. The evaluator attempts to determine these goals by contact with the participants when the program is in full operation. When the evaluator has identified the goals, he gives the impressions to the sponsors, informing them of his opinion as to whether the goals are being successfully met or achieved. In goal based evaluation, the goals are predetermined and clarified. The program is usually aimed towards meeting these goals and their specifications. The current effects or final outcomes are then compared to the original intents. Sinclair (1975) describes it as a process of comparing what is currently observed with what is currently desired.

Summative evaluation

Evaluation models can be perceived as having another set of special characteristics. Most models of evaluation can be classified as being summative or formative in nature. Until recently, most evaluation models or procedures applied to training programs were mostly summative in format. A typical example of summative evaluation reporting would be the report described by Benne, Bradford and Lippitt (1964) where they had participants of a T-Group account on their degree of pleasure or dissatisfaction with the training they had just completed. Thus, in this way, summative evaluation

can be a helpful tool for the training practitioner and others in determining the degree of satisfaction of the participants towards their training program as well as having further training workshop areas identified for future improvement.

The main disadvantage of summative evaluation is that little can be done to improve the specific training that has just been completed. Furthermore, it is difficult to specify the particular areas needing improvement as experience has demonstrated participants' responses are usually highly biased in favor of the training program and are often emotionally charged and affective in content. Another disadvantage of summative evaluation is that it is costly because of the existing need to have large samples and highly sophisticated research designs.

It has also been observed that with this form of evaluation, the nature of the evaluation design, (especially if the training experience has been positive) encourages only positive feedback from the participants. Negative feedback or criticism often takes the form of "no response" or "unanswered questions" on evaluation forms. Weiss (1972) comments that participant opinions regarding a program or training workshop are helpful measures, but many people like or dislike a program for reasons unconnected with its goals. Thus, the original expectations or goals of the participants of a program may be misguided or be hazy and ambiguous. For further discussion on designing summative evaluation models,

see Airasian's article in Popham (1974), and see Scriven (1967).

Formative evaluation

The main purposes of formative evaluation according to Baker (1973) is to evaluate:

1. The program effects or goals.
2. What the program is doing and how the participants are performing.
3. How the program is operating, the adequacy of presentation, sequence of training, format and indicate the areas of inadequacy and where revisions are necessary.
4. The manner in which data is collected such that it does not unduly effect the program, the costs, the time and the emergency.

Therefore, the advantage of formative evaluation is that not only does the evaluator have a method of receiving feedback on the strengths and weaknesses of the program or training design, but s/he can forward specific data to the training practitioners or co-ordinators enabling adjustments and corrections to be made to strengthen and improve the ongoing training program. A second advantage of formative evaluation is that it is possible to obtain information on how well the participants are performing. This information can be shared with the participants providing feedback for them on areas within their training which may require more emphasis and focus. This gives the participants a measure of how well they are doing. Periodic measure-

ment during training also helps to motivate the participants and adds incentive, thus increasing learning (Verner and Booth, 1964). Finklestein (1971) adds that evaluation provides direction as well as the rate of progress of both the participant and the program. Which method of evaluation the evaluator selects is determined by many factors, some of which are: the length of time provided for the evaluation by the sponsor, the budget, the urgency of obtaining the information, and the point in the program or training which the evaluation is required. Other factors influencing also are: the point at which the evaluator is requested to enter the program to start the evaluation process as well as the purpose of the evaluation, need of the data required, and the depth of information demanded or required.

Stake (1972) emphasizes that priority should be placed on determining the primary intent of an evaluation. This he suggests will provide the evaluator or investigator with the guidelines for selection of the appropriate design/model to be used in the evaluation.

Models for evaluation

The following section will review briefly six models of evaluation that are popularly used. A summarized chart of these models and others which time and space prevent reviewing can be found in Appendix A. The chart is presented in an informative, comparative and convenient format.

Alkin's (1969) model for evaluation is primarily a

system assessment model which is designed to provide specific information on five areas of program planning. The stages of his model are:

1. Systems assessment: determines the range and specificity of goals and objectives appropriate for the particular situation to be assessed.
2. Program planning: provides data and information for decision-makers. Which alternatives will best meet previously identified needs. This may take the form of internal evaluation where programs are examined to determine their unproductive segments and to what extent are these programs achieving the desired objectives. It may also involve external evaluation of programs to be implemented where research data on similar programs are used as predictors of present program outcomes.
3. Program implementation: is the program being introduced to the participants in the manner in which it was intended and is it reaching those persons for whom it was intended?
4. Program improvement: looks at how the program is generally functioning, how are the objectives being achieved and what unanticipated outcomes are being produced?
5. Program certification: this area requires the making of judgements about the program's worth and its potential application to other situations; as a result, additions, deletions, modifications and corrections are common activities.

For a more detailed, comprehensive review and explanation of Alkin's model see Alkin (1969), Weiss (1972), and Worthen and Sanders (1973).

Campbell, Dunnete, Lawler and Weick (1970, p. 7) provide a systems-training approach model for evaluation. Basically, they outline their approach in seven steps:

1. Determine training needs

2. Choose theoretical approach(es) appropriate for the type of skills, knowledge, activities or behavior which must be taught.
3. Clearly define the objectives of the training effort.
4. Design and develop the training program to meet the objectives.
5. Identify the individuals to be trained.
6. Conduct the training experience.
7. Determine whether the training has met the objectives by comparing the outcomes with the objectives intended.

The Fortune/Hutchinson Evaluation Methodology (1973) is primarily a model which provides the decision-maker with adequate information for decision-making. The following is a brief summary of a very long and elaborate process.

1. Negotiation of contract: the evaluator provides a detailed outline of the evaluation methodology to the sponsors or organizers. It is then determined if the methodology will meet the needs of those desiring the evaluation. Contact persons and key program decision-makers are identified at this phase.
2. Identification of enterprise: the program, the training or the enterprise to be evaluated is identified.
3. Elimination of misunderstanding: in this stage, rapport and mutual understanding between the evaluator and responsible directors is developed. The examination of the evaluation plan is carried on to identify errors in the design.
4. Identification of resources: here the available resources are identified and made available for use in the evaluation by those sponsoring the evaluation procedure.
5. Identification of all decision-makers: the orga-

nizational power structure and their power priority are identified as the evaluator must know who he must work with -- when, and for what purpose.

6. Contract preparation: the evaluator and the sponsor develop and agree on a written contract as a record of agreements and as a clear commitment for reference and referral.
7. Needs assessment: use Fortune/Hutchinson model for needs assessment to determine and clarify the goals and intents of the program or training to be evaluated.
8. Matching goals: the goals are matched to specific training, purposes and activities by the evaluators.
9. Operationalization: each goal is broken down into directly measurable and observational components. These same components are tested for completeness and then prioritized.
10. Development of observable techniques: instrumentation is developed such that they are used directly, unobstrusively and under natural conditions. If instruments are not available they are developed.
11. Implementing measurement: the instruments that have been selected or developed for data recording are applied to the situation to be evaluated.
12. Data analysis: information or data gathered is now organized, categorized and analyzed in preparation for feeding back to sponsors or participants.
13. Data reporting: data is fed back or reported to sponsors or participants indicating techniques used, operationalized activities and components used to measure each goal and program past evaluated.
14. Evaluating the design: this is a process of evaluating the evaluation design looking in particular to what extent decisions were made from the data provided. Were adequate amounts of data provided and in time for sponsor or participant need fulfillment? Finally, are there pressing needs

still left unmet.

15. Redesign: this is a process of checking back with the appropriate decision-makers or power structure and mutually deciding to make specified program or training adjustments and refinements. Then the testing of the redesigned parts takes place with decisions following to adapt the adjustments or again redesign as appropriate.

The next design presented is one of the more popular evaluation models known in the field of evaluation. Schmuck and Runkel, et. al. (1972), describe Stufflebeam's C.I.P.P. (Context, input, process and produce) evaluation model for clarifying decisions for program improvement. As the model's name suggests, there are four different approaches to evaluation at various periods or stages of time. The approaches are:

1. Context evaluation: this evaluation occurs prior to an intervention or evaluation attempt. This consists of the evaluator performing an assessment of any conditions existing in the system or organization which could have effects on the potential design of the intervention and its outcome. Context evaluation is divided into two modes, the contingency mode and the congruence mode.
 - a. Contingency evaluation: evaluates the opportunities and pressures outside the immediate systems to promote improvement within the system. It also probes the future for societal needs, values, and trends that may influence the system.
 - b. Congruence evaluation: compares the actual and intended system performances. Its function is to monitor the system to determine whether or not the goals are being achieved as intended and to provide an ideal model for the system to follow.
2. Input evaluation: the purpose of input evaluation

is to provide information for determining how to utilize the available program resources to meet the intended program goals. This is accomplished by identifying and assessing the relevant capabilities of the client, or sponsoring body or agency, identifying the strategies for achieving program goals, identifying and creating designs for implementing the chosen strategies.

3. Process evaluation: is used to predict or detect defects in the procedural design or its implementation. It provides information for programmed decisions and maintains a record of the procedure as it occurs. Process evaluation assesses the short-term effects of an intervention.
4. Product evaluation: the function of product evaluation is to measure and interpret attainments not only at the end of the project, program or training cycle, but to measure the attainments as often as necessary during the project, program or training term. This would include devising operational definitions of objectives, measuring criteria associated with objectives of the activity and comparing these measurements with predetermined absolute or relative standards.

An extended and more comprehensive description of the C.I.P.P. Model may be found in Popham (1974) and Schmuck and Runkel et al. (1972).

Suchman (1970) in Weiss (1972), states that evaluation is decision-making, programming, communicating, controlling, reappraising all of which operates in a continuous cycling process. Thus, his evaluation process model is a consecutive series of demonstration programs set up as a one-shot effort, each time attempting to gain greater control over the stimulus and its administration with the intention of improving the program each time it is attempted. The first phase of evaluation would be the "Pilot program". The de-

sired program or training would be presented on a trial and error basis, hoping for a "learn from experience" result. There is a need for careful selection of objectives, exploration of strategic factors and maintaining enough flexibility in the program or training to keep the demonstration useful and ongoing. The evaluation procedure would be quick and easy with the emphasis on providing feedback as to whether the desired outcomes have been achieved.

The second form of Suchman's evaluation model would be the "model program". This would be a result of a series of pilot programs based on what has been learned through piloting. A "Model program" would theoretically have a greater chance of success than a "pilot program"; therefore, the focus would be on demonstrating the program's or training's success through designing an experimental version while attempting to prove the program's worth and testing the sponsors's hypotheses. For example, that A, B, and C can achieve X, Y, and Z while ensuring the A, B, and C have been put into effect under desirable conditions. This naturally would take the form of a well controlled experiment, which would be well designed and highly structured in program input. The criteria for program training effectiveness would be clearly defined, and the instruments would be constructed to have a high degree of validity and reliability. The limitations would be that the findings could not be generalized to other standard operating pro-

grams.

The final demonstration program discussed by Suchman is the "Prototype model". This model is the result of the previous two models, the "Pilot program" and the "Model program". The findings of these two models will guide the prototype model into determining what can be practically and realistically done in terms of large scale effort with the available resources.

Tyler's (1942, 1958) model for evaluation is probably one of the earliest and most well-known and utilized. His model for evaluation can be summarized in seven steps:

1. Establish broad goals and objectives.
2. Classify the objectives.
3. Define objectives in behavioral terms.
4. Create or locate situations in which achievement of the defined objectives can be shown.
5. Develop and select measurement techniques such as found in Buros (1972).
6. Collect student, learner or participant performance data.

"Appendix A" will present a description of various other evaluation approaches in a summarized chart form for easy reference and referral. The models compared have been presented by such well-known evaluation methodologists as Alkin, Hammond, Provus, Scriven, Stufflebeam and Tyler.

A well selected evaluation model provides a good itinerary for planning the evaluation process in workshops.

The evaluator who has left this phase of the methodology to chance or circumstance has sentenced the evaluation to failure or invalidity. The evaluator who takes time and care in selecting the appropriate model for use, greatly increased his/her chances for success by providing the evaluation with focus, increased clarity and a specific strategic approach to problem solving and decision-making.

3.2 Application

Models for evaluation

The selection of any evaluation model is primarily dependent on what it is the program or training is trying to do. Usually the first question a training director or program director asks is "Is the program or training working and producing the desired effects?" The second question s/he asks is "If it is not, why not?"

In attempting to measure outcomes of programming and training workshops, the choice by the evaluator for the type of evaluation is limited to goal based. Furthermore, experience has shown that sponsoring organizations and participants attending programming or training, first want to know, "Is what is being done working?" -- which is a summative approach to evaluation. A more formative approach arises when the authorities who are financing or sponsoring the program or training, question, "why is it, or why is it not working?"

Briefly, then, it seems that the first line of concern in evaluating programming and training workshops is goal based and summative. This appears to be the thinking of Campbell, Dunnette, Lawler and Weick (1970) when they propose a simplistic evaluation model which they entitle a "Systems Training Approach". It is worthy of the novice evaluator's consideration and is outlined by the following steps:

1. A designated authority or body determines the training needs.
2. A theoretical approach (or a number of approaches) is chosen appropriate to the skills or needs identified to be taught.
3. Specific objectives of the training effort are clearly defined.
4. A training program is developed to meet these objectives.
5. The individuals to be trained are identified and selected.
6. The training program is conducted.
7. An evaluation is performed to determine whether the training program met with the desired objectives.

The advantage of this "Systems Training Approach" model by Campbell et al. (1970) is that it leaves it up to the discretion of the evaluator and other authorities as to how each step is performed and who is to be involved in what decision making process and at what stage of design.

Suchman's (1970) "Demonstration program models" would also be helpful in evaluating program and training work-

shop outcomes. In review, the levels of program development are:

1. The "Pilot Program"
2. The "Model Program"
3. The "Prototype Program"

In reviewing these program models, it is felt that the most useful for consideration within this methodology.

Program Models for Evaluation

Pilot Program

Experience demonstrates that many evaluation attempts are of a "trial and error" nature, carried on by inexperienced evaluators. There are specific outcomes desired as a result of a program or training session. If one technique appears to work (or work better than another); then, this method is adopted until a better one is found. The "Pilot program" model utilizes much of this concept, but adds a constant feedback and revision component to the experimentation. In other words, feedback and revision are structured by building them into the model so that are not left to haphazardly occur due to chance. The second important component that is built into this model's design is flexibility. Model flexibility is maintained by providing adequate opportunity for testing various factors and their effects on achieving the desired outcomes through a "learn

from experience" basis. The primary focus of this approach is not one of success or failure, but one of being able to learn enough to be able to eventually develop a program that can be evaluated in a more consistent manner. In using this approach, the novice should be made aware that the value and effectiveness of the "Pilot program" rests on careful selection and clearly defined objectives as well as careful consideration of all strategic factors involved since a true experimental design is not used to assist in doing this.

Model program

The nature of the approach of the "Pilot program" makes it necessary for its repeated application before any specific conclusions or suspicions can be assumed. The "Model program" is a direct outcome of repeated testing and the drawing of certain assumptions. This program can be considered a "second stage" program which encompasses a carefully controlled experimental design unlike the "Pilot program" and has the program input both well designed and highly structured allowing for variable measurability. An important difference the "Model program" has is that it has closely matched experimental and control groups for pre- and post testing. This feature helps to legitimize this model as a truly experimental model which attempts to control extraneous factors from contaminating the outcomes.

The Rossi (1972) and the Suchman (1970) approaches to evaluation are particularly appropriate for evaluating training workshops due to their highly experimental, trial and error, learn from experience approaches. Furthermore, the use of these models does not commit the use to either a highly sophisticated experimental model which will be costly and difficult to apply. It also does not commit the novice evaluator to having to deal with many confounding factors which would be far beyond his/her present skill and expertise.

C H A P T E R I V

4.0 SELECTING RESEARCH DESIGNS

4.1 Theory

Purpose of the Research Design

A research design is a structured plan or strategy that may be used to evaluate and measure the effects of a program or training workshop. It can also be used in the measurement of the degree of accomplishment of intended goals. A research design has two basic purposes (Kerlinger, 1973), that is, "to provide answers to research questions and to control the variance of variables." Kerlinger states that the research design sets the framework for adequate testing of the relations among variables. In addition, the design will suggest the observations that are to be made, how to make them, how to analyze the observations, which statistical concepts are to be used, and finally; it outlines the possible conclusions to be drawn from the statistical analysis.

Selecting the appropriate research design is a process requiring the full attention, knowledge and expertise of the investigator. Similarly, the evaluator/investigator must have an awareness of other factors such as participant availability, internal resource staff availability,

the prevailing climate or environment of the enterprise to be evaluated, the priorities or urgent goals or variables to be dealt with and most important, the existence of sufficient budgets to complete the evaluation project.

Variables

The research design the evaluator chooses primarily depends on the variables to be evaluated. Weiss (1972) suggests that one method that has been employed to help decide which variables are to be measured is what is known as a "Path Analysis" model. With the "Path Analysis" model the evaluator constructs a model of intended processes of the program or training and then tries to identify the means and the steps by which the program is intended to work. It may be described as a chain of predicted events which illustrates the strengths of the various linkages between these events; thus, demonstrating their relative importance. If the predicted sequence breaks down or does not work out, this tells the evaluator that further planning and organizing is required in constructing the model.

Wolf, in Popham (1974) classifies all variables into four categories: independent, dependent, control and supplemental. Kerlinger (1973) classifies variables as independent, dependent, active, attribute, continuous and categorical and intervening or construct.

Wolf's explanation of his four classifications of variables are:

1. Independent variable: in training this variable is the presumed cause in any endeavour (also known as the treatment variable). Examples would be a program, a skill, a teaching method or instructional material.
2. Dependent variable: this variable is presumed to depend on the effect of the treatment or independent variables. It represents an outcome or objective of an action. In training workshops, it is referred to as the goal or objective to be obtained.
3. Control variable: is an item of information obtained about a workshop participant before he/she entered the program or training, for example IQ scores, scholastic ability and instructor ratings.
4. Supplemental variable: a variable which does not fit into any one of the above mentioned categories in training and research.

Kerlinger provides a more comprehensive and detailed definition of his classifications of variables. He says the most useful way to categorize variables is by using the terms dependent and independent which is really a classification of the uses of the variables more than anything else. The independent variable commonly referred to as "X" is presumed the cause of the "dependent variable" usually referred to as "Y". The "X" or "independent variable" is the variable manipulated within the training workshop while the "Y" the "dependent variable" is not manipulated. It is the variable predicted to result in

some intended training outcome. The "Y" variable varies with the variation in the "independent variable" "X". If one were to lay two axes at right angles to each other, the "X" variable would be represented by the horizontal axis, while the "Y" variable would be represented by the vertical axis. It is possible to have a variable being independent in one study or training workshop and dependent in another.

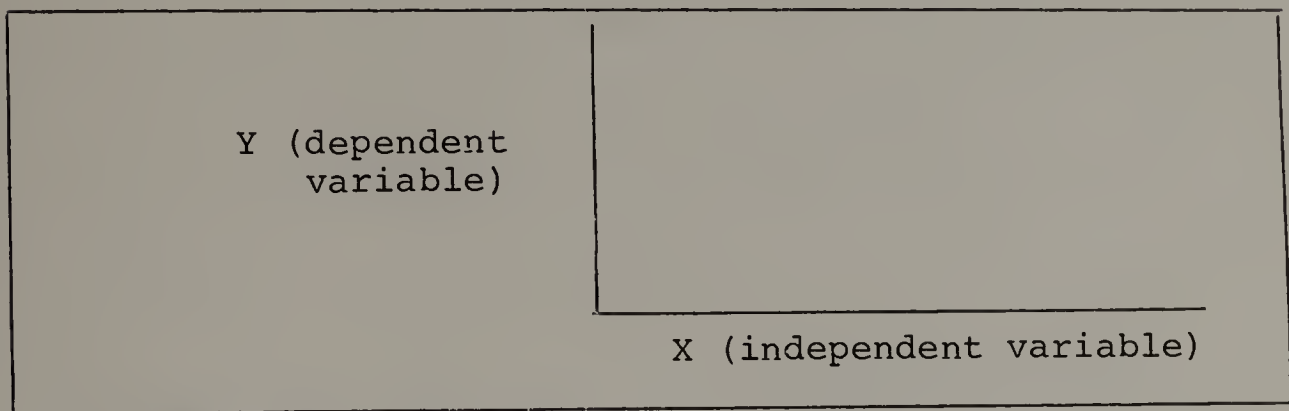


Figure 2. Independent and dependent variable axes.

Kerlinger also refers to the manipulated variable ("X") as the "active variable". The active or manipulated variable to which Kerlinger refers would be the training or treatment variable in a workshop environment. This variable would normally be represented by the specific skill or training being taught or applied in the workshop. The "Y" variable which cannot be manipulated, is called the "attribute variable".

"Intervening" or "construct variables" are terms that were invented for accounting of the internal and directly unobservable psychological processes that account for behavior that can't be seen, heard or felt (an in-the-head variable). In training workshops these variables would be ones which would be unidentified, but influencing the reactions to the skills or training being taught or applied in the workshop. Examples would be hostility, anxiety and motivation.

Suppose a workshop was offered in "empathy training" where nurses were to be exposed to training which focussed on changing specific behavior in the field of "nursing care" through the use of a well known behavioral model. In this example, the behavior to be changed (nursing care) is the independent variable "X" which is manipulated by exposure to the active or treatment variable (a well known behavioral model). The predicted or resulting behavior "Y" is known as the dependent variable. Intervening variables such as anxiety, hostility or motivation may be at work positively or negatively influencing the predicted resulting behavior.

Use of Control Groups

When an evaluator applies a control group to a research training design, he is attempting to rule out

variables that are possible "causes" of the effects s/he is studying or evaluating other than the variables that s/he has hypothesized to be the causes (Kerlinger, 1973). Effects of including a control group within a research training design according to Selltiz et al. (1962), is to ensure as far as possible, the validity of inferences made on the basis of the experiment and to increase the sensitivity of the experiment. Yet, a controlled experiment is not always possible (Weiss, 1972). In many training workshop settings, it is difficult to set up control groups due to many factors. One factor that prevents the use of a control group is that often there are not enough potential training participants available to form both an experimental group and a control group. Another factor preventing control group usage is that many sponsored training workshops are organized for everyone eligible and interested; therefore, difficulty exists in persuading authorities and participants to form a control group. In addition, some sponsors may be reluctant to deny the training experience to their employees or potential participants resulting in no control group as part of the research design.

A control group provides two specific advantages to the investigator or evaluator. First, it provides a basis for comparison of two groups; one which has been

exposed to the treatment or training variables, and one which has not. Any change between the group not exposed (control group) to the variables and the group exposed (experimental group) is attributed to the treatment or training. Second, in the case of a pretest and posttest, it provides a baseline for the before and after comparison (Belasco and Trice, 1969).

Control groups are selected in much the same manner as experimental groups and often they are chosen almost simultaneously. Two common methods of selection are through the process of randomization and matching. Weiss (1972) explains that the potential participants and the control group are drawn from the same group or population and assigned randomly (by chance in flip of a coin, or alternate selection) to either the control group or the experimental group. In "matching" the participants and the control group are matched on characteristics assumed relevant to the outcome, for example, age, sex, and economic level.

Recent literature has highlighted numerous negative responses and comments in relation to the use of "matching" as a means of selecting control groups. It has been suggested that "matching" lulls the investigator into a false sense of confidence that s/he has effectively controlled potentially confounding variables. Thus, this method of

selecting a control group is not highly recommended for use by the evaluator.

Some dangers do exist in the method of selecting control groups. One danger is that the evaluator may find that practitioners wish to assign participants to the experimental and control groups according to who they feel most needs the training or who needs it the least. For example, a principal may assign a new graduate teacher to attend a training program while assigning a more experienced and senior teacher to the control group for that training program. Another danger according to Weiss is that after random assignment, participants may drop out of the control group leaving the remainder of persons unrepresentative of the original population. Often, control groups may feel cheated, rejected, and angry due to their assigned role and the lack of attention and training exposure. This may result in the refusing to cooperate or to complete the required instrumentation for the collection of the evaluator's data. Thus, the evaluator must continually keep a sensitive ear to the control group as much as s/he does to his/her experimental or treatment group.

Weiss suggests a brief design that has not been highlighted or used to any great extent in the field. This design could assist evaluators in attempting to solve

some of the inherent problems associated with responses between experimental and control groups. Weiss suggests that in a situation where new programs or training workshops are being introduced over a period of time, the delayed recipients of the program or training can become the controls for those persons who receive the program or training early. This reduces the responses of persons in feeling neglected, cheated, or angry and increases general satisfaction.

Contamination

Any person attempting an evaluation of a training program or training session must be aware of the possibilities of contamination of the research design. "Contamination" may occur from three main sources. Belasco and Trice (1969) state that the action of obtaining the measure of criterion (pretesting) before the training experience serves to change the subject's attitude, perceptions and attentional sets towards the training and training materials. The second source of contamination is the effect of the passage of time and the occurrence of uncontrolled events related to the training participants between the pre and post testing times. The social milieu to which a person returns during an extended training session for instance, may prejudice his view of the training. The

third source of contamination mentioned by Belasco and Trice is the manner in which the data is collected. The solicitation of data from a few sources, or from peers, superiors or subordinates tends to lay the basis for contamination factors. The greater the number of sources there are from which data is retrieved, the less is the chance of contamination. Belasco and Trice finish by stating that merely sending persons to a training program may shift their perceptions even without the training effect.

Experimental Validity

Airasian (1973), Campbell and Stanley (1963), Kerlinger (1973), Lehmann and Mehrens (1971) and Weiss (1972) speak of two forms of experimental validity--external and internal. "External validity" according to Lehmann and Mehrens answers the following questions:

1. How far can we generalize our findings?
2. To what population samples, situations, events, or variables can this observed effect be generalized?

"Internal validity", considered more important in most evaluative studies than external validity, answers these questions:

1. Did the treatment really make a difference?
2. Can we be sure it was the treatment and only the treatment that resulted in the difference in performance between the groups?

An ideal research design would be one in which both types of validity, internal and external, are present.

In a training workshop, external validity can be jeopardized by the following four factors (Campbell and Stanley, 1963):

1. Reactive or interaction effect of testing: a pretest may increase or decrease the respondents' sensitivity or responsiveness to the experimental variable; therefore, the results obtained for a pretested population would be unrepresentative of the total sample or universe from which the experimental respondents were selected.
2. Interaction effects and selections of treatments: the selection and method of treatment provides for another area contributing to jeopardizing external validity.
3. Reactive effects of training or experimental arrangements: this would preclude generalizations about the effect of the experimental variable upon persons not exposed to it in nonexperimental settings.
4. Multiple treatment interference: the effects of prior treatments usually cannot be eliminated and thus provide a confounding effect on the experiment.

Internal validity is jeopardized in training workshops by the following eight factors:

1. History: events occurring between the first and second measurement in addition to the experimental variable "X".
2. Maturation: the processes of growing older, hungrier, more fatigued for example, have a tendency to contaminate the findings of the experiment.
3. Testing: the effects of taking tests upon scores of a second testing is another area providing contamination.

4. Instruction: changes in calibration of measuring instruments, or changes in observers or scorers used may provide changes in the obtained results.
5. Statistical regression: operating where groups have been selected on the basis of their extreme scores (for example high and low scores).
6. Bias: resulting in a differential selection of respondents for the comparison groups.
7. Experimental mortality: a differential loss of respondents from the comparison groups.
8. Selection-maturation interaction: in certain of the multiple-group quasi-experimental designs the effect might be mistaken for the effect of the experimental variable.

Airasian (1972) comments that threats to internal validity can be controlled or eliminated by the use of identification of a comparison group and the random assignment of subjects to participating groups. Further, he states that with short duration training workshops and programs (for example, one hour or one day), the threat of history and maturation can be eliminated with the use of a control group as these factors likely will have little effect. When the training or treatment in the workshop is insulated or placed in a controlled environment and participation is mandatory, or when the subjects are from the middle ranges of a distribution on a measuring instrument, history, mortality and regressive factors can be discounted. Airasian feels that when unobtrusive and nonreactive measures represent the sole model of data

collection, then no control group is needed to eliminate the threat posed by the testing factor. Often particular threats to internal validity are controlled by the nature of the treatment, its intended clientele, its duration or the conditions under which it is applied.

Experimental Samples

Sampling is the process by which an evaluator or researcher selects a portion of the population or universe which is representative of that population or universe for experimentation (Kerlinger, 1973). The sample size, according to Kerlinger, should be as large a sample as possible as very small samples cannot depend on any one mean (average) as an estimate of the population value. Selltiz et al. (1962) state there are two distinctive types of sampling "probability" and "non-probability". In probability sampling, one can specify for each element of the population the probability that it will be included in the sample. The major forms of probability sampling would be (a) random sampling, (b) stratified random sampling, and (c) cluster sampling.

- (a) Random sampling is a method of selecting the sample such that each member of the population or universe has an equal chance of being selected. It is an unbiased sample.
- (b) Stratified sampling is one of the most generally used forms. The population sample is stratified into men, women, blacks, whites for example and then the sample is selected.

- (c) Cluster samples are the most used method in surveys. It is the successive random sampling of units or sets or subsets.

Another method of sample selection which is under question as to whether it is probability sampling or not is systematic sampling. Kerlinger (1973) explains that the first element of the sample is randomly chosen from the number "I" through "K". Thus, subsequent elements are then chosen every Kth interval. Selltitz et al. (1962) and Kerlinger (1973) present a second distinctive type of sampling called "non-probability" sampling in which there is no way of estimating the probability that each element has of being included in the sample and no assurance that every element has some chance of being included. The major forms of non-probability samples are accidental samples, quota samples and purposive samples.

1. Accidental samples: one takes the cases that fall to hand continuing until the required size is obtained. In this approach the evaluator may take the first one hundred people he sees on the street who are willing to be interviewed. There is no way of evaluating the bias of such samples (other than doing a parallel study by complete census or probability study). Accidental samples are a common occurrence in working environments. Very often the sample population to be trained is selected on a variety of criteria such as who needs it most, who is most deserving of attending or who is most/least popular and who would be willing to attend.
2. Quota samples: there are provisions to guarantee the inclusion in the sample of diverse elements of the population and to make sure that the diverse elements are in the same proportions they appear in the population. The basic goal of

"Quota" sampling is to choose a replica of the population one wants to generalize. This method is a valuable one for training as a representative sample can be trained on a trial basis. If successful the training can be further applied to various groups in the population.

3. Purposive samples: the assumptions behind this form of sampling is that with good judgment and appropriate strategy, one can hand-pick the cases that will satisfy one's special needs. This form usually picks the typical cases of the population assuming the errors will be balanced out. This form of sampling is used for training workshops when testing a new form or plan in order to begin to determine if it would be a valuable addition to satisfy specific predetermined needs.

Selecting Research Designs

Many an evaluation or research design has never been operationalized due to the demands and limitations of time, money and resources. In selecting any research design for an evaluation model these factors cannot be ignored by those evaluating, as eventually they will have to be dealt with or resolved. If left until later in the process of evaluation, these factors may limit the desired scope of the evaluation or, even more unfortunate, contribute to its premature closure and termination.

Some of the designs suggested for use in this section have been selected for their simplicity in design and easy operationalization, applicability to training workshops and for those relatively inexperienced in the fields of research and evaluation methodology. Other designs suggested may be more closely related to the ideals of

Campbell and Stanley (1963), Kerlinger (1973) and Selltitz et al. (1962). One basic observation that can be made of the following designs is that they can provide the novice with sufficiently reliable data on which s/he can base specific, concrete decisions in relations to the direction in which the evaluation is moving.

Soft Technique Research

Rossi (1972) expresses the concern echoed by many researchers in the field of the expense and difficulty endured in designing and setting up controlled research designs. As previously stated, it is not always possible to obtain sufficient consent of organizations to randomly assign persons to experimental and control groups for properly controlled experiments; or even to carry them out when such consent is obtained due to factors of interaction, mortality, and history, as well as other factors mentioned earlier in this section. Rossi forwards the idea that some of these design difficulties related to selection and application of control groups can be avoided by more careful thought of what constitutes a control group. If this was done, Rossi feels that a research design could include a control group that would experience what he terms a "placebo effect". Instead of a control group not experiencing the treatment or training variable, Rossi

suggests that the control group be exposed to some other form of treatment or training that is designed to be of some value to the persons or group. This technique, he refers to as "soft research". Naturally, this approach would go a long way towards resolving some of the problems of expense and the difficulty experienced in setting up research designs.

Rossi considers that evaluation research should be done in two phases--a "Reconnaissance phase" and an "Experimental phase". The Reconnaissance phase would be the phase where soft correlation designs were used to screen out programs worthwhile for further investigation. The "Experimental phase" would be the phase where more powerful controlled experimental designs are used to evaluate the differential effectiveness of a variety of programs which were illustrated as having sizable effects in the first phase.

There are three stages of consideration for an evaluation approach using the "soft technique" method. If massive effect is expected to be the result of some treatment or training, Rossi proposes that it is not necessary to have a control group when the desire is to have complete remission of all symptoms in each and every individual subject exposed to the treatment or training. Thus, if one hundred percent results are not

observed, the evaluator knows the treatment/training was not totally effective. He concludes that if this exposure to the treatment or training shows no effect using the soft methods of evaluation; then, it is very unlikely that the more difficult and precise experimental model would show significant effects. If the correlational design used does show some effects from the application, naturally it is not clear whether they resulted from the treatment or training variables or other factors. Thus, one should consider the soft technique for evaluation as a first option, then discard the treatments which show no effects while keeping those with opposite characteristics to be tested with more controlled and powerful experimental designs.

Pretesting and Posttesting Research Design

In order to obtain an accurate measure of change or progress of participants being trained in a workshop environment, it is necessary to determine a position or baseline from which all participants are to be measured. This baseline provides the evaluator with a point of reference on which to compare the outcomes of the training. Caution must be exercised when using a pretest measure as it is not uncommon for pretesting to bias or contaminate the workshop participants through such things as

the wording of the pretest questionnaire, the focus of the questionnaire or the sequencing of questions. This has resulted in the development of various research designs which omit the pretesting element, but provide major emphasis on how participants are selected from a population where prior knowledge of their orientation or knowledge of the training is known. Still, the validity and reliability of such research designs poses questions for some researchers.

One-shot Case Study and One Group Pretest, Posttest Design

Rossi's approach lays the case (contrary to the opinion of Campbell and Stanley, 1963) for the use of pre-experimental designs such as the "One Shot Case Study" or the "One Group Pretest, Posttest Design" when the evaluator wishes to determine the massive effect of a particular treatment or exposure to training.

The One Shot Case Study does not control for the influence of history, maturation, selection and mortality for internal invalidity, nor does it control for external invalidity of the interaction of the selection process and the treatment. But, if all desired effects are achieved during the training, the researcher will be satisfied even if s/he does not know exactly in detail what the casual relationships were.

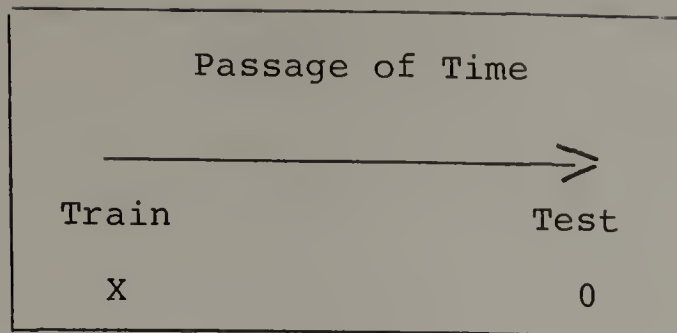


Figure 3. One shot case study

The One group pretest, posttest design does control for selection and mortality for internal invalidity, but, other variables such as history, maturation and testing are not controlled. Similarly, if results are one hundred percent affirmative, the researcher at this time may not be concerned with the "whys".

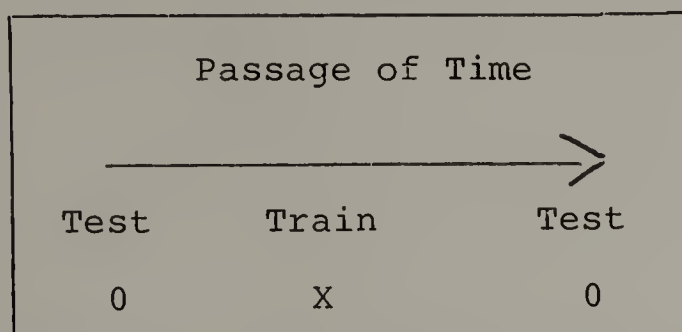


Figure 4. One group pretest, posttest design.

Ex Post Factor Design

It is not unusual for an evaluator to find himself/her-self contracted for an evaluation procedure after a particular program or training session has been well established and operating for a significant period of time.

Obviously, it is impossible for the evaluator to attempt to implement a true experimental design with control and experimental groups as well as measures for ensuring internal and external validity in the situation. The evaluator has no other alternative but to turn to an "Ex post facto design" (a very large proportion of today's research occurs in an ex post facto context, especially in psychological studies).

In Ex post facto research the investigator cannot manipulate or assign subjects or treatments because the treatments or training (independent variable(s)) have already occurred (Kerlinger, 1973). The investigator starts by observing the dependent variable (outcome) and retrospectively studies the independent variables (treatments, training activities) for their possible effects on the dependent variable.

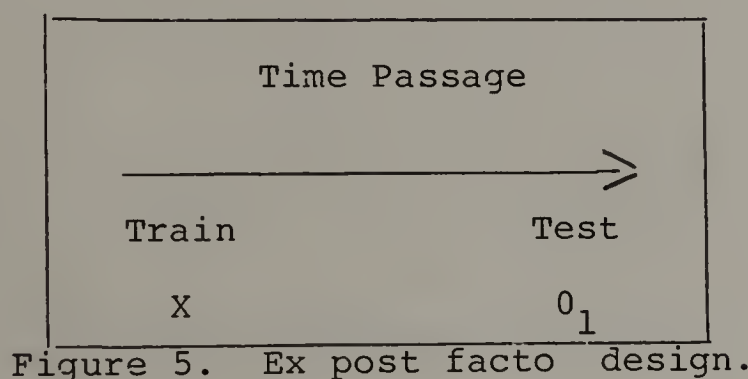


Figure 5. Ex post facto design.

The investigator tests (or observes) the participants in Figure 5 at 0_1 attempting to speculate which training activities at "X" were responsible for the outcomes found

at O_1 . Hence, the investigator may begin to develop a degree of speculation of the causal relationship between "X" and " O_1 ". Then, the investigator may wish to take the speculated treatment or training activity of "X" and attempt to see if its application to a similar group under similar or the same conditions, again results in the same observations received prior to O_1 . The major weaknesses of this design are (1) the inability to manipulate the independent variable, (2) the lack of power to randomize the selection of participants and (3) the risks of drawing improper conclusions from the speculations and observations. An advantage of Ex post facto research is that it can provide speculative data which can later be tested more thoroughly for more decisive conclusions under truer experimental design conditions.

Comparative Research Design

The "Comparative research design" can be used with a variety of known traditional designs such as Pretest, posttest design, Posttest design only, or Pretest, posttest control group design. The Comparative design compares groups of individuals who have been exposed to different types of programming or training, different levels of programming or training or various combinations of programming and training. Belasco and Trice (1969)

feel that with this design the investigator or evaluator can assess the relative value of the various different combinations of workshop training, treatment and programming being applied. Belasco and Trice conjecture that the most important advantage of this design is that it permits simultaneous evaluation of several different change experiences which ultimately increases the evaluation returns. But, they further state that the primary disadvantage of the Comparative design is that the investigator finds it impossible to tell whether the results can be attributed to the program or whether the participants were equally as well off without the change experience or training. Adding a traditional "pure zero treatment" or control group may be one way of handling this problem. Even so, this design has the same problems found in other more traditional designs--that of control and contamination.

The training practitioner would find this design particularly applicable to programs in the area of social action as it allows the investigator to have knowledge of the comparative benefits of different kinds of programs, rather than how effective a program was or if it was better than the training a control group lacked. A further benefit of this design is that it yields the following two kinds of information (Weiss, 1972):

1. The increase in generalization of results
2. The increase in specification of which strategy and which conditions have better effects with different kinds of participants.

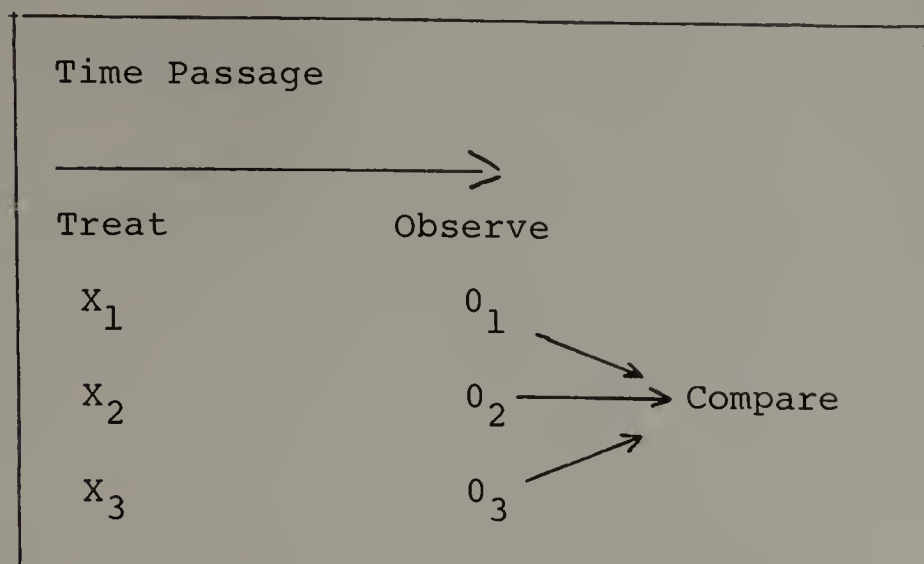


Figure 6. Comparative design.

Own Control Group Design

Miles (1969) presents an interesting design that has not been extensively used by evaluators and researchers. This is the "Own control group" design and can be used when a control group is difficult to find. The evaluator will use the program or training participants to serve as their own control group. The participants attending the training or program are measured some time before the training; then they are measured again after a time lapse, just prior to attending the training; finally, they are measured a third time after the training at an interval comparable to that between the first two measures, has elapsed. If changes between measures two

and three are greater than those between one and two; then, the inference is that the training has caused these changes. With this design, the evaluator must insure that the instrument is not subject to serious practice effects. To achieve this, the evaluator locates a random sample of other persons in the system who agree to respond to the instrument at two different times.

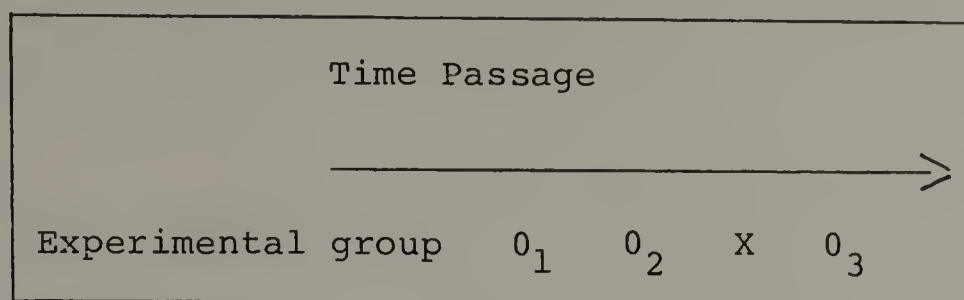


Figure 7. Own control group design.

Time Series Design

This design advocated by Campbell and Stanley (1963), Kerlinger (1975), Miles (1969), Selltiz et al. (1962), and Weiss (1972) is one of the most attractive of the quasi-experimental series of designs as it allows for the study of progressive behavior of individuals or groups over time. This should be an attractive design for those involved in operating long and short term training experiences.

Several observations of the group to be evaluated are taken prior to the training workshop or program intervention to establish a baseline performance level.

After the observations have occurred, the training or program is given. The outcomes of the training workshop or programming are observed and analyzed statistically to determine what effect has taken place.

The ideal application of the "Time series" design is on a long range basis as suggested previously, but some factors have discouraged its use over extended periods of time. One of the major factors encountered is in keeping the participants involved and interested in the experimental design for repeated testing over extended periods of time. Experimental and control group attrition is a common occurrence with extended "Time series" designs due to the participants loss of interest and other natural factors such as change of geographic location, mortality, maturation and environmental contamination. In addition, this design is an expensive one to maintain for long periods of time specifically the follow-up costs of repeated testings and data collection. If these and other complicating factors can be controlled and effectively dealt with; then, the "Time series" design becomes an extremely valuable design which can provide specific information on the effectiveness and success of training over extended periods of time.

This design is actually an extension of the Pretest, posttest experimental design and is an excellent design

for extended training (over one year). It is also good for training programs that last between fifteen and twenty days.

This design does not control for history and the findings can only be generalized to those groups subject to repeated testing. Campbell and Stanley (1963) feel that maturation, instrumentation, regression and selection are ruled out as contaminating factors. Mortality can be ruled out only if data is collected by the individual participants themselves. Another variation of this design is the multiple time series design which may have advantages over the time series design (Weiss, 1972).

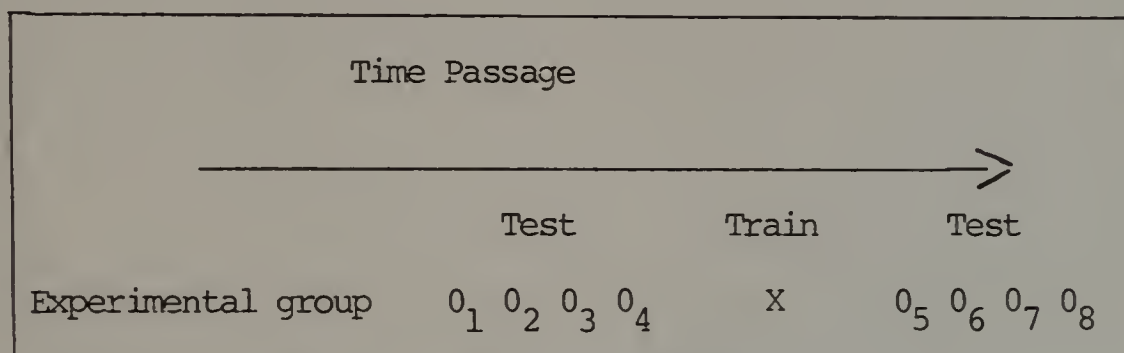


Figure 8. Time series design.

Nonequivalent Control Group Design

Probably, the "Nonequivalent control group design" is the design that is most commonly in use today by researchers and evaluators. The Comparison group design previously mentioned is very similar, but it lacks the randomized control group of this design. Campbell and

Stanley (1963) and Weiss (1972) advocate the use of this design as it controls the main effects of history, maturation, testing and instrumentation. This design, not to be confused with the Pretest, posttest control group design, does not assign participants randomly from a common population to the experimental and control groups, but rather, gathers participants from naturally assembled groups though not similar enough to do away with pretest treatment (pretest can be set aside if randomization of participants is done well, that is within the limits of confidence stated by tests of significance). When using the Nonequivalent control group design, training or a form of treatment (a program) is assigned randomly to one group and is under the investigator's control. A group consisting of individuals with similar characteristics is used as a control. The major problem existing with this design is how to make the comparison group as similar as possible to the experimental group. This design suffers from not being able to control the effects of regression, when matching is used on the basis of pretest scores, rather than using randomization as the method of selection. Generally speaking, selection and interaction of selection contributes to the contamination of the outcomes while the design controls the majority of the other factors contributing to contamination mentioned earlier. The

basic problem around the selection variable is more specifically related to "self selection". As Weiss points out, "people who choose to enter a program have different motivation factors than those that are assigned". This factor has been overcome at times by selecting all participants from volunteers. Other controls have been found in those persons who might have joined the program if they knew of it, or controls have been obtained from those people with similar characteristics as the participants, but who live in inaccessible locations preventing them from registering or attending the program or training workshop.


	Time Passage 		
	Measure	Train	Measure
Experimental Group	O ₁	X	O ₂
Control Group	O ₁		O _X

Figure 9. Nonequivalent control group design.

Pretest, Posttest, Control Group Design

Campbell and Stanley (1963), Lehmann and Mehrens (1971), Miles (1969), Suchman (1970) and Weiss (1972) see this traditional experimental design form as having good internal validity, thus, controlling for history, maturation, testing, instrumentation, selection, mortality and interaction

of selection and maturation through good research practices and sensitivity of application. After application of the training and then measuring the experimental group, the investigator finds some change. The same change does not appear in the control group (which has not attended the program or training); therefore, the investigator can infer with reasonable certainty that it was the training or the program which caused the change in the experimental group.

The major disadvantage of this design is the introduction of the control group. Belasco and Trice (1969) state: as soon as a control group is introduced to any research design, it usually brings with it contamination, such as the influence of a testing tool. This design also assumes that before training, both the experimental and control group have a comparable baseline with respect to the measures being used. Unless of course, randomization has been implemented in assigning the participants to experimental and control groups.

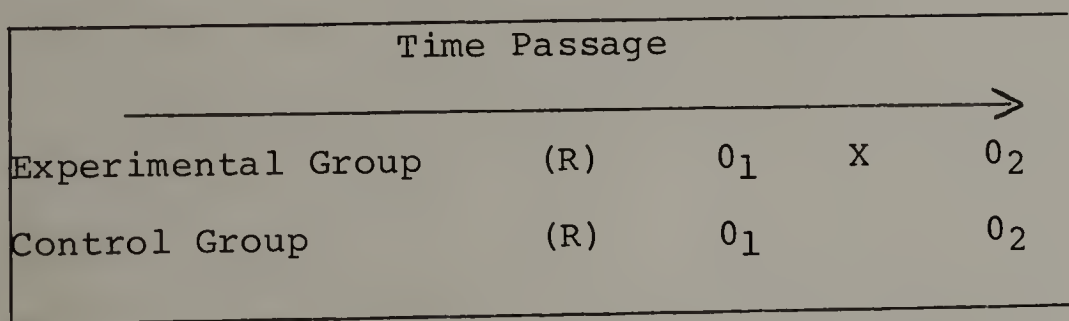


Figure 10. Pretest, posttest, control group design (with random selection "R").

The answers to the following questions will assist the researcher in selecting an appropriate research design for his/her training workshop:

1. Are massive effects expected from the training?
If so, there may be no need at the present for a control group.
2. Does the training, the number of participants or the sponsors allow for the use of a control group?
3. Does the potential participant population and the environmental conditions allow for the use of random selection and assignment of participants to experimental and control groups?
4. Is the training long term or short term?
5. Is more than one training variable being tested and measured?
6. How many participants are actually available for training? When? For how long?
7. At what point in time is the evaluator engaged for evaluation? Before the training? During the training? After the training?
8. Which potential design can best control for the factors present in the training workshop in relation to internal and external validity?
9. Which design can be most readily applied to the specific environmental conditions of the training workshop to be evaluated?

Some constraining forces which would limit the choice of research design are:

1. Time allotted for the evaluation study.
2. The budget and commitment of those involved.
3. The skills of the evaluating person(s).
4. The specific variables being evaluated.
5. The environmental conditions presiding at the time

of evaluation.

6. The specific time and duration of the evaluator's presence for evaluation.
7. The resources available to all for evaluation (manpower, secretarial services, participants).
8. Has similar research been carried on before?

Advantages		Disadvantages
One-Shot Case Study	Inexpensive. Tests for the need of further experimental designs.	Does not control for influence of history, maturation, selection and mortality or external validity.
One-Group Pretest-Posttest Design	Controls for selection and mortality--if results 100%, may not need further testing or research.	Does not control for effects of history, maturation, and testing.
Ex Post Facto Design	Provides speculative data which later can be more thoroughly tested under truer experimental conditions.	No internal and external validity. Unable to manipulate independent variable. Unable to randomly select workshop participants. Risks of drawing improper conclusions exist.
Comparative Design	Allows for the simultaneous comparison of various types of training and treatment as well as simultaneous evaluation of several different change experiments. Thus, increasing evaluation returns.	Can't determine if results are due to the program and training or other factors. Subject to various factors of contamination. Lacks randomized control group advantages.

Figure 11. Some advantages and disadvantages of selected research designs.

	Advantages	Disadvantages
Comparative Design (continued)	Applicable to social action programs, increases generalizability of results, specifies which training or conditions work better with different kinds of participants.	
Own Control Group Design	Useful when unable to find control group or when control group unavailable.	Must insure instrumentation not subject to practice effects. Must find random persons to respond to the instrument at two different times in order to overcome practice effects. Unable to determine if testing or instrumentation have effected evaluation.
Time Series Design	Allows study of progressive behavior of groups or persons over time. Good for long range and some short range training. Controls maturation, instrumentation,	Does not control for history. Findings can only be generalized to training groups subjected to repeated testing.

Figure 11. (continued) Some advantages and disadvantages of selected research designs.

	Advantages	Disadvantages
Time Series Design (continued)	<p>regression and selection. Mortality controlled if participants collect their own data.</p>	
Nonequivalent Control Group Design	<p>Uses randomized control group, thus controls main effects of history maturation, testing and instrumentation. Gathers participants from naturally assembled groups while still using pretest. Treatment and training assigned and controlled by investigator. Design controls most factors of contamination.</p>	<p>Need to make control group as similar as possible to experimental group. If matching used (usually undesirable) can't control regression factor, effects selection and interaction of selection, contributes to contamination.</p>
Pretest, Posttest Control Group Design	<p>Good internal validity.</p>	<p>Introduction of control group usually introduces contamination such as</p>

Figure 11. (continued) Some advantages and disadvantages of selected research designs.

Advantages	Disadvantages
Pretest, Posttest Control Group Design (continued)	testing tool influence. Assumes that both experi- mental and control group have equivalent baselines before training. Need for randomized assign- ment of participants to experimental and control groups.

Figure 11. (continued) Some advantages and disadvantages of selected research designs.

This selective review of research designs is only a very small sample of what is available in evaluation and research literature. As the presentation and discussion of the previous designs has been brief and selective, the reader may wish to read authors such as Belasco and Trice (1969), Campbell and Stanley (1963), Kerlinger (1973), Miles (1968), Popham (1974) and Weiss (1972) for greater detail and study of these examples and of other alternative evaluation designs available for use by the researcher or evaluator.

4.2 Application

Research Designs

When designing this methodology, a search of current literature was carried on to discover research designs which would be practical, logical and relatively easy to apply to training workshops by novice evaluators. The approach of Rossi (1972) and Suchman (1970) appear to fill this prescription. Rossi's "Reconnaissance phase" and Suchman's "Pilot program" model essentially do similar things. They suggest an approach to selecting and carrying out a research design which allows for the application of a variety of variables which are tested under prescribed conditions to see if the results compare with the outcomes desired. In applying either of these approaches, the evaluator can determine whether a truer and more sophisticated

research design is necessary for the training at hand without becoming involved in the costly and more complicated process of setting up an experimental design with control groups and more sophisticated sampling process.

The selection of the Rossi or Suchman approach to evaluation appears to be ideal for use in training workshops which are set up under experimental conditions to test or train persons in special techniques or skills. These two methods permit a low key and inexpensive approach to evaluation while avoiding the more expensive and elaborate need of using control groups. Once the researchers determine which variables or training methods have the greatest potential and impact, then, more elaborate and truer experimental designs can be utilized for determining resulting cause and effect relationships of the training variables being applied.

The selection of Rossi's or Suchman's method would provide a variety of choices of research designs available to the evaluator for workshop evaluation. These designs have been categorized by various names such as pre-experimental, non-research or elementary research designs. Specifically some of the designs available to the evaluator would be:

"One Shot Case Study"

This design discussed by Campbell and Stanley (1963)

is not difficult to apply. A training group is selected to be studied once by the evaluator. It is subjected to specific treatment, training or programming which is believed to have specific change effects or outcomes. The resulting outcomes are compared to the expectations of the researcher by casually observing the results and comparing them to the memories the researcher has from similar comparable situations. If the resulting outcomes match the researchers expectations usually further research will not follow. If they do not, most likely a "learn by experience" or a "trial and error" approach will be used until an acceptable level of outcomes is achieved. The major weakness of this design would be its lack of reliability and validity and the opportunity to generalize the findings.

The advantage this design brings with it, is that it is easily applied by the novice, inexpensive to operationalize and apply in various situations, and finally, it saves time, money and energy absorbed by more sophisticated designs which may yield the same information. This design provides the advantage of quickly providing the evaluator with specific information as to whether the expected outcomes of the training are being achieved.

One Group Pretest Posttest

Similarly this design attempts to do what the pre-

viously described design does, but it is an improvement over the One shot case study due to the addition of a pretest which establishes a baseline for the training group being tested. If a satisfactory degree of desired change takes place between the pretest and the posttest; then it is unlikely that further research will take place. This design also lacks reliability and validity of its findings (Campbell and Stanley, 1973; Rossi, 1966).

Ex Post Facto Design

This design has all of the inherent weaknesses of not being able to control the factors of reliability and validity due to the speculative nature of obtaining the data. It is included at this point only because many evaluators are not called upon to do an evaluation of a training workshop until after the program and its goals have been conceived and operationalized. Then, it becomes the problem of the evaluator to try to speculate what aspect or degree of the treatment or training has resulted in what is presently being observed in the group under evaluation. To attempt to evaluate a training workshop after the training and goals have been conceived and operationalized is a unrewarding task due to the poor quality of data likely to be obtained. However, as this or similar situations frequently arise in the evaluation field, it is necessary to include this model with all its inadequa-

cies as a suggested starting point for the novice in attempting to come to grips with the common problem. Caution should be exercised in the interpretation of the resulting data due to the speculative nature of its origin.

The data resulting from this approach would be compared with data from another training workshop test case to see if the outcomes or causal relationships would be the same using the same treatment or training under the same conditions. If so, a truer experimental design would be recommended to draw more decisive conclusions on the outcomes (Campbell and Stanley, 1963; Kerlinger, 1973; Rossi, 1966).

Time Series Design

Campbell and Stanley (1963), Kerlinger (1973), Selltiz et al. (1962) and Weiss (1972) have much to say in favor of the Time series design when it is used for evaluating both long and short range training or programming. Its primary advantage to the evaluator is that it establishes a baseline measure for the group being evaluated and follows with numerous sequenced testing or measures which determine the degree of change at various stages of progression. Although this design can be adapted for evaluating training workshops and programming, difficulty is often experienced by researchers in gaining the long range commitment of the group participants necessary for complet-

ing the evaluation measures.

A further advantage of a time series design allows the evaluator to determine the reactive effect of the treatment variable, if any, over a designated period of time and if so, at what point does it take effect or begin to occur. Time series designs are often presented with the difficulty of not being able to control the effects of external learning other than what has been presented in the treatment or training. This tends to cloud the issue as to whether any resulting changes were due to the training or exposure to other uncontrolled learnings external to the workshop. The longer the time period, the greater the problem as the time factor becomes a variable. In order to counter this effect the researcher must take this into account when analyzing the data, or demonstrate that the experimental manipulation was greater than the external or extraneous influences (Kerlinger, 1973).

Another advantage of using a "Time series" design in a training workshop is to help counteract possible reactions due to the "Hawthorne Effect". The time span in this designs allows for the diminishing of this effect on the workshop participants.

Much of the research being carried out today lacks good long term follow-up; specifically, in the training workshop field. As a result many training activities and methods are seldom modified or changed to increase their

effectiveness. For example, T-groups have been extensively used in recent years to improve the quality of teaching and teacher interpersonal performance. Little research was ever carried on to determine the most desirable methods or approaches to achieve this goal. Similarly, there was little research carried on to determine the effectiveness of the methods used when applied on the job. Recent research is suggesting the T-group approach is not readily transferrable to the job situation unless the total system was exposed to the extensive training which in most cases would be unfeasible.

Comparative Design

The comparative design has gained favor from the ability of researchers to use it to measure the same training, programming or treatment on multiple grounds under the same conditions, then seeing if the outcomes are comparative. This design can also be used to test the effect of various methods of training, programming or treatment of similar grounds under identical conditions to determine if the results have similar outcomes. As stated earlier, this design would have appeal for those persons requiring a comparison of effect of various training methods or programming such as an organization who is testing the effects or specific training methods to improve communication. This design can be greatly strengthened by the addition of

a control group which will be discussed in the next section.

Experimental Design

Both Rossi (1966) and Suchman (1970) advocate the use of the "True experimental design" once the need for further research methods have been established by soft research and demonstration programs. Rossi refers to this as the "Experimental phase" and Suchman calls this his "Model program". This more sophisticated evaluation approach uses more powerful experimental designs in further evaluating those programs or variables which have shown promise; as a result, a better quality of research data is obtained. There are many designs that can be considered, depending on the preferences of the evaluator and the variables being measured, or the conditions under which the evaluation takes place. It would be most practical to utilize a traditional pretest, posttest control group design as it has good internal validity and reliability. The introduction of contamination becomes a major factor through the use of the control group. But, careful application of the design and the establishment of comparable baselines for experimental and control groups, or the assignment of participants to experimental and control groups using randomization techniques, can help to overcome this problem.

Three further suggestions may be considered to help the control group problem. The first is suggested by Miles and is described in Chapter II (p. 55) as the "Own Control Group Design". In this design the experimental group is used both as the experimental and the control group and is measured twice prior to treatment (O_1 and O_2); then treated, and again measured after treatment (O_3). It is inferred that training has been the cause of any change if the changes between O_2 and O_3 are greater than the changes between O_1 and O_2 .

The second suggestion concerning the use of control groups has been suggested by Rossi. He suggests that rather than providing "zero treatment" to the control group, other relevant treatment or training can be applied to the group and comparisons between the groups can be then taken as usual. He refers to this as "placebo treatment". This avoids the problems of reactions resulting from denying and withholding treatment, training or programming from participants. Weiss (1972) simply suggests dividing the potential training participants into two groups. One receives the training now, the other later. The group who is to receive training later acts as the control for the group who is presently receiving the training.

It is very difficult to advise the novice evaluator to use one design or another without thoroughly knowing or

understanding what is to be evaluated and under what conditions or circumstances the workshop is to occur. Some have attempted to solve this problem by frequently utilizing a favored design and modifying the situation to fit the design. Although this practice is more frequently in use than desired, Tripodi, Fellini and Meyer (1969) question the ethics of such a move as well as the validity of the resulting research.

Follow-up

No research design is complete without some attempt at follow-up research in order to determine the lasting effects of the training or to become aware of significant changes which have occurred as a result of the exposure to the training workshop. The concept of follow-up can not be too greatly emphasized, as often, due to budget limitations, time restrictions and participant availability, it is sacrificed from the overall design (that is, if it were included originally). As a result, the data obtained may often suffer from the lack of true accuracy. In turn, the decisions which are based on this information may fall short of achieving the desired results anticipated.

When follow-up procedures are omitted, posttest results and interpretations should be viewed with caution by the novice evaluator as the apparent results may change with the passage of time. This may go undetected due to

the absence of follow-up research. Two further advantages may be attributed to the use of follow-up research. They are:

1. Follow-up research can serve to further confirm posttest results, thus giving a greater validity and reliability to the posttest findings.
2. Follow-up procedures can demonstrate the extent of the learning and its specific application to the back home/on the job situation. This is invaluable to further planning and to future workshop training designs.

C H A P T E R V

5.0 DATA COLLECTION AND INSTRUMENTATION

5.1 Theory

Introduction

Once the evaluator or investigator has selected the appropriate research design for the purpose at hand, s/he must decide what type of data needs to be collected and how it should be collected. Weiss (1972) suggests that there is a variety of sources and types of data that can be gathered. Some of the sources she suggests are interviews, questionnaires, observations, ratings (by peers, colleagues, experts), psychometric tests of attitudes, values, personality, beliefs, preferences, tests of information, interpretation, skills and application of knowledge. Selltiz et al. (1962) suggests that various projective techniques and available records may also be substituted for some of the observation methods presently in use. But, the most popularly used methods in the field are observation, interviewing and the questionnaire. Dimock (1973) describes the three methods as a process of observing the behavior of the trainee or program participant, having them report on the program or training through interview, self report or questionnaire and finally, testing the learning of the respondent by the use of a questionnaire.

Observation

Observation is a primary tool of scientific inquiry according to Selltitz et al. (1962) when it serves a formulated research purpose, when it is planned systematically, when it is recorded systematically and when it is related to more general propositions, rather than being used as a curiosity and finally, when it is subjected to checks and controls of validity and reliability. Observation can be used to obtain many types of data of which one primary source is behavior. Selltitz et al. feel that the advantage observational techniques have over other data collection techniques are that they make it possible to record behavior as it occurs, allowing for immediate observation of responses and consequences of the acts observed. The use of the interview, which will be more thoroughly discussed in the next section, permits a system of checks and clarification when used in conjunction with the observation method.

Another advantage of observational methods is that often data can be recorded which would never have been collected through other methods such as the interview and questionnaire. Generally, the observations made of individuals at work, at play and in everyday life are reliably accurate even though they are aware that they are being observed and may even try to disguise their actions. The

longer the opportunity for observation, the greater the consistency and reliability of the information observed. On the other hand, observation methodology does not go without its limitations. It is not always possible for the evaluator or investigator to predict when certain behavior will occur and then be present to observe and record it. Often unforeseeable factors interfere even with the most organized and planned observation attempts. These limitations place observation methods in a category of being an expensive procedure. Interviewing, if possible, may be a more economical choice. In addition, there are many types of behavior that are not accessible or readily observable for the investigator, such as sexual behavior or spontaneous family crisis.

A variety of research purposes may be secured by this method of evaluation. Observation methodology may be used to gain specific information that can be later tested by other techniques, and either qualify or help interpret other findings obtained by other methods. Observation is also used as the primary method of data collection in studies designed to provide accurate descriptions of situation or in an everyday, real life situation. Selltiz et al. point out that the observational procedures may range from a very flexible, free flowing observation process to a highly structural use of formal observational instruments designed to record predetermined categories or de-

scriptions of behavioral actions planned in advance. The flexibility of this method may also be extended to the role of the observed/investigator/evaluator. S/he may have the option of observing from a participant's role or a non-participant's role. That is, the observer, himself/herself participates actively in the group being observed, or s/he is defined as a member of that group while keeping his/her participants to a minimum. Other choices that are available to the investigator or that s/he may elect to be an observer who is not part of the group or s/he may decide that his/her presence remain unknown to some or all of the group members being observed. To summarize, the degree of structuring of the observation approach and the degree of participation of the observer/investigator tends to vary with the purpose of the evaluation study.

Four key questions should be considered by the evaluator regardless of the purpose of the evaluation if observation techniques are to be used. They are:

1. What should be observed?
2. How should observations be recorded?
3. What procedures should be used to try to assure the accuracy of observation?
4. What relationship should exist between the observer and the observed, and how can such a relationship be established?

The following areas are suggested by Selltiz et al. as areas of possible choices for observation:

1. The participants.
2. The setting.
3. The purpose (what has brought people together).
4. The social behavior (a) stimulus or event initiating it, (b) what appears to be its objective, (c) toward whom or what is behavior directed, (d) what are the qualities of the behavior (intensity, duration, persistence), (e) what are the effects, (f) what is the form of the activity entailed in the behavior.
5. The frequency and the duration of the training or experimentation.

Recording Observations

Whether the evaluator intends to use an unstructured or structured approach for evaluating training workshops, the question always arises as to how records should be kept. Should they be in writing, or recording tape, videotape or on film? Each method brings with it certain limitations dependent on the abilities and resources of the investigator, the adaptability and use in specific environments or possibly just the bias of the evaluator. A second question is always evident when recording observations; that question is "when should recordings take place?" Should they be before, after or during the actions being observed?

Again the answers depend on many factors known to the observer because of his contact and awareness of the group, the environment and the situations. Selltitz et al.

cautions the observer who uses the unstructured approach to observation in that his/her involvement in the situation may lessen the sharpness of observation because s/he may become accustomed to certain kinds of behavior and eventually closely identify with those s/he is observing. This danger is not as prevalent when observations are being made of groups using a structured observation approach. With this approach, predetermined behavior or aspects of group activity are observed. The role of the observer is more defined and structured and is more often limited to non-participation while recording of information is done using predetermined categories and choices.

Kerlinger (1973) suggests two primary modes of observation. He says that the investigator can watch what people do and hear what they say, or that the investigator can ask people about their own actions and the behavior of others. The method of obtaining information is by experiencing some action, phenomenon or situation or by having someone tell the investigator what happened. Kerlinger (1973) sees the observer as both a strength and a weakness in that s/he must digest the information observed; then, attempt to make inferences about constructs. This process is subject to the bias and the prejudices of the observer and can be quite incorrect in the final analysis. Kerlinger (1973) is in agreement with Selltitz et al. when he also states that he feels the observer has little last-

ing effect on the situation being observed and recorded, and that both groups and individuals quickly adapt to the observer's presence and act as usual. In summary, there is no one best method of recording observations although some procedures yield certain kinds of data that others cannot. It is usually in the final judgement of the evaluator or investigator what approach should be used and which method is likely to yield the types of data required in the final analysis. For further discussion on observation techniques, see Runkel and McGrath (1972).

Reliability and Validity of Observations

Reliability of observations made by an investigator may be increased by ensuring, prior to carrying out the observations, that clear definitions of the kinds of behavior required exist, and that they are written down and understood by those recording the observations. Ensuring that those observing have a high degree of judgment and confidence about themselves and the assignment further ensures the likelihood of reliability. Reliability is also increased by selecting the observers in relation to the behavior they are to record. Finally, careful training of the observers through practice, role playing, question and answer sessions further ensures reliable results (Selltitz et al., 1962).

Kerlinger (1973) suggests that reliability of obser-

vations can be ensured by the agreement of multiple observers and multiple observations while using time sampling and through the use of sophisticated statistical methods, namely "analysis of variance".

Validity, according to Kerlinger (1973), is far more difficult to achieve in relation to observations. Validity may suffer with increased observer interpretation but, he goes on to say, if the variables being measured are embedded in a solid theoretical framework, there should exist a relation (construct validity). If repeated testing of variables under the same conditions and situations demonstrates continued reliability of findings, then the evaluator may attempt to consider the observations have validity.

Observation methods have proven to be good for observing behavior, but less effective in providing data about people's perceptions, beliefs, feelings, motivations, anticipations or future plans. To obtain such information, the interview and the questionnaire have been highly utilized (Selltiz et al., 1962).

Questionnaires and Interviews

Observational methods discussed in the section prior to this are primarily used to describe and understand behavior as it happens (Selltiz et al., 1962). As the investigator uses either the questionnaire or interview

method or both, s/he will be aware that they depend almost entirely on the participant or trainee's account of what was experienced or what they were subjected to. Most commonly, the investigator is not present and the trainee's account will be verbal and limited to what they are prepared to share.

The data collected may or may not be interpreted by the investigator/evaluator and in some cases it is taken at face value while in other cases, it may be analyzed in relation to some knowledge or theory. Further, Selltiz et al. feel that the questionnaire can be considered a form of interview which takes place with the interviewer absent. The interview most commonly used is in the form of a series of predetermined questions which may be written down prior to the face to face encounter. At first glance, one may readily see the similarity and possible overlap of the two methods, but there are distinct and important differences between the two methods. For example, in the questionnaire, the data the evaluator/investigator obtains is specifically limited to the responses of the program participant or trainee to the prearranged questions. In the interview, the interviewer and the respondent (trainee, participant) are both present when the questions are asked and thus, there is opportunity for a greater flexibility in obtaining the information. Selltiz et al. add that

the interviewer has the opportunity to observe both the subject and the total situation to which the subject is responding.

The following is a list of further advantage of questionnaires:

1. They are less expensive than interviews.
2. It requires less skill to administer a questionnaire; interviewing demands training and skill.
3. Questionnaires can be readily administered to large groups of people simultaneously; interviews cannot.
4. Questionnaires can often be mailed; interviews cannot.
5. In comparison and with a given amount of funds, it is usually possible to cover a wider area and obtain information from more people by means of a questionnaire than by interview.
6. Uniformity from one situation to another is insured through the impersonal nature of the questionnaire, its standardized working and question order and its standardized instructions to respondents; interviewing situations are rarely uniform from one interview to the next.
7. The anonymity of the questionnaire may provide for freer answers and greater confidence of the respondents to express themselves.

The following is a list of the advantages of interviews:

1. The interview does not demand a degree of literacy to ensure the gathering of data as does the questionnaire.
2. Interviews can be used with almost all segments of the population.
3. Potential respondents are often more prepared to

cooperate in an evaluation or study when all they are required to do is talk.

4. There are many factors that influence the percentage of negative returns of questionnaires than of personal interviews.
5. The interview has greater flexibility as there can be interaction between the interviewer and interviewee to clarify, to probe deeper for truer responses or simply to rephrase questions for clearer respondent recording; in a questionnaire all interpretations and responses are final.
6. The interview provides opportunity for the interviewer to see how the respondent responds to a question as well as what part he responds to.
7. The interview also provides opportunity for additional relevant information to be recorded by the interviewer, for example non-verbal behavior responses to questions.
8. The interview provides opportunity for the investigator to set or create an environment conducive to obtaining the desired information while the questionnaire does not provide for this.

It should be noted that the questions content of both questionnaires and interviews can be organized as such to focus primarily on ascertaining only factual information, ascertaining beliefs about what the facts are, and ascertaining feelings or discovering standards of action. Further, focus of the content can also be directed towards finding data on past and present behavior, and conscious reasons for beliefs, feelings, policies and behavior (Selltiz et al., 1972).

Types of Questionnaires and Interviews

Interviews and questionnaires vary in form and struc-

ture dependent on their use, the type and form of data desired and the personal bias of the investigator. The interview has the greatest range of variance. It may be in a standardized format rigidly adhered to where the questions and responses of those being questioned are predetermined. It also may be so unstructured that a free flowing atmosphere prevails where the responses of those interviewed and the questions asked may neither be pre-planned or predetermined. Even though, the questionnaire does have some variation. The range of the structure is far more limited.

Standardized Interviews and Questionnaires

When an investigator/evaluator uses the standardized format of the interview or questionnaire, s/he presents the questions with the exact same wording and order to each respondent, program participant or trainee. This procedure ensures that all respondents will be replying to the same question. The difference between standardized interviews and questionnaires may lie, however, in the degree of structuring of the questions used (Selltitz et al., 1962). The structuring may be in the form of "fixed alternative" answers (closed answers) or a free flowing answer known as the "open ended" answer. In "fixed alternative" or "closed answer" questions, the respondent is required to respond to a set of stated or written alter-

natives. The alternatives could be answered in the form of a "yes" or "no". Another form would be to select one choice of a set number of choices which best describe the respondent's feelings, behavior and opinion.

The fixed alternative or closed questionnaire is standardizable, simple to administer, quick and relatively inexpensive to analyze (Selltitz et al., 1962; Kerlinger, 1973). The fixed alternative question ensures that the responses will be usable in the final analysis, and answered within the given frame of reference within the limits of the inquiry. Fixed alternative questions may also require the respondents themselves to make a judgment about their attitude, rather than leaving this up to the interviewer and risking interviewer bias or differences of opinion. This may or may not be desirable depending on the nature of the question. Fixed alternative questions are preferred when possible alternatives are known, such as income, age and education. Benjamin (1969) adds that closed questions or fixed alternative questions are valuable for specific information when that is desired. If the closed format is used by the investigator; then, the questions should be stated succinctly, remaining clear, and understandable. Kerlinger (1973) comments on the self administered instruments such as the questionnaire, by suggesting that the closed item type of format is more

conducive to validity and provides more uniformity of stimulus. If it is completed anonymously by the respondent; then, a higher potential for honesty and frankness of responses can be encouraged. Since, the questionnaire can be easily administered to large numbers it is also seen as an advantage, but an even greater advantage is that it is far more economical, costing a fraction of the cost of the interview. An excellent example of a self administered, standardized questionnaire can be found in "Appendix D" in the format of a "Workshop Evaluation Questionnaire" developed by Hambleton (1976).

Unstructured Questionnaires and Interviews

The "open ended" question or unstructured format, encourages the respondent to answer freely, the questions posed by the interviewer. In this case the respondent is not given a choice of answers to choose from. Benjamin (1969) further categorizes the open question format. He says questions can be asked in a "direct" or "indirect" manner. He says that open questions are direct questions and that they can be made more open by stating or asking them directly. This can yield a rich source of data which can be later analyzed and coded for categorization by the evaluator.

The "open ended" question format is difficult and expensive to analyze. The categories must be built by

those persons or coders who are attempting to tally the responses. While gathering considerable additional data that is often helpful in the analysis of the responses, the open ended question may gather too much data causing confusion in the analysis and fatigue of the respondent and interviewer. A preference of most evaluators (dependent on the situation and circumstances) is to use a combination of both fixed alternative questions and open ended questions. Dimock (1970) suggests using the open ended approach to build the categories; then, to use the fixed alternative approach to tally the opinions and attitudes.

Benjamin (1969) summarizes with a comparison between the fixed alternative (closed) question and open-ended question approach.

<u>Fixed Alternative (Closed) Question</u>	<u>Open-ended Question</u>
1. Narrow	1. Broad
2. Limits to specific answers.	2. Allows for full scope of answers.
3. Curtails perceptual field.	3. Invitation to widen perceptual field.
4. Demands cold facts only.	4. Solicits views, opinions, thoughts and feelings.
5. May circumvent the widening and deepening of contact.	5. May widen and deepen contact.
6. May discourage rapport.	6. May encourage rapport.

The interview varies in format also. Dimock (1970a, 1970b) suggests that interviews can be classified as focused or non-directive. The primary function of the focused interview is to focus attention upon a given experience and its effects. The investigator knows in advance what it is that he wishes to cover during the interview. The interview is clearly in the hands of the investigator.

The non-directive interview has the initiative of the interview clearly in the hands of the respondent. The respondent is encouraged to express his/her feelings without direct suggestions or questions from the interviewer. The specific function of the interviewer is to get the respondent to talk and provide the catalyst to keep him/her talking. For further detail and greater discussion of questionnaires and interviews, it is suggested that the reader consult Kahn and Cannell (1956, 1969).

Sociometry

Selltiz et al. (1962) see sociometry as being concerned with social interactions among any group of people. They add, that with the sociometric method data collection is geared to obtaining information about the interaction or lack of interaction among members of any group. The content or type of these interactions may be of any one of a variety of social behaviors, for example, who sits next

to who, who eats with who, or who buys for who.

Sociometry is seen by Selltitz et al. as a questionnaire or interview approach. Kerlinger (1973) classifies it as an observation method while Dimock (1970a, 1970b) does not limit it to one approach or another. The difficulty in classifying this approach for data gathering arises from the fact that the method uses all three primary approaches to collect information: observations, interviewing and questionnaires.

Kerlinger (1973) identifies three basic forms of sociometric analysis: sociometric matrices, sociograms and sociometric indices. He feels that perhaps the sociometric matrices contain the most important possibilities and implications for the behavioral researcher. Kerlinger defines sociograms as being diagrams or charts of choices made in groups. Sociometric indices are defined as simple numbers calculated from two or more numbers yielded by sociometric data which indicate sociometric characteristics of individuals and groups. Sociometric matrices, states Kerlinger, are rectangular arrays of numbers or other symbols usually indicating individual's choices (see Appendix B). From these can be discovered cliques in groups, communication and influence channels, patterns of cohesiveness, connectedness and so on. Kerlinger adds that sociometry is a simple, economical and naturalistic

method of observation and data collection. Selltitz et al. (1962) state that in studies of reliability of the sociometric approach, reliability is based on repeated tests which have indicated that the indices arrived at are really quite stable. This gives the sociometric approach a good reliability rating. Dimock (1970a, 1970b) has written two excellent monographs explaining and demonstrating the function and use of the sociometric approach to behavioral sciences.

Popularity of Data Gathering Techniques

The questionnaire is most commonly used for data gathering in training workshops. The advantages previously cited in this section are the reasons for its popularity in use. In particular, its tendency for standardizing questions and answers makes it a useful tool for comparing training groups and/or various training techniques and components.

Observation techniques are the second most popular method for gathering data on training workshops. Very often the questionnaire is combined with observation techniques, providing further dimensions in the data gathering process. Sensitivity-training and T-grouping have heavily relied on both of these methods to probe into the interpersonal depths of participant's behavior in order to encourage greater interpersonal growth.

Sociometrics have also found considerable popularity in use, in interpersonal relationship training and the social group work field. Sociometrics depend heavily on observation techniques as well as considerable use of instruments, plotting charts and modified rating scales in interpersonal and group data gathering. A greater use and popularity of sociometrics for providing feedback would be useful and beneficial in providing new data for learning in various training workshop environments. Yet, there seems to be reluctance to do so due to its highly interpersonal and behavioral aspects and the difficulty in obtaining measurable specifications using this method of data collection.

The interview is seldom used for data gathering in training workshops due to its high costs and time consumption required. Occasionally, the interview will be used as evaluative tool at the end of a workshop or sometimes after (for example, one month after). Then, only a limited number of interviews are carried on. The interviewees are usually randomly selected and are only a small representative sample of the population that has been trained. Naturally, the validity and the reliability of the data obtained under these post-training conditions is questionable due to participant contamination.

Projective Techniques

For the purposes of this paper it is only necessary

to mention the projective technique method in passing as it will have limited applicability for the novice evaluator/investigator. Selltiz et al. state that projective methods were first devised by psychologists and psychiatrists concerned with the treatment of patients suffering from emotional disorders. This has continued to be its major use. Usually, the projective method involves giving a subject a stimulus that will have a particular meaning to the subject which ultimately will indicate the subject's particular patterned view or perception of his surroundings and his/her response to it.

For further reading and discussion on data collection methods see Kahn and Cannell (1956), Lehmann and Mehrens (1971), Miles (1969), Schmuck and Runkel (1972), Warwick and Linninger (1975) and Worthen and Sanders (1973).

Attitude and Rating Scales

During an evaluation process, the investigator or evaluator often finds that it is not enough to just determine whether goals and objectives have been achieved or whether a person's behavior has changed in one direction or another. It is often even more important for the evaluator to determine to what degree are changes occurring, or the degree things are being accomplished. This is the primary purpose and function of attitude and rating scales (Selltiz et al. 1962). These scales permit the investi-

gator to assign participants or trainees to numerical positions allowing him/her to make distinctions of degree possible. The scale positions generally only indicate the order of the positions with respect to the characteristic or feature being measured. Most attitude and rating scales provide primary ordinal measurement, that is, it does not imply the distances between positions on that scale.

The ordinal scale tells whether the workshop trainee has more or less or the same amount of the characteristics being evaluated than another trainee for example. A nominal scale, as implied earlier in the paragraph, only implies whether the workshop trainee has characteristic or not. It does not indicate the degree of possession. An interval scale allows for the characteristics evaluated to be arranged in terms of greater, equal, or less, with all the units of measurement being equal. For instance, the distances between item No. 1 and No. 2 are equal to the distance between item No. 2 and No. 3 on this scale (Selltitz et al., 1962). Kerlinger (1973) defines a scale as a set of symbols or numerals constructed such that the symbols or numerals can be assigned by rule to the individuals (or their behaviors) to whom the scale is being applied. The assignment is indicated by the individual's possession of whatever the scale is supposed to measure. Kerlinger adds that tests are a form of scale, but scales are not necessarily tests, as a test suggests competition

METHOD	DESIGN	DEPTH	FLEXIBILITY
INTERVIEW	EASY	Much Most	<u>Most</u>
QUESTIONNAIRE	<u>Difficult</u> Preparation takes time	<u>Less</u> Depends on length and complexity	<u>Less</u>
OBSERVATION AND SOCIOMETRY	<u>Varies</u> (Depends on how systematic)	<u>More</u> Can't deny behavior	<u>Varies</u>
RECORDS AND DOCUMENTS	<u>Easy</u>	<u>Varies</u>	<u>Least</u>
PROJECTIVE	Requires some care	Much Most	<u>Much</u>

Figure 12. Summary of methods of data collection including comments.

METHOD	QUANTIFI- CATION	TIME TO DO	EFFECTS OF DOING	OTHER
INTERVIEW	<u>Difficult</u> (content analysis)	<u>Much</u>	Builds trust, may change people's views	Open to bias
QUESTION- NAIRE	<u>Easy</u>	<u>Least</u>	Less trust building - may change views	more objective
OBSERVATION AND SOCIOMETRY	<u>Varies</u> (Systematic tabulation easy. Narrative record hard.)	<u>Much</u>	Possible problems if observers seen as intruders or evaluators	Do observers participate?
RECORDS AND DOCUMENTS	<u>Least</u>	<u>Little</u> (but varies with complexity of records)	<u>Least</u> no effect	Public vs. Real data non-reactive
PROJECTIVE	Near Impossible	<u>Little</u>	Generates energy seen as childish	Requires some trust

Figure 12. (continued) Summary of methods of data collection including comments.

and success or failure. Many scales are not designed for competitiveness or to measure pass/fail. Their function is to determine if a characteristic exists or not, and in some cases to what degree.

According to Kerlinger, achievement tests are an example of the competitive style of scale. Their intent is to measure the present proficiency, mastery and the understanding of both general and specific areas of knowledge. For the most part they have been used to measure the effectiveness of instruction and learning.

Rating Scales

In using a rating scale Selltitz et al. (1962) explain that the evaluator or rater places the person or object being rated at some point along a continuum or in one of an ordered series of categories. A numerical value is attached to the point or category. They further add that scales differ in the fineness of the distinctions they permit and in the procedure involved in assigning persons or objects to positions. Some of the more common types of rating scales are graphic rating scales, itemized rating scales, comparative rating scales, and self-rating scales (see Appendix C).

The evaluator who chooses to use rating scales must be cautioned and made aware of some of the pitfalls they may befall their application. Selltitz et al. (1962) sug-

gest that because of the element of judgment used in the rating by the rater, an element of bias may be introduced. As examples: "halo effects", "generosity errors" or "contrast errors" may be unknowingly brought into effect by the raters and evaluators. These errors can be reduced by more careful training of raters and recorders. If raters and evaluators are made aware of their bias, this often serves to reduce the error. The error of generosity can be reduced by giving neutral descriptive terms for the scale position rather than evaluative ones. The halo effect can be reduced by having various ratings of a given person made independently by different raters or with the same raters at different times, but the rater is unaware that it is the same person.

Reliability of rating scales can be increased by providing clear definitions, specifications of reference groups, and wherever possible define scale points with illustrations.

Attitude Scales

In an effort to reduce error, attitude scales have been carefully standardized and constructed. With an attitude scale, the respondent does not directly describe himself/herself in terms of his/her position on the dimension in question, but rather s/he expresses his/her agreement or disagreement with a number of statements relevant to it (Selltiz et al., 1962). On the basis of these re-

sponses, s/he is assigned a score.

It may be noted that attitude scales differ in the method of construction, method of response and basis for interpreting the scores. The interest of the investigator in an attitude scale is not in the response to each item, but in the total score attained by the respondent to all the items. Two criteria are generally used in selecting the items for a scale. The first criterion is that the items must elicit responses that are psychologically related to the attitude being measured; and the second is that the scale differentiates among people who are at different points along the dimension measured. The way in which the attitude scale differentiates among individuals depends on the construction of the scale and the method of scoring.

Three main types of attitude are commonly used: differential scales (Thurston Scale), summated scale (Likert Scale) and cumulative scales (Guttman Scale). Kerlinger (1973) states that a differential scale has each item assigned a scale value indicating the strength of the attitude of an agreement response to that item. Each item differs in scale value and the scaling procedure finds the assigned scale values. In the Thurston type scale, the items on the scale are so selected that the intervals between them are equal. The lower the scale value with the

differential scale, the more positive the attitude.

In the summated scale, there is a set of attitude items which are considered approximately equal in "attitude value" and to which the respondent is asked to answer with degrees of agreement and disagreement (intensity). The scores of the scale are added (or summed or may be averaged) to give the attitude score, placing the respondent on a continuum of the attitude in question.

Usually a seven point scale is used indicating the degree of agreement and disagreement and is viewed as the scales most useful to behavioral research (Kerlinger, 1973). It should be mentioned that the Likert scale has found considerable popular use in training workshops, specifically in management and administrative fields. This scale is frequently used to compare attitudes of various organization sections or levels to various criteria such as styles of leadership, problem solving approaches, communication patterns and administration practices.

The Guttman or cumulative scale consists of a relatively small set of homogeneous items which are unidimensional (measuring only one variable). This scale is also referred to as a cumulative scale due to the relation between items and the total scores of individuals. For example, if four persons are asked three arithmetic questions, the person who would get question "A" correct would

very likely get question "B" and "C" correct. The person who misses question "A" but gets "B" correct is likely also to get question "C" correct. The person who misses question "C" is not likely to get question "A" or "B" correct. Thus, a pattern of responses can be predicted by knowing a person's total score when using a cumulative scale.

Further discussion and reading on these three types of scales may be found in the writings of Likert (1964), Kerlinger (1973), Selltiz et al. (1962) and Severy (1974).

Numerous other approaches to scaling procedures have been developed through the years. They can neither be categorized as differential, summated or cumulative. Two such approaches are the "Q-Sort" and the "Semantic Differential". For the purposes of this paper, and the degree of expertise required of the investigator to use either of these methods, the author will only briefly review these two procedures in passing.

The Q-Sort is similarly constructed as that of the previously mentioned differential or Thurston scale, in that it has equal appearing intervals. The respondent is presented with a large number of statements believed to be relevant to the topic under investigation. The respondent is asked to sort these statements into piles according to the criterion. The statements are usually sorted in nine or eleven piles (Selltiz et al., 1962). The pur-

pose of the Q-Sort is to get a picture of an individual's own view or attitude toward the object being considered. The extent a respondent agrees with the statement is the criteria used for sorting. The patterns into which the statements are sorted constitutes data for analysis. There is also a limit to the number of cards placed in each pile.

The semantic differential (Maguire, 1973; Selltitz et al., 1962; Severy, 1974) is a method of measuring the meaning of an object to an individual. The subject is asked to rate a given concept on a series of usually seven point bipolar rating scales (Appendix "E"). The responses are used to determine whether for the individual the two concepts are alike or different. A profile is made of the meaning of each concept by drawing lines between the points checked on each of the scales for a given concept. Three examples of bipolar word sets used in bipolar scales would be fair-unfair, large-small, and active-passive.

Due to the degree of difficulty in using and constructing the Q-Sort and semantic differential method, it is recommended that the novice evaluator consider using simpler rating scales mentioned previously. Often these scales may be used as is, or they can be adapted for specific use, or modified to the users need. For the beginner, this would be the more practical and wiser approach. Kerlinger (1973) presents a more detailed and comprehensive coverage

of these scaling methods for the person desiring more advanced discussion and knowledge.

5.2 Application

Introduction

There presently exists a wealth of instruments that can be easily adapted for immediate use by the evaluator. The evaluator's energy could then be spent reviewing and selecting available materials most suitable for collecting the needed information. Three good sources which provide a wide range of instruments presently being used in the field are "Measuring Human Behavior" by Lake and Miles (1973), "Instrumentation in Human Relations Training" by Pfeiffer and Heslin (1973) and "Measures of Social Psychological Attitudes" by Robinson and Shaver (1973). Each instrument presented by these authors provides a complete description of the instrument, a complete instrument sample, specific instructions for its administration and use, its scoring procedure and interpretation, as well as a brief critique of its advantage and disadvantages.

Use of Questionnaire

Selecting and Adapting Prepared Instruments

As the potential evaluator is probably well aware, the selection of a questionnaire for use, involves more than scanning a few tests or research studies and randomly

picking a questionnaire which appears suitable for direct application. The following list of suggestions and questions by Selltitz et al. (1963), Miller (1974) and Ten Brink (1974) have been designed to assist the evaluator in this difficult task.

1. Obtain a thorough grasp of the area to be studied as well as a clear understanding of the objectives of the study and the nature of the data needed.
2. Just gather the data needed using the criteria of selection of how it is to be used and analyzed.
3. When readapting questions for use, keep them in the language of the respondents.
4. Pick words and terminology that would have the same meaning to all recipients of the instrument.
5. Avoid long questions.
6. Don't assume respondents have first had factual information on the topic.
7. Establish the frame of reference from which the respondents are being asked to respond.
8. Protect the respondent's ego when using questions of a sensitive nature.
9. If the information desired is of an unpleasant orientation, give the respondents opportunity to express their positive feelings so that they are not put into an unfavorable light when answering the question.
10. Decide whether personal or impersonal questions will obtain a better response.
11. Limit the question to a single idea or reference.
12. Decide whether you need direct questions or indirect questions or an indirect question followed by a direct.

13. Decide also whether the questions required should be open or closed or general or specific.
14. When deciding the type of questionnaire to be used, consider the subject matter, the sample of people to be reached, mode of administration and the kind of analysis and interpretation intended.
15. Avoid questions that have ambiguous wording, biased and leading questions or questions objectionably phrased.

Once the tentative questions have been selected (it would be a good idea to select twice as many questions as actually needed, allowing for a good choice number for the final draft of the questionnaire), the evaluator should begin to organize a sample of the intended questionnaire. The following suggestions have been proposed for this as a check list.

1. Start with easy questions which may interest the respondent in answering. Research has found that questions such as age, marital status and occupation, negatively affect the average respondent.
2. Sequence questions general to specific, or easy to difficult, attempting to avoid conditioned responses of respondents by prior questions.
3. Leave personal questions to later in the questionnaire, avoiding uncontrolled emotional responses which may negatively affect the answering process.
4. Open ended questions usually require most thought and writing; thus, should be kept to a minimum and placed at the end of the questionnaire.
5. Attempt to secure a sequence of questions which will be natural and easy for the respondent to answer (in addition to making sense).
6. Examine the sample page of the instrument noting

the appearance, spacing, ease of response, flow of the items, and how data will be recorded.

7. Individual questions should be examined for "double barrelled" responses, terseness and effect.

Revising Questionnaire

Once the draft of the questionnaire has been formulated; then, it is time to revise it so that any technical defects, biases or blind spots can be eliminated. If the evaluator can find persons who would be familiar with the questionnaire and can use them as a "sounding board" or "helpful critique" it would be most beneficial at this stage.

Pretesting Questionnaire

Prior to final use, all questionnaires should be pretested at least once in order to project tentative respondent's reactions in completing the instrument. If many adjustments are necessary in the pretesting, there should be further pretests until all the necessary changes meet with satisfaction of those concerned. The pretesting should be carried out on a representative sample population duplicating the conditions for administrative as closely as possible. Responses to the pretesting should be done in the form of personal interviews. If care has been taken in constructing the instrument and following the advice of those providing positive criticism; few

personal interviews should be necessary.

Final Editing of Questionnaires

The last step in this series for preparation of data gathering instruments is the final editing of the instrument. Here, the evaluator ensures that the questionnaire has the desired context, order, form, sequence of questions, spacing arrangement, general appearance and specific procedures or instructions for its use. The final area, the instructions for use, should be clear and precise for both the respondent and the person(s) administering the instrument.

The following points should be considered by the evaluator prior to the administration of the instrument to the proposed recipients.

1. Is the administrator of the questionnaire both familiar and experienced in its application?
2. Is the environment conducive to the administration of this instrument?
3. Have provisions been made for follow up of absent recipients?
4. Do the recipients have a clear understanding of the purpose, the scope, the content and the ultimate use of the information desired?
5. If applicable, has confidentiality been assured and restated?
6. Are the recipients of the instrument voluntary or have they been coerced or directed to participate in its application as part of the training workshop design?

Use of Interviewing

Of the three methods of data collection being emphasized in this section, the interview seems to be the most demanding and the most difficult to perform as it utilizes all of the skills of the questionnaire and observation techniques. Questions must be formulated ahead in preparation for the interview (taking into consideration the questions suggested in constructing the questionnaire). Ten Brink (1974) feels that evaluators must utilize all of these skills when observing in order to interpret the responses, as well as providing additional data for their interpretation. The evaluator must also have skills in interviewing to provide a comfortable rapport which will encourage the training respondent to relax and voluntarily share the desired information. Selltiz et al. (1962) describe interviewing as an "art" for which they provide the following suggestions for use by persons inexperienced in its use.

1. Create a friendly atmosphere by using positive brief and casual statements when interviewing a respondent.
2. Proceed to the questions quickly (lowers anxiety and lessens suspicions).
3. The interviewer should only answer legitimate questions and they should be answered honestly.
4. If it is necessary, identify oneself beyond what would be normally required of an interviewer; for example, it is not necessary for a respondent to be familiar with the interviewer's life history, but it may be necessary for the interviewer to

share his experience, qualifications and credentials.

5. The interviewer should always be friendly, or courteous, unbiased, and conversational.
6. The interviewer should avoid showing responses of disapproval or surprise at the answers of the respondent.
7. The interviewer should avoid giving personal opinions.
8. The interview should be kept focussed but friendly.
9. Each question asked by the interviewer should be given as it is worded. Changes open the possibility of bias, misinterpretation and the change of intended meaning.
10. Do not explain questions, as this also leads to potential bias.
11. The questions given should be asked in the order in which they have been written.
12. Responses by the respondent of a "Don't know" category may be honest responses, but they may also be hiding attitudes.
13. The interviewer should try to get verbatim reporting as much as possible.
14. Finally, Selltitz et al. remind the interviewer that bias in interviews often results from the respondents' perception of the interviewer, the interviewer's perception of the respondent or from questions in the interview which pose a threat to either the interviewer or the respondent.

Setting the Environment

Benjamin (1969) advises that as the evaluator begins preparation for the interview; the environment where the interview is to take place should be of prime importance.

One should consider the room in which the interview is to take place. Attempts should be made to ensure that it is non-threatening, free of noise, non-distracting and temperature controlled.

What is normally part of the room should stay that way. The room should be arranged prior to the interview such that both interviewer and interviewee will be comfortable with the arrangement. The ultimate goal of the arrangement of the room is to provide a conducive environment for communication. Benjamim adds that external conditions such as interruptions and interferences have no place in the interview environment and should be avoided at all costs as they are destructive to concentration, rapport, understanding, effective communication and may even block trust and understanding. Telephone calls can be held or monitored by other staff and interruptions can be avoided if planned accordingly.

Starting the Interview

The behavior of the interviewer according to Benjamin should be in a manner which allows the interviewee to freely and thoroughly explore his/her feelings and opinions. The more human an interviewer appears to the interviewee, the greater the chances that s/he will share their true responses, feelings and opinions. In opening the interview with the respondent, a good starting point is for the inter-

viewer to clearly and honestly state what has led to the request for him/her to meet with the interviewer. This action sets a tone of honesty which is intended to model further responses towards openness. The interviewer must try to help the respondent feel that the interviewer is ready and willing to listen to them and what they have to say is felt to be important by those interviewing.

Two cautionary notes should be made at this point. The first is that persons inexperienced in interviewing often become so overly concerned with what they have to say that they find it hard to listen and absorb what is going on in the interview. This makes the interviewee anxious and uncomfortable and produces an adverse effect on the interview as a whole. The second note is that the interviewee initially may not know why he is there to be interviewed. S/he may decide to fight, be uncooperative or may be imagining several reasons why s/he is there. This may lead to confusion and anxiety which blocks the interviewing process. An evaluator should act to clarify such situations prior to starting the interview by being well prepared and open with the interviewee.

The Interview Structure

The interviewer should start all scheduled interviews on time to ensure harmony as well as provide for timely closure. Once an interviewer has gathered the necessary

information, or the interview starts to drag and the behavior of the interviewee is not conducive to continue, the interview should be concluded. A guideline for an interview which provides adequate time for all functions to be performed would be thirty to forty-five minutes. A helpful aid to ending on time is to inform your interviewer of the time perimeters when commencing the interview. As time comes for termination, politely do so. Time is often lost in an interview when the interview is allowed to stray or lose focus. Interviews that run overtime are often found to be suffering from the loss of focus and thus, need to run overtime to get the required data. During closure, the interviewer should not allow new material to be introduced; rather, a second interview should be scheduled. As the interview begins to reach the last ten minutes, the interviewer should begin termination which will allow for some last minute introduction of some important aspects. A good closure statement is "Well our time is just about up."

The style of closure will depend on the interview, the interviewee and the interviewer. Other appropriate closures might be "Thanks for coming in" or "This meeting has been a fruitful one for both of us, I think". Whatever the choice of statement, it should be short and to the point. A further point worth mentioning related to time is

that it is a good policy for the interviewer to allow several minutes between each interview to change from Mr./Ms. "A" to Mr./Ms. "B" and focus on the new interview. It could also be an ideal opportunity to record extra notes or impressions during this period.

Listening and Humanness During the Interview

The most important skill in interviewing is listening. To be most effective the interviewer must provide his/her full attention to the responses of the interviewee. The interviewer cannot be preoccupied as it is his/her responsibility to be totally aware of what is said, how it is being said, as well as being aware of the tone and the accompanying gestures and expressions employed by the interviewee. It is also of extreme importance (in some cases) to be aware of what is not being said or what is being held back.

In relation to humanness, Benjamin feels the interviewer should try to employ as much natural humanness as possible as a model for the interviewee. He feels the interviewer is more than a puppet and a technician within his/her role of interviewer. In his opinion, if the interviewer is remote and cold so goes the interview and interviewee. Similarly, if the interviewer is cautious and wary, so will be the interviewee.

Interview Responses

Five responses may help the new interviewer encourage

the interviewee to talk with greater freedom and verbosity. They would be the acknowledged Mm-hm, the restatement, the clarification statement, the reflective statement and the explanatory statement. Their purposes are as follows:

1. Mm-hm: This usually indicates permissiveness to the interviewer to go on or continue what s/he is saying.
2. Restatement: This serves as an echo allowing the interviewee to hear what s/he has just said in order to help them continue or look deeper.
3. Clarification: The interviewer uses this method to check on his understanding of what the interviewee has just said or to further clarify statements for the interviewer.
4. Reflection: This technique serves as a mirror in which the interviewee can see his/her own feelings and attitudes reflected.
5. Explanation: The interviewer may use this as a lead in structuring the interview. This should be a neutral response to the interviewer's statements and questions.

Recording the Interview

Recording and notetaking are always a problem for the novice interviewer. Some literature warns of all sorts of pitfalls and reactions to open recording of information during an interview. Many feel much of this has been overplayed. Recording and/or taking down of relevant data by the interviewer during the interview is mostly dependent on the circumstances and environment in which it occurs; the openness of what is being recorded and why; and for what

ultimate purpose will the information be used.

If note taking is made an integral part of the interview, and is a "given", yet, done with discretion, there is usually little or no reaction. The interviewer must remember that recording of information helps the interviewer maintain focus, remember, and gather information of greater accuracy. Recording and note taking does not always occur during the interview. Depending on the method of data collection, selected information may be collected for later use during or after the interview, either in full or in partial heading jottings. In some cases, the interviewer may also find it desirable or necessary to use total recall.

Each interviewer will develop his/her own style of interviewing and recording. If the interviewer is relaxed and comfortable with the method, so likely will be the interviewee. If the notes or recordings are intended for research purposes, this should be stated at the onset of the interview by the interviewer so that confidentiality can be dealt with effectively and not interfere with the interview process.

Tape recording and other methods employed such as video-tape should be treated in the same manner as mentioned of recording and note taking. These methods are often useful for assisting both the interviewee and inter-

viewer in recall, research and later learnings. It has been found that, usually after about five minutes, the presence of mechanical and recording devices are usually forgotten by the interviewee. To assist in this process it helps to have these instruments and other methods of recording are done in a semi-obtrusive manner.

Three "Don'ts" of Benjamin should be highlighted in closing this discussion on the interview process. They are:

1. Don't turn recording and note taking into a cross examination.
2. Don't let recording and note taking interfere with the flow of the interview.
3. Don't be secretive about recording or note taking lest it arouse anxiety and curiosity.

Use of Observation Techniques

Very little more can be said about observation techniques than what has already been reviewed. Like questionnaire construction and interviewing, observation is a skill to be improved only by repeated practice. What has been previously said about interviewing can be applied to observation because interviewing skills do include basic observation techniques. Most observation opportunities require some reference to particular interpretation of behavior and verbal and non-verbal responses. Specially designed instruments such as observation report forms, skill rating surveys, post meeting observation forms, friendship charts,

sociograms and post or recall observation forms can be utilized in collecting data as required. Dimock's monographs (1970a, 1970b, 1973 and n.d.) provide for greater detail and explanation than space allows in this methodology. It is highly recommended that the evaluator wishing to increase his/her skill in observation should freely consult these easily read and inexpensive monographs. Sample observation questionnaires can be found in Appendix "F". A further series of books which provide an anthology of classroom observations instruments are entitled "Mirrors for Behavior: Research for Better Schools" by Simon and Boyer (1967).

Use of Scales

Scale construction is an expensive, time consuming adventure which reaches far beyond the perimeters of the intent of this methodology. It is the frequent comment and advice of researchers that before attempting to construct one's own scale, one should check out others in existence for suitability. For this methodology the evaluator will heed this advice with one exception. Mention has been made earlier of the "Likert Scale" being used in needs assessment. Further, Kerlinger (1973), feels that this type of scale is probably one of the most useful ones for the behavioral scientist; therefore, it would seem appropriate to briefly describe the considerations and processes in-

volved in its construction (Selltitz et al. 1962). For an example of a modified "Likert Scale", see Appendix "G".

Likert Scale Construction

1. The investigator assembles a large number of items considered relevant to the attitude being investigated (either clearly favorable or clearly unfavorable).
2. The items are administered to a group of subjects representative of those with whom the questionnaire is to be used. The subjects respond by checking one of the categories of agreement or disagreement.
3. The responses of various items are scored in such a way that a response indicative of the most favorable attitude is given the highest score.
4. The responses must be scored consistently in terms of the attitudinal direction they indicate.
5. Each individual's total score is computed by adding his/her item scores.
6. The responses are analyzed to determine which discriminate most clearly between the high scores and the low scores on the total scale.
7. Internal consistency (each item is related to same general attitude) is determined by eliminating those which do not correlate with the total score/or do not elicit different responses from those who score high or score low on the total test.

Scoring Likert Scales

Since the "Likert Scale" is a cumulative scale, the evaluator would be interested in the total scores of the responses to the items in the scale and would score the items as follows (Likert, 1967; Severy, 1974):

1. Assign a scale value to each response.

2. Since the scale is a five point scale, the items are scored by providing 1-5 points, respective of the predetermined value of the item chosen.
3. Determine all positively worded items and all negatively worded items and score.
4. Add up all the positively worded items and call them "A".
5. Add up all the negatively worded items and multiply by the number of items times the number of categories plus one.
6. Subtract the result of the multiplication from the score on each of the negative items which results in a subtotal called "B". Then, add "B" to the score for positive items "A". The result is the scale score.

The construction and scoring of other scales including the Semantic Differential and the Q-Sort should be left for the expert or advanced researcher. However, if the potential evaluator wishes more information in this area s/he is advised to consult the literature in the field or specifically the references listed in Chapter II.

As a closing note for this section, it should be pointed out that many of the basic principles involved in constructing questionnaires are frequently used in constructing various types of scales.

C H A P T E R V I

6.0 ANALYZING DATA

6.1 Theory

Introduction

Some knowledge and understanding of statistical procedures is necessary for those novice evaluators who wish to attempt to evaluate training workshops. It is not uncommon for persons, specifically training practitioners who are not familiar or experienced in the field of evaluation research, to want to avoid study and involvement in the area of statistics. Unfortunately, for those interested in evaluation, it is unavoidable. Miller (1974) feels that statistical analysis provides the evaluator opportunity to study and to describe precisely, averages, differences and relationships of results obtained from experimentation treatment, or training. From these results, answers to specific questions or hypothesis may be obtained. Statistical analysis may also assist the novice evaluator in gaining new insights or greater familiarity with specific training methods or techniques used in workshops. In addition, statistical procedures can provide data as to the frequency that certain results are achieved or the frequency they occur

in association with other training variables. Finally, the use of statistical procedures may provide information on the causal relationships between variables and the opportunity to test out various hypothesis (Selltitz, 1959).

In the field of research two primary stages of statistical analysis exist. They are descriptive statistics and inferential statistics. For the scope of this study, the emphasis will be placed on descriptive statistics and less on the more sophisticated and advanced inferential statistics.

Descriptive Statistics

Wolf (1974) describes descriptive statistics as a computation of various measures from a set of scores or observations so as to describe or characterize the sample. Warwick and Linninger (1975) describe these as simple statistics or techniques most often used for describing the characteristics of a sample. The following statistical processes can be included in this category: frequency distributions, measures of central tendency (mode, mean, median), graphs (histograms, polygrams, positive and negative skews, scatterplots and normal curves), variability (range, standard deviation and average deviation), variance (error, within group and between group), correlation and regression co-efficients.

A brief description of each of the major statistical procedures follows.

Frequency Distributions

Brown, Amos and Mink (1975) view frequency distributions as a set of ordered scores and their corresponding frequencies. Warwick and Linninger (1975) say that frequency distribution show the characteristics or categories for a variable under consideration, for example, age, sex, income and attitudes together with the number of percentage of sample cases falling into each category. Frequency distributions are most useful for survey data consisting of categories which can't be treated as numbers such as race, marital status, occupations and attitudes. A frequency distribution can be graphically illustrated by a frequency polygon or a histogram. A special case of frequency distribution is the proportion in which only a single characteristic or attribute is expressed as a fraction of the total or 1.00 (for example, .25 and .60). It may also be presented as a percentage which is most common.

Measures of Central Tendency

The three most common measures of central tendency used in descriptive statistics are the mean, the mode and the median. Of these, the mean is the most popularly

used. These measures provide an indication of the "central tendency" of a set of scores.

Mean The mean is the average of all the scores added together and divided by the number of scores. In research, the means of different groups are compared to study their relationships; for example, which group A_1 , A_2 , or A_3 under condition "X" has the greatest productivity? The mean is also greatly influenced by extreme scores (Brown, Amos and Mink, 1975; Kerlinger, 1973) and when these extreme scores are plotted on a curve, the curve is said to be a "skewed curve". A positively skewed curve (Figure 13) has the majority of the scores located at the left side of the curve, with the tail of the curve extending to the extreme right where a few scores are indicated.

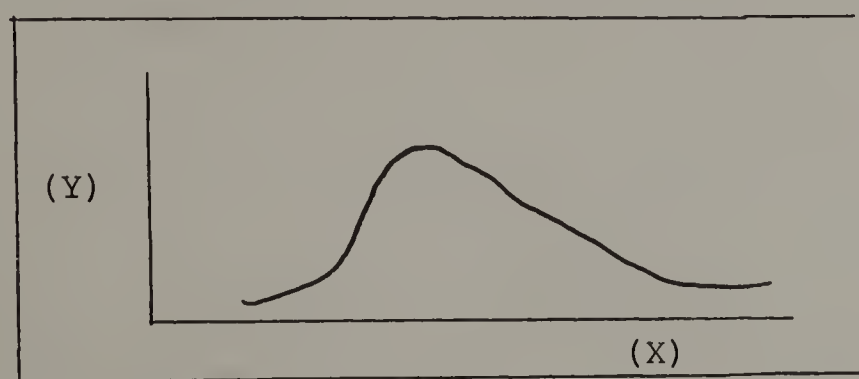


Figure 13. Positive skewed curve.

A negative skewed curve (Figure 14) has the scores massed at the right end of the curve with only a few scores located on the left.

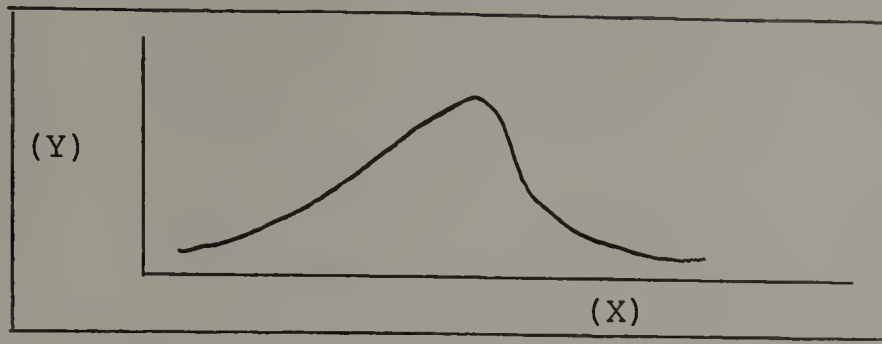


Figure 14. Negatively skewed curve.

Median. The median or the middle score as it is sometimes referred to, is a point on a scale of measurement above which are exactly half of the cases and below are the other half of the cases. For example, cases 4, 6, 8, 10 and 12. The median is 8, but it cannot be obtained until the cases or measures are placed in ascending order from the lowest to the highest. By counting up the scale, the point is selected above and below which there are an equal number of cases. If an even number of cases exist, average the middle two scores (Brown, Amos and Mink, 1975).

Mode The mode is the value or score that occurs with the most frequency, for example, among the seven scores 2, 3, 2, 7, 2, 8, 2 "2" is the mode.

In training and research the mode is often used as a simple inspectional average to show quickly the centre of concentration of a frequency distribution. It is generally used only with a large number of cases, but if the training cases are small in number, several scores will have the same frequency. The mode is often used in preference to a mean or median as it shows what is the number occurring most frequently. Modality is also used to describe the shape of some distributions; for example, a histogram or frequency distribution which has two peaks is said to have a bimodal distribution. More than two is called multimodal (Brown, Amos and Mink, 1975).

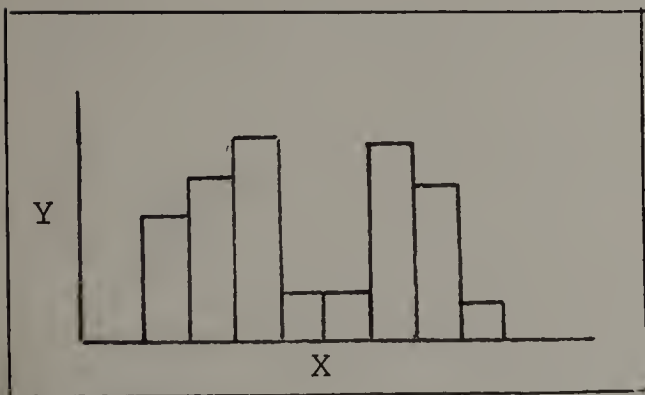


Figure 15. Bimodal histogram graph.

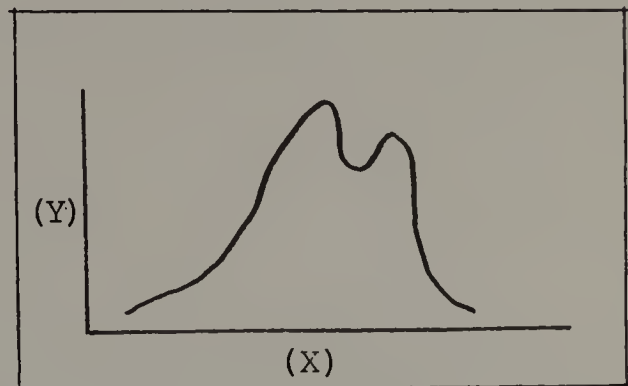


Figure 16. Bimodal polygram graph.

Graphs

Graphs come in many shapes and forms. Kerlinger (1973) describes a graph as a two dimensional representation of a relation or relationship. It pictorially exhibits its sets of ordered pairs in a way no other method can.

Graphs can demonstrate if the relationship is positive, negative or linear. Some samples of graphs are histograms, polygrams, and scatterplots. Graphs can be very useful to workshop participants as they are more easily interpreted by the participants, thus, the data is more readily understood to everyone's benefit.

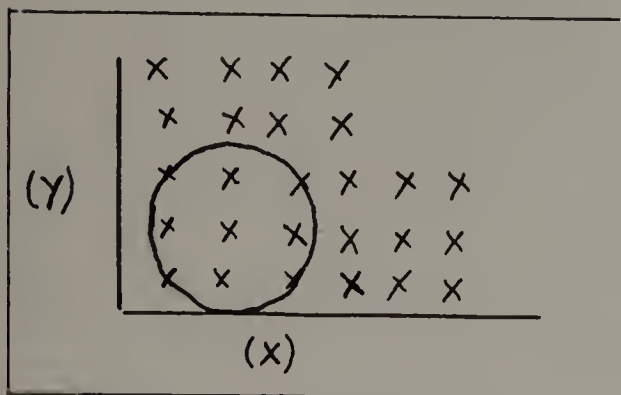


Figure 17. Scatterplot graph.

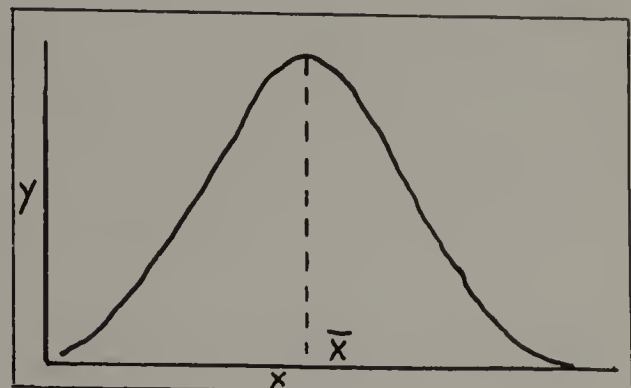


Figure 18. Normal curve graph.

Variability

Another way of describing a group is to have some index of how much variability exists. Some common measures of variability are the range, the standard deviation, the average deviation and the quartile deviation (Brown, Amos and Mink, 1975; Selltitz et al., 1962).

Kerlinger (1973) describes the range as the difference between the highest and lowest measures of a set of measures. As Brown, Amos and Mink (1975) describe it:

"the range of a set of scores is the distance between midpoints of the lowest and highest scores". The range

is often used in tests of statistical significance in analysis of small groups of twenty or less. The size of the range will depend on the size of the sample.

Another index of variability is the standard deviation. The standard deviation is a measure of how much a person or group (or training group) deviates from the group mean. Groups that would deviate very little from each other would have a small standard deviation, while groups that differ greatly would have a large standard deviation. Thus, two groups subjected to specific training variables could be compared by using the size of their standard deviations. The standard deviation is a more stable measure of variability than the range, as the range only considers the highest and the lowest scores, while the standard deviation considers a number of scores (Brown, Amos and Mink, 1975). Selltiz et al. (1962) speak of the average deviation and the quartile deviation. The average deviation like the standard deviation, measures the average distance of individuals from the group mean. The quartile deviation shows the points within which the central half of the cases fall.

Variance

The variance is the square of the standard deviation and is used in many kinds of statistical analysis.

In the training research, the means of different training groups are compared to study the relations; for example, which group, A_1 , A_2 or A_3 , under directive leadership, has the greatest productivity (Brown, Amos and Mink, 1975; Kerlinger 1973). Kerlinger states that when using ordinary scores, "variance is the measure of dispersion of a set of scores or how much are the scores spread out". It also describes how much the scores are different from each other. There are three commonly used types of variance: between group variance, within group variance, and error variance. Between group variance in a workshop is due to differences between groups of individuals and is often caused by active manipulation of the independent or training variable by the experimenters and trainers. Within group variance is due to differences between individuals within their training groups. The mean would be the average variations within the training group. Error variance is due to the fluctuation of varying measures due to chance.

Kerlinger adds that variance is a summary of whole sets of scores. Studying sets of numbers are too unwieldy; therefore, it is necessary to reduce the sets in two ways. The first way is to calculate the averages or measures of central tendency and the second way is to calculate the measures of variability (or variance).

Correlation and Regression Coefficients

Brown, Amos and Mink (1975) state that many variables or events in nature are related to each other. As the sun rises, the day warms up; as children age, their thinking is more complex. Such relationships are called correlations. Correlations can be classified as being "positive" or being "negative". If the increase in one variable coincides with an increase in another variable, the variables are said to have a "positive correlation". Other variables are said to have a "negative correlation" when an increase in one coincides with a decrease in another. For example, the greater the interpersonal concern of individuals for each other, the lower their work productivity. When a high correlation between two variables exists, it is possible to predict the values of one variable from those of the other. The most popular numerical measure of correlation is the "product moment correlation coefficient" which is symbolized by the letter "r". The letter "r" measures the degree to which the relationship between the two variables can be represented by a straight line. The value of "r" ranges from 1.00 for a "perfect positive" linear relationship (straight line) through 0.00 for no linear relationship to -1.00 for a "perfect negative" linear relationship. Scattergrams are used most often to represent the positive or negative associa-

tions between two variables (Warwick and Linninger, 1975). The minus sign (-) indicates a negative correlation or an inverse relationship and the plus sign (+) indicates a positive correlation or a direct relationship.

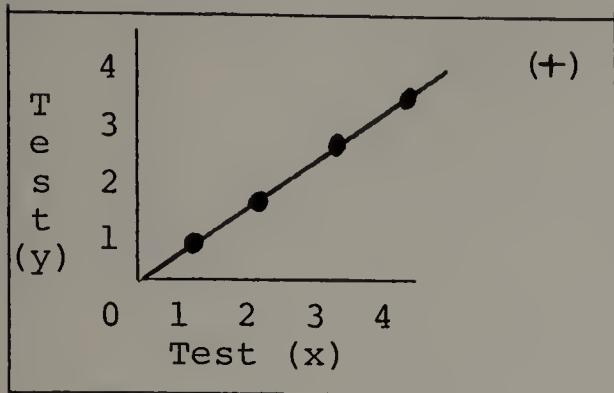


Figure 19. Positive (+)
Correlation
Scattergram

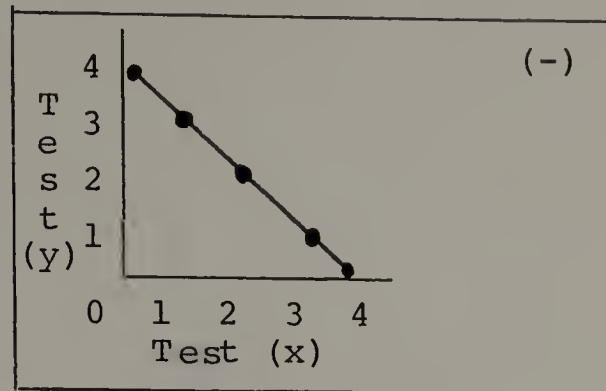


Figure 20. Negative (-)
Correlation
Scattergram

The "regression coefficient" is a "cousin" to the "coefficient of correlation". The primary difference is that the "regression coefficient" calls for the form and the relationship to be specified, while the "correlation coefficient" deals only with the degree of association.

Inferential Statistics

Inferential statistics are more complicated and sophisticated processes to apply as they involve learning things about unseen persons or in predicting the future. They infer unknown data from known data. In as much as these statistical procedures go beyond the scope of this

methodology, the author will merely mention a few of these statistical procedures in passing. They are "significance testing", "analysis of variance", "chi-square" and the "t-test".

Significance Testing

The most common method of significance testing is called testing the "null hypothesis" (Brown, Amos and Mink, 1975; Kerlinger, 1973; Selltiz et al., 1962; and Warwick and Linninger, 1975).

This method begins with the hypothesis that no difference exists between the population groups being analyzed and that any differences observed in the same arise from chance variations. To demonstrate that there are no significant differences, it is necessary to reject the "null hypothesis". The statistical method used to carry out such a test makes use of the concepts of "confidence levels" and "confidence intervals". If the differences observed in the data are greater than that which could be expected to occur by chance, the "null hypothesis" would be rejected and the differences would be assumed to be real or statistically significant.

The "confidence level" refers to the probability that a given statement is correct (the chances one is willing to take on the estimate). The "confidence interval" is the range around the population value within

which the estimates from the samples can be expected to be at a given confidence level.

Analysis of Variance Analysis of variance or "ANOVA", according to Brown, Amos and Mink (1975), is used primarily to test the significance of differences between the means of groups. Kerlinger (1973) sees it as a method of identifying, breaking down, and testing for statistically significant variances that come from different sources of variation.

Chi-square The simple and most useful statistical test is the chi-square test. Chi-square answers the question: Is the fact that the observed is different from what the evaluator expected, more likely due to chance or does it more likely represent the actual population differences? Whether the frequencies are significantly different than expected is determined by the test chi-square.

T-test The t-test, a third common statistical procedure of inferential statistics, simply tests for the differences between means of groups that are statistically significant (Kerlinger, 1973).

Reliability and Validity

Reliability of a test or an instrument or statistical procedure is indicated when the same results are obtained time after time using the same instrument, test or statistical procedure under the same conditions. The three most common methods of measuring reliability are "test-retest", "split half" and "equivalent forms".

"Test-retest" reliability is established by correlating the scores on the same test given at two different times.

"Equivalent forms" reliability is established by the correlation between equivalent forms of the same test given at the same time.

"Split-half" reliability is established by correlating the scores on two halves of the same test given at the same time.

Validity is expressed as the extent to which a test measures what it is supposed to measure. There are four types of commonly used validity: predictive, concurrent, content and construct.

As validity testing is a process that should be left to the more advanced and knowledgeable researcher, this paper will not discuss its use, but will suggest that the reader who desires more information be referred to such

texts written by Kerlinger (1973), Selltiz et al. (1962) and Warrick and Linninger (1975).

The novice evaluator may wish to refer to other specific books that have been written on the specific use statistical procedures in data analysis. Among the more recent authors are Anastasi (1968), Brown, Amos and Mink (1975), Kerlinger (1973), Lehmann and Mehrens (1971), Miller (1974), Popham (1974), Selltiz et al. (1962), Severy (1974), Ten Brink (1974) and Warwick and Linninger (1975).

6.2 Application

Descriptive Statistics

Frequency Distributions

Frequency distributions can be useful in evaluating training workshops by providing data for the evaluator as to how many participants responded in such a manner when "X" variable was applied compared to the number of participants who responded in a different manner when "Y" variable was applied. For example, a frequency distribution could be used to determine the number of students out of 40 students who responded in a hostile manner towards a group leader when s/he used a directive style of leadership compared to the number of students who

responded in a friendly manner when s/he used a high relationship style of leadership.

Measures of Central Tendency

Mean In a training workshop, the mean is often used to determine the average response or score a training group achieves when a treatment or training variable has been applied to the group. Each participant's score can be compared with the mean, or the mean of the group can be compared with the mean of other groups who have received similar training or treatment under the same conditions. For example, if workshop training had been taking place to increase workshop participants skills in problem solving, a group mean or average could be taken of the ability of participants to solve a given problem. The mean score could be obtained from another group receiving similar training and who were given the same problem. The two mean scores could then be compared to determine which group had the overall best average performance.

Median In a training workshop, the median can be used to represent the average score of a distribution resulting from a treatment variable so that the effect of one or more extreme scores can be minimized.

The median can be useful in a training environment by indicating how many extremely high scoring scores are

there compared to how many extremely low scoring scores there are in response to the training being applied. For example, if workshop participants were being trained to become more aware of disruptive behavior and they were being scored on a scale of 1 to 10 with 10 being highly aware and 1 being unaware, and if three persons scored 1, two persons scored 3 and three persons scored 5, and 5 persons scored 10 the mean score would be $(3 \times 1, 2 \times 3, 3 \times 5 \text{ and } 5 \times 10 \div 13)$ approximately 6, or (5.69). The median score would indicate that on this distribution that there were 5 cases that scored above 6 and that there were 8 cases that scored below 6. This would indicate a smaller number of people responding well to the training than responded poorly. Thus, an indication for further investigation would be for the evaluator to determine by how much, and why did the majority respond poorly.

Mode The mode could be used in training workshops to indicate participants preferences to specific training methods or content. It can also be used to determine the most frequently reoccurring score on a test. In the example cited for the median, the participants rates of scores were: 3 scored 1, 2 scored 3, 3 scored 5 and 5 scored 10. The mode of this distribution would be 10.

Variability

Range The range may be used in a training workshop environment to indicate statistical significance. For instance, if a group leader wished to determine how different or how similar workshop participants responded to her/him using a lecture format for leadership training, compared to an experiential type format for leadership training; the range would provide the leader with that information. The difference between the highest score of participants responses and the lowest score would be an index of their responses of similarity or difference.

Standard deviation The standard deviation is a descriptive statistical measure of variability that can be used to compare one group with another. It measures the average distance of individuals from the group mean.

Correlation Coefficients

Positive and negative correlations can be quickly and readily observed as demonstrated in Figure 19 and 20 when a scattergram is used. An example which demonstrates the use of a correlation coefficient in training would be an investigation as to whether the leadership style of the group leader, positively or negatively effects his/her group productivity. If the leader is democratic and

group productivity falls off, it is said to be negatively correlated. If productivity should increase when his/her style is democratic; then, it is said to be positively correlated.

Inferential Statistics

Introduction

At the onset of this research study it was indicated that the prime interest of this chapter would be to provide the novice evaluator with a basic knowledge of elementary statistical procedures. This has been accomplished. It was also stated that "inferential statistical analysis" concepts would be left for the more advanced researcher; therefore, the description of the application of inferential statistical procedures will be limited to the three following examples.

Analysis of Variance An example of use or application of this procedure would be, when workshop leaders wish to determine which of a number of methods of decision making was most effective in a workshop situation when used by men or when used by women. An analysis of variance test could be applied to see if the separate means of the two groups differ significantly from each other.

Chi-square A chi-square test could be used in a workshop situation where behavior samples were taken of the participants in an attempt to study individual differences between male and female decision-making behavior. The results of one observation could show that male and female behavior were about the same, but males were more aggressive. But, the recorded experience of group leaders indicated that female decision-making behavior was more aggressive. In this instance a chi-square test could be applied to determine the significance of deviations or leader observations from the expected results.

T-test A t-test can be used in a situation where the leadership style of the participants involved is measured before and after training to determine if the training workshop content influenced their leadership styles in any way. A mean score of the workshop participants on the pretest is taken as well as a mean score of the participants on the posttest situation. The mean of the pretest score is subtracted from the mean of the posttest score. In this way a "mean difference score" is obtained and submitted to a t-test to ensure or test for statistical significance.

Specific references to training workshops have not been made in describing every statistical procedures mentioned as it was felt to be unnecessary as the under-

lying theme of this study is directed with the training workshop in mind. Furthermore, as so many of the methods of statistical analysis described here were very elementary in nature, it is assumed that the novice evaluator will have little difficulty in applying and relating them to the training workshop field.

C H A P T E R V I I

7.0 REPORTING OUTCOMES AND GIVING FEEDBACK

7.1 Theory

The Report

Reporting back information gathered from an evaluation design is often viewed as a matter of routine procedure. This stage of the evaluation methodology should be viewed as critically as any other stage. As great care and consideration should be employed in selecting, preparing and planning the report as should be taken when determining the goals or selecting the appropriate research design. If an organization, sponsor or program participant is going to accept the information collected and put it to appropriate and positive use; then, the data must be presented in a clear, understandable and organized fashion, taking into consideration the needs of those being reported to, as well as being sensitive to their situation and feelings.

The report should generally include where the goals and concerns originate, the way goals and items were generated, how the items were refined, the choice of subjects, the reliability and validity of the work; a brief description of the actual utilization of instruments,

scales, data gathering devices in most meaningful areas and the final results obtained (Severny, 1974). Weiss (1972) cautions the evaluator/investigator that care must be taken to ensure that the appropriate information should be shared with the appropriate persons so that it can be put to the best of use and insure the confidentiality of the process, as well as the investigator's credibility. She also suggests that by presenting useful comparisons, the data has a better chance of being understood and later utilized. Weiss says that the data must be prepared and reported at the time that the decisions and data are needed. If not, the report will join the ranks of previous reports, on the shelf and gathering dust. The investigator, no matter how intelligent, verbose and capable, must be able to communicate the evaluation outcomes in a language that can be understood by those being reported to. If at all possible Weiss adds, report the data personally for the greatest impact effect. Selltitz et al. (1962) report that from their experience in forming reports, the client or sponsor receiving the report will perceive the scope of the report with greater clarity if samples of typical questions and answers used in the evaluation were included in the report. See "Appendix H" for sample tables used in the reporting and feedback of data. Other useful information (within the perimeters of confidentiality)

which help sponsors or participants get a better feeling for the evaluation would be: "Who were the subjects studied, how many were in the sample, how long was the study, how were interviewers or staff members selected and trained, what techniques were used and what were their levels of confidence?"

In forming the evaluation report Selltiz et al. suggest the following thoughts should be considered:

1. The evaluator/investigator should consider what is to be reported, how, and how is it related.
2. When preparing the report outline, check for things omitted.
3. Report all data in understandable terms and diagrams (graphs, polygrams, distributions, and tables for example).
4. It is important to label the diagrams clearly, with a brief title mentioning the subject matter.
5. All explanations and qualifications should be footnoted.

The evaluator on completion of the first draft of his/her report should ask the following questions of the report:

1. Is it clear and grammatically correct?
2. Does it say what is intended?
3. Could any points be expressed more simply?
4. Does it fit together?

The final point Selltiz et al. suggest is to have the report proof-read and friendly criticized.

Communicating Results

In reporting data, Anastasi (1968) stresses that it is always desirable to take into account the characteristics of those persons receiving the information. The question of who receives what information and how much always should be a priority concern for the evaluator. Safeguards of confidentiality and complete clarity of information should be observed when sharing information with a third party to prevent misinterpretation. A third party should not be the recipient of information unless under the written permission of the first and second party. On occasion the sharing of information to a third party may be at the discretion of the investigator/evaluator if it is intended to help some participant or trainer who is not capable in helping themselves in the opinion of the investigator (and other professionals or authorities present). When sharing information and data, the investigator must always be aware and sensitive to the reactions of the recipient, especially if s/he are learning about their own assets and shortcomings. A general guide for the investigator is to share only that information or data requested, thus preventing information overload and unpredicted reactions.

Professional Ethics

Anastasi (1968) further suggests that those persons

involved in collecting data from evaluations and research, should always regard the information with the highest degree of confidentiality and ethical standards, while respecting the participants and their desires. A helpful guide for the novice, she suggests, is the "Ethical Standards of Psychologists" (Anatasi, 1968, pp. 627-635).

Feedback

A form of reporting, but much less formal and more dynamic than the written report for providing information to recipients is a method known as "feedback". Brown (1972) states that in providing "feedback" (data reporting) both the context of communications (diagnosis and the relationship among communicators seem to improve through the mutual sharing of information. People become more personal and less threatened; thus, the data is freer and undefended. Also resulting from freer data is a clear and better understanding by the organization of the collected data providing for their better analysis and interpretation. In addition, feedback often results in increased involvement by the participants and trainees in their own organization.

Schmuck and Miles (1971) view feedback as having three main components: data, meetings and process analysis. The data component involves a process of sharing the data

outcomes, which usually collaborates previous feelings or disconfirms them; thus, the disconfirmation process is a powerful force for change. The sharing of data often stimulates questions of reappraisal and focuses on actual problems. As a result, change goals become focused and people become highly motivated. The second component of feedback is the meeting. Meetings serve to bring various groups (both fringe and direct) together which generally stimulates talking, sharing of ideas and collaboration. Pressures toward conformity grow. Schmuck and Miles say that attitudes towards tasks become more favorable as well as pressures to clarify one's position in relation to the goals also grows. Generally, people tend to acknowledge data, rather than avoid it, but there is a danger that group pressures may force people to go along and conform unwillingly to data they do not own.

The final component of feedback mentioned by Schmuck and Miles (1971) is process analysis. During the process of data study, participants/trainees have an opportunity to realistically study in detail the processes of what has been happening during the program and training, as well as their involvement in it. The trainees or participants have an opportunity to look at their own behaviors in relation to the goals for achievement and the desired outcomes seen by all parties.

The participants/trainees have opportunity to suggest to the investigators alternative processes or goals and new ways of accomplishing things. This process allows new issues to be raised and the opportunity to reevaluate the priorities and consider new emerging ones.

In summary, there is general consensus by those evaluators active in the evaluation field that the content, quality and format of the report, as well as the quality of the presentation of the report are factors which determine the influence the report will have on those sponsoring it. Similarly, further evaluation and research can be dependent on these factors. Just as the first steps in the evaluation methodology, "Needs assessment and goal clarification" was key in launching the evaluation process, "Reporting outcomes and giving feedback" rates equal importance when ensuring that positive actions will follow as a result of the report's presentation.

7.2 Application

Providing Feedback

As providing helpful feedback is critical in the reporting process, the following are some general guidelines for giving feedback which can be helpful to the potential evaluator (N.T.L., 1971, 27-28):

1. It is descriptive rather than evaluative. By describing one's own reaction, it leaves the individual free to use it or to use it as he sees fit. By avoiding evaluative language, it reduces the need for the individual to react defensively.
2. It is specific rather than general. To be told that one is "dominating" will probably not be as useful as to be told that "just now when we were deciding the issue you did not listen to what others said and I felt forced to accept your arguments or face attack from you."
3. It takes into account the needs of both the receiver and giver of feedback. Feedback can be destructive when it serves only our own needs and fails to consider the needs of the person on the receiving end.
4. It is directed toward behavior which the receiver can do something about. Frustration is only increased when a person is reminded of some shortcoming over which he has no control.
5. It is solicited, rather than imposed. Feedback is most useful when the receiver himself has formulated the kind of question which those observing him can answer.
6. It is well timed. In general, feedback is most useful at the earliest opportunity after the given behavior (depending, of course, on the person's readiness to hear it, and support available from others, etc.).
7. It is checked to insure clear communication. One way of doing this is to have the receiver try to rephrase the feedback he has received to see if it corresponds to what the sender had in mind.
8. When feedback is given in a training group, both giver and receiver have opportunity to check with others in the group the accuracy of the feedback. Is this one man's impression shared by others?

For further views of reporting and feedback processes see Schmuck and Runkel (1972), and Worthen and Sanders (1973).

Writing the Report

The following guidelines have been suggested by Selltiz et al. (1962), as a guide for the researcher who must write a report and wishes it to be as worthy as the research to be reported.

1. The first recommendation suggested is to make adequate time available so that the report is not rushed and time can be spent in putting it together in a logical way.
2. The report should present a clear communication of results as they occurred and the measure utilized in obtaining them.
3. The report should be directed to a specific audience previously agreed upon, however, very often this will be the sponsor.
4. When the report is written consideration should be given as to what audience is being reported to. Will it be the program participants or the administrators or both?
5. The following two questions should be asked by the report writer when formulating the report:
 - I. What does the intended audience want or need to know about the study?
 - II. How can this information be best presented?
6. The report should contain: (a) a statement of the problem studied or the situation raising concern, (b) a description of the procedures used in the study design, a description of the nature of the sample chosen, how it was

selected, a description of the data collection techniques and a description how the independent variables were manipulated (if there was an experiment), (c) a statement of outcomes or results, and (d) a statement of the implications drawn from the outcomes or results.

7. The report should contain illustrations, diagrams and tables to help emphasize and clarify procedures used and outcomes which resulted.

It is felt by most evaluators that the report should be as brief as possible (not sacrificing impact and understanding), precise, and in closure contain a short organized summary of the proceedings and their outcomes for easy reference.

Severy (1974) adds that in his opinion, the report should also include:

1. A description where the concern originated.
2. A way that the items studied were generated.
3. How the items were refined.
4. How the participants were chosen.
5. A description or statement on the reliability and validity on the work done in the evaluation.

The opinions and suggestions of these authors serve to emphasize the importance that should be placed on the reporting and providing of feedback of evaluation outcomes.

Reporting and Feedback in Training Workshops

The information which has been covered thus far in this chapter is primarily directed towards reporting in

a formalized way to a formal sponsor or organization. In training workshops much of the reporting and feedback occurs informally and spontaneously as the training proceeds, although much of what has already been said is applicable to the training situation. The following considerations will be helpful to the evaluator who is requested by a training group to share his/her expert views and opinions.

1. Information shared with the group in training should be in a supportive and helpful manner.
2. Information which is shared should be factual and not speculative, biased or detrimental to the future functioning of the group.
3. The evaluator, before sharing data, should do it in cooperation with those directing the training.
4. The evaluator should have a good feeling for the needs of the group before s/he shares any information with them.
5. The evaluator should remember that his/her role is to only suggest alternatives and not that of the decision maker.
6. Anyone who is in a position to evaluate is often seen by training participants as an outsider and often seen as a threat to those in training.
7. Often, when an evaluator realizes this, with good intentions, he moves to rectify the situation. Then, unless he is very careful, the danger exists that s/he may be co-opted by the system, or s/he begins to identify with the training program goals. S/he starts to form personal relationships with the training participants and develops specific biases as s/he gets caught up with the routine daily operations and becomes ineffective as an evaluator.

C H A P T E R V I I I

8.0 EVALUATING THE EVALUATION

8.1 Theory

Introduction

A procedure that is seldom suggested or written about is the process of evaluating the evaluation process that has just taken place. It is felt that the inclusion of this procedure provides for a crucial check point in time where the investigator and sponsor can determine if the evaluation has achieved what was desired. This slows down the decision-making process allowing for reconsideration of all essential steps and data. Harrison (1968), presents a series of questions that can be used in this process. Some of these are:

1. Was there a flow to the evaluation design which provided for a natural sequence and adequate bridging from one step to the next?
2. Did the design create a climate for learning?
3. Was the evaluation designed in response to the real learning needs of the person(s) for which it was intended?
4. How will this be done?
5. Is there a plan for follow-up and back home support?
6. Did the evaluation accomplish what it set out to do?

Further Checkpoints in Evaluating Evaluation Designs

Benedict (1971) lists questions as checkpoints in evaluating an evaluation design. Some would be:

1. Is the evaluation providing data for the decision making needs relative to the identified enterprise?
2. Is the evaluation efficient, complete and focussed?
3. Can the evaluation be evaluated in terms of its component parts, goal identification, evaluation methodology, research design and so on?

Worthen and Sanders (1973) present the thinking of Guba and Stufflebeam (1968) and Stufflebeam et al. (1971) in the following criteria for judging evaluation studies.

1. How close does the data obtained relate to the objectives of the study?
2. What priorities are placed on the information collected and the components presently evaluated?
3. How comprehensive was the design of the evaluation study?
4. Does the investigator have credibility with his audiences? Will they act on his/her recommendations?
5. Will the reports be finalized and available when needed?
6. What are the cost benefits of the study?

Belasco and Trice (1969) further add to the list of questions with the following:

1. Was there development of adequate yardsticks which were reliable and relevant and which could measure progress toward achieving stated objectives?

2. Were the yardsticks aptly applied in relation to the time span implied by the goals?
3. Were there established at least two sources of information to evaluate the situation in terms of the goals to be evaluated (minimizing bias)?
4. Was there specification and examination of those underlying personality and situational factors which help explain the changes identified.

Finally, Schmuck and Runkel (1972), to add to this list ask these questions:

1. Was the evaluator/investigator noncoercive, placing no demands on participants and trainees to change?
2. Did the evaluator/investigator consider the feelings of all involved?
3. Were behavior descriptions objective and non-judgmental?
4. Was the report and feedback timely in that the data was presented close to the time of the events taking place.
5. Was the feedback and reporting focused on things that could be changed or done differently?

8.2 Application

Evaluating the Evaluation

The basic point to be stressed at this phase of the evaluation methodology is to avoid rushing into decision making processes before both the sponsor and the evaluators have taken a close look at the results of their evaluation efforts to see if the data now on hand is what was

originally intended or needed at the onset. The process is relatively simple and much like comparing experimental outcomes to the original hypothesis. If the degree of achievement or satisfaction of the outcomes of the evaluation satisfactorily fulfills the original need or goal identified in the "goal clarification or need assessment" stage, then, the evaluator and sponsor are ready to move forward into the next phase of decision making. But, if the original goals or needs are not satisfactorily fulfilled or the information obtained is not adequate enough to be able to determine the degree of satisfaction or achievement, then, either the evaluation process must be checked to see why it did not provide the required information, or the effectiveness of the programming or training is put into question. An "Evaluation Design Checklist" is included in "Appendix 1" to assist the evaluator in identifying some areas in the evaluation design which may require further modification or attention.

Further investigation of the design, the method of application and the outcomes are necessary in order to determine the cause of the unanticipated results.

Finally, a decision must be made by the evaluator and the sponsor to continue or not. If the information (although not up to expectations) received is still applicable and reliable enough to continue the decision will be made to advance.

C H A P T E R I X

9.0 ACTION DELIBERATION AND REDESIGN

9.1 Theory

Decision Making

After the evaluator has taken into consideration the quality of the completed evaluation, and the validity and reliability of the resulting information, s/he is confronted with making or recommending some action decisions. Schmuck and Runkel (1972) suggest the following considerations when forecasting consequences of intended actions.

1. Do the probabilities for success outweigh the needed labor required for implementation?
2. Try to imagine all possible things that might go wrong--simulate them, role play them to get reactions and feedback using outside people.
3. Involve those directly effected, in critiquing plans.
4. Anticipate the barriers from the environmental sources.

Action plans can be identified and operationalized through brainstorming ideas, listing materials and resources, placing ideas and actions in time sequence, estimating dates for starting actions, planning for periodic evaluation of action effectiveness as they are implemented, and preparing to revise action sequences as

they unfold in relation to available information.

Beck, Raynor, Raynor and Schraggle (1975) call this phase "Action Consulting". According to these authors there are three phases: generating alternatives, deliberation, and decision-making. Generating alternatives involves the evaluator/investigator and others, (the participants, trainees or sponsors) generating alternatives from the data obtained from the evaluation. The data is used to identify alternative goals, objectives, and experiences. Deliberation involves these same people in weighing the anticipated consequences of the alternatives generated. The decision-making phase involves the participants/trainees, sponsors or client system choosing the best action based on their deliberations.

The evaluator concerned with making quality decisions must expend all effort to gather all available resources and information which will provide sufficient data for identifying the avenues or choices available. Next, the evaluator must attempt to foresee or predict the paths of least resistance and highest payoff in achieving what is to be desired. Finally, the evaluator must set up his/her possible choices in sequence or priority; then, test each for satisfaction. This process is known as generating alternatives. It is the most critical area in the decision making process due to the

fact that the more alternative actions one generates, the greater the chances are for success. The inexperienced or novice evaluator must resist the temptation of jumping too quickly on a few tentative action choices, as this often results in an "all or nothing" situation where the evaluator usually ends up with "nothing".

A method which prevents an "all or nothing" situation from occurring by generating a more than adequate supply of alternatives, (and has been found very successful), is called "brainstorming". Generally, the theoretical steps for operationalizing "brainstorming" procedures are as follows:

1. Think of as many alternative actions or choices as possible related to the situation or problem that is being presently considered.
2. Without discussion, list these alternatives on a piece of paper in the form they are given.
3. Continue this listing until all alternatives and suggestions have been exhausted.
4. Allow each alternative to be explained or clarified by those who suggested it. All other persons listen intently to understand the essence of each suggestion.
5. Questions may be asked for further explanation, but all judgmental, biased or prejudicial remarks are to be withheld.
6. All suggestions of choices or alternative actions are to be examined by everyone to eliminate repetitions, overlapping, and non-applicable choices and alternatives.
7. Choices which are similar, can be combined.

8. These choices and alternatives are written clearly and precisely to indicate their meaning and intent.
9. All involved may suggest adjustments or modifications to any item suggested, provided it receives the ratification of the majority present.
10. Priority choices are made for initial action by consensus if possible (majority vote if all else fails).
11. The final priority list suggests the possible alternative action choices to be tested in finding the solution.

Models for Decision Making

Delphi Method

The "Delphi method" of determining needs has been thoroughly explained in Chapter II. It is suggested that this method can be adapted for use for decision making by projecting tentative outcomes or their inhibiting consequences for any one action being contemplated by the evaluator. The same steps may be followed for operationalizing the "Delphia method" as earlier explained, but replace the projection of goals and needs with that of outcomes and blocks or inhibiting consequences of potential actions.

Cost Analysis and Projection

Another method of decision making which attempts to foresee the future and assists in the process of choosing

alternative actions, is cost analysis. Haller (1974) says that for every choice there has to be information as to the cost of the action resulting from the decision. The evaluator/investigator and sponsor must determine if the action is worth the expenditure, if the budget can afford the action, or is there a less costly way of doing it without greatly losing the value outcome? Many evaluators have no expertise or knowledge in projecting evaluation costs of programming, but it can be a helpful skill if the investigator has this knowledge and can assist those sponsoring the evaluation with cost based decisions. There is no one set of procedures that can be recommended for use that can be readily and generally applied for a cost analysis in all cases.

Haller presents some general considerations related to costs. He says that the concept of cost is useful because it provides a criterion for choosing among available alternatives. Further he says an estimate of the costs of any choice is an estimate of the benefits forgone as a consequence of that choice.

Costs are benefits lost and are inextricably tied to decisions. In order to accurately estimate the cost of one decision, the evaluator is placed in a position of having to define the consequences of at least one other alternative.

On measuring costs, Haller says that cost analysis consists of essentially identifying, measuring and evaluating alternatives but not necessarily in dollars. In determining the cost of an evaluation program, the investigator should develop a list of required resources to operate the program. S/he could then describe some alternative uses for the same resources. In addition, he may want to estimate the value of the alternatives listed, or s/he may want to attached a dollar value or expenditure figure to lists of required resources needed to operationalize the program.

Haller notes that some programs can be measured in dollars and others cannot, but when programs can be measured in dollars, dollars provide a convenient, generalizable and comparable estimate of the operational costs of a program.

When evaluating any evaluation program, the following resources have to be considered: time, space, equipment and supplies. The cost and availability of time, space, equipment and supplies has to be determined before any estimate of program costs can be made. In addition, further costs must be included for research and development (resources required to develop the program to the stage where it can be introduced into the system), investment (costs necessary to impliment the program, (for

example, equipment and staff training) and operating costs (recurring costs required to operate the program over a period of time). Haller presents a useful structure for costing program evaluation as follows:

<u>Research and Development</u>	<u>Investment Costs</u>	<u>Operating Costs</u>
1.1 Time	2.1 Time	3.1 Time
1.2 Space	2.2 Space	3.2 Space
1.3 Equipment	2.3 Equipment	3.3 Equipment
1.4 Supplies	2.4 Supplies	3.4 Supplies

In summary of cost analysis, Haller suggests:

1. A cooperative involvement between the evaluator/investigator and the sponsoring client system so there will be a relationship as a base for clear communication for knowledge of future behaviors of factors influencing program costs.
2. Develop a structure that comprehensively describes the resources necessary to carry out the evaluation (time, space, equipment and supplies).
3. Consider only relevant costs, those that are affected by the decisions under consideration.
4. Costs may be measured in dollar expenditures, dollars, other quantifiable units or in non-quantifiable subject judgments concerning the consequences of a decision. All four types can be used depending whatever procedure is most appropriate to the activity.
5. Try to get the decision stated clearly in the form of two or more alternative courses of action from which a choice is to be made.
6. A rule of thumb for projecting future costs is to allocate the evaluator's time in proportion to the magnitude of the impact of a given category of costs on the total program.

7. Alternatives are usually compared in terms of their total average or marginal costs. "Average costs" are involved when comparing two or more alternative procedures for attaining some goal. "Marginal costs" are the incremental costs of producing one additional unit of some good or service.

Redesign

In the event that the evaluation methodology is not obtaining the desired effect or outcomes, the evaluator must consider reviewing the process and its individual steps to determine why. Each phase and the encompassing decisions throughout the methodology must be reviewed in terms of potential weaknesses or misinterpretation in light of any new information and any new awareness of the situation being evaluated. The redesigning and the re-planning permits the continuance of the evaluation process and can be considered a checkpoint to midcourse corrections.

9.2 Application

Cost Projection

A practical model for the inexperienced evaluator to use in the projecting of costs and considerations in planning is one presented by Rouse (1972) entitled "An Activity Planning and Implementing Concept in a Goal Oriented System". This model presents four basic areas

that the evaluator must consider when attempting to decide the success or determine consequences of specific actions from potential alternatives which have been generated.

The four basic areas are: a) expenses, b) costing and fee setting, c) start up consideration, and d) goal accomplishment considerations. The specifics to be considered under each area, aiding in effective choice of actions, are the following:

A. Expense:

1. Full time staff
2. Supervisory staff
3. Leadership
4. Facilities
5. Transportation
6. Equipment and supplies
7. Promotion
8. Travel
9. Administration
10. Other

B. Income:

1. Program fees
2. Subsidy

C. Costing and Fee Setting:

1. Previous operating experience

2. How do standards and training effect costs?
3. Does the "cost of living" effect costs?
4. What are fees of comparable programs in other units, institutions, organizations or locals?
5. What is the value worth of the training program?
6. What are the sources and extent of funds?

D. Start Up Considerations:

1. Dates (start/finish)
2. Hours
3. Promotion
4. Registration procedures
5. Number and length of sessions intended.

E. Goal Accomplishment Considerations:

1. Intensity of supervision
2. The number and type of leaders needed
3. The type of facilities required
4. The extent and type of leader training required
5. Type of equipment needed
6. The time required
7. The content plans if applicable
8. Identification of applicable standards
9. Identify support systems
10. Capacity
11. Identify and state performance standards

This model will also provide a useful planning tool and checklist for the inexperienced evaluator attempting to organize his/her initial evaluation procedure.

Cost Analysis

The process of cost analysis is a resource that can provide specific information for the evaluator and sponsor for the continuance, termination or modification of training and programming. The cost, utilization and availability of staff and resources have a close relationship to the potential success or failure of any training workshop or educational program.

The application of methods such as the "Delphi Method" and "Cost Analysis" assist the evaluator and the sponsor in planning future actions in the best interests of the intended training outcomes. If all has gone well and the projections and the testing of possible alternatives has indicated that no major problems appear to exist, then, the action alternative which has the greatest potential for success should be adopted for use. On the other hand, if the costs analysis and other methods of projection indicate potential problems, then, replanning or redesign considerations will be implemented.

Redesign

In attempting to redesign any segment or phase of

an evaluation process, the evaluator must ask the following questions:

1. In light of present knowledge and new information, is the present step, design, model, the best available in achieving the desired outcomes?
2. Does each step or phase of the methodology fall into its logical sequenced order?
3. Is there any part of the design which can be improved in order to obtain the desired results?
4. Where are the weakest areas in the evaluation methodology? How can they be improved?
5. Is the design still relevant to any change in situation that has occurred? If not, which situation(s)?
6. Does the sponsor or organization still support the idea of evaluation?
7. In view of present knowledge, can the evaluation process be continued in total? In part? Modified, or discontinued?

Very often this stage of the evaluation methodology results in generating new thinking as well as viewing the achieving of goals and objectives from new perspectives. It is not uncommon for an evaluator to discover that what was voiced originally as a priority need or goal, has now been replaced by new priorities as a result of changing situations, attitudes and new knowledge.

Thus, the cycle has completed a full turn. New needs and priorities must be examined in terms of commitment and authenticity. Each step of the methodology must be applied with the thought of how best to obtain the data

which will indicate the extent of the success or failure in achieving the intended outcome desired. The evaluation cycle stops when the sponsors and the evaluator decide they have obtained adequate results for their efforts, or the training workshop comes to an end. The information obtained from the evaluation process in relation to newly emerged needs and priorities can be used by the sponsors for future planning within their organization, or as a basis for future training workshops.

C H A P T E R X

10.0 SUMMARY, RECOMMENDATIONS, SIGNIFICANCE AND LIMITATIONS OF THE STUDY

10.1 Summary of the Study

As most training workshops are poorly evaluated or not evaluated at all, it was felt that the development of an evaluation methodology which was both easily understood and readily applicable would serve to encourage training practioners to start to evaluate their training workshops or increase the quality of evaluations presently being done.

The primary purpose of this study was to develop an evaluation methodology which could be easily used by those inexperienced in evaluation, but active as leaders of training workshops. A second purpose of this study was to provide a resource itinerary of evaluation steps for those persons or organizations who wished to engage an external evaluator for the evaluation of the workshops and not do it themselves. The resource itinerary would provide the necessary information as to what specific skills are needed by the evaluator to be hired and what systematic thinking processes and considerations must go on when carrying on the evaluation process.

The study describes eight basic steps to evaluating training workshops. They are: needs assessment and goal clarification, choosing evaluation models, selecting research designs, data collection and instrumentation, analyzing data, reporting outcomes and giving feedback, evaluating the evaluation and action deliberation and redesign.

The first step of the evaluation methodology assessing needs and clarifying goals, reviews the importance of setting specific goals and objectives for training workshops by the sponsors and the participants prior to training. Numerous assessment models were reviewed with suggestions for applying specific models under four common field situations encountered by evaluators.

The second step of the evaluation methodology called "Choosing evaluation models" reviewed various approaches to performing an evaluation as well as suggesting and specifying numerous models which could be utilized in an evaluation process by the novice.

Step three, "Selecting research designs" emphasizes the importance of selecting the most appropriate research design when evaluating training workshops. A comprehensive list of tentative designs were reviewed, including mention of numerous factors which influence their validity and

effectiveness in application. Six designs were forwarded as being appropriately useful in evaluating training workshops.

"Data collection and instrumentation" step four, mentioned various data collection techniques which are popularly used in the evaluation field, but highlighted the most popular which are, observation, questionnaires and interviews. This section also covered, adapting prepared instruments for workshop use, as well as pre-testing and editing of instrumentation, and the construction and scoring of a "Likert scale". Directions were also presented for the techniques of arranging, setting up and carrying out personal interviews.

The following step, step five "Analyzing data" has its primary emphasis on explaining numerous examples of descriptive statistical methods for use by the novice evaluator. Mention is made of "inferential statistical methods", but this topic was left for further pursuit by the more advanced evaluator. Examples of the application of various statistical procedures to training situations were also presented in this section.

In the sixth step of the evaluation methodology, "Reporting outcomes and giving feedback", the importance of presenting a well organized, complete and comprehensive report was emphasized with specific guidelines and

instructions as to content and presentation. Specific mention was made of professional ethics related to reporting and communicating results. As feedback is an informal method of reporting data or information within training workshop environments, comments were introduced emphasizing various considerations and principles used in providing feedback.

"Evaluating the evaluation", the seventh step, presented a series of checkpoints suggested by professional evaluators for determining the quality and completeness of an implemented evaluation process.

The final step in this methodology, step eight entitled "Action deliberation and redesign" presented various models and methods which could be used for deciding on and carrying out action plans. Two models were outlined for potential use by the novice evaluator; they were "cost analysis" and "cost projection" models. As partial redesign of some parts of an evaluation process are a common occurrence, a series of questions were included as guidelines for the novice evaluator when confronted with this possibility.

10.2 Recommendations for Additional Research

The test of any theory or methodology is its specific application to experimental conditions. Thus, the

strengths and weaknesses of this evaluation methodology can only be determined by its application to training workshops in educational settings. In doing this, it is hoped that it will serve to encourage further use of appropriate evaluative designs to determine training value and effectiveness by practitioners. Naturally, any application and field testing of this methodology will provide further information on the need to make additional adjustments to the methodology, thus increasing its applicability and effectiveness.

The development of an anthology of evaluative models, designs, and instruments to further supplement this methodology would be a valuable contribution as it would provide the evaluating practitioner with convenient and additional resources to which s/he can quickly refer. Some examples of such designs, models and instruments might be observation report forms, interviewing schedules and questionnaires.

Encouragement should be given to the creation of alternative evaluation methodologies by other researchers, as opposing views and different approaches often give rise to further discussion and creative thinking. This will all contribute toward strengthening the methodology and enhancing the thinking behind it.

Also, thought could be given to using this evaluation methodology as a basis of developing a methodology which practitioners would find applicable to other fields such as business, industry and areas involving social action.

Furthermore, this methodology could be applied to "Leadership Training Workshops", "Communication Workshops", "T-Group and Encounter groups", "Therapy groups" as well as a host of other workshop type environments which use various forms of training to improve skills or change behavior.

The researching of this study revealed a great volume of evaluation literature available for the researcher to sort through. An area of further study would be a compiling and documenting of the most relevant aspects of this literature for easy access and referral for use by future evaluators.

Finally, it would seem to be both timely and appropriate to recommend the planning and development of a series of evaluation workshops which would be specifically designed to train interested practitioners in developing skills and expertise in evaluative methods and technology. Such workshops would elevate the status and importance of evaluation within the training field as well as providing the field with practitioners skilled in evaluating training programs. The end result would be

that more training programs would be properly evaluated, thus improving both their quality and their worth.

10.3 Significance of the Study

The presentation of this study has provided potential evaluators with a specific series of instructional steps for implementing an evaluation as well as providing a selected review of the pertinent literature available in the evaluation field. The evaluation methodology has been designed such that it may be used in its entirety and applied to training workshops, or each step may be used independently and applied as an individual unit depending on the needs of the evaluator, and the demands of the workshop being evaluated.

In addition, this methodology can be used by those persons or organizations who have little knowledge of evaluation, but would require sufficient information on evaluation in order to hire a competent evaluator. The format of this study divides each step of the evaluation process into a chapter which is further subdivided into sections on theory and application. This provides an abundance of basic information and instruction for those choosing to evaluate a training workshop themselves or for those wishing to hire a competent evaluator.

It is intended that this evaluation methodology will provide the incentive for training practitioners to become more involved in evaluating their training efforts as well as upgrading the quality and success of training workshops. Thus, this evaluation methodology has attempted to provide the new evaluator with a source of knowledge and direction for evaluating training workshops such that through its use the evaluator may determine which training is effective, which is not and thus, provide the needed assistance in formulating further policy and planning for future training workshops.

10.4 Limitations of the Study

Field Test

The true potential of this evaluation methodology has not yet been fully determined as the opportunity for a thorough field testing has not taken place. Thus, the strengths and weaknesses of this study have not been explored under the rigors of a workshop training program. Only then can the proper modifications, adjustments and corrections be made to better adapt it for the function it has been specifically designed to perform.

Instrumentation

A second limitation of this study is the absence of

valid and reliable instrumentation designed for evaluating training workshops. If such instrumentation had been available, it would have provided additional depth and resources to the evaluation methodology for the potential user.

Selective Theory Review

The third limitation existing for this study is related to the vast quantities of information existing in the field for the various areas of this research study. On many occasions selective choices were made for inclusion or exclusion of information due to time and space limitations. For example, in the area of needs assessment in Chapter II a few selected models were chosen for inclusion from an overwhelming number in the field. As a result many interesting and useful models were omitted from the discussion. Similar choices were necessary for Chapter IV on research designs and Chapter VI on analyzing data. None the less, these choices were made, thus keeping the evaluation methodology functional and applicable for the novice evaluator.

A P P E N D I X

COMPARISONS OF CONTEMPORARY EVALUATION MODELS ON SELECTED CHARACTERISTICS

	STAKE	SCRIVEN	PROVUS	HAMMOND
Types of Evaluation	(1) Formal vs. informal.	(1) Formative summative. (2) Comparative non-comparative. (3) Intrinsic-payoff. (4) Mediated.	(1) Design. (2) Installation. (3) Process. (4) Product. (5) Cost.	(1) Instructional dimension. (2) Institutional dimension. (3) Behavioral dimension used for describing programs.
Constructs Proposed	(1) Data matrices description (intentions & observations) & judgment. (2) Processing descriptive data contingency among antecedents, transactions, outcomes, congruence between intentions & observations. (3) Bases for forming absolute & relative judgments.	(1) Distinction between goals (claims) & roles (functions). (2) Several types of evaluation.	(1) Discrepancy concept. (2) Feedback & revision of objectives and/or program.	(1) The application of evaluation design to existing program. (2) Decisions about adequacy of current program in relationship to the objectives. (3) Feedback from (2) leads to innovation (4) Application of evaluation to innovation itself. (5) Notion that feedback could continue.
Criteria for judging evaluation	(1) Should be panoramic, not microscopic. (2) Should include descriptive & judgment data. (3) Should provide immediate relative answers for decision-making. (4) Should be formal (e.g. objective, scientific, reliable).	(1) Should be predicted on goals. (2) Must indicate worth. (3) Should have construct validity. (4) Should be wholistic program evaluation.	(1) Term involvement. (2) Assume one-to-one correspondence between design & solution. (3) Compare performance against standards as a tool for improvement & assessment. (4) Periodic feedback.	(1) Related to behavioral objectives. (2) An on-going process. (3) Provides feedback on goal achievement for program modification. (4) Uses local personnel, and is part of local educational program.

	STAKE	SCRIVEN	PROVUS	HAMMOND
Implications for design	Very general structure. Matrices should be included in design.	<p>(1) Look at many factors.</p> <p>(2) Be involved in value judgments.</p> <p>(3) Require use of scientific investigations.</p> <p>(4) Evaluate form within (formative) or from without (summative).</p>	<p>(1) Provide continuous evaluation (feedback loops).</p> <p>(2) Provide relevant & timely information for making decisions.</p> <p>(3) Provide cost-benefit analysis.</p> <p>(4) Involvement of evaluation in program development.</p>	<p>(1) Use of multi-variate structure-focus on interactions of dimensions.</p> <p>(2) Generate empirical research.</p> <p>(3) Necessity for inclusion of local personnel.</p>
Definition	Describing and judging an educational program.	Gathering & combining performance data with weighted set of goal scales.	Comparing performance against standards.	Assessing effectiveness of current & innovative programs at the local level by comparing behavioral data with objectives.
Purpose	To describe and judge educational programs based on a formal inquiry process.	To establish & justify merit or worth. Evaluation plays many roles.	To determine whether to improve, maintain, or terminate a program.	To find out whether innovation is effective in achieving expressed objectives.
Key Emphasis	Collection of descriptive & judgmental data from various audiences.	Justification of data gathering instruments, weightings, & selection of goals. Eval. model combining data on different performance scales into a single rating.	Identifying discrepancies between standards & performance using team approach.	Local program development.

	STAKE	SCRIVEN	PROVUS	HAMMOND
Role of Evaluator	Specialist concerned with collecting, interpreting, & judging descriptive & judgmental data.	Responsible for judging the merit of an educational practice for products (formative & summative).	A team member who aids program improvement & counsels administration. He should be independent of the program unit.	Consultant who should provide expertise in data collection. He is also a trainer of local evaluators (program personnel).
Relationship to Objectives	Examination of goal specifications & priorities. Identification of failures & successes. It is up to the evaluator to assist in writing behavioral objectives.	Look at goals & judge their worth. Determine whether they are being met.	Agreement of evaluation team & program staff on standards. Comparison of performance against standards to see whether a discrepancy exists.	Evaluation focuses on the definition & measurement of behavioral objectives.
Relationship to Decision-Making	Descriptive and judgmental data result in reports including recommendations) to various audiences. Judgments may be based on either absolute or relative standards.	Evaluation reports (with judgments explicitly stated for producers or consumers) used in decision-making.	Evaluation staff collects information essential to program improvement & notes discrepancies between performance & standards. Every question involves a criterion (C), new information (I), & a decision (D). Eval. provides the new information.	Evaluation is the source on which to base decisions about instructional, & behavioral dimensions.
Contributions	(1) Provides a systematic method for arranging descriptive & judgmental data, thus em-	(1) Discriminates between formative (ongoing) & summative (end) evaluation.	(1) Provides continuous communication between program & evaluation staff through feedback	(1) Makes use of local personnel who can carry on evaluation process once initiated.

	STAKE	SCRIVEN	PROVUS	HAMMOND
Contributions (Con't)	<p>phasizing inter- & intra-relations between them.</p> <p>(2) Considers both absolute & relative judgment.</p> <p>(3) Requires explicit standards.</p> <p>(4) Generalizability of the model.</p>	<p>(2) Focus on direct assessment of worth, focus on value.</p> <p>(3) Applicable in diverse contexts.</p> <p>(4) Analysis of means & ends.</p> <p>(5) Delineation of types of evaluation.</p> <p>(6) Evaluation of objectives.</p>	<p>loops.</p> <p>(2) allows for program improvement as well as assessment either at early stages or at end.</p> <p>(3) Acknowledges alternative procedures in adjusting objectives & in changing treatment.</p> <p>(4) Forces explicit statement of standards.</p>	<p>(2) Considers inter-action of several dimensions & variables.</p> <p>(3) Provides feedback on program development & revisions: stresses self-evaluation.</p> <p>(4) Requires specification of behavioral objectives.</p>
Limitations	<p>(1) Inadequate methodology for obtaining information on key constructs.</p> <p>(2) Some cells of design matrix overlap, some distinctions not clear.</p> <p>(3) Possibility of leading to internal strife withing program, value conflicts possible.</p>	<p>(1) Equating performance on different criteria & assigning relative weights to criteria creates methodological problems.</p> <p>(2) No methodology for assessing validity of judgment.</p> <p>(3) Several overlapping concepts.</p>	<p>(1) Demands a lengthy time commitment, may be expensive to carry through.</p> <p>(2) Inadequate methodology for establishing standards.</p> <p>(3) Requires large, expert, well-articulated staff.</p> <p>(4) Designed for complete evaluation; partial evaluation not considered.</p>	<p>(1) Difficulty of quantifying data involving several dimensions & variables.</p> <p>(2) May be complex & time-consuming to set up.</p> <p>(3) Possible fixation of evaluation on the "cube".</p> <p>(4) Neglects judgmental dimension.</p> <p>(5) Motivation problem in local personnel.</p>

	STUFFLEBEAM	ALKIN	PERSONAL JUDGEMENT (e.g. ACCREDITATION)	TYLER
Types of Evaluation	(1) Context. (2) Input. (3) Process. (4) Product.	(1) Systems assessment. (2) Program planning. (3) Program Implementation. (4) Program improvement. (5) Program certification.	(1) Self study. (2) Visitation. (3) Annual reports. (4) Evaluation panels.	Pre-post measurement of performance.
Constructs Proposed	(1) Context eval. for planning decisions. (2) Input eval. for programming decisions. (3) Process eval. for implementing decisions. (4) Product eval. for recycling decisions.	Evaluation of educational systems vs. evaluation of instructional programs, five areas of evaluation.	Use of content specialists as judges.	(1) Statements of objectives in behavioral terms. (2) Teaching objectives are pupil-oriented. (3) Objectives must consider pupil's entry behavior, analysis of our culture, school philosophy, learning theories, new developments in teaching, etc.
Criteria for Judging Evaluation	(1) Internal validity. (2) External validity. (3) Reliability. (4) Objectivity. (5) Relevance. (6) Importance. (7) Scope (8) Credibility. (9) Timeliness (10) Pervasiveness. (11) Efficiency.	Information provided to a decision-maker should be effective & not confusing or misleading. Appropriate evaluation procedures should be used for different decisions.	(1) Reflects interests of program administrators. (2) Standard criteria often used.	(1) Behavioral objectives clearly stated. (2) Objectives should contain references not only to course content but also to mental processes applied.

	STUFFLEBEAM	ALKIN	PERSONAL JUDGEMENT (e.g. ACCREDITATION)	TYLER
Implications for Design	(1) Experimental design not applicable. (2) Use of systems approach for evaluation studies. (3) Directed by administrator.	Evaluation domain determined by the decision-maker, the objects of evaluation vary along a continuum from discrete definable objects to complex systems.	(1) Involvement of professional community. (2) Quick feedback.	(1) Need to interpret & use results of assessment. (2) Develop designs to assess student progress.
Definition	Defining, obtaining, & using information for decision-making.	The process of ascertaining the decision areas of concern, selecting appropriate information, & collecting & analyzing information.	Focusing attention on processes of education using professional judgment. Development of standards of standards for educational programs.	Comparing student performance with behaviorally stated objectives.
Purpose	To provide relevant information to decision-makers.	To report summary data useful to decision-makers in selecting among alternatives.	To identify deficiencies in the education of teachers & students relevant to content & procedures, self-improvement.	To determine the extent to which purposes of a learning activity are actually being realized.
Key Emphasis	Evaluation reports used for decision-making.	Evaluation reports used for decision-making.	Personal judgment used in evaluating process of education, self-study.	Specification of objectives & measuring learning outcomes of pupils
Role of Evaluator	Specialist who provides evaluation information to decision-makers.	Specialists who provide evaluation information to decision-makers.	Professional colleagues who make recommendations-a professional judge.	Curriculum specialist who evaluates as part of curriculum development & assessment.

	STUFFLEBEAM	ALKIN	PERSONAL JUDGEMENT (e.g. ACCREDITATION)	TYLER
Relationship to Objectives	Terminal stage in context eval. is setting objectives, input eval. produces ways to reach objectives; product eval. determines whether objectives are reached.	Range & specificity of program objectives determined in systems assessment; program planning produces ways to reach objectives; program improvement provides data on the extent to which objectives are being achieved; program certification determines whether objectives are reached.	Self-study judgments are based on sets of predetermined criteria.	Evaluation implies attainment of behavioral objectives stated at the course.
Relationship to Decision-Making	Evaluation provides information for use in decision-making.	Evaluation provides information for use in decision-making.	When deficiencies are found, program revisions are requested, thus correcting standard conditions, corrective process built in.	Actual pupil performance data will provide information for the decision-maker to use on strengths & weaknesses of a course or curriculum.
Contributions	(1) Provides a service function by supplying data to administrators & decision-makers charged with conduct of the program (2) Is sensitive to feedback.	(1) Provides a service function to administrators & decision-makers. (2) Allows for evaluation to take place at any stage of the program. (3) Wholistic.	(1) Is easy to implement, team can observe & make judgment. (2) Has little lag time between observations made, data collected, & feedback. (3) Breadth of variables noted is large. (4) Leads to self-study habit & self-improvement.	(1) Is easy to assess whether behavioral objectives are being achieved. (2) Is easy for practitioners to design evaluative studies. (3) Checks degree of congruency between performance & objectives; focus on clear definition of objectives.

	STUFFLEBEAM	ALKIN	PERSONAL JUDGEMENT (e.g. ACCREDITATION)	TYLER
Limitations	<p>(1) Little emphasis in value concerns.</p> <p>(2) Decision-making process is unclear, methodology undefined.</p> <p>(3) May be costly & complex if used entirely.</p> <p>(4) Not all activities are clearly evaluative.</p>	<p>(1) Role of values in evaluation unclear.</p> <p>(2) Description of decision-making process incomplete.</p> <p>(3) May be costly & complex.</p> <p>(4) Not all activities are clearly evaluative.</p>	<p>(1) Objectivity & empirical basis are questionable.</p> <p>(2) Attention to process of education not balanced by attention to consequences.</p> <p>(3) Replicability is questionable.</p>	<p>(1) Tendency to oversimplify program & focus on terminal rather than on-going & pre-program information.</p> <p>(2) Tendency to focus directly & narrowly on objectives, with little attention to worth of the objectives.</p>

A P P E N D I X B

A SOCIOMETRIC MATRIX - FRIENDSHIP CHART

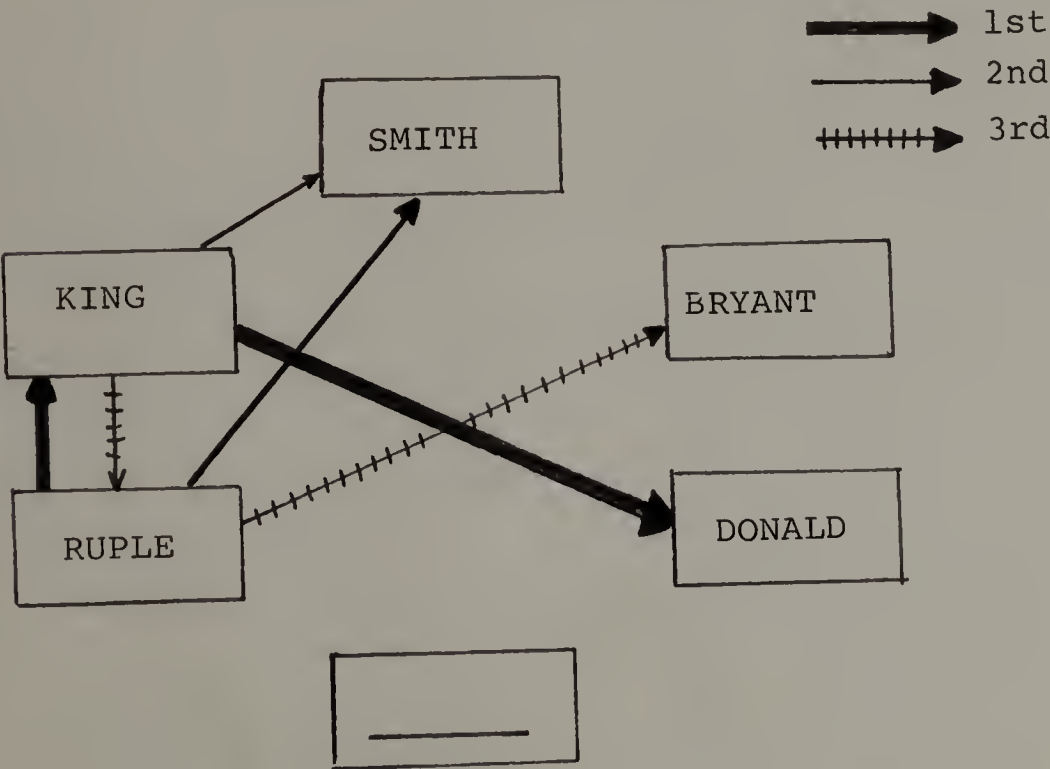
FRIENDSHIP CHART
(in-going choices)

Group Advisor Scudder
Group Anzacks
Date May 2nd
Criteria for choices Sleep mates
First choice = 5
Second choice = 3
Third choice = 2

CHOSEN

C
H
O
O
S
E
R
S

Last Names	1. Ruple	2. King	3. Smith	4. Bryant	5. Donald	6.	7.	8.
1. Ruple		5	3	2				
2. King	2		3		5			
3. Smith								
4. Bryant								
5. Donald								
Totals								



LEADER BEHAVIOR

	always	often	occasionally	seldom	never		always	often	occasionally	seldom	never
1. I make my attitudes clear to the group						2. I do personal favors for subordinates.					
3. I try out my new ideas with the group.						4. I do little things to make it pleasant to be a member of the group.					
5. I rule with an iron hand.						6. I am easy to understand.					
7. I speak in a manner not to be questioned.						8. I find time to listen to subordinates.					
9. I criticize poor work.						10. I mix with subordinates rather than keeping to myself.					
11. I assign subordinates to particular tasks.						12. I look out for the personal welfare of individuals in my group.					
13. I schedule the work.						14. I explain my actions to subordinates.					
15. I maintain definite standards of performance.						16. I consult subordinates before action.					
17. I emphasize the meeting of deadlines.						18. I back up subordinates in their action.					
19. I encourage the use of uniform procedures.						20. I treat all subordinates as equals.					
21. I make sure that my part in the organization is understood.						22. I am willing to make changes.					
23. I ask that subordinates follow standard rules and regulations.						24. I am friendly and approachable.					
25. I let subordinates know what is expected of them.						26. I make subordinates feel at ease when talking with them.					
27. I see to it that subordinates are working up to capacity.						28. I put suggestions made by my group into action.					
29. I see to it that the subordinates' work is coordinated						30. I get group approval in important matters before acting.					
TOTAL						TOTAL					

A P P E N D I X D

A SELF ADMINISTERED STANDARDIZED QUESTIONNAIRE
"WORKSHOP EVALUATION QUESTIONNAIRE"

WORKSHOP EVALUATION FORM

The purpose of this brief questionnaire is to provide you with an opportunity to indicate your reactions to the workshop and to suggest ways in which the workshop could be redesigned to improve its usefulness. It is not necessary for you to indicate your name on the questionnaire. Thank you in advance for your cooperation.

1. What was the topic, date, and location of the workshop?

Topic: _____

Date: _____

Location: _____

2. Listed below are the goals of the workshop, as defined by the workshop leader. Please indicate the extent to which you think the workshop was successful in achieving each goal. Indicate your answer to each goal by circling the number corresponding to your answer.

1= Very Successful; 2= Successful; 3= Somewhat Successful; 4= Unsuccessful

<u>Goal</u>	<u>Successful</u>	<u>Successful</u>	<u>Somewhat Successful</u>	<u>Unsuccessful</u>
a. _____	1	2	3	4
b. _____	1	2	3	4
c. _____	1	2	3	4
d. _____	1	2	3	4
e. _____	1	2	3	4
f. _____	1	2	3	4

3. Listed below are statements that are often used to describe specific skills and characteristics of a workshop leader. Please indicate the extent to which you think the workshop leader needs improvement. Indicate your answer to each statement by circling the number corresponding to your answer.

1= No Improvement Needed 2= Little Improvement Needed 3= Considerable Improvement Needed 4= Not a Necessary Skill

<u>Skill/Characteristic</u>	<u>No Improvement Needed</u>	<u>Little Improvement Needed</u>	<u>Considerable Improvement Needed</u>	<u>Not a Necessary Skill</u>
a. Knowledge of the topic.	1	2	3	4
b. Use of effective teaching methods.	1	2	3	4
c. Sensitivity to needs of audience.	1	2	3	4
d. Clarity of presentation.	1	2	3	4
e. Answering questions.	1	2	3	4
f. Involving group in the learning process.	1	2	3	4
g. Use of a variety of teaching methods.	1	2	3	4
h. Generation of interest and enthusiasm for the material.	1	2	3	4

4. Listed below are several aspects of the workshop experience. Please indicate the extent to which you were satisfied with each of the features. Indicate your answer to each statement by circling the number corresponding to your answer:

1= No Improvement Needed 2= Some improvement Needed
3= Considerable Improvement Needed

<u>Component</u>	<u>No Improvement Needed</u>	<u>Some Improvement Needed</u>	<u>Considerable Improvement Needed</u>
a. Workshop facilities	1	2	3
b. Workshop organization	1	2	3
c. Use of visual aids	1	2	3
d. Use of handouts	1	2	3
e. Workshop publicity	1	2	3

5. How would you judge the length of the workshop? (Circle one)

1. Too long 2. Too Short 3. About right

6. Overall, how valuable was the workshop to you? (Circle one)

1. Very Valuable 2. Valuable 3. Somewhat Valuable
4. Worthless

7. How helpful do you feel the skills developed in this workshop will be to you in your future work? (Circle one)

1. Very Helpful 2. Helpful 3. Somewhat Helpful 4. Not Helpful

8. Overall, how would you rate the workshop leader? (Circle one)

1. Excellent 2. Very Good 3. Good 4. Fair 5. Poor

9. Overall, how would you rate the workshop experience? (Circle one)

1. Excellent 2. Very Good 3. Good 4. Fair 5. Poor

A P P E N D I X E

A SEMANTIC DIFFERENTIAL SCALE

LPC

Name _____

Think of the person with whom you can work least well.
He may be someone you work with now, or he may be someone
you knew in the past.

He does not have to be the person you like least well,
but should be the person with whom you had the most
difficulty in getting a job done. Describe this person
as he appears to you.

Friendly	: 8 : 7 : 6 : 5 : 4 : 3 : 2 : 1 :	Unfriendly
Rejecting	: 1 : 2 : 3 : 4 : 5 : 6 : 7 : 8 :	Accepting
Helpful	: 8 : 7 : 6 : 5 : 4 : 3 : 2 : 1 :	Frustrating
Unenthusiastic	: 1 : 2 : 3 : 4 : 5 : 6 : 7 : 8 :	Enthusiastic
Lot of Fun	: 8 : 7 : 6 : 5 : 4 : 3 : 2 : 1 :	Serious
Tense	: 1 : 2 : 3 : 4 : 5 : 6 : 7 : 8 :	Relaxed
Distant	: 1 : 2 : 3 : 4 : 5 : 6 : 7 : 8 :	Close
Cold	: 1 : 2 : 3 : 4 : 5 : 6 : 7 : 8 :	Warm
Cooperative	: 8 : 7 : 6 : 5 : 4 : 3 : 2 : 1 :	Uncooperative
Supportive	: 8 : 7 : 6 : 5 : 4 : 3 : 2 : 1 :	Hostile
Boring	: 1 : 2 : 3 : 4 : 5 : 6 : 7 : 8 :	Interesting
Quarrelsome	: 1 : 2 : 3 : 4 : 5 : 6 : 7 : 8 :	Harmonious
Self-Assured	: 8 : 7 : 6 : 5 : 4 : 3 : 2 : 1 :	Hesitant

Efficient	: 8 : 7 : 6 : 5 : 4 : 3 : 2 : 1 :	Inefficient
Gloomy	: 1 : 2 : 3 : 4 : 5 : 6 : 7 : 8 :	Cheerful
Open	: 8 : 7 : 6 : 5 : 4 : 3 : 2 : 1 :	Guarded
Pleasant	: 8 : 7 : 6 : 5 : 4 : 3 : 2 : 1 :	Unpleasant

A P P E N D I X F

SAMPLE OBSERVATION QUESTIONNAIRES

GROUP PARTICIPATION SCALE

For each person in the group assign only the one question that most appropriately describes his behavior.

1. Who puts group suggestions into operation?
2. Who pushes new ways of doing things?
3. Who reacts unfavorable to everything the group members want to do?
4. Who never does anything?
5. Who has a hard time putting things across?
6. Who can't seem to get the point of what the group is doing?
7. Who urges orderly methods of doing the job?
8. Whose advice do group members most often take?
9. Who is a good follower?
10. Who tries to work without a plan?
11. Who never listens to what others say?
12. Who gripes a lot, but says little that is constructive?
13. Who gives information on how to do things?
14. Who sometimes says or does good things in the group?
15. Who has little to offer the group?
16. Who just sits without doing anything?
17. Who encourages slow workers to greater effort?

18. Who likes to be told what to do?
19. Who knows how to get things done?
20. Who changes his mind often when his suggestions meet opposition?
21. Who seems half-hearted about what he does in the group?
22. Who tries hard to do a good job?
23. Who usually agrees with what is said?
24. Who helps members most with their thinking about group suggestions?

The following are samples of Observation Report forms that can be adapted and used by evaluator.

Observation Report Form

1. To what extent has (name person or persons) behavior changed in staff meetings since the training program terminated?

None	Some	A Fair Amount	Greatly
------	------	------------------	---------

2. Describe briefly some of things which s/he is now doing differently?
3. What changes have you noticed in the way other people have reacted to him/her during this time?
4. What changes have occurred in your relationship with this person during this time?

PROFILE OF ORGANIZATIONAL CHARACTERISTICS

Organizational variables	SYSTEM 1	SYSTEM 2	SYSTEM 3	SYSTEM 4	
How much confidence and trust is shown in subordinates?	Virtually none	Some	Substantial amount	A great deal	1
How free do they feel to superiors about job?	Not very free	Somewhat free	Quite free	Very free	2
How often are subordinate's ideas sought and used constructively?	Seldom	Sometimes	Often	Very frequently	3
What dominant use made of rewards, 1 threats, 2 punishment, 3 rewards, 4 involvement?	1, 2, 3, occasionally 4	4, some 3	4, some 3 and 5	5, 4, based on group	4
Where is responsibility for achieving organization's goals?	Mostly at top	Top and middle	Fairly general	At all levels	5
How much cooperative teamwork exists?	Very little	Relatively little	Moderate amount	Great deal	6
What is the usual direction of information flow?	Downward	Mostly downward	Down and up	Down, up, and sideways	7
How is downward communication accepted?	With suspicion	Possibly with suspicion	With caution	With a receptive mind	8
How accurate is downward communication?	Usually inaccurate	Often inaccurate	Often accurate	Almost always accurate	9
How well do superiors know problems faced by subordinates?	Not very well	Rather well	Quite well	Very well	10
At what level are decisions made?	Mostly at top	Policy at top, some delegation	Broad policy at top, more delegation	Throughout but well integrated	11
How much subordinates involved in decisions related to their work?	Almost never	Occasionally consulted	Generally consulted	Fully involved	12
How much does decision-making process contribute to motivation?	Not very much	Relatively little	Some contribution	Substantial contribution	13
How are organizational goals established?	Orders issued	Orders, some comments invited	After discussion, by orders	By group action (except in crisis)	14
How much covert resistance to goals is present?	Strong resistance	Moderate resistance	Some resistance at times	Little or none	15
How concentrated are planning and control functions?	Very highly at top	Quite highly at top	Moderate delegation to lower levels	Widely shared	16
How often is informal organization existing the formal one?	Yes	Usually	Sometimes	No---same goals as formal	17
What are cost, productivity, and other control data used for?	Policing, punishment	Reward and punishment	Reward, some self-guidance	Self-guidance, problem-solving	18

A P P E N D I X H

SAMPLE TABLE FORMAT FOR REPORTING DATA

Means and Standard Deviations* for Importance and Enjoyment
Scores on Each Concept in the Two Instructional Programs

	Individualized Instruction Program				Traditional Program			
	Enjoyment		Importance		Enjoyment		Importance	
	M	SD	M	SD	M	SD	M	SD
Teachers	5.5	1.3	6.8	1.4	5.0	1.3	6.4	1.4
Science	6.1	1.5	6.3	1.5	4.8	1.5	5.7	1.5
Student Freedom in Class	6.5	1.5	6.9	1.5	6.2	1.5	6.6	1.5
School	5.8	1.6	7.4	1.5	4.7	1.7	6.6	1.7
Testing	3.7	1.5	5.9	1.5	3.4	1.5	5.3	1.6
Individualized Instruction	6.0	1.4	6.3	1.4	5.3	1.4	6.2	1.4
Resource Center	5.2	1.4	6.5	1.4	5.5	1.3	6.3	1.3
Enrichment Program	5.4	1.3	6.0	1.3	5.5	1.3	5.8	1.3

* Approximate

A P P E N D I X I

AN EVALUATION DESIGN CHECKLIST

Instructions for the Use of "Evaluation Design Checklist"

The "Evaluation Design Checklist" has been designed to quickly provide evaluators with information directly related to evaluating their evaluation designs. The checklist is divided into five primary categories. Design, Objectives, Evaluator role, Measures and The Report. Within each of these categories there are related questions which can be answered by the evaluator in the two columns to the right.

Each question is designed to be answered with either a "yes" or a "no" response using a checkmark (✓) in the appropriate column. Frequent "no" responses or a majority of "no" responses should be treated by the evaluator as a warning that further modification or changes are necessary within the evaluation design and process being carried on.

Evaluation Design Checklist

	YES	NO
<u>Design</u>		
1. Is there flow to the evaluation design which provides for a natural sequence and adequate bridging from one step to the next?		
2. Does the design create a climate for learning?		
3. Has the evaluation design been designed in response to the real learning needs of the person(s) for which it is intended?		
4. Is there a written plan for follow-up and back home support?		
<u>Objectives</u>		
5. Does the evaluation accomplish what it set out to do?		
6. Is the evaluation efficient, complete and focused?		
7. Can the evaluation be evaluated in terms of its component parts, goal identification, evaluation models, research design and so on?		
<u>Evaluator Role</u>		
8. Were the evaluators noncoercive, placing no demands on participants and trainees to change?		
9. Did the evaluators consider the feelings of others involved when carrying out the evaluation?		
10. Will the evaluator's audience act on his/her recommendations?		

<u>Measures</u>	YES	NO
11. Was there development of adequate yardsticks which were reliable and relevant?		
12. Were these yardsticks capable of measuring progress toward achieving stated objectives?		
13. Were the yardsticks aptly applied in relation to the time span implied by the goals?		
14. Were there at least two sources of information to evaluate the situation in terms of the goals to be evaluated (minimizing bias)?		
15. Was there specification and examination of those underlying personality and situational factors which help explain the changes identified?		
<u>The Report</u>		
16. Was the feedback and reporting focused on things which could be changed or done differently?		
17. Was the report and feedback timely such that the data was presented close to the time of the events taking place?		

REFERENCE

- Airasian, P.W. Designing summative evaluation studies at the local level. In W.J. Popham, (Ed.), Evaluation in education: Current application. Berkeley: McCutchan Publishing, 1974.
- Airasian, P.W. & Madaris, G.F. Functional types of student evaluation. Measurement and Evaluation in Guidance, 1972, 4, 221-233.
- Alkin, M.C. Evaluation theory development. In C.H. Weiss (Ed.), Evaluating action programs: Readings in social action and education. Boston: Allyn and Bacon, Inc., 1972.
- Argyris, C. Organization and innovation. Homewood, Illinois: The Dorsey Press, 1965.
- Baker, E.L. The technology of instructional development. In R.M.W. Travers (Ed.), Second handbook on research on teaching. Chicago: Rand McNally, 1973.
- Beck, J., Raynor, M., Raynor, T.E., & Schragle, S. Evaluation: Definition and methodology. Unpublished manuscript, May 1975. (Available from 3501 Northcliffe Montreal, H4A 3K8, Quebec, Canada).
- Beckhard, R. Organizational development: Strategies and models. Reading, Massachusetts: Addison-Wesley Publishing Company, 1969.
- Belasco, J.A., & Trice, H.M. The assessment of change in training and therapy. New York: McGraw-Hill Book Company, 1969.
- Benedict, L. Practical guide for evaluation. (Capitol Region Education Council Project). Unpublished manuscript, University of Massachusetts, 1971.
- Benjamin, A. The helping interview. Boston: Houghton Mifflin Company, 1969.
- Benne, K.D., Bradford, L.P. & Lippitt, R. The laboratory method. In L.P. Bradford, J. Gibb & K.D. Benne, T-Group and laboratory method: Innovation in re-education. New York: John Wiley and Sons, 1964.

- Bloom, B.S. Some theoretical issues relating to educational evaluation. In R.W. Tyler (Ed.), Educational evaluation: New roles, new means. Sixty-eighth year book, Part II. Chicago: National Society for the Study of Education, 1969.
- Bradford, L.P., Gibb, J.R., and Benne, K.D. T-group and laboratory method: Innovation in re-education. New York: John Wiley and Sons, Inc., 1964.
- Brown, F.L., Amos, J.R. and Mink, O.G. Statistical concepts: A basic program. New York: Harper and Row, Publishers, 1975.
- Brown, L.D. Research action: Organizational feedback, understanding and change. Journal of Applied Behavioral Science, 1972, 8, 697-711.
- Buchanan, P.C. Crucial issues in organizational development. In G.L. Lippitt, L.E. This, & R.G. Bidwell, Jr. Optimizing human resources. Reading, Massachusetts.: Addison-Wesley Publishing Company, 1971.
- Burke, W. World of organizational development. Seminar presented at Sir George Williams University, Montreal, 1972.
- Buros, O.K. The Buros mental measurements year book. Highland Park, New Jersey: The Gryphon Press, 1976.
- Campbell, D.T. Administrative experimentation, institutional records and non-reactive measures. In J.C. Stanley (Ed.), Improving experimental design and statistical analysis. Chicago: Rand McNally Publishing Co., 1967.
- Campbell, D.T. and Stanley, J.C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally Publishing Co., 1963.
- Campbell, J.P., Dunnette, M.D., Lawler, E.E. III. and Weick, K.E. Jr., Managerial behavior, performance and effectiveness. New York: McGraw-Hill Book Co., 1970.
- Clark, P.A. Action research and organizational change. New York: Harper and Row Publishers, 1972.
- Cronbach, L.J. Validation of educational measures. In P.H. Du Bois (Ed.), Towards a theory of achievement.

- ment measurement. Proceedings of the 1969 Invitational Conference on Testing Problems. Princeton, New Jersey: Educational Testing Service, 1970.
- Culbert, S.A. Using research to guide an organizational development project. Journal of Applied Behavioral Science, 1972, 8, 203-236.
- Delbecq, A.L., and Van de Ven, A.H. A group process model for problem identification and program planning. Journal of Applied Behavioral Science, 1971, 9, 466-492.
- De Phillips, F.A., Berliner, W.M. and Cribbin, J.J. Management of training programs. Homewood, Illinois: Richard D. Irwin, Inc., 1960.
- Dimock, H.G. How to analyze and evaluate group growth. Montreal: Sir George Williams University, 1970 (a).
- Dimock, H.G. How to observe your group. Montreal: Sir George Williams University, 1970 (b).
- Dimock, H.G. How to plan staff training programs. Montreal: Sir George Williams University, 1973.
- Dimock, H.G. Planning for training. Montreal: Centre for Human Relations and Community Studies, Sir Georges Williams University, n.d.
- Ebel, R.L. Evaluation and educational objectives. Journal of Educational Measurement, 1973, 4, 273-279.
- Finklestein, E.L. Assessment and evaluation handbook. Albany, N.Y.: Title No. 1, The University of the State of New York, the State Education Department Office of Research and Evaluation, 1971.
- Fordyce, J.K. and Weil, R. Managing with people: A manager's handbook of organizational development methods. Reading, Massachusetts: Addison-Wesley Publishing Co., 1971.
- Fortune, J. and Hutchinson, T. The Fortune/Hutchinson evaluation methodology. Amherst, Massachusetts: University of Massachusetts, 1973.
- Gagné, R.M. Behavioral objectives. Educational Leadership, 1972, 29 394-396.

- Guba, E.G., & Stufflebeam, D.L. Evaluation: The process of stimulating, aiding and offsetting insightful action. Paper presented at the 2nd Phi Delta Kappa National Symposium for Professors of Educational Research, Boulder, Colorado, November 1968.
- Haller, E.J. Cost analysis for educational program evaluation. In W.J. Popham (Ed.), Evaluation in education: Current applications. Berkley: McCutcheon Publishing Co., 1974.
- Hambleton, R.K. Evaluation for human services. Seminar presented at the University of Massachusetts, Amherst, 1976.
- Hambleton, R.K. Workshop Evaluation Questionnaire, School of Education University of Massachusetts, Amherst, 1976.
- Harrison, T. Designing training. New York: National Council Y.M.C.A., Veranna Litho., January 1968.
- Hogben, D. Curriculum development and evaluation: The need to look beyond behavioral objectives. Teachers College Record, May 1973, 74 529-536.
- Jenks, S.R. An action research approach to change. Journal of Applied Behavioral Science, 1970, 6, 131-150.
- Kahn, R.L. & Cannell, C.F. The dynamics of interviewing: Theory, techniques and cases. New York: Wiley, 1957.
- Kahn, R.L. & Cannell, C.F. Interviewer's manual. Anne Arbor: Survey Research Centre, Institute for Social Research, University of Michigan, 1969.
- Kerlinger, F.N. Foundations of behavioral research. New York: Holt, Rinehart and Winston, 1973.
- Kindall, A.F. & Gatzka, J. Positive program for performance appraisal. Harvard Business Review, 1963, 41, 153-167.
- Lake, D.G., Miles, M.B. & Earl, R.B. Jr. Measuring human behavior. New York: Columbia Teacher's College Press, 1973.
- Lehmann, I.J. & Mehrens, W.A. Educational research: Readings in focus. New York: Holt, Rinehart and Winston, 1971.

- Lewin, K. Group decisions and social change. In E.E. Maccoby, T.M. Newcomb, and E.L. Hartley (Ed.), Readings in Social Psychology. Third Edition: Holt, Rinehart and Winston Incorporated, 1958.
- Likert, R. The method of constructing an attitude scale. In M. Fishbein (Ed.), Readings in attitude theory and measurement. New York: John Wiley and Sons, 1967.
- Lippitt, G.L., This, L.E. & Bidwell, R.G. Jr. Optimizing human resources. Reading, Massachusetts: Addison-Wesley Publishing Company, 1971.
- Mager, R.F. Preparing instructional objectives. Palo Alto Fearon: 1962.
- Maguire, T.D. Methodology for structuring attitudes. American Educational Research Journal, Fall, 1973, 10, 295-306.
- McGill, M.E. & Horton, M.E. Jr. Action research designs for training and development. Washington: National Training and Development Service Press, 1973.
- McGregor, D. A critique of performance appraisal. In Human Side of Enterprise, New York: McGraw-Hill, 1960.
- Miles, M.B. Learning to work in groups. New York: Teacher's College Press, Columbia University, 1969.
- Miller, D.C. Handbook of research design and social measurement. New York: David McKay Co., 1970. National Training Laboratories, Institute Applied Behavioral Science. Laboratories in Human Relations Training: Book of Readings. Washington, D.C., 1971 27-28.
- Pace, C.R. Evaluation perspectives: 1968. Los Angeles: Centre for the Study of Evaluation, University of California at Los Angeles, 1968.
- Payne, D.A. Some old and new wives tales concerning curriculum evaluation. Educational Leadership, Association of Supervision and Curriculum Development, January 1973.
- Pfeiffer, J.W. & Heslin, R. Instrumentation in human relations training: A guide to 75 instruments with

- wide application to the behavioral sciences. Iowa City: University Associates, 1973.
- Philips, B.S. Social research: Strategy and tactics. New York: MacMillan Publishing, 1966.
- Popham, W.J. Must all objectives be behavioral? Educational Leadership, 1972, 29, 605-607.
- Popham, W.J. Evaluation in education. Berkeley, California: McCutchan Publishing, 1974.
- Provus, M. Evaluation of ongoing programs in the public school system. In R.W. Tyler (Ed.), Educational evaluation: New roles, new means. Sixty-eighth Yearbook, Part II. Chicago: National Society for the Study of Education, 1969.
- Reed, H. Modified needs assessment model. Seminar presented at the University of Massachusetts, Amherst, Spring, 1975.
- Riecken, H.W. Memorandum on program evaluation. In C.H. Weiss (Ed.), Evaluating action programs: Readings in social action. Boston: Allyn and Bacon, 1972.
- Robinson, J.P. & Shaver, P.R. Measures of social-psychological attitudes. Ann Arbor: Survey Research Centre, Institute for Social Research, University of Michigan, 1973.
- Rossi, P.H. Booby traps and pitfalls in evaluation of social action programs. In C.H. Weiss (Ed.), Evaluating action programs: Readings in social action and education. Boston: Allyn and Bacon, 1972.
- Rouse, R. An activity planning and implimentation concept in a goal oriented system. Seminar presented at Sir George Williams University, Montreal, 1972.
- Runkel, P.J. & McGrath, J.E. Techniques for observing and recording behavior. From Research on human behavior: A systematic guide to method. New York: Holt, Rinehart and Winston, 1972.
- Schmuck, R.A. & Miles, M.B. (Eds.), Organizational development in schools. Palo Alto, California: National Press Books, 1971.
- Schmuck, R.A., Runkel, P.J., Saturen, S.L., Martell, R.T. & Derr, C.B. Handbook of organization development in schools. Centre for Advanced Study of Educational

Administration, University of Oregon, National Press Books, 1972.

- Scriven, M. The methodology of evaluation. In R.W. Tyler, R.W. Gagné & M. Scriven (Eds.), Perspectives of curriculum evaluation. A. E. R. A. Monograph Series on Curriculum Evaluation, 1967 1, 39-83.
- Scriven, M. Goal-free evaluation. In E.R. House (Ed.), School evaluation: the politics and process. Berkeley: McCutchan Publishing, 1973.
- Scriven, M. Evaluation perspective and procedures. In W.J. Popham (Ed.), Evaluation in education: Current applications. Berkeley: McCutchan Publishing, 1974.
- Selltiz, C., Jahoda, M., Deutsch, M. & Cook, S.W. Research methods in social relations. New York: Holt, Rinehart and Winston, 1962.
- Severy, L.J. Procedures and issues in the measurement of attitudes. (ETS Report 30). Princeton, N.J.: Educational Testing Service. December, 1974.
- Simon, A. & Boyer, E.G. (Eds.), Mirrors for behavior. Philadelphia: Research for Better Schools, 1967.
- Simon, J.L. Basic research methods in social science: The art of empirical investigation. New York: Random House, 1968.
- Sinclair, R. Model for curriculum evaluation. Seminar presented at the University of Massachusetts, Amherst, Massachusetts, 1975.
- Stake, R.E. The countenance of educational evaluation. Teachers' College Record, 1967, 68, 523-540.
- Stake, R.E. Objectives, priorities and other judgement data. Review of Educational Research, 1970, 2, 181-212.
- Stake, R.E. An approach to the evaluation of instructional programs. Paper presented at the American Educational Research Meeting, Chicago, April 1972.
- Stake, R.E. & Denny, T. Needed concepts and techniques for utilizing more fully the potential of evaluation. In R.W. Tyler (Ed.), Educational evaluation:

- New roles, new means. Sixty-eight Yearbook, Part II. Chicago: National Society for the Study of Education, 1969.
- Stufflebeam, D.L. The relevance of the CIPP evaluation model for educational accountability. Journal of Research and Development in Education, 1971, 1, 19-25.
- Stufflebeam, D.L. Toward a characterization of curriculum evaluation. Source unknown.
- Suchman, E.A. Action for what? A critique of educational research. In R. O'Toole (Ed.), The organization, management and tactics of social research. Cambridge, Massachusetts: Schenkman Publishing Company, 1970.
- Ten Brink, T.D. Evaluation: A practical guide for teachers. New York: McGraw-Hill, 1974.
- Tracy, W.R. Evaluating training and development systems. New York: American Management Association, 1968.
- Tripodi, T., Fellini, P. & Meyer, H.J. The assessment of social research. Itasca, Illinois: F.E. Peacock Publishers, 1969.
- Tyler, R.W. General statement on evaluation. Journal of Educational Research, 1942, 35.
- Tyler, L. & Klein, M.F. Evaluation within the context of curriculum development and instructional materials. In G.D. Borcich (Ed.), Evaluating educational programs and products. Englewood Cliffs, N.J.: Educational Technology Publications, 1974.
- Van Maanen, J. The process of program evaluation: A guide for managers. Washington, D.C.: The National Training and Development Service Press, 1973.
- Verner, C. & Booth, A. Adult education. New York: Centre for Applied Research in Education, 1964.
- Vroom, V. Methods of organizational research. Pittsburgh, University of Pittsburgh Press, 1964.
- Walberg, H.J. (Ed.), Evaluating educational performances: A sourcebook of methods, instruments and examples. Berkeley: McCutchan Publishing, 1974.

Warwick, D.P. & Linninger, C.A. The sample survey: Theory and practice. New York: McGraw-Hill, 1975.

Weiss, C.H. Evaluating action programs: Readings in social action and education. Boston: Allyn and Bacon, 1973.

Weiss, C.H. Evaluation research: Methods of assessing program effectiveness. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Welsh, W.A. & Hambleton, R.K. On the use of goals in evaluation: A study of selected issues. Paper presented at American Educational Research Association Meeting, Washington, 1975.

Wolf, R.M. Data analysis and reporting considerations in evaluation. In W.J. Popham (Ed.), Evaluation in education. Berkeley: McCutchan Publishing Corporation, 1974.

Worthen, B.R. & Sanders, J.R. Educational evaluation: Theory and practice. Worthington, Ohio: Charles A. Jones Publishing Company, 1973.

