

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

AN EVALUATION OF A
MANAGEMENT TRAINING PROGRAMME

A thesis presented in partial fulfilment
of the requirements for the degree
of Doctor of Philosophy
in Psychology at
Massey University

Judith-Anne Brook

1982

93-0831

ABSTRACT

The evaluation of a management training programme for research scientists and senior technicians formed the basis of this applied research project. Before, during and after measures were used to assess the knowledge, skills, attitudes and work behaviours of trainees and control group members in order to compare the changes which occurred as a result of the programme. Both formative and summative evaluation were incorporated into the experimental design and careful attention was paid to internal and external validity issues by the use of matched control groups, multiple measures of the dependent and independent variables, including both objective and subjective methods, self-reports and reports by trainees' superiors. Repeated measures of knowledge, skills, attitudes and work behaviours were taken utilizing a longitudinal design and some replication was conducted with other groups of trainees at other times. The plan of approach involved an initial analysis of training needs, examination of the input to training including course content and training techniques, immediate and longer term evaluation of subjects' reactions to the programme and an assessment of the outcomes of training. Outcome evaluation comprised measures of effectiveness at both immediate and intermediate levels with measurements taken immediately and at three, six and twelve months following training. Effectiveness was assessed in relation to the goals determined during the preliminary analysis of training needs, the focus shifting gradually during the course of evaluation from an emphasis on learning effects to changes in work behaviour and performance. Formative evaluation in the form of feedback to course controllers was provided throughout while the summative evaluation consisted of a final reporting of the effectiveness of the training programme over a twelve month period and the consistency of results from a subsequent sample of trainees. A model was suggested to enable the formative and summative aspects of evaluation to proceed simultaneously within a single study. Finally, by monitoring the effects of environmental factors, the writer was able to begin to distinguish the moderator variables operating at various levels of evaluation and move towards construct validation of a theoretical model of training and evaluation.

ACKNOWLEDGEMENTS

I wish to thank Professor George Shouksmith for his supervision and constructive criticism throughout this project. I would also like to acknowledge the support given by other members of the Psychology Department at Massey University, particularly Mr Mike Smith whose interest in the project and advice as second supervisor was much appreciated, and Dr Dave Clarke who provided assistance and suggestions in the initial stages. Sincere thanks are due to Ms Margaret-Mary Oulaghan and Ms Gail Tyson who had the difficult task of typing the final manuscript, and members of the Psychology Department office staff, including Ms Alison Stewart, who assisted with the preparation.

Finally, I would like to thank Dr Richard Brook for his assistance with the analysis of data, and other members of my family for their tolerance and patience.

TABLE OF CONTENTS

	<u>Page</u>
Abstract	ii
Acknowledgements	iii
Chapter 1. The evaluation process	1
1.1 Introduction	1
1.2 Brief history and definitions of evaluation	2
1.3 Some experimental design problems in evaluation research	5
Chapter 2. A theoretical model of evaluation	13
2.1 The Wortman model	13
2.1.1 Formative and summative evaluation	15
2.1.2 Validity issues	16
2.1.3 Political considerations	17
2.1.4 Ethics and confidentiality	18
2.1.5 Randomization and quasi-experimental design	20
2.1.6 Measurement of variables	21
2.2 A decision - theoretic approach	22
2.2.1 The place of value judgements	25
2.2.2 Subjectivity	26
2.2.3 Participant - observation	27
Chapter 3. The technology of training evaluation	29
3.1 Introduction	29
3.2 A practical framework for evaluation of management training	31
3.2.1 Problem-centred training	33
3.2.2 The cycle of evaluation	33
3.2.3 The discovery approach to evaluation	37
3.3 Criterion development	38
3.4 Training evaluation methodology	43

	<u>Page</u>	
Chapter 4.	Introduction to the present research	50
	4.1 Background to the evaluation study	50
	4.2 Aims of the present research	57
Chapter 5.	Preliminary studies	65
	5.1 Introduction	65
	5.2 Identifying training needs	66
	5.2.1 Subjects	66
	5.2.2 Measuring instrument	68
	5.2.3 Procedure	70
	5.2.4 Results	72
	5.2.5 Discussion	75
	5.2.6 Inter-rater reliability	81
	5.2.7 Conclusions	86
	5.3 Construction of attitude scale	86
	5.3.1 Subjects	88
	5.3.2 Attitude scale development	88
	5.3.3 Concluding comments	90
Chapter 6.	Overview of experimental design	92
	6.1 Introduction	92
	6.2 Subjects	93
	6.3 Methods	97
	6.3.1 Context evaluation	98
	6.3.2 Pre-test measures	98
	6.3.3 Reactions evaluation	99
	6.3.4 Immediate outcome evaluation	99
	6.3.5 Intermediate outcome evaluation	101
	6.3.6 Ultimate outcome evaluation	102
	6.3.7 External validity and reliability of scales	103
	6.4 Concluding comments	103
Chapter 7.	Evaluation of immediate outcomes	104
	7.1 Introduction	104

	<u>Page</u>	
7.2	Reasons for attending courses	105
7.3	Expected areas of organizational and personal change	107
7.4	Trainees' assessment of sessions	110
7.5	Evaluation of course topics	112
7.6	Training techniques preferred	122
7.7	Changes in attitudes towards the section leader role	125
Chapter 8.	Follow-up studies	130
8.1	Introduction	130
8.2	Evaluation of intermediate outcomes	130
8.2.1	Effects of work performance after three months	130
8.2.2	Areas of organizational and personal change	137
8.2.3	Allocation of time to topics	140
8.2.4	Indirect measures of effectiveness	142
8.2.5	Perceived understanding and relevance of topics	145
8.2.6	Changes in behaviour after six months - trainees' report	150
8.2.7	Changes in behaviour after six months - superiors' report	160
8.2.8	Organizational climate	165
8.2.9	Twelve month follow-up study	173
8.3	Summary of findings	176
Chapter 9.	Case studies	180
9.1	Introduction	180
9.2	Aim of studies	180
9.3	The repertory grid technique	181
9.4	Method	182
9.5	Results and discussion	185
9.5.1	Case study A	187
9.5.2	Case study B	191

	<u>Page</u>
9.5.3 Case study C	195
9.5.4 Case study D	198
9.5.5 Case study E	201
9.5.6 Case study F	204
9.6 Conclusions	207
 Chapter 10. Replication studies	 210
10.1 Introduction	210
10.2 Aim	213
10.3 Subjects	213
10.4 Procedure	213
10.5 Results	214
10.5.1 Reasons for attending courses	214
10.5.2 Training techniques preferred	215
10.5.3 Attitude scale	216
10.5.4 Performance after three months	218
10.5.5 Allocation of time to topics	219
10.5.6 Indirect measures of effectiveness	220
10.5.7 Understanding of topics	222
10.5.8 Organization climate questionnaire	225
10.5.9 Effects of age, tenure and leadership experience	 227
10.6 Discussion	227
10.6.1 Reasons for attending courses	228
10.6.2 Training techniques preferred	228
10.6.3 Attitude scale	228
10.6.4 Performance after three months	229
10.6.5 Allocation of time to topics	230
10.6.6 Indirect measures of effectiveness	231
10.6.7 Understanding of topics	231
10.6.8 Organization climate questionnaire	232
10.6.9 Effects of age, tenure and leadership experience	 232
10.7 Conclusions	232

	<u>Page</u>
Chapter 11. The effectiveness of the training courses	235
11.1 Introduction	235
11.2 Context of training	235
11.3 Ultimate objectives of training	237
11.4 Training input	239
11.5 Immediate outcomes of training	241
11.6 Intermediate outcomes of training	242
11.7 Organizational environment	248
11.8 Reactions to training	251
11.9 Training techniques	255
11.10 Conclusion	256
Chapter 12. Conclusions concerning evaluative research	257

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
2.1	Wortman's model of evaluation research	14
2.2	Process evaluation as a feedback loop	15
3.1	Training/evaluation control system	32
8.1	Changes in understanding of topics at three times, pre-test (t_1), immediate post-test (t_2) and three months (t_3)	149
8.2	Average number of behaviour changes (1) attempted and (2) implemented, by trainees and controls	153
8.3	Number of behaviour changes made in each behavioural category expressed as a proportion of subjects in the group	159
8.4	Summary profiles of organizational climate for government and semi-government organizations	169
8.5	Summary profiles of organization climate for leaders and non-leaders	169
9.1	Dendrogram of elements for Case A	187
9.2	Dendrogram of constructs for Case A	189
9.3	Dendrogram of elements for Case B	192
9.4	Dendrogram of constructs for Case B	193
9.5	Dendrogram of elements for Case C	195
9.6	Dendrogram of constructs for Case C	196
9.7	Dendrogram of elements for Case D	198
9.8	Dendrogram of constructs for Case D	199
9.9	Dendrogram of elements for Case E	202
9.10	Dendrogram of constructs for Case E	203

<u>Figure</u>		<u>Page</u>
9.11	Dendrogram of elements for Case F	204
9.12	Dendrogram of constructs for Case F	206
10.1	Changes in understanding of topics at three times, pre-test (t_1), immediate post-test (t_2) and three months (t_3)	223
10.2	Summary profiles of organizational climate for government and semi-government organizations	226
10.3	Summary profiles of organizational climate for leaders and non-leaders	226
11.1	The cycle of evaluation including moderator variables	236
12.1	General criterion model of managerial effectiveness	259
12.2	Decision-making model of evaluation	266

LIST OF TABLES

<u>Table</u>		<u>Page</u>
3.1	Training evaluation schemes of Hamblin and Warr, Bird and Rackham	34
4.1	The seven phases of the present evaluation study	59-64
5.1	Subjects who took part in precourse interviews	68
5.2	Number of responses in ten categories of management by four sub-groups of respondents	74
5.3	Number of responses in five categories of personal and group relationships by four sub-groups of respondents	81
5.4	The frequency of incidents placed in categories by the two judges	85
6.1	Subjects involved in main evaluation study categorized by groups (rows) and types of employing organization (columns)	94
6.2	Biographical characteristics of subjects in all groups	95
6.3	Average ages, tenure and leadership experience of trainees and control group members	96
7.1	Pre-test responses to question concerning reasons for attending courses	106
7.2	Subjects' rank ordering of areas of organizational change expected	109
7.3	Subjects' rank ordering of areas of personal change expected	109
7.4	Changes in perceived understanding and relevance following training	115

<u>Table</u>		<u>Page</u>
7.5	Medians and chi-squares of results obtained on the perceived understanding and relevance scales	118
7.6	Number of subjects who indicated preference for each of six types of training methods	123
7.7	Means and standard deviations of attitude scale scores obtained by experimental and control group subjects	127
7.8	Results obtained when attitude scale scores are compared for different groups	128
8.1	Number of trainees who indicated changes in work behaviour	132
8.2	Number of behaviour changes specified by individual course members after three months	133
8.3	Subjects' rank ordering of areas of organizational change following training	138
8.4	Subjects' rank ordering of areas of personal change most affected by the training courses	139
8.5	Subjects' preferences for time allocation to course topics	141
8.6	Changes in trainees' perceptions of understanding and relevance of topics studied during training	147
8.7	Number of subjects who indicated behavioural change following training courses	153
8.8	Retrospective change behaviour reported by subjects six months after training period	154

<u>Table</u>		<u>Page</u>
8.9	The relation between immediate measures of attitude change and longer term change score at six months	155
8.10	Relationships between work behaviour and age, tenure and experience of Group A and B subjects	157
8.11	Correlation between frequency of change and age, tenure and experience	158
8.12	Table showing number of "Yes", "No" and "Don't Know" responses made by fourteen controlling officers about their subordinates' work behaviour	164
8.13	Medians and ranges of ratings made by subgroups of subjects (training Groups A and B, combined)	171
8.14	Group means and standard deviations of training effectiveness on three scales	175
10.1	Modal rankings of reasons for attendance given by training group members	215
10.2	Attitude scale means and standard deviations for groups A, B C and D at times, before (t_1) and immediately after (t_2) training	217
10.3	Number of trainees who indicated change	218
10.4	Number of changes specified by individual course members after three months	219
10.5	Subjects' preference for amount of time allocated to course topics	220
10.6	Subjects who read books related to course topics	221

<u>Table</u>		<u>Page</u>
10.7	Subjects who discussed course with work colleagues	221
10.8	Subjects who maintained contact with other course members	222
10.9	Number of positive changes in understanding of course topics for groups A, B, C and D	224
10.10	T statistics obtained using Friedman's test on subjects' rank ordering of level of understanding of topics	224
10.11	Medians and ranges of ratings made by Group C and D trainees on the climates within their organizations	225
10.12	Relationship between work behaviour and age, tenure and experience of Group C and D subjects	227
Ia	Number of responses, mean score and range on each criterion	268
Ib	Rank-ordering of session assessments on each criterion	269

LIST OF APPENDICES

<u>Appendix</u>		<u>Page</u>
I	Results of session assessment forms of section 7.4	268
II	Analysis of repertory grid data of chapter 9	270
III	Precourse interviews and questionnaires	275
IV	Session assessment form	280
V	Immediate post course questionnaires	281
VI	Three month follow-up questionnaire	284
VII	Six month follow-up questionnaire for trainees and control subjects	287
VIII	Six month follow-up questionnaire for controlling officers	289
IX	Twelve month follow-up questionnaire	293
X	The Kelly repertory grid	294

CHAPTER 1.

THE EVALUATION PROCESS

1.1

INTRODUCTION

The aim of this study was to assess the effectiveness of a management training programme designed for senior scientists and technical staff employed by several industrial and scientific organizations in New Zealand. As an applied research project it provided the opportunity to examine the special characteristics and problems associated with conducting social experimentation within a field setting. An initial examination of the literature revealed that, while there were a number of evaluation models available, there was little agreement on the best way to conduct an evaluation study. A recent review article stated that "Evaluation is a set of theoretical and practical activities without a widely accepted paradigm" (Glass and Ellett, 1980).

In the first three chapters of the report a review of the literature is given. The writer concentrates on those aspects which reflect her theoretical orientation and at the same time provide the best solutions to the practical problems which confronted her. The limitations of the chosen model of evaluation are carefully documented and the constraints of the practical situation are reported at each step. Nevertheless, the writer has chosen to work within this particular framework, acknowledging the practical difficulties, because it promised to provide the most definitive answers to questions concerning both the overall effectiveness of the training programme and the nature of the modifications needed in future programming. The ultimate purpose of the evaluation was to assist the programme organizers to bring the outcomes of training more in line with organizational objectives.

The review begins with a brief historical perspective on evaluation research.

It has become increasingly common in recent years for those who finance, operate and participate in programmes of social change to demand evidence of their effectiveness. Consequently, a variety of intervention techniques, including educational and training schemes are now routinely evaluated in an attempt to estimate their effects, both intended and unintended. Moreover, a number of theorists have begun to look critically at the concepts underlying the practice of evaluation research itself for, in order to justify the considerable time and money involved, it must be shown to possess a sound scientific and theoretical basis.

Writers like Riecken and Boruch (1974) have discussed the multiple facets of this area of applied research and have provided us with a comprehensive survey of the present state of the art. Riecken (1977) observes that the history of evaluation research can be traced back to 1950 when Gordon Allport wrote that "the post-war institutions were entering a period of critical self-examination", but indeed, isolated examples of the experimental investigation of social problems can be found as early as the 1930's. From 1950 until the middle of the 1960's it seems that the main thrust of evaluation in the United States came from major governmental organizations in such areas as public health, education, military training and community development.

Suchman (1967) has presented a summary of evaluation research extending into the mid-60's. He makes a number of significant points concerning its concepts, experimental design, methodology and implementation, among which the following are probably the most relevant to the present thesis. First, he distinguishes between *evaluation* which involves making judgements about the worth of a particular intervention and *evaluative (evaluation) research* which utilizes the methods and techniques of classical experimentation for the purpose of making an evaluation. He notes that, while the former is similar to the common-sense usage of the term *evaluation*, the latter refers to application of principles and procedures of the scientific method to control for bias and subjectivity.

He strongly recommends the latter procedure but he concedes that applied and basic research, while possessing a common underlying logic, each demand their own adaptations in terms of research design and methodology. Indeed, each individual applied project must adapt to the prevailing conditions of the situation being investigated. Since evaluation research is essentially a social enterprise which normally takes place under natural conditions in a field setting the researcher must exercise special ingenuity to "reach the best possible compromise between the demands of science and the realistic conditions of research" (Suchman, 1967, p.161).

This does not imply in any way that scientific controls should be neglected and Suchman enumerates at least six essential steps:

- (1) Identification of goals to be evaluated.
- (2) Development of measurable criteria specifically related to these goals or objectives,
- (3) Analysis of the problems with which the activity must cope.
- (4) Description and standardization of the intervention procedure.
- (5) Measurement of degree of change that takes place by means of setting up a controlled situation to determine the extent to which these objectives and any negative side effects, are achieved.
- (6) Determination of whether the observed change is due to the intervention activity or to some other cause.
- (7) Some indication of the durability of the effects.

He does not entirely rule out "nonscientific" methods for evaluation, but stresses that evaluative research must be more than mere subjective judgement.

The period from the early 1960's to 1974 is regarded as the period of modern social experimentation when major government action programmes as well as treatment of mental illness, education, nutrition and the like were all subject to scientific investigation. This also marks the period when modern social experimentation began to

replace quasi-experimentation and serious attempts were made to isolate causal factors underlying the observed changes (Riecken and Boruch, 1978).

As we have seen, Suchman (1967) makes a distinction between *evaluation* and *evaluative research*. Other writers like Wortman (1975), Riecken (1977) and Weiss (1972) use the terms *evaluation* and *evaluative research* more or less interchangeably. Riecken (1977) provides the following definition of the term evaluation:

"*Evaluation* is the measurement of desirable and undesirable consequences of an action designed to achieve some objective that the actor values. *Action* means a conscious attempt to change individual or group behaviour or psychological state in a valued direction - the negative case being the prevention of negatively valued change A *programme* of action includes one or more treatments, which are the sets of operations or specific steps undertaken to produce the desired effects, and these effects can be called the *objectives* of the programme". Similarly, Weiss (1972) sets out the purpose of evaluation research as the measurement of effects of a programme against specific goals as a means of contributing to subsequent decision-making about future programming. These definitions contain the concept of impact evaluation which Boruch and Gomez (1977) describe as a estimation, in the least equivocal and biased way, of the relative effects of a programme on its target group. Such definitions of the evaluation process imply the use of control groups, competing treatment groups or some other standard of comparison. Another term, *summative evaluation*, is also used to describe research which emphasizes whether the programme accomplishes its objectives (Scriven, 1972). However, as Riecken says, in practice, evaluation may entail either more or less than is suggested by these terms. A full scale evaluation may include an initial survey of the demand for intervention, including a systematic analysis of institutional or organizational needs. These needs and requirements then become the objectives against which the effectiveness of the programme can finally be judged. In addition, evaluation may be extended to cover a detailed examination of the means or techniques used and the complete range of programme operations. This may consist of a simple listing of the facilities provided, topics covered, time allocation and techniques employed, or a precise esti-

mation of the scope, coverage and adequacy of techniques, according to predetermined standards (Riecken, 1977). This type of evaluation, aimed at providing corrective feedback to programme organizers is known as *formative evaluation* (Scriven, 1972).

As it is generally unwise to rely upon the popular meaning of a term in common usage as the basis for scientific study, the present writer proposes to adopt Riecken's broad definition of the term evaluation as a starting point. In referring to the evaluation of a specific programme or project, the writer employs the term evaluation or evaluation study while reserving the term evaluation research for the discussion of the principles of evaluation in more general terms. Moreover, in this writer's view, it is necessary to distinguish between formative and summative evaluation as distinct facets of the process and to specify clearly the procedures and techniques related to each. When these latter terms are used in the text, the definitional distinctions of Scriven (1972) are accepted.

1.3 SOME EXPERIMENTAL DESIGN PROBLEMS IN EVALUATION RESEARCH

Evaluation as a form of applied research is subject to the normal tensions and conflicts generated between the pure and applied fields. Bernstein (1975), discussing validity issues in evaluative research, observes that while the evaluator is expected to supply valid and reliable information which will enable policy-makers to formulate better decisions, she labours under serious constraints not experienced by the theoretical research scientist. Evaluative research is carried out in the world of action which imposes its own set of rules. This presents a problem in terms of research design and there are a number of writers who argue that methods traditionally employed for conducting pure research are entirely inappropriate in this context (Guttentag, 1973). The present writer favours the arguments of other researchers like Riecken and Boruch (1978) who contend that the constraints are similar to those encountered in many other areas of applied psychology. These writers maintain that the evaluator must make certain demands of the client organization, on the one hand, while being prepared to negotiate and compromise, on the other, provided that the concessions do not seriously undermine the integrity of the research design. They reject the idea of adopting a totally decision-oriented model, for

example, at the expense of continuing to develop more valid and reliable methods for measuring change. They stress that good data, both objective and subjective, is needed to make sound decisions regarding the effectiveness of the intervention and this applies to the measurement of both independent and dependent variables. The problem of developing valid criteria of effectiveness is well known and it is at this point that consultation and cooperation between all the groups involved is vital if criterion measures are to be relevant to organizational goals.

Both Suchman (1967) and Riecken (1977) have emphasized that, within the context of the experimental paradigm, evaluation may take the form of either quasi-experimental (before and after measurements) or better still, true experimental design employing randomization and untreated control or comparison groups. The choice depends on the stage of programme development at which the evaluator enters. Ideally, entry should occur at the beginning or early enough so that the following preliminary activities can be carried out prior to the evaluation proper:

- (1) Defining the target population of those who will participate in the programme.
- (2) Obtaining base-line measures of dependent variables such as pre-treatment performance.
- (3) Pilot testing the measuring instruments and treatments.
- (4) Initiating randomized sampling procedures for assigning subjects to experimental and control groups or other methods of treatment assignment.

In practice, this ideal set of conditions rarely exists and the evaluator must be content with something less, usually a form of modified experimental or quasi-experimental design which may involve a matching process for obtaining an untreated control group and a combination of before-and-after measures conducted with the experimental group and with a limited number and type of controls. Frequently, retrospective data only are available, thus placing further limitations on the conclusions which can be drawn. There are many

advantages associated with the use of experimental or quasi-experimental designs which become practicable when evaluation is part of the original intervention plan, and they are all aimed at better experimental control and the elimination of bias. These include the control of participant motivation to give socially desirable responses or to engage in behaviour change simply in response to being a member of a specially treated group (the Hawthorne Effect). Likewise, specification and manipulation of the independent variable can be more tightly controlled under proper experimental conditions. These types of control require the use of multiple control groups.

Both the researchers cited approve the use of mixed strategies for the collection of data and for the measurement of the dependent variable. They advocate collecting subjective, judgemental and clinical information as well as strictly objective data. The present writer agrees that such methods improve the explanatory power of the data and provide the richness and depth of meaning required to understand the complex issues involved.

Bernstein (1975) has enumerated the central problems of research design and data analysis for evaluation. They are described here because they are typical of the problems encountered in the present research.

(1) Research questions.

Evaluation usually begins with the independent variable, for example, the treatment programme. The evaluator is asked to assess how it affects a set of vaguely defined goals. This is clearly an inductive rather than the traditionally deductive approach commonly adopted by basic research in the social sciences where one starts with a dependent variable and generates and tests hypotheses about factors which predict that variable (Dubin 1976, p.18)

(2) Selection of the appropriate population.

In evaluative research the task is to utilize and supplement the available sample so as to rule out potential threats to internal

and external validity (Campbell and Stanley, 1966). Unfortunately, the researcher rarely has complete power to select participants to form the experimental and control groups.

(3) Selection, assignment and maintenance of subjects.

Boruch (1975) stresses the feasibility and importance of randomization. However, it is very rare for the evaluator to achieve complete freedom to assign subjects randomly to even one experimental and one control group much less to have access to sufficient numbers of subjects for multiple control groups.

(4) Multiple points of assessment and reporting.

Bernstein points out that for research which has practical consequences, it is necessary to obtain immediate, intermediate and longer range estimates of performance. The organization may be more interested in long term effects of the treatment than in the immediate outcomes.

(5) Alternative and corrective devices for defective designs.

As already stated, optimum conditions, such as early entry, power and control on the part of the evaluator rarely prevail and a variety of quasi-experimental devices must be utilized to "patch-up" a somewhat less than ideal experimental situation.

Two final discrepancies between applied and theoretical research need to be mentioned. These are the temporal ordering of events and the communication of results to the different target audiences. Wortman (1975) demonstrates that the priorities of the theoretician are different from those of the evaluator. Internal validity, construct validity, conclusion validity and lastly external validity is the natural order for the theoretical researcher, while for the evaluator, the most important is internal validity followed by external, conclusion and construct validity. This need for re-ordering of priorities in itself can pose problems to the evaluator and may become critical when results must be interpreted for a variety of audiences including practitioners and theoreticians.

The Evaluation of Management Training

The above comments are equally applicable to all types of evaluation, including the evaluation of management training with which the present research is specifically concerned. In this context, a report by Grant and Anderson (1977) serves as a starting-point. They argue that the training of employees is sufficiently important to rank with other major programmes of social reform and if the goals of an occupational training programme include the development of more satisfied, productive and affluent individuals then we can expect a wide range of direct and indirect benefits to follow. However, if these expectations are not met by the programme then the large expenditure of time and money is unjustified. In order to make informed management decisions concerning the training of employees it becomes imperative to assess training impact and if the quality of these studies is to be improved then evaluation should be thought of as an integral part of the total training programme. It should include both formative as well as summative evaluation, preliminary surveys by non-experimental methods as well as a quantitative measurement of outcomes, periodic programme monitoring procedures and longer term follow-up estimates of effectiveness measured against external criteria. The authors of this article cite a number of evaluation studies which illustrate some of their recommendations.

The evaluation exercise becomes particularly complex and difficult when we are considering the training of management personnel. Measuring the effectiveness of management training poses a special set of problems in addition to those associated with the evaluation of other types of training. All the difficulties of establishing suitable criteria for job performance are magnified for managerial positions and conventional methods of jobs analysis cannot be applied in the same way. In a sense, every managerial role is unique. In trying to define the characteristics of a "good" manager one finds that critical features include such factors as decision-making skills, effective interpersonal relationships and the ability to adapt personal leadership style to the particular task. Measures of these attributes are notoriously unreliable. A major problem associated with the assessment of managerial performance is the relatively unstructured work environment experienced by managers compared with employees engaged in production

work. In fact, a person in the managerial role must be able to cope with a wide variety of different tasks demanding an extensive range of talents, skills and behaviour.

Moreover, when the researcher tries to observe and measure managerial behaviour she must cope with the irregularity with which the different situations arise. For example, if a time sampling approach is adopted it is found that some behaviours occur infrequently or not at all. When they do occur their duration may be variable or extremely short. Since the managerial role is concerned primarily with people rather than with objects, the measurement problem is entirely different in nature from the measurement of, say, mechanical skills. Undeniably, the impact of interpersonal skills is much more difficult to assess than the interaction between man and his physical environment.

As one means of dealing with the problems of evaluation of management training, Burgoyne (Burgoyne, 1973; Burgoyne & Cooper, 1975) suggests that since evaluation is rarely an end in itself but rather a part of a larger attempt to make organizational decisions, changes or choices, it must be defined in broad rather than in narrow terms, and he suggests that the action research orientation is particularly appropriate. Action research differs from discipline research in two important respects:

- (1) Results generated are used to influence the situation which is being researched.
- (2) The action researcher monitors the changes brought about by his intervention in the research situation.

Although Cherns (1969) has emphasized this second feature of action research Burgoyne acknowledges that in management development, the term is commonly used to refer to the first condition, namely, influencing the situation. Thus in an evaluation study the repeated measures may be used for two purposes, for feed-back to programme controllers and trainers and as indices of effectiveness. Burgoyne quotes Rackham, Honey and Colbert's (1971) method of training interpersonal skills where expert observation and recording of the frequency

and type of interaction between group members is utilized to both influence the experimental situation and to monitor the changes brought about by the intervention. This, according to Burgoyne, represents economy of research effort but has the disadvantage that further consequences of change, as a result of feedback, are not monitored by the researcher.

This example is further evidence of the tension which exists between the two types of research. It frequently requires the evaluator, on the one hand, to make value judgements regarding the amount of control she demands and, on the other, to live with a measure of ambiguity. It seems that this type of dilemma is an inevitable consequence of evaluative research and one which must be resolved if the formative evaluation function is to be maintained. In fact, the present writer sees a very close relationship between action research and formative evaluation to the extent that the latter can be fully implemented only within an action research context. However, as Bonoma (1977) comments, such negotiations occur within a power structure created by the interaction of the various groups concerned, including the sponsors, evaluators, administrators, trainees and training personnel. In a similar vein Argyris (1976, p175) draws a distinction between mechanistically and organically oriented research. The latter encourages client participation in defining and modifying goals, designing instruments and choosing research methods and strategies of change, while at the same time it provides feedback to facilitate the desired changes and development.

The present writer favours an action research approach to evaluation, judging it to be the most effective means of coping with the complexity and multidimensionality of the criteria to be examined as well as going some way towards meeting the needs of the client organization. However, it must not be permitted to detract seriously from experimental rigor and objectivity. The design must be sufficiently controlled to permit variables to be operationally defined, hypotheses made explicit and testable and statistical analyses performed to achieve both practical and theoretical goals. In the next chapter the writer will describe a model which she believes, fulfils these basic requirements even when, as usually happens, it must be

modified in order to accommodate the practical situation. This is contrasted with a second model, the decision-theoretic approach which in this writer's view is less adequate as a complete model of evaluation.

CHAPTER 2

A THEORETICAL MODEL OF EVALUATION

2.1

THE WORTMAN MODEL

In a paper published in 1975, Wortman presented an ideal model of evaluation which emphasizes those aspects of experimental design characteristic of the scientific hypothesis-testing approach including internal and external validity, randomization and the use of control groups. Moreover, the model adopts a systems orientation by establishing a series of feedback loops from each successive stage of the evaluation process. Wortman identifies six major processes. They are summative evaluation, formative evaluation, construct validity, internal validity, external validity and conclusion validity, and he describes how they are integrated into a descriptive model of evaluative research. His model, illustrated in Figure 2.1 shows how the evaluation process can provide a complete system of evaluation and feedback to all of the parties involved.

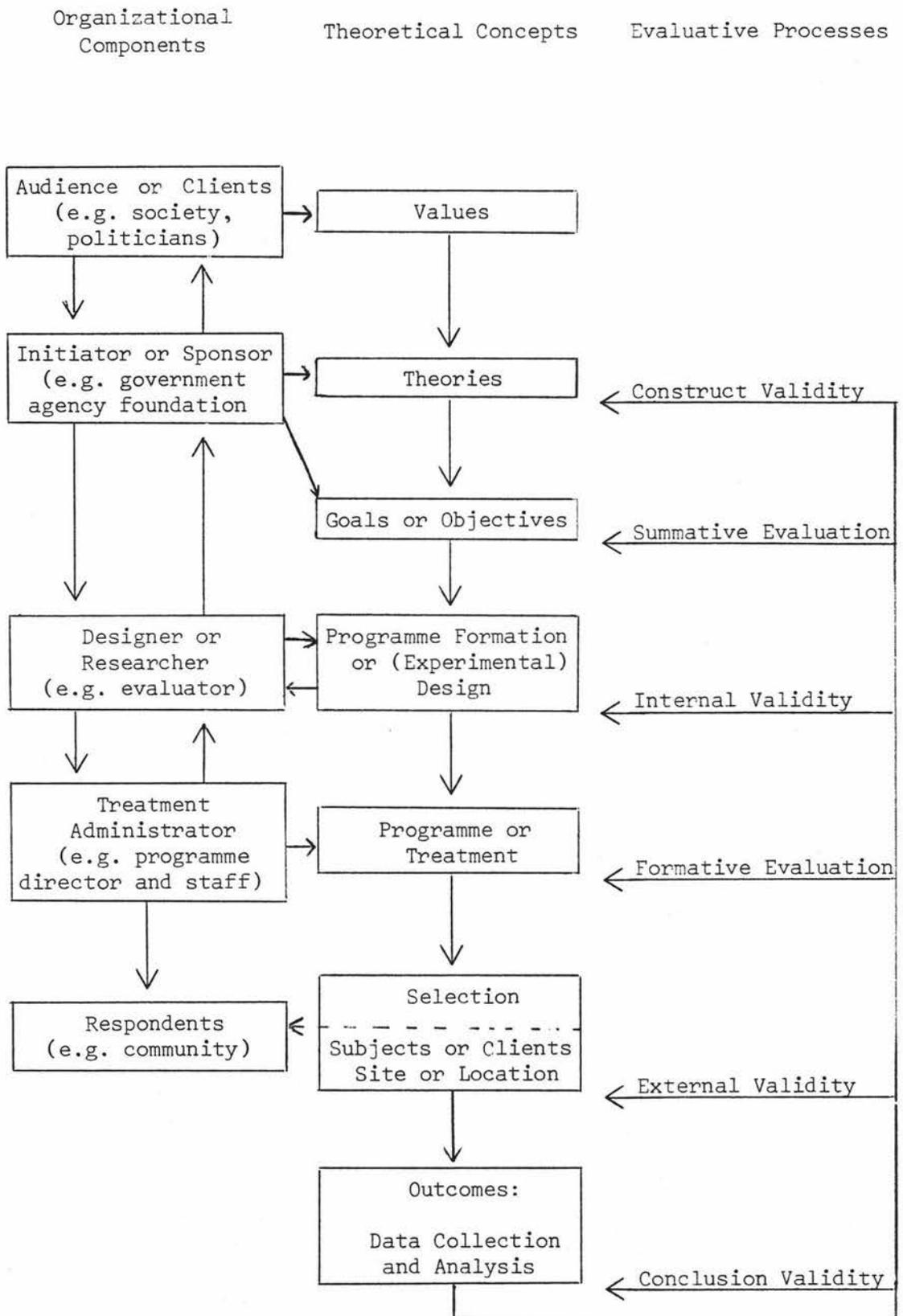


Figure 2.1 - Wortman's model of evaluation research

A key figure in this research paradigm as it operates within the field setting is the evaluator who is frequently a professional psychologist. In an effort to define the role of the evaluator within this general model Williamson, Prost and George (1978) have described the role as one of "providing feedback on system functioning" in accordance with the process model, illustrated in Figure 2.2.

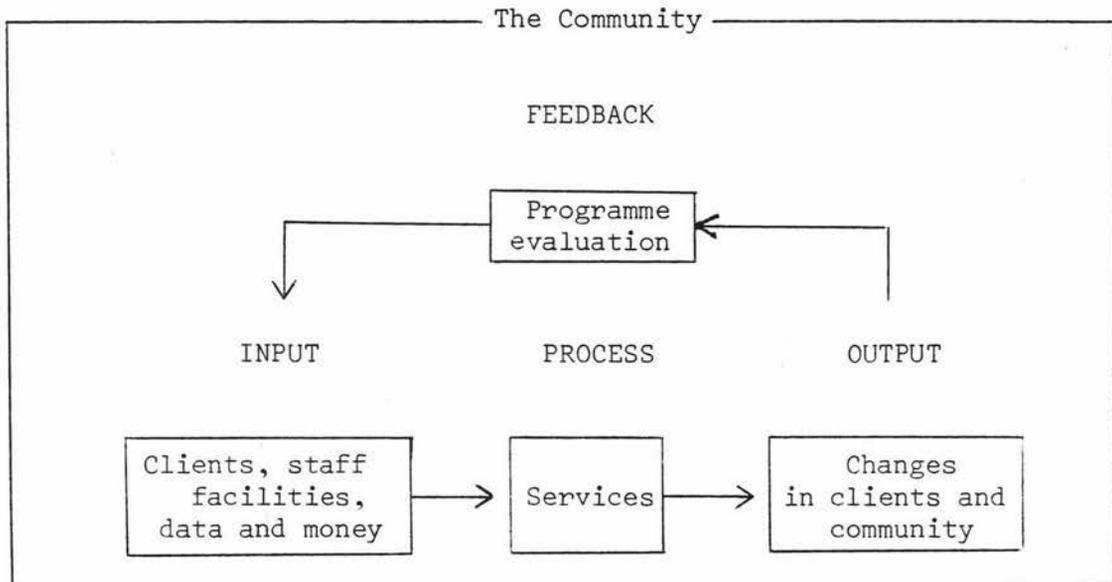


Figure 2.2 - Process evaluation as a feedback loop

Williamson et al. (1978) have translated the evaluative processes in the Wortman model into six stages of evaluation:

- (1) General Effectiveness (unrelated to specific goals and a minimal level of evaluation).
- (2) Means-Ends Analysis (formative evaluation).
- (3) Internal Validity (including conclusion validity).
- (4) Goal-Outcome Congruence (Summative evaluation).
- (5) External Validity.
- (6) Construct Validity.

2.1.1 FORMATIVE AND SUMMATIVE EVALUATION

The concepts of formative and summative evaluation were defined by Scriven (1972) and have been adopted by Wortman (1975).

Formative evaluation is concerned with the *process* of intervention which, in this case, is training and it is most important in the early stages of programme development. It involves testing, analyzing, evaluating, modifying and retesting the independent variable and is likened by Wortman to the pilot testing phase of experimentation.

Formative evaluation focuses specifically on the process variables and considers the degree of match between programme emphasis, such as training techniques, and the identified programme goals. It includes "means-ends" analysis in the Williamson, Prost and George (1978) terminology. Whenever such ongoing monitoring and feedback is supplied to trainers, programme organizers or participants it can be regarded as a type of formative evaluation even though actual modifications may be delayed until after the initial evaluation has been completed. Clearly formative evaluation is most readily implemented within an action research framework where the feedback and monitoring can occur continuously.

Summative evaluation, in contrast, is wholly concerned with the operationalization of goals and analysis of programme performance in relation to those goals (goal-outcome congruence). Hence the emphasis is on outcome variables. Williamson et al. refer also to "general effectiveness" which is the measurement of change, unrelated to any specific goals and is a first and minimal level of evaluation.

2.1.2 VALIDITY ISSUES

The Wortman model includes four types of validity essential to the evaluation process. Internal and external validity are familiar concepts and Campbell and Stanley (1966) and Cook and Campbell (1976) have examined at least twelve "threats to validity", together with a series of quasi-experimental designs to eliminate or control them. *Internal Validity* is concerned with making correct inferences about the causal relationships between treatment and response variables thereby ruling out other alternative explanations such as prior history or natural maturation processes. Experimental control in the form of control groups, representative sampling strategies, random assignment of treatments and subjects are means of eliminating such systematic

biases. *External Validity*, on the other hand, refers to the extent to which the observed results can be generalized to other populations, settings and treatment variables and this involves both representative sampling and replication at other times and places, with different subjects and measures.

Wortman adds two additional types of validity to these. *Conclusion validity* is a type of internal validity but, rather than arising from non-randomization, it is the result of errors introduced by small sample size, differences in the administration of training or treatment techniques, unreliable measuring instruments or inappropriate statistical tests. Wortman proposes standardized treatments and replication as two methods for combating such threats to conclusion validity. *Construct validity* refers to the operationalization and implementation of theoretical constructs. It involves accurately translating the theory into independent and dependent variables and establishing the correct conceptual linkages between them. Again, it is by means of randomly assigning subjects to experimental and control groups, by controlling for experimenter and response biases and by using multiple measures of treatment and response variables that alternative explanations for observed effects can be ruled out.

2.1.3 POLITICAL CONSIDERATIONS

On the left-hand side of the diagram, Figure 2.1, Wortman presents the organizational components of the feedback model. The present writer would endorse the general feeling that the mounting of an evaluation experiment is a "political act" (Wortman 1975) and an activity which takes place within a social context involving a wide variety of interest groups (Bonoma, 1977).

These various parties may include programme sponsors, the evaluator, administrators, training staff, trainees and their work colleagues. The evaluator can recommend changes suggested by the results of the evaluation but the change itself is generally initiated from other sources even in an action research context. Wortman provides a useful description of the roles of these various parties and the complex relationships between them.

Bonoma (1977) observes that evaluation, by its very nature, is a critical judgement and one of the aims must be to minimize the defensiveness of participating parties. This defensiveness can be reduced by stressing the formative function of the evaluation and by placing emphasis on improving the training programme. Cooperation rather than rivalry is essential from the earliest stages of programme development and the writer would extend this collaboration to subjects as well in certain circumstances. Although this solves many managerial problems, it does not eliminate the possibility of a biased assessment of treatment effects which is a constant threat to the evaluator.

Referring again to the process model, Figure 2.2, it is clear that the conduct, analysis and application of evaluation data all form part of a closed feedback loop which, in turn, is part of a much larger system incorporating other levels of the organization and of the surrounding environment. However, as Wortman points out, there is ultimately a single, discrete decision which has to be made concerning the effectiveness of the total programme and this implies a summative evaluation. As changes in behaviour may occur slowly, a longitudinal study is required with numerous measures taken over a period of time. The long-term nature of such a project creates its own problems. While a short-term study may not allow sufficient time for the intervention to achieve its goals, in a long-term study the impact of feedback tends to become diffused. The adoption of a systems approach enables the evaluation to be tailored to the situation and it is more likely that both formative and summative evaluation may be accomplished within the same framework. Thus, for the present researcher the systems approach of Wortman is the preferred model and an action research orientation is chosen because it is more conducive to the formative aspect of the evaluation process.

2.1.4 ETHICS AND CONFIDENTIALITY

It may be supposed that many of the ethical problems, especially those associated with possible harmful side effects, are less of an issue when we are considering something like management training. However this writer has found that it is not possible to foresee all possible side effects and there are some unanticipated outcomes which

may be undesirable. These include dissatisfaction with self, dissatisfaction with the job or loss of freedom to make future career choices. Policymakers, training staff and researchers must be alert to these possibilities. Riecken and Boruch (1978) maintain that those who develop programmes should, before initiating them, be satisfied in their own minds that the programme benefits will outweigh any foreseeable adverse side effects. The present writer shares the opinion that these sorts of issues inevitably involve value judgements and can best be handled by open and systematic discussion.

Another contentious point in many delivery programmes is the failure to provide the benefits of treatment or training, if indeed they do prove to be beneficial, to the control group. It is suggested that one way around this problem is to offer alternative treatments which are equally good or to institute a system of delayed treatment for the group used initially as controls (Riecken and Boruch, 1974, p.249). It is an established principle that in any experiment involving human beings, the informed consent of the subjects must be obtained but because subjects' prior information about the experiment can seriously undermine the validity of results, decisions need to be made concerning how detailed the information is to be. Finally, it is imperative that privacy and confidentiality be maintained at all costs. This is especially important when subjects are asked to divulge information or opinions which they may not wish to share with their superiors or other members of the organization. Thus, all the usual means for protecting the privacy of the individual must be implemented, but at the most basic level, it is the evaluator herself who is responsible for the good physical security of the data collected and the protection of individual anonymity and confidentiality.

Maintaining anonymity of the subjects is less difficult when, as in the present case, most of the interest is directed towards group results rather than the performance of individual subjects but as many recent texts in experimental psychology remind us, all data collected in behavioural research should be considered "privileged communication" (Matheson, Bruce and Beauchamp, 1974, p.198). The present writer stresses the importance of maintaining an ongoing dialogue with the subjects involved by means of interview and letter as well as

eschewing any form of deception. Moreover, the informed consent to participate in the testing must be obtained from the subjects even when that means a degree of sample attrition in certain phases of the experiment. Finally, some form of periodic face-to-face contact between experimenter and subjects not only promotes openness and honesty but provides an excellent opportunity for debriefing and the correction of any misunderstanding or mis-information.

2.1.5 RANDOMIZATION AND QUASI-EXPERIMENTAL DESIGN

As we have seen, many authors support the use of randomized experiments for evaluation purposes. By shifting the source of error from systematic to random, one can reduce the plausibility of alternative explanations and increase confidence that the observed effects can be attributed to the training or treatment. The aim is to maximize the equivalence between experimental and control groups and only randomization is completely satisfactory in this respect. Boruch and Rindskopf (1977) argue strongly for the feasibility of this type of design and quote numerous examples of its successful implementation. However, in field experiments, there is always the possibility that randomization may be improperly implemented. As a less favoured alternative, they suggest the use of quasi-experimental techniques providing that the researcher has sufficient administrative power to structure the evaluation accordingly, pointing out that estimates obtained by quasi-experimental means should be regarded as tentative and subject to possible competing explanations. They do admit, however, that it is not uncommon for a good researcher to use a variety of methods in combination during the course of an evaluation.

Wortman, generally speaking, supports this view-point, reinforcing the notion that randomization is neither unethical nor impossible in many cases. However, he warns against assuming that even randomized experimentation will be trouble-free for there may be problems in gaining co-operation to implement multiple treatments non-selectively. Moreover, randomization without deception of subjects and experimenter, the double-blind situation, produces experimenter effects and a reactive situation which in itself introduces new biases (Rosenthal, 1966). As an alternative he suggests more "participatory"

experiments. With the type of subjects in the present evaluation, this became a feasible option. Also relevant to the present research are the difficulties of small sample size and the resultant loss of reliability of measurement (Cantril, 1944).

2.1.6 MEASUREMENT OF VARIABLES

On this critical topic, the writer adopts a stance which is similar to that of a number of previous researchers whose view-points are summarized here. Boruch and Gomez (1977) argue that in an evaluation experiment, it is necessary to establish from both theory and data, firm linkages between response and treatment variables. The operationalization of these variables for a management training programme, as for other similar interventions, is a complex task. It involves finding means of assessing both treatment conditions and response variables. The choice lies between using standardized tests which have been field-tested, are relatively stable and well validated or developing specialized instruments which are likely to be more responsive to programme effects and more relevant to the group and to the situation under study. Riecken and Boruch (1978) point out that standardized tests are designed to maximize individual differences but tend to be less sensitive to group differences and therefore to the influence of the treatment programme. On the other hand, developing specialized instruments is a difficult, expensive and time-consuming exercise and requires adequate lead-time for piloting and pre-testing measures plus a lengthy follow-up period for validation procedures. A thorough developmental programme of this type is rarely possible but there are a number of useful approaches which can be adopted as a practical compromise. One example is the practice of utilising a mixture of standardized and specialised measuring instruments where commonsense suggests that one or other is appropriate.

The use of repeated measures, a common practice in this type of research, means that the instruments are more sensitive to treatment programme effects than to individual effects since repeated measures are more precise for detecting group changes but less adequate for individual changes. Other devices suggested by these authors include unusual, non-reactive or unobtrusive measures of the dependent variable and the replication of the tests with other groups

at other times which, as Wortman says, not only furnishes reliability data but can also provide information on external validity where different groups of subjects are used. Because of the uncertainty of the assumed linkages between treatment and response variables, especially in the area of management training these writers suggest that multiple indications of all variables should be gathered wherever possible.

One of the problems associated with measures of response is their insensitivity to floor and ceiling effects which, according to Riecken and Boruch, decrease the power of the test to detect overall changes. This is something that the researcher must be aware of particularly where the sample is not homogeneous. Another is the problem of attrition where control group members may tend to drop out faster than experimental subjects. This is largely a motivational problem. To prevent such changes in group composition, efforts can be made to encourage co-operation by additional face-to-face contact and reminder letters especially in the latter stages of the study.

Riecken and Boruch (1974 and 1978) refer to the use of criterion-referenced measures in the form of specific behaviours as an aid to maximizing the relevance of the response variable. They advocate that treatment effects be evaluated by objective, observable behaviour or actions as far as possible and that appropriate behavioural statements be developed for questionnaire protocols, lessening the distortions caused by subjects' imperfect memory and inaccurate recall of events. They recommend that when direct measures of behaviour are not feasible then one must use non-behavioural measures which "closely reflect behaviours under study". Moreover, non-behavioural measures may be supplemented and combined with behavioural measures. They conclude that in common with other social experiments the measurement of variables in evaluation involves a complex trade-off between ease of measurement, conceptual appropriateness and overall convincingness.

2.2

A DECISION - THEORETIC APPROACH

The writer has chosen to contrast the model described in the foregoing discussion with an alternative model proposed by Guttentag (1973). Guttentag's Decision - Theoretic Approach, based on the multi-

attribute method of Edwards (1971) consists of two stages:

- (1) Determining and placing in order of importance all the values or utilities (positive and negative) of each possible course of action as well as deciding upon the overall utility of the programme.
- (2) Making judgements about the utility (probability of success) of the various courses of action.

Guttentag recommends the use of Bayesian statistics to achieve this. By employing Bayesian methods of analysing the data the various subjective hypotheses of the decision-maker are compared with each other rather than attempting to make a direct estimate based on the null hypothesis. Under this scheme the value dimensions are first of all assigned weightings appropriate to their order of importance, a form of goal setting. According to Guttentag, these value dimensions or criteria may be purely objective but whether objective or subjective are "measured" on a common scale (0-100) from the minimum to maximum plausible value.

The overall utility of any action U_i , is then calculated by combining these subjective utility estimates, U_{ij} in the following manner:-

$$U_i = \sum_j W_j U_{ij}, \text{ where } W_j \text{ is the value attributed to each dimension.}$$

The model has very little to say about the use of objective data other than to suggest that subsequent to the above procedure, objective data should be gathered to replace subjective judgements wherever possible. As this occurs after the utility decisions have been made one wonders how, in practice, this information is integrated into the decision-making strategy at all. Nor is there any explicit method suggested of establishing the validity of such data. The chief advantage of the decision-theoretic approach, in the writer's opinion, is that it forces the value judgements of the different groups out into the open for discussion. It may overcome the problem afflicting other approaches such as the experimental approach, that while they appear to be value-free, in fact, they have merely failed to make explicit the value premises leading to the final decision.

Guttentag, quite rightly, raises issues concerning the dissatisfaction which many people have felt with earlier efforts to evaluate social programmes but incorrectly, in the writer's opinion, blames the use of the classical research paradigm. She correctly identifies evaluation as a judgemental process which ultimately must be made by decision-makers other than the evaluator and points out that programme goals are never as clear-cut and specific as the researcher would like them to be. Furthermore, as explained in an earlier section, the evaluator does not normally have complete control over the formulation of her own hypotheses but rather is engaged in an inductive research activity where evidence is accumulated leading to conclusions about the effectiveness or otherwise of a programme already in existence. Certainly, as Guttentag suggests, the programme may be in a state of development or, at least the sub-goals may change during the course of the evaluation but the present writer contends that it is part of the evaluator's job to differentiate between formative and summative evaluation. In so far as a summative evaluation of outcomes is called for, then the evaluator should ensure that at least a minimum level of stability be maintained throughout the evaluation period. Where this is not feasible then it may be possible to provide a series of evaluative statements during the life-cycle of the programme. Guttentag argues further that the evaluation cannot fulfil most of the Campbell and Stanley (1966) criteria for validity because randomization is often impossible but she ignores many of the options provided by various quasi-experimental designs.

Finally, Guttentag notes that classical procedures of statistical inference are violently biased against the null hypothesis and, "in evaluation research, even if the assumptions of classical design could be met then statistical tests of significance could lead to faulty inferences". She uses this as an argument for the exclusive use of Bayesian statistics ignoring all other techniques from classical statistics which would be equally applicable. She accuses evaluators, therefore, of adopting inferior strategies such as studying superficial aspects of the programme or employing descriptive methods leaving the final decisions to the organizers. Although she offers the decision-theoretic approach as an alternative to the traditional model of evaluation, critics like Wortman and Muirhead (1977) and Apsler (1977) argue that the decision-theoretic model merely concentrates on the planning,

decision-making phases of evaluation research and is therefore simply one aspect of the total process. Like Bonoma (1976), the present writer disagrees with the notion that evaluation research is a unique pursuit, distinct from normal social psychological enquiry or that traditional scientific methods do not possess the experimental tools necessary to cope with it. In fact, Guttentag's attitude tends to promote a restricted view of field experimentation and to deny the diversity which earlier researchers like Campbell and Stanley (1966) and Cook and Campbell (1976) have introduced by means of quasi-experimental design.

Moreover, Bayesian statistics, while they do lead to stronger conclusions, demand more assumptions about prior probabilities which in themselves require subjective decisions and therefore could vary with the researcher. The present researcher prefers the use of non-parametric tests which, while demanding fewer basic assumptions, are nearly as powerful and do not depend on the researcher's prior beliefs.

2.2.1 THE PLACE OF VALUE JUDGEMENTS

A more convincing argument in favour of Guttentag's decision-theoretic approach and one which the present writer would endorse is to be found in her plea to allow the value judgements of the decision-makers to be integrated into the data gathering process and fed back into the system at frequent intervals. This is formative evaluation. However, this seems to be logically part of the planning, objective-setting and monitoring functions of evaluation rather than a complete model. As such, the sequence of events suggested by Guttentag has a useful contribution to make. The steps she suggests are typical of a more general decision-making model and therefore may be seen as one aspect of evaluation. The model does not specify how one obtains the objective, accurate and unbiased data needed to make good decisions.

Guttentag's method does describe, however, how important value dimensions can be generated by group discussions and collaboration between all the parties concerned, so that evaluative research will be more closely linked to the values of the decision-makers. This should be the aim of all goal-setting activities in evaluation and is one of the purposes of the preliminary interviews in the present research.

Smith and Kendall (1963) first suggested consultative procedures for determining the dimensions of managerial-type jobs and since then it has been used widely in behavioural expectation scaling and retranslation techniques for the development of behaviourally-anchored rating scales. In his defence of the experimental paradigm Apsler (1977) observes that most programmes have been funded to meet some quite specific needs and it is rather the procedures for meeting these needs that have been less well-defined. Apsler warns, as well, against ignoring the unanticipated outcomes which neither the evaluator nor the decision-maker may be in a position to forecast but which may, subsequently, be incorporated into a more complete model of the type that Wortman suggests. It seems to this writer that neither the decision-theoretic approach nor Bayesian statistics per se hold the key to better evaluative research but rather more concentration on good criterion development by paying careful attention to the needs or objectives of the programme together with more adaptable and flexible approaches to both classical and quasi-experimental design and statistical analysis.

Certain aspects of this adaptability and flexibility are illustrated in two techniques discussed by Glaser and Backer (1972, 1973) and considered briefly here under the headings of *subjectivity* and *participant-observation* in evaluation research.

2.2.2 SUBJECTIVITY

Weiss (1972), after reviewing ten evaluation projects observed that it is not always possible to adopt rigorous quantitative approaches such as structured interviews, objective measuring instruments and "true" experimental interventions in evaluation programmes and one must look for viable alternatives or supplementary means of collecting evaluative data. The so-called clinical approach to programme evaluation adopts a holistic orientation, perceiving the programme as an entity with complex interrelationships, which attempts to measure the reactions, thoughts, beliefs and perceptions of all those concerned with the project. It adopts the view-point that a long-term project, in particular, is a dynamic rather than a static entity where ongoing feedback on changes which occur becomes important information for course controllers and evaluators. Moreover, the judgements and opinions of the people involved in the programme, be they staff, administrators, trainers or

trainees, are some of the most meaningful data. Glaser and Backer (1972, 1973) note that this type of subjective orientation may be useful for obtaining follow-up data where, for example, in a management training programme, the changes in management style may have become so diffuse after a year that objective measurement alone may not provide an adequate criterion of success.

Another example of this technique occurs in Kirkpatrick's (1967) report on the evaluation of management trainees where subjective measurement techniques were incorporated within a standard experimental design. In this case the researcher used questionnaires which included both objective and open-ended questions.

Glaser and Backer end with a caution, noting that subjective measures should be used sparingly and always augmented with other evidence, less subject to bias and perceptual distortion. They emphasize the importance of ongoing consultation between evaluators and all other parties involved not only to define criteria but also as a "political move" to help reduce defensiveness and to defuse some of the tensions associated with evaluation. However, Weiss (1972) warns that such consultation may also serve as a source of bias.

2.2.3 PARTICIPANT - OBSERVATION

When the main purpose of the evaluation is to provide constructive, corrective feedback, that is formative evaluation, participant-observation is a particularly suitable technique. Glaser and Backer (1973) point out that this needs to be combined with other more objective and definitive forms of evaluation to facilitate decisions which must be made about whether programmes should be continued, modified or curtailed.

Participant-observation is described as, "a systematic sharing of the life activities and interests of a group of persons". It studies a process or an environment by observing and experiencing it in depth and it is a naturalistic approach to data collection. They list a number of advantages and disadvantages of this approach when used in evaluation research:-

Advantages

- (1) It enables information to be obtained by direct observation rather than by post hoc reports.
- (2) It provides a comprehensive and complete picture of the programme including the richness and detail of first-hand observations.

Disadvantages

- (1) Sometimes the recording of events occurs at a later date and is therefore subject to memory biases and distortions.
- (2) It is difficult to quantify the perceptions and observations gathered.
- (3) There may be loss of objectivity if the observer takes on the biases of the programme staff and recipients.
- (4) The presence of an observer may cause staff or subjects to behave atypically.
- (5) It is time-consuming and expensive.

Subjective techniques deserve careful consideration because they tend to lend more flexibility than traditional experimental methods. It is not only an attempt to deal with the practical problems associated with evaluative research but also an acknowledgement of the dynamic quality and multidimensionality of the phenomena under study and an attempt to develop a more complete model to explain relationships between variables.

The present writer concedes that this emphasis on subjectivity takes the researcher onto very dangerous ground and many theoretical researchers would prefer to avoid it altogether. This is a luxury which the applied scientist cannot afford and she must make decisions which will minimize the loss of precision while optimizing the gains in practical utility. These matters are pursued further in the next chapter where we consider the more technical aspects of evaluation to be found in the writings of several British and European researchers.

CHAPTER 3

THE TECHNOLOGY OF TRAINING EVALUATION

3.1

INTRODUCTION

In a review of evaluation methodology, Burgoyne and Cooper (1975) note that it is at the stage of choosing the methodology for an evaluation research study that one must endeavour to integrate the sort of conceptual, technical and practical issues which we have been discussing so far. The two critical questions posed by Campbell, Dunnette, Lawler and Weick (1970, p.271) about training evaluation are:-

- (1) What kind of information about training effects can we obtain?
In other words, what are the criteria or objectives of training?
- (2) How should we go about obtaining it? Here we must consider both experimental design and the measurement of variables.

Most writers agree that the purpose of evaluation research is to gather information to improve the quality of decisions made in the practical world. However, Glass and Ellett (1980) explain that there is, at present, a wide variety of approaches to this task. As we have seen, people like Wortman, Boruch and Riecken adopt a pure research approach, emphasizing strict adherence to experimental design after the fashion of Campbell and Stanley (1966) utilizing control groups, randomization, and representative sampling in order to ensure internal and external validity. Some members of the European school such as Hamblin (1974), Warr, Bird and Rackham (1976) and Burgoyne and Cooper (1975) tend to offer a more pragmatic framework for the evaluation of training. They propose a trade-off and a compromise between increasing quality of information and the cost, duration and utility of the research. Like Guttentag, they are mindful of the realities of the world of the decision-maker but their solution to the problem is somewhat different. Nevertheless, their emphasis is on the utility of evaluation in the sense of providing information at the time when it is most needed.

Burgoyne and Cooper (1975) like many of their American counterparts propose a systems approach to evaluation aiming to provide an

effective feedback link of two types. Firstly, they seek to obtain information which will check, test and improve basic assumptions. A second aim is to obtain up-to-date information on the present situation. While the first is a theoretical, the second is a practical goal. The latter is designed to aid immediate and on-going decision-making. As the authors explain, the two aims have different but not necessarily incompatible requirements. These are:

- (1) Reliable, valid, controlled studies of enduring utility (summative evaluation and construct validity).
- (2) Immediate and continuous feedback fulfilling a monitoring function (formative evaluation).

The choice between which of these should be emphasized raises questions about whose values, the sponsors, the training administrators, or the trainees, are to guide the research and who should control the project. The point is made that the researcher-evaluator cannot hope to avoid all value judgements in order to maintain scientific objectivity, partly because there is no scientific evidence to support the notion that managerial effectiveness is a single global concept (Campbell et al., 1970, p.282). Thus the objective criteria chosen for any training evaluation study are related to the particular task and situation at hand and the values of the organization are as important to the researcher as they are to the decision-maker.

Burgoyne and Cooper raise other important issues concerning the subjects in the study. In particular, there is the question of whether the self-reports of trainees are able to provide valid and reliable data. The controversy of the subject as "patient" or "agent" is reminiscent of the subjectivity issue discussed by Glaser and Backer (1973). A subject in the patient role is passive and reactive while a subject in the agent role is active and purposive.

The arguments presented by Burgoyne and Cooper suggest, to this writer, that there is no fundamental conflict between the Wortman model of evaluation and the methodology of these researchers from the European school. Thus the present writer decided to attempt to adopt the practical framework of Hamblin (1974) and Warr et al. (1976) within the general philosophical and theoretical approach espoused by Wortman, Boruch and Riecken as discussed in Chapter 2. In the remainder of this chapter the two questions concerning the type of information and the method for

obtaining it are discussed in relation to a number of practical examples of evaluation which have influenced the direction of the present research.

3.2 A PRACTICAL FRAMEWORK FOR EVALUATION OF MANAGEMENT TRAINING

Hamblin (1974) and Warr, Bird and Rackham (1976) have offered a scheme which will enable the evaluator to answer the basic pragmatic questions posed by programme administrators. Hamblin (1974, p.9) comments that it is not a question of whether or not training *can* be evaluated but rather *how* it should be evaluated. In some cases, it may not be appropriate for a formal evaluation to occur at all. For economists, accountants and even some administrators, the only relevant evaluation is a financial one but for many organizations involved in service and research and, indeed, for some profit-making organizations there are other important objectives such as employees' career development and work satisfaction. Even in cases where the ultimate objectives are economic ones, there may be several intermediate steps leading to the economic goal. Rightly or wrongly, there may be an assumption that, for example, better human relations practices or communication will have financial spin-offs in the long run. It must also be remembered that in management training, different people including trainers, administrators, the trainees, their superiors, colleagues and subordinates all have different goals. The pragmatic questions to be answered in a training-evaluation system are, according to Hamblin:

- (1) What needs to be changed?
- (2) What procedures are most likely to bring about this change?

These two questions should be considered before training begins, but in fact, may be vague and ill-defined at that stage.

- (3) What evidence is there that change has occurred?

The answer to this question is, in the Wortman model, the first step and minimal level of evaluation. It is unrelated to any specific goals and is simply an assessment of general effectiveness. When we analyse

this third question we see that there are a further three questions to be answered.

- (4) How can we improve the criteria by which training is evaluated or more specifically, what are the *right* training objectives?
- (5) How can we improve our methods for gathering the information on the basis of which we evaluate training?
- (6) How can we use the evaluation to improve the training?

These questions are implicit in the latter steps of the Williamson, Prost and George (1978) sequence and include internal validity, goal-outcome congruence, external validity and construct validity. Basic to the whole process is the delineation of clear, objective criteria or goals of training and evaluation. However, Hamblin (1974, p.10) also explains that the latter three questions concerning evaluation are interrelated and that it is pointless to select criteria which are unmeasurable or which cannot be used to improve training. He says that evaluation is not an "end in itself" but is a means of improving training. It is the "art of the possible" and the correct approach is the one that is feasible and useful in a practical sense.

It follows from these remarks that one of the purposes of evaluation is the control of training by a process of collecting, analyzing and evaluating information leading to decision-making and action. As well as this controlling restraining function, evaluation should facilitate and create the freedom necessary for further action. The cycle suggested by Hamblin and illustrated in Figure 3.1 is a continuous process, whatever the point of entry. This cyclic conception of the control system underlies the work of Rackham, Honey and Colbert (1971) who describe the purpose of evaluation as that of creating a feedback loop, a self-correcting system.

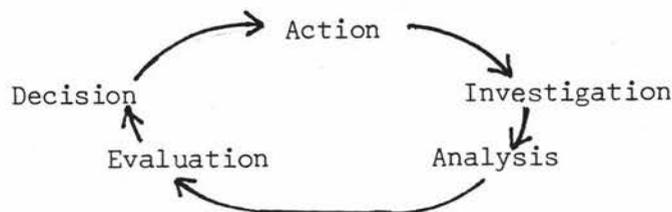


Figure 3.1 Training/Evaluation control system

In short, evaluation is an integral part of the training system, constantly interacting with it and thus should be tailor-made to fit the training programme. None of these concepts in the present writer's judgement are in conflict with those of Wortman (1975) or of Williamson et al. (1978).

3.2.1 PROBLEM - CENTRED TRAINING

As we have said, central to the whole process is the formulation of valid, objective, measurable and relevant criteria. What Hamblin calls problem-centred training is the type most suited to the in-service managerial and supervisory training fields and the present study falls within this general category. It is also the type where training and training evaluation are most closely linked and where the demands for intergroup co-operation are greatest. Warr, Bird and Rackham (1976) complain that, all too often, training people try to translate operative training ideas and methods into the management field where they fail to cope with its special problems.

The present writer agrees that to obtain the necessary criteria for measurement there should be careful pre-training investigation of job behaviour but this need only involve an identification of problem areas, not necessarily a total job analysis. To identify these training needs both the potential trainees and their immediate supervisors must become actively involved and as organizational objectives are also important other members of the organization may contribute as well. Warr et al. (1976) refer to this phase as *Context Evaluation* which they describe as,

"obtaining and using information about the current operational context in order to determine training needs and objectives".

3.2.2 THE CYCLE OF EVALUATION

Within the cycle of evaluation (Hamblin, 1974) there are different types of training evaluation and there are several possible models to follow. The one chosen here is a combination of that suggested by Hamblin (1974) and by Warr, Bird and Rackham (1976). The system offered by the latter authors is identified by the initials CIRO. CIRO stands for Context evaluation, Input evaluation, Reactions evaluation and

lastly, Outcome evaluation. These outcomes may be either immediate, intermediate or ultimate level outcomes. In reference to questions about what must be changed and what procedures are needed to bring about this change, Warr et al. specify a second type of evaluation, namely, *Input Evaluation*. This is similar to Williamson, Prost and George's Means-Ends analysis and involves a consideration of training resources and techniques or, "obtaining and using information about possible training resources in order to choose between alternative inputs to training". (Warr et al., 1976, p.20).

Hamblin (1974) is principally concerned with collecting information about the changes caused by training and the five different levels of these training effects. This phase is described by Warr et al. as *Outcome Evaluation*. It involves obtaining and using information about the outcomes of training in order to improve subsequent training. The various levels within this type of evaluation constitute an hierarchy of training outcomes. The schemes of Hamblin and Warr et al. are compared in the following Table.

Table 3.1
Training evaluation schemes of Hamblin (1974)
and Warr, Bird and Rackham (1976)

	<u>Hamblin</u>	<u>Warr, Bird and Rackham (CIRO System)</u>
	-	Context
	-	Input
Level 1.	Reactions	Reactions
		Outcome
Level 2.	Learning	Immediate
Level 3.	Job Behaviour	Intermediate
Level 4.	Organization	Ultimate
Level 5.	Ultimate Value	

Each level of training effects corresponds in operational terms to a level of objective-setting and of evaluation and assumes a cause-and-effect chain of events, linking the levels together.

The present author chose to adopt the terms immediate, intermediate and ultimate outcomes because they implied the time sequence which she wished to convey to the client. Moreover, using the term *learning* at level 2 as in Hamblin's scheme, may be interpreted as merely referring to the acquisition of knowledge, while true learning effects take place over a much longer time period. At the intermediate level, the emphasis shifts increasingly to on-the-job behaviour while the ultimate level of evaluation is associated and must be linked with more global and long-term organizational goals.

Outcome evaluation, which includes the latter stages of the Wortman model requires, according to Warr et al. careful pre-training preparation with the evaluator attending to the following points:-

- (1) Defining training objectives which is accomplished as part of context evaluation.
- (2) Selecting and constructing some measure of these objectives. This activity assumes major importance as evaluation is fundamentally a problem of measurement. Some of these issues will be discussed later.
- (3) Taking measurements at the appropriate time. *When* this should occur depends largely on the nature of the training objectives. Whether criteria should be considered long-term or short-term depends on the specific context. In any case, measurements must be taken before training as well as afterwards.
- (4) Assessing the results and using them to improve later training. The evaluator then draws conclusions and feeds them back to the organization as an aid in the planning of future training. Level 1 training effects in Table 3.1 are referred to as "reactions" by both Hamblin and Warr et al. who define them as, "obtaining and using information about trainees' expressed current or subsequent reactions in order to improve training". (Warr, Bird and Rackham, 1976, p.20).

The writer agrees with Hamblin, in this instance, that trainees react to many things, trainers, subject-matter, training methods, training

setting and to one another, particularly when the course is participative. Members' reactions may be influenced by the trainers, fellow-trainees, course content and external events, past, present and future. Moreover, reactions, by virtue of their spontaneous nature, are complex and changeable so that one must be selective and decide on the particular type of reactions to be measured. Reactions are usually evaluated by obtaining trainees' opinions and are heavily subjective. Despite these difficulties, since information on trainees' reactions can be extremely useful, it should be collected systematically.

Level 2, the immediate outcomes or learning outcomes in Hamblin's terminology concentrate on changes in knowledge, skills, and attitudes which must occur before new behaviour can emerge. These three inter-related target areas can be assessed as soon as the training course has been completed.

Level 3, the intermediate or job behaviour level, consists of changes in the employees' work behaviour which must be achieved if ultimate objectives are to be met. Here we are particularly interested in the transfer of learning to on-the-job behaviour and it is necessary to formulate these objectives in behavioural terms.

Although Hamblin does not consider that it is always possible or worthwhile to go beyond the job behaviour level of outcomes, sometimes later follow-up can reveal organizational effects. Consequent changes in organizational climate may be measured subjectively by means of questionnaire or interview data or by objective measures of work output, labour turnover, number and frequency of staff meetings, amount of interaction and so on. As far as ultimate outcomes are concerned, it is difficult and often impossible to isolate causes of changes which happen to the entire department or organization so Warr et al. do not recommend evaluating specific training programmes at this level. Hamblin suggests that if these ultimate criteria are financial ones then sometimes cost-benefit or cost-effectiveness techniques can be used but in the case of human benefit criteria a satisfactory solution to the measurement problem has not yet been found. It is probably impossible to measure all levels, including levels 4 and 5, within a single study. While conceding the wisdom of the above arguments the present writer

maintains that some tentative statements can and should be made relating ultimate outcomes to organizational goals where a longitudinal study of a year or more has been carried out. The client organization frequently expects, and is possibly entitled to, the best indications for longer term effects that the evaluator is able to give but clients should understand the severe limitations imposed on such forecasts.

3.2.3 THE DISCOVERY APPROACH TO EVALUATION

Hamblin (1974) speaks of the discovery approach in contrast to the scientific approach to evaluation. While the scientific approach requires that the experiment be controlled, rigid and inflexible to permit questions of causality to be answered and that training be kept separate from evaluation to avoid contamination, the discovery approach encourages flexibility and an integration of training with evaluation. Moreover, the scientific approach involves multiple measures of experimental and control groups but the practical difficulties of obtaining matched controls to fulfil the requirements of, for example, a Solomon (1949) four-group design are usually prohibitive. Hamblin recommends therefore placing more emphasis on the monitoring aspect of evaluation where the links between successive stages in the cycle are more important than a summative evaluation, and he strongly recommends the more discursive and exploratory methods of the discovery approach for the evaluation of training. This appears to parallel Wortman's formative evaluation and certain aspects of Guttentag's decision - theoretic approach. Finally, Hamblin and Warr, Bird and Rackham offer suggestions for the development of tailor-made measuring instruments appropriate to the different types and levels of evaluation but they stress the need for them to be adaptable to the particular study in hand. While acknowledging the advantages of the flexible and adaptable approach to evaluation just described, the present writer would continue to strive for an experimental design which allowed hypotheses concerning training effects to be tested as rigorously as possible. The evaluator must endeavour to reconcile the demands of a scientific investigation with the equally important requirement of providing the sort of information which has sufficient depth, insight and utility to allow the client to make good decisions in the real world.

3.3

CRITERION DEVELOPMENT

The first step in Warr, Bird and Rackham's (1976) CIRO system is the investigation of the training context. Over the years, there have been continued efforts to develop classification systems for managerial-type positions. The examples cited here were chosen because they have guided the steps taken by the present researcher to clarify training objectives and define the context of training.

Campbell, Dunnette, Lawler and Weick (1970, p.272) summarize the various types of criterion information that has been used in the past to evaluate managerial training. They are:-

- opinions of various interested parties and reactions of persons involved in the training.
- attitude changes brought about as a result of training, measured subjectively by means of questionnaires.
- objective examination of knowledge such as quizzes of course content.
- actual post-training performance, on the job such as managerial turnover, rate of communication and performance ratings by peers, subordinates or superiors.
- situational tests like the In-Basket (Frederiksen, Saunders and Ward, 1957).

In regard to the measurement of attitude change, researchers have been concerned with attitudes towards supervisors and subordinates, towards different leadership methods and to the organization itself.

In answer to questions about who provides the criterion information, Campbell, Dunnette, Lawler and Weick (1970) list the following; trainees, their superiors, peers and subordinates. This information is obtained before training starts, immediately afterwards and at intervals following training. The training objectives examined may be either short-term or long-term. Subsequent on-the-job behaviour is emphasized but

interest focusses on *changes* in behaviour from before to after training rather than on absolute levels. Of major importance is the degree of transfer of learning to on-the-job performance. In this regard, it is recognized that retention and transfer is affected by forces operating within the organizational environment.

Although serious attempts have been made, in recent years, to devise systematic methods of job analysis, for example, Fleishman (1967), McCormick, Jeanneret and Mecham (1972), Shouksmith (1978), taxonomies for managerial-type jobs have generally had to be designed for the particular organization concerned as leadership behaviour tends to be situation-specific (Vroom, 1976, p.1537). Campbell, Dunnette, Lawler and Weick (1970) distinguish between three levels of outcomes of the managerial role:-

- (1) *Behaviour* refers to what people do in the course of their work, such as planning research projects, interviewing, writing reports, encouraging junior staff.
- (2) *Performance* is behaviour that can be measured in terms of its contribution to the goals of the organization such as personnel assessment and selection, communication of research results to clients and to the scientific community, the organization and management of work groups or sections.
- (3) *Effectiveness* refers to the ultimate goals of the organization for which the individual is only partly responsible. This may involve enhancing the status and productivity of the organization, advancing the causes of science and improving the level of technology in society.

The critical difference between performance and effectiveness is that the latter is partly a function of the individual's behaviour and partly of other factors such as the economy, governmental policy, administrative constraints, and so on. The above authors claim that as evaluators, the psychologist should be trying to measure and predict the major dimensions of performance rather than effectiveness which is too subject to wider contextual variations.

One attempt to develop a standardized managerial job taxonomy was that of Tornow and Pinto (1976). Following on from the pioneering work of Hemphill (1960) who obtained ten dimensions for executive positions using the Executive Position Description Questionnaire, Tornow and Pinto's goal was to develop a behaviourally based management job taxonomy which was independent of worker traits, abilities and other individual difference variables. Items from management literature and data obtained from interviewing 433 executives, when subjected to factor analysis and cluster analysis procedures revealed 13 independent job factors. The final form of the researchers' Management Position Description Questionnaire described job behaviour for executive and management positions in terms of responsibilities, demands and activities.

Other researchers, as well, have attempted to classify different types of managerial jobs. Stewart (1975) similarly emphasized the behavioural demands of the managerial position over a wide range of very diverse jobs in industry, commerce and the public service. In an early evaluation of the effectiveness of a human relations training course for supervisors, Stroud (1959) defined training objectives in terms of job behaviour using a questionnaire consisting of the "consideration" scale of the Leadership Behaviour Description Questionnaire, LBDQ, (Fleishman, 1957a) plus the critical incident technique (Flanagan, 1954). Trainees and controls were asked to describe four types of incidents that had occurred on the job, two successful and two unsuccessful. Hand and Slocum (1972), in a study to evaluate the effectiveness of a managerial human relations programme, sought to detect changed attitudes and whether or not the changed attitudes were reflected in organizational behaviour. Their study involved an experimental and control group over a period of two years. The criterion variables of self-awareness, sensitivity to the needs of others and leadership style were measured by the Leadership Opinion Questionnaire, LOQ, (Fleishman, 1957b) while behaviour was measured by performance ratings of superiors to see whether learning was transferred to the job situation.

In order to assess the convergent and discriminant validities of scales designed to measure managerial behaviour, Lawler (1967) and House and Rizzo (1972) adopted a Campbell and Fiske (1959) multitrait-multi-method approach to construct validity. House and Rizzo found that eight

of the nineteen scales had satisfactory convergent and discriminant validities when the Organization Description Questionnaire was validated against a set of eight criterion variables. Similarly, Lawler, using a multitrait-multirater approach obtained good convergent and discriminant validities from the managerial ratings of different groups of raters (superiors, peers and subordinates). In this case the judges were rating research scientists on four separate traits. In a second sample, superior, peer and self-ratings were obtained on top-level managers in a manufacturing industry. Again there was some evidence for convergent and discriminant validity. Methods other than factor analysis have been used to determine the dimensions of managerial-type jobs. There are examples, of both cluster analysis (Tornow and Pinto, 1976) and multidimensional scaling (Smith and Siegal, 1967; Brown, 1967).

A procedure which has received considerable attention and has the objective of trying to measure and predict the major dimensions of managerial performance is the method of scaled expectations (Smith and Kendall, 1963). It involves several steps:

- (1) Step one employs the critical incident technique (Flanagan, 1954) which requires people from within the organization to generate examples of behaviour illustrating components of performances for the job in question.
- (2) The specific behavioural incidents are submitted to a qualitative cluster analysis by other members of the organization or by the experimenter (Campbell, Dunnette, Arvey and Hellervik, 1973) and sorted into homogeneous categories.
- (3) The third step involves retranslation, the process whereby the tentatively defined performance dimensions and incidents are re-assessed by other personnel. Incidents are retained only if they can be reliably placed in their original dimensions.
- (4) Each surviving incident is scaled by members of the organization on a dimension representing "good" to "poor" performance.

Smith and Kendall's aim was to develop sets of unambiguous anchors for rating scales based on the users' own judgements and language. They

point out that raters will have different perceptions because even though most executive or administrative jobs have an official title they are likely to differ either in level of performance required or in the performance dimensions which are considered to be important. Nevertheless the raters can be expected to share a common core of experience and values concerning behaviour on-the-job and it is this commonality on which Smith and Kendall base their method. The scales concentrate on the actual observed behaviours or behavioural expectations. These latter are inferences or predictions from observations and in the development of the original behaviourally anchored rating scales (BARS) evaluations were made by judges who were similar to those who would eventually make the judgement. Both the terminology used and the behavioural qualities included were determined by a process of discussion, cross-checking and revision by the raters themselves.

Blood (1974) has shown that this method of behavioural expectation scaling is also suitable for purposes other than the development of rating scales. The first step in the procedure has been employed frequently to gather information for job analyses (Blum and Naylor, 1968, p.495). In these cases as in the present study where the aim is to determine training needs by a problem-centred approach, it is not necessary to eliminate items by retranslation because the aim is to obtain a large number of descriptive job statements rather than a limited set of performance dimensions.

Borman (1972) suggests that performance evaluation scales should be developed from more than one perspective and should include information provided by both superiors and peers. This not only broadens the range of behaviours that can be assessed, but where a wide divergence between several organizational groups is observed, it may be possible to pinpoint important areas of disagreement and lack of clarity in organizational policy. Additional advantages of using such a scheme are the opportunity it affords for active participation by people within the organization who will use the information and the fact that the resulting dimensions and items are written in the language of the members. (Campbell, Dunnette, Arvey and Hellervik, 1973)

From a review of previous studies, such as these, the present writer has identified five features which are important in the development of criteria for managerial training:

- (1) The analysis of training needs should revolve around a particular managerial position or organization.
- (2) The actual behaviours of the managers are the important criteria at intermediate levels of evaluation.
- (3) A critical incident technique is a particularly suitable method of obtaining behavioural information.
- (4) The dimensions of the managerial role may be obtained by extracting basic factors by statistical methods or by means of clinical judgement.
- (5) Construct validity may be ascertained by multitrait-multimethod or multirater approaches.

3.4

TRAINING EVALUATION METHODOLOGY

There are a number of issues which have been confronted by previous researcher-evaluators and which have application in the present situation. In discussing the plan or procedure to use in evaluation research, Campbell, Dunnette, Lawler and Weick (1970) state that, in general, it must aim to provide results which are relatively unambiguous and amenable to causal interpretation. As suggested earlier this requires that answers be found to questions relating to experimental design. Such questions are related to the use of after-only or before-and-after measurements, the use of one or more control groups and placebos and the need for replication. In fact, it requires a consideration of all the usual threats to validity discussed by Campbell and Stanley (1966). In addition, we must consider the feasibility of conducting "true" experimentation in the field setting along with the possibilities of quasi-experiments, patched-up designs, longitudinal experiments, and so on. Further, decisions must be made at an early stage about the most appropriate statistical tests to use. Campbell et al. (1970) raise the problem of over-reliance on statistical tests of significance

in decisions made about the effectiveness of training. They claim that statistical significance is but a minimal step and other questions such as the practical significance or utility to the organization must be considered as well. Also important is the logical analysis of the training programme or means-ends analysis. This includes questions about whether the content of the training programme is relevant to the goals of training. The researcher must study the effects of the environment, for example, the reward structure of the organization, which may affect the long-term outcomes of training. Lastly, Schein (1965) has made the point that management training may mean training for an uncertain future in which managers must be able to respond to unpredictable changes within the organization and the larger society.

Burgoyne and Cooper (1975) note that the timing of the measurements must be carefully planned and they discuss the advantages and disadvantages of using pre- and post-measures. They stress the need to employ follow up measures weeks or even months after training in order to check on long term effects. In regard to the use of control groups, they point out the advantages of employing a comparison group in preference to a matched control group. Members of a comparison group receive an alternative type of treatment rather than no treatment at all (Friedlander, 1967). Again, it is often the circumstances surrounding the research which ultimately determine the answers to such questions. Burgoyne and Cooper distinguish between the validation of external and internal criteria. External criteria refer to behavioural changes observed in carrying out the organizational role and are therefore the intermediate or longer-term outcomes of training, while internal criteria are directly related to training content and are assessed during or immediately after the learning experience. The links between these two sets of criteria are considered to be especially important by the present evaluator.

Two final requirements of training evaluation are noted by the above authors. Firstly, they recommend that the training techniques themselves should be evaluated. In the terminology of Williamson, Prost and George (1978), this is means-ends analysis. Secondly, working within a systems context, the evaluator must be alert to possible conflicting objectives of the training programme and the organizational

environment. This is illustrated by the difficulties trainees experience in applying their new learning back home in the work situation. Burgoyne and Cooper (1975) warn against the danger of total reliance on self-report measures of the attitudes and perceptions of subjects. They stress the need to investigate the behavioural correlates of learning. They advocate the use of a variety of indices of change such as standardized questionnaires, self-report attitudinal measures *and* behavioural measures. The latter, they claim, have proved to be by far the best predictors of change and post-training differences.

Dalziel, MacWilliam, Strong, Hayes and Lunt (1972) consider that it is necessary to evaluate participants' *expectations* prior to training. By expectations, they mean reasons for attendance and anticipations about the objectives, form and content of the training programme. As a type of "means" assessment, they suggest the use of session assessment forms where each session during the day is evaluated on several scales such as new information learned, amount of interest generated, amount of participation and applicability. Areas of organizational and personal change anticipated and fulfilled were also investigated in the Dalziel et al. (1972) study, as were the subjects' reactions to different teaching methods. They show that precourse questionnaires can serve the dual purpose of measuring pre-test levels of performance while at the same time help to set course objectives and subsequent behavioural objectives. They mention the use of histograms of results for providing easily assimilated feedback to programme organizers and suggest that more work could be done in supplying feedback about individual performance during group sessions.

Another question in their study was aimed at future behaviour to assess the more lasting effects of the course and consisted of asking trainees what they intended to do with the new learning acquired. This is a form of personal goal-setting and was followed up later when subsequent behaviour was assessed and used as a criterion measure.

An in-house management course for middle managers in a large engineering company was investigated by Ashton and Gibbon (1974). The study was divided into two major components:

- (1) A formal research function, designed to measure as rigorously as possible certain specified changes in the participants' knowledge, skills and attitudes.
- (2) A more flexible monitoring function, designed to provide general and 'ad hoc' information about certain aspects of the day to day running of the course.

In the case of the former, the authors state that the development of instruments, the research design and the objective-setting phase should ideally have begun three months before the beginning of the course but this was impossible for administrative reasons. This is a feature common to many such projects.

Ashton and Gibbon were presented with four general statements which expressed in global terms the main objectives of training. The evaluators' first task was to translate these objectives into valid measuring instruments and to administer them to all relevant people. This required a more detailed break-down of the original four statements and selection of measuring instruments. An experiment was then designed within the limitations of the practical situation. A variety of measures were chosen such as interviews, questionnaires, performance ratings and the Kelly Repertory Grid, (Kelly, 1955). Initial measurements were made and control groups were chosen from a shortlist of prospective candidates for future training courses.

As well as this, the monitoring function of evaluation was fulfilled, with the more flexible objective of providing up-to-date information on programme functioning using the session assessment forms, trainees' reactions and open-ended exploratory type questions.

A management development course for senior officers in the Health Services in London provided further examples of evaluation techniques. Jones and Huczynski (1975) acknowledge that major methodological inadequacies in their study stemmed from the fact that it was limited to immediate, short-term effects and that it assumed that the training input was homogenous and common to all trainees. Burgoyne and Cooper (1975) have shown that trainees are not passive recipients or patients

but active agents and are technically free to modify or even reject any or all of the training programme being offered. This may give rise to unanticipated outcomes which should be measurable. Thus trainees' reactions are important and the evaluator must maintain an open mind to possible outcomes. Furthermore, at management level the objectives themselves are likely to be broad and complex. These include flexibility and adaptability to change. In the early stages part of the evaluator's job is to articulate priorities for the objectives and for this reason considerable time must be spent on the initial interviews. Burgoyne and Cooper (1975) describe the "focused" and the "diffuse" effects of training. While the focused effects produce specific, recognizable changes in behaviour, skills and knowledge, like improved planning of work or sectional organization and may be derived from lectures, formal discussions and instructional sessions, the diffuse effects are longer term, more durable attitudes and perceptions of self and others. They influence the trainees' values and belief system, his self confidence and self awareness and are more likely to be derived from interaction with others and more experiential training methods. These often turn out to be highly valued by participants and such reactions are most easily detected in post-course interviews. The present writer agrees that the evaluator must be alert to these important but less tangible effects.

In another management development evaluation, Baynes (1975) identified several important outcome variables of the management training process:

- (1) Change in knowledge about facts, functions and techniques related to the organizational role.
- (2) Attitudes towards the organization and its members and the manager's own role and actions.
- (3) Skills like decision-making, communication and ability to motivate subordinates.

He also stressed the salience of the work environment.

Morris (1972) studied a training scheme for senior managers in an international computing firm. This is an illustration of evaluation

carried out before, during and after training and the author made the point that it is necessary to exploit as many ways of measuring training effectiveness as possible. He also reinforced the idea of clarifying objectives with the decision-makers and of collaborating to test out those objectives. The study measured knowledge pre- and post-training and perceptions of trainees by means of reaction level tests. Performance measures were carried out during training and allocation of training time and relevance to job were assessed. Individual session assessments were used and data on trainees' reactions were gathered by free comment and by face-to-face interviews, as well as by means of structured interviews. Trainees were asked to recall goals set at the end of training and were subsequently questioned about how much their work behaviour had changed and how durable were these changes. Further, they were asked for specific proof of these changes using a critical incident technique. Questionnaires were designed for continuity of certain questions to accommodate a repeated measures procedure. To anticipate some of the problems of analysis, structured questions were used with a simple mode of response like placing ticks in boxes. These were counter-balanced by more open-ended items requiring more complex content-analytical scoring techniques. Efforts were made to involve management at all stages of the project.

Williams and Berger (1972) have drawn some useful general conclusions about evaluation of management development from their work with the Industrial Training Board. They agree with Campbell, Dunnette, Lawler and Weick (1970) that it is an "impossible task" to evaluate training in terms of ultimate criteria. They suggest, rather, measuring the effects of introducing particular variables. They describe a three-stage system of training, consisting of the following sequence:

- (1) The period before the learning experience when anticipations, feelings and expectations are assessed.
- (2) The teaching or learning phase when trainees receive new skills, precepts, knowledge and understanding and attempt to assimilate by insight and consideration, their possible application to the job.
- (3) The time afterwards, back on the job when a supportive atmosphere is essential.

They see the course itself as a small but important segment of the three-phase system. Thus before and after measures are essential. Williams and Berger state that learning is frequently measured by:

- (a) asking observers for their opinions
- (b) asking course members to rate their own learning
- (c) content analysing members' responses to open-ended questions concerning what they have learned from the programme.

These authors make the observation that the findings rest heavily on the types of measures used and many researchers in the past have cast doubt on the validity of self-rating scales of the sort suggested above (Miles, 1965; Berger, 1968; Berger & Berger, 1969; Lawler, 1967). In this regard Kimber (1970, p.277) says that such measures,

"provide a pointer to those parts of the course which need immediate alteration (but) are practically useless if used as the sole means of evaluating training".

As an alternative to self-rating, Williams and Berger (1972) used responses to the following two open-ended questions:

"What specific things have you learned?"

"What improvements in your work do you anticipate over the next few months?"

Finally, Williams and Berger suggest longer term measures, for example, the extent to which course learning relates to improved job performance plus follow-up measures of work performance with trainees and with their bosses. They found that end of training motivation is related to subsequent behaviour change.

From these latter studies the present writer took special note of the suggestions relating to the timing of data collection and the variety of techniques to employ. The main restrictions on this data collecting phase is cost, time and the willingness of the subjects to cooperate. Out of all the possible approaches and ideas reviewed the present evaluator selected those which were most applicable to her own situation, bearing in mind the restrictions mentioned above. In the next chapter we examine the events which preceded the implementation of the project and the general aims which emerged.

CHAPTER 4.

INTRODUCTION TO THE PRESENT RESEARCH

4.1

BACKGROUND TO THE EVALUATION STUDY

The study began with the writer's appointment as research associate in a project to evaluate an in-house management training programme designed for scientists and senior technical staff in a government department and several semi-governmental organizations in New Zealand. Training courses had been instituted two years previously for the purpose of preparing selected members of staff for administrative and managerial roles associated with their positions as section, group or project leaders engaged in scientific and industrial research. Three courses had already been completed when the present evaluation took place. Although the programme had been instigated by the government department, several smaller organizations which were partially funded by private industry had participated from the beginning contributing a minority of the trainees to each course. According to those responsible for the training programme, it had been initiated in response to difficulties experienced in finding people from within each of these organizations with suitable background and skills to fill administrative positions and to a general dissatisfaction with existing management training courses which were available to their staff.

Even at section leader level which was equivalent to a middle management position there was considerable unease concerning the ability of the highly qualified men and woman trained in the various scientific disciplines to manage effectively the people and the resources within their sections. Neither their previous training nor their personal interests and inclinations had prepared them for this sort of leadership role. It was recognised that section leaders, first of all, needed training in all aspects of staff appraisal and evaluation since the major responsibility for this task fell to them. Secondly, since they were responsible for the allocation of available funds and resources to individual researchers and projects within their own sections, they needed training in handling such matters. Above all, there was a strong feeling that the type of training offered should be relevant to the work performed and should enable the trainees to relate their own performance to that of the entire organization.

The aims of training quoted to the present evaluator by programme organizers were:

- (1) To introduce course members to some of the principles of management and administration which they would find useful in developing their leadership potential.
- (2) To give them a broad understanding of science administration.

In addition to these general aims, each of the topics studied was described by means of a brief statement and a very sketchy summary of the content, for example,

- (1) Resource Allocation: To illustrate the factors taken into account in allocating resources to a variety of activities and the information required for the assessment of such activities.

Content - Examples of individual three-year research plans.

- (2) Marking (Personnel Appraisal and Assessment): Familiarization with the science and science technician promotion systems.

Content - Means of assessment, promotion criteria, determination of orders of merit and priority for promotion, correlation meetings, promotions and appeals.

There was general agreement that these objectives were inadequate as they were too vague and imprecise but at this time there were no clearly defined specifications for the section leader role.

Topics covered in the three previous training courses had included Organization and Delegation, Forecasting, Planning and Control, Resource Allocation, Reporting and Marking, Interviewing and Selection, Motivation, Communication, Personal and Group Relationships and Leadership. The topic supervisors were generally chosen from outside of the organizations but some experienced administrators from within were used as resource persons, conducting lectures and seminars on the main topics. Other senior administrators gave talks on selected subjects of a more general nature. All of the topic supervisors and guest speakers were chosen for their knowledge and expertise in course related areas.

The first course had taken place on two separate weeks in consecutive months in 1975. The practice of splitting up the course in this way had presented organizational difficulties and was subsequently discontinued.

The topics Organization and Delegation, Forecasting, and Planning and Control had consisted of an introduction by topic supervisors followed in each case by an assignment involving a written report. Personnel Management including Interviewing and Selection, Motivation, Communication, Personal and Group Relationships and Leadership had been taught by outside educators from a tertiary institution. Several sessions were devoted to explaining in some detail the Reporting and Marking (assessment) system used by the major organization. Other sub-topics included information on support services available and a working party set up to investigate some current problems within the organizations.

Each new intake of trainees was organized into three syndicate groups. These consisted of seven members selected from different geographical locations, branches and disciplines of the organization. Each syndicate appointed a chair-person and secretary for their group and these positions changed from one topic to the next. Towards the end of the training course these elected officers presented the group reports on behalf of their syndicates to a combined meeting of all trainees. Other syndicate members were allocated various tasks, such as chairing and hosting meetings for guest speakers and arranging social functions.

A number of changes were made in successive courses along the following lines:

- (1) The working party and the session on support services were omitted. Both of these had proved unsuccessful and some of this material was subsequently incorporated into other topics.
- (2) A brief introductory session on the first night of the course was presented by the course controller who made an effort to introduce syndicate members to each other at that time.
- (3) In courses four and five (1977) a Management Forum was introduced towards the end of the two week programme in which top officials

from the main organization met with the trainees in a combined session to discuss matters which had arisen during the training programme. Six people were delegated to collect questions and comments for discussion and to present them at this meeting.

- (4) The topics Forecasting and Planning were presented in separate sessions but later were combined for the purposes of the assignment exercise.
- (5) The topic Resource Allocation was revised for courses four and five and a preparatory exercise was completed by all trainees prior to the course. During the course each syndicate functioned as a work team and allocated certain resources to team members in a simulated work situation. Following this the three syndicate chair-persons took part in a role play with the support of their team members together with administrators from head office. In this session the syndicate was required to justify the decisions which they had made.

During its two years of operation, course controllers had attempted only a very informal type of evaluation. For each of the earlier courses records of sessions had been kept by an observer appointed as secretary. It was generally felt that trainees had gained a favourable impression of the training programme. Trainees were subsequently interviewed and attempts made to gather critical comments after two, six and twelve months. The investigation was informal and the data collected was not amenable to formal analysis. It was noted, however, that the strong rapport established during the training courses was quickly re-established at these follow-up interviews. As a result of these informal evaluations a report had been written on the difficulties of constructing such a course. It was decided, on the basis of this initial attempt to gain information on the effectiveness of the programme, to proceed with a more formal and systematic type of evaluation.

In terms of the definition discussed earlier, evaluation up to this point had consisted of subjective judgements about the worth of the training intervention, a procedure which Suchman (1967) considered to be seriously lacking in control of bias and subjectivity and which

Riecken (1977) rejected because it lacked the clearly defined objectives necessary for measuring the outcomes of an attempt to bring about change.

The present evaluator began work shortly before the commencement of the two 1977 courses. From the outset, she was confronted with the problem of a seriously restricted preparation period and it was clear that her first task was to set priorities so that the most important preliminaries could be accomplished within the limited time available. In this respect the problems were similar to those listed by Suchman (1967), Guttentag (1973), Bernstein (1975), Riecken (1977) and Riecken and Boruch (1978) all of whom mention the compromises and adjustments which must be made in an applied situation. From her reading of the literature and mindful of practical constraints, the writer decided that the most important tasks were:

(1) The clarification of goals or objectives of training and evaluation. Previously, goals had not been specified in sufficient detail and this was one of the reasons why earlier attempts to evaluate the programme had failed. The objectives of training as described by the course controllers were too general to allow behavioural or other changes to be identified if and when they did occur during the training or post-training period. In spite of the point made by Guttentag (1973) that training programmes are frequently in a state of evolution and therefore goals are constantly changing and are never likely to be as clear-cut as the evaluator would like them to be, it was felt that some further criterion development was necessary in order to meet the demands of the Wortman (1975) concept of summative evaluation. In addition to this, the behavioural goals identified could be used as a pretest measure of performance, a strategy which would represent considerable economy of research effort (Burgoyne, 1973).

(2) From a practical point of view, there were a number of questionnaires and scales to be developed and distributed to subjects before the beginning of the first course for the current year. A similar situation existed in the Ashton and Gibbon (1974) evaluation of a management course for engineers where general goal statements had to be analysed and converted into valid measuring instruments and administered within a restricted time-frame.

(3) Initial interviews with the majority of people involved in the study were essential. For a start, the contributing organizations all of whom were government or semi-government departments required certain procedures to be followed in obtaining permission and making contact with various members of those organizations. The evaluator's credibility had to be established at all levels within the hierarchy and matters of policy and ethics dealt with at the outset. Moreover, as an outsider, the evaluator was well aware of her own lack of knowledge about the functioning of the organizations. Ideally such deficiencies should be corrected prior to any efforts to evaluate the management training programme.

Equally important were the subjects who were to undergo training and whose performance was to be examined. It was essential to the success of the venture that their commitment of time and effort be negotiated well in advance. Weiss (1972), Glaser and Backer (1973) and Guttentag (1973) have all demonstrated the value of adopting a more clinically oriented approach to programme evaluation while at the same time emphasizing the reality of the pitfalls involved. In addition to these arguments for an extensive interviewing programme, it was an excellent opportunity to obtain assessment of pretraining performance from a wide range of people. This could include self-reports by trainees and reports from their immediate controlling officers, subordinates and peers. In effect, it was possible to achieve within the lead time available to the evaluator a limited preliminary study which concentrated on:

- (1) Clarification of goals by means of an analysis of training needs.
- (2) Development of pretest scales and questionnaires.

It may be noted that the evaluator did consider using the first training group of the year for a more complete pilot study but it was decided that in view of the small numbers available, one half of the subjects could not be set aside for this purpose. Moreover, the relatively short time-lapse between the two courses was insufficient for data to be collected and analysed before the next course began. Instead, the writer used most of the lead time to contact and interview a wide range of people, including administrative staff, course controllers, divisional directors, trainees, controlling officers or other members of the work

units as well as almost all of the scientists and technicians who were involved in the study. As a result of this activity she was able to obtain many different points of view from people at various organizational levels. This information included perceived training needs and course objectives from a variety of perspectives as well as longer-term goals of the participating organizations. The next step was to integrate this information and to define, for the purpose of the evaluation, the goals of the science management training programme.

The analysis of this criterion information was carried out by the researcher, her colleagues and members of the participating organizations who were not directly involved in the training course so as to maintain a level of control over the courses to be evaluated. The aim of this phase was the formulation of valid, objective, measurable and relevant criteria, and is characteristic of what Hamblin (1974) calls the problem-centred approach. This then enabled the evaluator to develop scales and questionnaires to measure attitudes, knowledge, skills and behaviour relevant to the job of section or group leader. At this point it was decided to adapt to the present situation a number of scales described by researchers like Hamblin (1974) and Warr, Bird and Rackham (1976). This procedure had the advantage of utilizing some pretested scales with a minimum of modification so that even though a thorough pilot study was not possible, the researcher was reasonably confident that the scales were valid and reliable measures of the variables of interest. Following the suggestions of Riecken and Boruch (1978) the present study utilized a mixture of standardized tools and others designed more specifically for the project. The interviews also served the function of alerting the trainees' controlling officers to the criterion behaviours which they would be asked to assess some time after the trainees had returned to work. Moreover, they provided the opportunity for organizational members to express these behaviours in their own terminology so that when the questionnaires were constructed the researcher could employ the personal constructs of those who would use the scales rather than her own wording and ideas. Finally, the interviews provided the opportunity to explain the purpose of the research to the subjects taking part. All of these aims are listed under Phase I in Table 4.1.

The research to be reported here is concerned with the investigation of various aspects of the science management training programme and the changes observed in trainees over the period before, during and after training occurred. Firstly, we are interested in the amount and nature of the observed changes and whether they are related to the training programme itself or to some other factors operating within the total environment of the organizations concerned. Secondly, we look for changes in knowledge, attitudes, skills and behaviours exhibited by subjects after they have undergone the period of training and when they have returned to their normal work settings. Within this general framework a number of investigations are carried out each with separate aims and each designed to test certain aspects of relevant attitudes, knowledge, skills or behaviour associated with the section or group leader role. The relevance of these attitudes, knowledge, skills and behaviour to the work performed by the section leaders was determined by a careful examination of training needs. Since the course controllers were primarily interested in the overall effectiveness of the training programme and whether or not their training objectives were being met, part of the evaluation was directed towards establishing the specific goals of training and then assessing whether these goals were being achieved. This is the part of the evaluation process referred to by Scriven (1972) and by Wortman (1975) as summative evaluation and by Williamson, Prost and George (1978) as general effectiveness and goal-outcome congruence. The third major objective of evaluation which is strongly recommended by both Wortman (1975) and Guttentag (1973) is the provision of ongoing corrective feedback to course controllers and in some cases to trainees as well. Our previous discussion has shown how this requirement can pose the most serious threat to validity and experimental control. In the present study the researcher refrained from circulating this information among trainees during the period of evaluation in order to maintain a more stable data base. Nevertheless, the evaluator felt some responsibility to the organizational members particularly to course controllers to provide the type of monitoring function which would assist the continual development of the courses. In many cases, under central funding an organization cannot enjoy the luxury of waiting for two or three years to begin to make the needed changes in such a programme. Thus we

have the rationale for a formative type of evaluation. In this respect and to a limited extent only, the present study adopted the type of approach described by Cherns (1969) and Burgoyne (1973) as "action research" and by Argyris (1976) as organically oriented research.

Because the writer proposed to adopt a systems approach to the training-evaluation cycle, assessments were made of those aspects of the trainees' working environment which were considered likely to effect the outcomes of training and to modify its effectiveness. Basically the evaluator chose to work within the theoretical model of Wortman (1975) who adopts an experimental type of approach to evaluation set within a systems framework. Therefore, it is necessary to consider internal, external, construct and conclusion validity issues as well as causal and feedback linkages and the effect of other organizational variables (Figure 2.1., p.14)

While the Wortman model discusses the conceptual problems associated with evaluation, one must turn to the work of Hamblin (1974), Warr, Bird and Rackham (1976) and other European writers to answer some of the practical questions about how training can be evaluated, which criteria should be used, what sort of time sequence is appropriate and when various measures of change should be implemented. The latter group of researchers seek to explain the cycle of evaluation (Hamblin 1974) and the specific types of measures to be used at particular times during the cycle. Subsequent chapters describe the sequence of events followed during the seven phases of the present evaluation study and the reader is referred to Table 4.1 for a summary of these phases.

Table 4.1

The seven phases of the present evaluation study

PHASE I (INITIAL INTERVIEWS)

AIMS:	<u>Chapter</u>
(1) To determine training goals (criterion development) ...	5
(2) To develop questionnaires and other instruments to measure change	5
(3) Appraisal of pre-training performance	5
(4) To explain purpose of research to subjects and to establish rapport	5

SUBJECTS: *

Scale development group (n = 17)

Course A and B members, 1977 (n = 47)

Previous trainees, control group E (n = 11)

Future trainees, control group F (n = 19)

Controlling officers of above subjects (n = 23)

Course C and D member, 1978 (n = 45)

METHODS:

Conference and group discussion

Open-ended responses

Critical incident technique

Semi-structured question format

* Sample sizes are the maximum available in each subject group. The actual n's may be less for reasons stated in the text, see for e.g. p.105.

Table 4.1 continued

<u>PHASE II (PRE-TEST MEASURES)</u>	
AIMS:	<u>Chapter</u>
(1) To obtain biographical information	6, 10
(2) To establish base-line measures of knowledge, attitudes, skills and behaviour	6, 10
(3) To determine trainees' expectations of the courses	6, 10
 SUBJECTS:	
Course A and B members (n = 47)	
Previous trainees, control group E (n = 11)	
Future trainees, control group F (n = 19)	
Course C and D members, the following year (n = 45)	
 METHODS:	
Biographical questionnaire	
Attitude scale	
Questionnaire 1	
(1) Reasons for attending course	
(2) Expected areas of personal and organizational change	
(3) Preferred training techniques	
(4) Perceived understanding of topics	
(5) Perceived relevance of topics	

Table 4.1 continued

PHASE III (DURING COURSE MEASURES)

AIMS:	<u>Chapter</u>
(1) To assess trainees' immediate reactions to specific sessions	7
(2) To evaluate input, i.e. training techniques used ..	7
 SUBJECTS:	
Course A and B members (n = 47)	
 METHODS:	
Leipas Scale	

PHASE IV (IMMEDIATE POST-TRAINING MEASURES)

AIMS:	<u>Chapter</u>
(1) To obtain immediate reactions to course topics studied and training techniques used	7, 10
(2) To measure changes in knowledge and attitudes	7, 10
(3) To obtain information about trainees' future behaviour	7, 10
(4) To evaluate work setting or context	8, 10
 SUBJECTS:	
Course A and B members (n = 47)	
Course C and D members, the following year (n = 45)	
 METHODS:	
Questionnaire 2	
(1) General feelings and impressions about course	
(2) Behavioural intentions	
(3) Perceived understanding of topics	
(4) Perceived relevance of topics	
(5) Reactions to training techniques	
(6) Reactions to time allocation of topics	
Attitude scale	
Organization Climate Questionnaire	

Table 4.1 continued

PHASE V (INTERMEDIATE POST-TRAINING MEASURES:
3 - MONTH FOLLOW-UP)

AIMS:	<u>Chapter</u>
(1) To obtain longer-term reactions to course	8, 10
(2) To assess whether newly acquired knowledge was retained	8, 10
(3) To measure changes in behaviour as a consequence of training (transfer of training)	8, 10

SUBJECTS:

Course A and B members (n = 47)

Course C and D members, the following year (n = 45)

METHODS:

Questionnaire 3

(1) Perceived understanding of topics

(2) Perceived relevance of topics

(3) Opinion about time allocation to topics

(4) Changes in behaviour

i Areas of personal and organizational change

ii Books read

iii Discussions held

iv Contacts maintained

v Self reported behavioural change

Face-to-face interviews with 15 of the trainees to supplement behavioural information (group interviews conducted in some cases)

Table 4.1 continuedPHASE VI (INTERMEDIATE POST-TRAINING MEASURES:6 - MCNTH FOLLOW-UP)

AIMS:	<u>Chapter</u>
(1) To obtain self-reported changes in work behaviour	8
(2) To obtain superiors' reports on observed changes in trainees' work behaviour	8
(3) To assess changes in interpersonal relationships in the work situation	9

SUBJECTS:

Course A and B members (n = 47)

Future trainees, control group F (n = 19)

Controlling officers of above subjects (n = 23)

METHODS:

Questionnaire 4

(1) Self-reported changes in work behaviour

(2) Superiors' reports on observed changes in work
behaviour

Repertory Grid (interpersonal relationships)

Table 4.1 continued

PHASE VII (INTERMEDIATE POST-TRAINING MEASURES:12 - MONTH FOLLOW-UP)

AIMS:

Chapter

- (1) To evaluate effect of training on interpersonal relationships at work, individual work performance and performance of group or section 8

SUBJECTS:

Course A and B members (n = 47)

METHODS:

Questionnaire 5

- (1) Rating scale to measure effect on work relationships
- (2) Rating scale to measure effect on individual work performance
- (3) Rating scale to measure effect on organization of group or section
- (4) Open-ended statement

CHAPTER 5

PRELIMINARY STUDIES

5.1

INTRODUCTION

As our earlier discussion has shown, one of the major difficulties encountered in planning any type of managerial training evaluation is to determine the precise objectives of training. By reference to a clearly stated set of objectives, not only is it possible to make a more accurate evaluation of the success, or otherwise, of the training programme but it also provides a sense of purpose to those who are participating in the course. Warr, Bird and Rackham (1976) have stressed that adequate answers must be given to trainees' questions such as "What are we expected to learn?" and "How is this course supposed to improve my performance as a manager?", if they are to identify with the trainers' goals or even to be highly motivated to benefit from the learning opportunities provided. Within the terminology adopted by the present evaluator, the gathering of information about the operational situation for which the training programme was devised comes under the heading of context evaluation (Warr et al., 1976). In practice this involves determining the training needs of the people in the organization but, as previously explained, there are major difficulties involved in drawing up detailed job specifications for managerial positions. The variability of the management task and the wide scope for individual approaches to the job prevents it from being readily defined. This writer believes, nevertheless, along with many other researchers that even a relatively informal task analysis is better than none at all and every opportunity should be taken to increase the information available in order to provide a set of guidelines which may be used when choices are being made concerning the content and presentation of the training programme. Such an opportunity was afforded by the interviews conducted prior to the 1977 science management courses and this became the first aim of these preliminary studies.

The second aim was to construct questionnaires and other measures of change to be used in the main study. Most of the questionnaires containing direct questions (see later chapters) were devised by the

researcher herself or in consultation with members of the client organization. Items pertaining to biographical data, knowledge of course topic areas and work behaviours were discussed and tried out with course organizers and members of the scale development group (Table 4.1). Other measuring instruments like the Leipas Scale and the Organization Climate Questionnaire were adopted with slight modification from their original source. Again, the applicability of each of these was discussed with collaborators from within the organization. Most of the scale development time was spent on an attitude change scale constructed specifically for the present study and a detailed report of the procedure follows in section 5.3.

5.2

IDENTIFYING TRAINING NEEDS

Since the primary purpose of the initial interviews was to collect information from the trainees and their colleagues which would enable estimates to be made of any changes occurring in their performance following training this information was used first to establish criteria for judging the effectiveness of the current training courses. Subsequently it could be used in the formative sense and fed back to course controllers for the development of future training. Thus the three-fold aim of this part of the study may be summarized thus:

- (1) To determine a set of valid and reliable criteria for the evaluation of the management training programme.
- (2) To assist in establishing base-line measures of the dependent variables such as pre-training levels of performance.
- (3) To provide information which could ultimately be used by programme controllers for the development of future programmes.

5.2.1 SUBJECTS

In order to identify training needs both trainees and their immediate controlling officers were interviewed. These subjects included 38 management trainees (29 section leaders and 9 non-leaders), 10 trainees from previous courses (8 section leaders and 2 non-leaders), 18 employees who had not yet participated in the management training courses (12 section leaders and 6 non-leaders) and 20 controlling

officers. These people were derived from twenty one divisions of the government department and eight research organizations located in all of the main centres as well as several substations (Table 5.1). The following is a description of the subject categories:

- (1) Twenty controlling officers or someone designated as their substitute for the purposes of the interview who were asked to discuss the training needs (both general and specific) of members of their staff.
- (2) Thirty-eight people who were about to take part in the science management course for 1977. Twenty-nine of these were already occupying section leader, group leader, officer in charge of substation, director or deputy director positions while the remaining nine were scientists or technical officers who were not, at the moment, in direct leadership roles.
- (3) Eighteen people who had never taken part in the training programme and who were not scheduled to do so immediately but were matched as far as possible with respect to divisional affiliation and level in the organization. Twelve of these were already occupying one or other of the above leadership roles.
- (4) Ten people who had taken part in a previous science management course. Eight of these were currently occupying leadership roles of one sort or another.

Table 5.1

Subjects who took part in pre-course interviews

SUBJECTS	GOVERNMENT	SEMI-GOVERNMENT	TOTALS
Controlling Officers	18	2	20
Participants in 1977 Management Courses	34	4	38
Leaders	25	4	
Non Leaders	9		
Non Participants	16	2	18
Leaders	10	2	
Non Leaders	6		
Previous Courses	8	2	10
Leaders	6	2	
Non Leaders	2		
Total			86

5.2.2 MEASURING INSTRUMENT

The interview is a widely used and powerful method of gathering research data (Bouchard, Jr., 1976, p.368). As a technique for personnel assessment and selection it has been seriously questioned by a number of researchers (Carlson, Thayer, Mayfield & Petersen, 1971) but Cannell and Kahn (1953, p.330) have pointed out that when the focal data for a research project are the attitudes and perceptions of individuals, the most direct and often the most fruitful way of obtaining the information is to ask the individuals themselves. Observational methods are less likely to be useful for the measurement of attitudes and perceptions and are obviously unable to probe the past or to determine an individual's intentions for the future. The present interviews were designed to investigate past experiences and problem areas associated with the management of people, the beliefs and feelings associated with this task and the future intentions of people occupying leadership roles. Taken

by themselves, the pre-course questionnaires which of necessity limited the number of categories of response for ease of quantification could not tap the same level of depth or insight.

Limitations of the interview

- (1) The involvement of the individual in the data she is reporting and the consequent likelihood of bias.
- (2) The inability of the respondent to provide certain types of information. The interviewer is then thrown back on her own ability to make valid inferences from the personal information obtained.
- (3) Memory bias and the problem of recall. It is, therefore, necessary to carry out a longitudinal research design over a period of time, for example, before, during and at intervals after training when the ability of the respondent to recall is at a high level.

The major problems in interviewing stem from the inability or the unwillingness of the respondent to communicate. The interviewer cannot always apply an identical set of techniques because she is dealing with a varying situation. Thus the value of the interchange is dependent upon the skills of the interviewer and the flexibility achieved during the interview sessions within the limitations imposed by the need for standardization in the interests of reliability.

Respondent Motivation

In this case it was judged to be reasonably high because subjects had been prepared in advance, they understood the purposes of the interview (this was ascertained and clarified where necessary, prior to launching into the key questions) and they had some personal commitment to the goals of the evaluation project because they were interested in bringing about necessary changes in the course itself. Secondly, opportunity was offered to the respondent to talk about topics in which he/she was interested and involved, namely, his/her job and job performance. Furthermore, the subject was assured of confidentiality regarding his/her identity, and would not be threatened by the interview situation.

5.2.3 PROCEDURE

The procedure chosen in this case to determine the major dimensions of managerial performance was similar to the method of scaled expectations (Smith and Kendall, 1963; Blood, 1974) which consists of the following steps:

- (1) The critical incident technique (Flanagan, 1954) is used to get people from within the organization to generate examples of behaviour illustrating job components.
- (2) The specific behavioural incidents are submitted to a qualitative cluster analysis by other members of the organization or by the experimenter and sorted into homogenous categories.
- (3) Retranslation - the tentatively defined performance dimensions and incidents are re-assessed by other personnel and incidents retained only if they can be reliably placed in their original dimensions.
- (4) Each surviving incident is scaled by members of the organization on a dimension representing "good" to "poor" performance.

A fuller discussion of this method was presented in chapter 3.

Each interview (Step 1, above) took between one and one and a half hours to complete and was loosely structured around the following questions:-

- (1) For controlling officers -
 - (a) What do you consider are the particular benefits to be derived from the science management course for your staff?
 - (b) As you consider each member of your staff in turn, what are the particular changes you would hope to observe in them as a result of the training course?

- (2) For course participants -
- (a) Describe briefly the nature of your management position and what it involves.
 - (b) What do you hope to get out of the course, personally?
 - (c) What particular changes would you like to make in your own behaviour in your role of section leader?

These questions were adapted to suit different groups.

Apart from these rather broad outlines, the interviews were conducted in an entirely open-ended manner, interviewees being encouraged to comment freely on their own particular needs in the area of management skills and to state their personal opinions and feelings with a minimum of direction from the interviewer.

Critical Incidents (Flanagan, 1954)

Throughout the interviewing sessions, respondents were encouraged to provide specific incidents of their own or other people's behaviour which they considered to be examples of particularly good or particularly poor performance. These incidents were all recorded immediately by the interviewer and in this way she was able to accumulate a large pool of critical behaviours associated with the role of section leader.

Since the interviewer adopted a "problem centred" approach and questions were structured around the perceived needs of the trainees the responses focused on those aspects of their behaviour where, in their opinion, change or improvement was considered to be desirable. Thus it was possible to calculate the number and frequency of specific problems raised. Each problem was stated in reference to the respondent himself (in the case of course participants) or in relation to a staff member (in the case of the controlling officers).

A total of 222 behaviour incidents generated by this method were written on separate cards and these were subsequently sorted into

15 separate categories of managerial behaviour by the experimenter (Step 2), who then attached tentative labels to each of these categories. At this stage the information could have been referred to other members of the organization for retranslation (Step 3). However, an alternative method adopted in the present case was to obtain a second opinion by an independent judge who was not associated with the study nor with the organizations. He independently assigned the cards bearing the behavioural incidents according to his own categorization system (see section 5.2.6 for an analysis of these judgements).

Relationship between Interview and Pre-course Questionnaire

These interviews together with the pre-course questionnaires afforded a two-pronged approach to the estimation of pre-training levels of performance:

- (1) Direct approach - pre-course questionnaires were used to obtain quantifiable data (Appendix III)
- (2) Indirect approach - the interviewer adopted a funnel approach, using open questions requiring more general answers at the beginning followed up with questions requiring specific details and critical incidents later on.

The content of both approaches was determined by the research objectives. It may also be noted that the interview tapped a slightly different conceptual level than the pre-course questionnaire since the interviews were not conducted with special reference to the course content and indeed, in most cases, occurred before any of the course material had reached the respondent whereas the questionnaire was sent out later when the trainees had at least had an opportunity to study the programme and were aware of the topics to be covered during the course.

5.2.4 RESULTS

The 222 critical incidents collected by the above method all consisted of problems raised by the 86 respondents or behavioural incidents in which changes were considered to be desirable. They were sorted initially into 10 main categories by the experimenter and labelled as follows from the most frequently to the least frequently mentioned:

- (1) Personal and group relationships.
- (2) General administrative skills.
- (3) Information about the organizations, other related government departments and industry.
- (4) Forecasting and planning procedures.
- (5) Development of organizational skills.
- (6) Resource management.
- (7) Interaction between divisions.
- (8) Decision-making skills.
- (9) Reporting and marking procedures.
- (10) Report publishing.

From Table 5.2 it can be seen that of the 222 responses, 85 referred to the Personal and group relationships, followed by General administrative skills (30), Information (29), Forecasting and planning (24), Organizational skills (20), Resource management (16), Interactions between divisions (10), Decision-making skills (5), Reporting and marking (2) and Publishing (1). The subjects were then divided into four sub-groups:

- (1) Controlling officers
- (2) Government department section leaders
- (3) Government department non section leaders
- (4) Personnel from semi-government departments

It was found that while the controlling officers' order of priorities corresponded to the above pattern, there were slight variations between other sub-groups. For example, employees of the semi-government organizations placed Organizational skills ahead of Forecasting and planning and government section leaders put Information, Forecasting and planning and Resource management before General administrative skills. Those

people who were not presently in leadership roles also preferred Personal and group relationships before General administrative skills but, on the whole, there tended to be considerable agreement about important training needs.

Table 5.2

Number of responses in ten categories of management
by four sub-groups of respondents

	GOVERNMENT CONTROLLING OFFICERS (n = 18)	SEMI GOVERNMENT PARTICIPANTS (n = 10)	LEADERS (n = 41)	NON- LEADERS (n = 17)	TOTALS (n = 86)
1. Personal & group relationships	23	9	40	13	85
2. General administration	16	3	6	5	30
3. Information	12	1	12	4	29
4. Forecasting & planning	8	2	12	2	24
5. Organizational skills	6	6	6	2	20
6. Resource management	5	2	9	0	16
7. Interaction between divisions	3	0	6	1	10
8. Decision- making skills	3	0	0	2	5
9. Reporting and marketing	0	0	2	0	2
10. Publishing	1	0	0	0	1
TOTALS	77	23	93	29	222

5.2.5 DISCUSSION

In order to clarify the meanings attached to the 10 categories of managerial behaviour, a descriptive summary of each is given together with a sample of the actual comments made by respondents.

1. Personal and Group Relationships:

Since the general area gave rise to such a large number and variety of different problems, it was decided to divide it further into a number of subsections for examination. (Table 5.3). A primary distinction was made between relationships within the department or organization (internal relationships) and those outside, with industry, other government departments, the scientific community and the community at large (external relationships).

- (1) *External Relationships* involve presenting oneself and one's ideas to a wide range of people from other divisions, government departments, business and industry. The following examples are typical of the comments in this category:-

Public relations with manufacturers.

Management of industrial relations.

Managing contacts with people from different divisions and departments.

Working with people in industry; a greater awareness of the client population and structures within industry.

Communication with the news media.

Oral presentation of ideas to groups.

- (2) *Internal Relationships* require a set of skills which may be subdivided into a further five distinct types.

- a) *General Interpersonal Skills* involve gaining confidence and self assurance in dealing with people on the one hand, while increasing understanding, co-operation and tolerance on the other. They include the ability to handle interpersonal conflicts.

Examples are:-

Need to get along with others - understand what makes people "tick".

Tolerance towards subordinates.

More self assurance in relationships.

Sense of responsibility towards others in the department.

Needs confidence and effectiveness in dealing with people.

Handling interpersonal conflicts.

- b) *Interviewing and Selection* consist of developing interviewing techniques for appraisal and selection purposes. An example of this is:-

Selection of personnel and interviewing techniques.

- c) *Motivating Staff* may be described as encouraging staff within a section to set objectives, plan and execute scientific research and to work to their full capacity.

Examples are:-

Encouraging innovation and creativity and maintaining momentum to the completion of a project.

Motivating people to work to their full capacity.

Motivating and developing the interest of technicians in their work.

- d) *Communicating* includes teaching junior members of staff how to carry out their work and conveying ideas and information horizontally within and between sections and vertically through the department.

Examples are:-

Communication problems with staff.

Efficient ways of teaching people to do jobs.

Communicate problems and ideas upward as well as down.

- e) *Performing at Meetings* as both leader and participant is the fifth sub-type.

Examples are:-

Training in chairing and participating in meetings.

Running meetings and public speaking.

2. Administrative Skills

In this category we include efficient clerical procedures such as letter writing and all types of paper work.

Examples are:-

Basic principles of administration, neatness and tidiness.

More effective in formal letter-writing.

Encouragement to take on more administration.

Appreciation of the need for administration.

Learning about alternative administrative techniques especially as applied to science administration.

A system of dealing with filing.

A matter of "polish".

3. Knowledge and Understanding of Policies

Gaining knowledge and a better understanding of the mechanisms and policies of the organizations, other government departments and

industry is another important area of need.

Examples are:-

Learn how other more complex industries manage from people who are willing to admit their mistakes.

More information concerning State Services.

Broad understanding of the organizations.

Guidelines regarding policies for resource and finance allocation and promotions.

Open up mechanisms for upward and downward communication between Head Office and section level.

Departmental policies concerning basic and applied research.

4. Forecasting and Planning Skills

This implies long range planning of research within the total setting in relation to the needs and values of society and the overall policies and goals of the department or organization.

Examples are:-

Planning of projects in respect to the total goals of the organization.

Overall planning with other divisions.

Learn to take a hand in the overall policy-making of the department.

Meshing the objectives of research with respect to national problems.

Originality and ability to project into the future.

Formalized planning and the need to use government money well.

5. Organizational Skills

Organization and planning of day to day work of self and others including team and project work is involved here. This includes the setting of objectives and delegation of responsibility.

Examples are:-

Planning programmes of work for others.

Seeing shorter ways of doing things.

Organization and streamlining of work involving teams.

Learn to delegate work and become more effective thereby.

Control of projects.

Setting objectives, organizing time and commitment priorities.

6. Resource Management

This topic requires an understanding of the most effective ways of budgeting and allocating finances and other resources within the limitations imposed by the organization.

Examples are:-

Better working knowledge of budgeting and finance within the department.

Gaining expertise in cost-benefit analysis of research.

Need to understand limitations of financial resources.

Procedures for purchasing of stores.

7. Interactions between Divisions

This means making informal contact and learning to communicate and understand the problems of people in other divisions.

Examples are:-

Mixing with people in other divisions and understanding their problems.

Communicating about common problems.

Working with and understanding people from other divisions.

8. Decision Making Skills

Problem solving and decision making activities are both included here.

Examples are:-

Make better decisions taking all information into account.

Learning to recognize the problems.

9. Reporting and Marking

These terms are commonly employed to describe the system used by section leaders to assess the performance of their technicians.

Examples are:-

Promotions are a great problem.

Developing a more flexible approach to appraisal of personnel.

10. Report Publishing

This category refers to the publication of scientific research. An example of an incident from this category of work behaviour is:-

No good doing research without publishing.

Table 5.3

Number of responses in five categories of personal and group relationships by four subgroups of respondents

	GOVERNMENT CONTROLLING OFFICERS (n = 18)	SEMI GOVERNMENT PARTICIPANTS (n = 10)	LEADERS (n = 41)	NON- LEADERS (n = 17)	TOTALS
External Relationships	3	3	7	2	15
Internal Relationships					
1. General inter-personal skills	17	3	10	4	34
2. Interviewing and selection	0	1	9	3	13
3. Motivating staff	3	1	6	0	10
4. Communicating	0	1	4	3	8
5. Running of meetings	0	0	4	1	5
Totals	23	9	40	13	85

5.2.6 INTER-RATER RELIABILITY

After the subdivision of category one into six sections by Judge A (the experimenter), qualitative analysis of the 222 behavioural incidents produced 15 separate categories which were labelled as follows:

1. Administrative skills
2. Information about the organizations

3. Interactions between divisions
4. Forecasting and planning
5. Organizational skills
6. Resource management
7. Publishing
8. Conducting meetings
9. Decision making skills
10. Reporting and marking
11. General interpersonal skills
12. Communication
13. Interviewing and selection
14. Motivating staff
15. External relationships

The same behavioural incidents were subsequently assigned to 11 categories by Judge B (independent judge) who labelled them:-

1. Administration
2. Organizational issues
3. Planning
4. Self improvement
5. Budget and finance
6. Research issues
7. Human relations
8. Communication
9. Interviewing
10. Motivation
11. Public relations

There is clearly a close correspondence in the labels used by both judges. Table 5.4 is a contingency table where the rows represent the classification used by Judge A and the columns represent the classi-

fication used by Judge B. The numbers in the cells are the numbers of behavioural incidents assigned to these categories.

The writer explored the question of whether the two classifications (that of the experimenter and the independent judge) could be assumed to be statistically similar.

Brennan and Light (1974) considered the case where two observers classify objects into categories which are not defined in advance. They showed that the total number of observed agreements (A') which are pairs placed in the same category by both raters or in different categories by both raters, is given by

$$A' = \frac{n(n-1)}{2} + \sum_{i=1}^r \sum_{j=1}^c n_{ij}^2 - \frac{1}{2} \left(\sum_{i=1}^r n_i^2 + \sum_{j=1}^c n_j^2 \right)$$

where n = total number of statements (objects)

n_{ij} = Number of statements in the ij cell

n_i = row totals

n_j = column totals

With reasonably large n 's this is approximately a standard normal distribution so that in the present example, we have $A' = 21,128$.

The expected value of A' i.e. $E(A') = 19,424$

with a variance, $\text{var}(A') = 1432.9$

This gives a standard normal score of

$$Z = \frac{A' - E(A')}{\text{standard error}(A')} = 45.0 \quad (\text{significant at } p \leq .01)$$

Clearly, this is significant and indicates that there is considerable agreement in the classification by the two judges of pairs of incidents.

From Table 5.4 one can see that Judges A and B agree strongly on the following categories:

- (1) 1A (Administrative skills) and 1B (Administration).
- (2) 2A (Information about the organization) and 2B(Organizational issues).

- (3) 4A (Forecasting and planning) and 3B (Planning).
- (4) 5A (Organizational skills) and 3B (Planning).
- (5) 6A (Resource management) and 5B (Budget and finance).
- (6) 11A (General interpersonal skills) and 7B (Human relations).
- (7) 12A (Communicating) and 8B (Communication).
- (8) 13A (Interviewing and selection) and 9B (Interviewing).
- (9) 14A (Motivating staff) and 10B (Motivation).

The extra categories employed by Judge A who used 15 categories tended to be incorporated into one or other somewhat similar category used by Judge B who had only 11 categories at his disposal. For example, category 8A (Conducting meetings) and 9A (Decision making skills) were both included in 1B (Administration). 2B (Organizational issues) incorporated both 2A (Information about the organization) and 3A (Interactions between divisions).

On the other hand, there was little or no agreement on Judge A's fifteenth category, 15A (External relationships), which was not a separate category for Judge B.

The more subtle discriminations made by Judge A (the experimenter) may have been due to her much greater familiarity with the organizations concerned. Judge B knew nothing of the organizations at all. He had simply been presented with the stack of cards on which each of the critical incidents of behaviour had been recorded.

Table 5.4

The frequency of incidents placed in categories by the two judges.

		Categories of Independent Judge B.											Total
		1	2	3	4	5	6	7	8	9	10	11	
Categories of Experimenter Judge A.	1	24	-	-	5	1	-	-	-	-	-	-	30 = $n_{1.}$
	2	4	16	1	1	3	-	-	2	-	-	2	29 = $n_{2.}$
	3	1	7	-	-	-	-	-	2	-	-	-	10 = $n_{3.}$
	4	4	4	13	-	-	-	-	-	-	-	4	25 = .
	5	7	1	12	-	-	-	-	-	-	-	-	20 = .
	6	3	1	-	-	11	-	-	-	-	-	-	15 = .
	7	-	-	-	-	-	1	-	-	-	-	-	1 = .
	8	5	-	-	-	-	-	-	-	-	-	-	5 = .
	9	5	-	-	-	-	-	-	-	-	-	-	5 = .
	10	1	-	-	-	-	-	-	1	-	-	-	2 = .
	11	3	2	-	3	-	-	26	-	-	-	-	34 = .
	12	1	-	-	-	-	-	-	7	-	-	-	8 = .
	13	-	-	-	-	-	-	1	-	12	-	-	13 = .
	14	-	-	-	-	-	-	-	-	-	10	-	10 = .
	15	2	1	-	1	-	1	3	3	-	-	4	15 = $n_{15.}$
Total	60	32	26	10	15	2	30	15	12	10	10	222 = n	

$$= n_{.1} \quad n_{.2} \quad n_{.3} \quad \dots \quad n_{.11}$$

5.2.7 CONCLUSIONS

In this investigation the writer has attempted to summarize the training needs of the organizations as reported by the trainees, future trainees and their controlling officers. The critical incidents were categorized twice, once by the experimenter and once by an independent judge. Judge A identified 15 separate categories or dimensions while Judge B identified eleven. The classifications used were not significantly different even though the categories were not pre-determined.

The frequency with which incidents were freely mentioned by subjects during the interview sessions was used to indicate the priority they placed on certain problem areas. These behavioural statements were critical aspects of the job of section leader and provided detailed descriptions of each of the important job dimensions. Thus they could be assessed following training. They would also permit more objective goals of training to be set for future training courses but, as explained earlier, this second purpose could be implemented only after the present evaluation study had been completed.

5.3

CONSTRUCTION OF ATTITUDE SCALE

Triandis (1971) defines an attitude as "an idea charged with emotion which predisposes a class of actions to a particular class of social situations". Attitudes are inferred from the consistencies in response observed in the individual's behaviour in situations involving the social object. In this case the social object is the section leader and the various functions associated with that role. Triandis describes the three component parts of an attitude.

- (1) A cognitive component: This is the minimum requirement of an attitude and consists of an *idea* which is generally some category used by human beings in their thinking.
- (2) The affective component: This is the *emotion* which charges the idea. Thus a person "feels good" or "feels bad" when he thinks about the category - he evaluates the role of section leader.
- (3) The behavioural component: This involves a predisposition to action and is modified by the behavioural intentions of the

individual, habits, cultural norms, opportunity and ability to perform the particular behaviour and the desirability of the outcomes of action.

Although there is a strong tendency towards consistency among these components, there is the possibility of inconsistency and when change occurs the three components of attitude may change at different rates. Attitudes are changed through both direct and indirect means. While direct experiences with the attitude object usually change all of the components of the attitude, indirect experiences typically change the cognitive or behavioural components, since they are usually informational or normative in nature.

The most popular types of attitude measurement are the standardized verbal specific methods, which were developed by Thurstone (1928), Likert (1932), Guttman (1944) and Edwards and Kilpatrick (1948). These have been summarised by Edwards (1957a). In all these methods a number of opinion statements which refer to the attitude object are accumulated. They are then subjected to a screening procedure so that factual, ambiguous, confusing and excessively long statements can be eliminated. Most of the standardized methods of measurement use verbal responses and attempt to assess the degree of positive or negative affect associated with some psychological object (Thurstone, 1931). There is considerable doubt as to whether or not it is possible to obtain adequate measures of the affective component of attitudes with purely verbal methods of measurement (relatively pure methods of measurement of the affective component are physiological, such as pulse rate, galvanic skin resistance, heart-beat, and so on). As soon as words are used then some cognitive elements are introduced into the measurement (Dillehay, Bruvold & Seigel, 1969).

A number of attitude scales of the above types have been used with managers in business and industry to assess the effectiveness of training. The items generally centre around the problems of supervision, attitudes toward supervisors and subordinates, attitudes towards various leadership methods and attitudes toward the organization itself. Questionnaire type measures include, the Leadership Opinion Questionnaire (LOQ). This was developed by Fleishman (1957b) and is used by the subject

himself for the purposes of self-description while a companion instrument, the Leadership Behaviour Description Questionnaire (Fleishman, 1957a) can be used by subordinates, peers or superiors to describe the leadership behaviour of the individual in question. "How Supervise?" developed by File and Remmers (1948) is also a paper-and-pencil instrument which the individual can use to describe his attitudes toward a number of different aspects of the work environment and the supervisory or leadership role.

Because of the somewhat unique nature of the function of a section leader in science management, we decided to construct for our own purposes a questionnaire which related more directly to the role of section leaders.

5.3.1 SUBJECTS

A total of seventeen individuals, none of whom were involved in the main part of the evaluation took part in this second part of the preliminary study (see Table 41, p. 59). Attitude statements were obtained from eight scientists from four separate divisions of the main government research organization. A further nine individuals from different levels and job categories within the organization were involved in the second stage of development of the attitude scale. All of these latter individuals were familiar on a day to day basis with the section leader role.

5.3.2 ATTITUDE SCALE DEVELOPMENT

The development of this attitude scale required a special scaling technique which was adapted directly from a method used to construct a questionnaire to measure versatility in machine operators (Smith and Smith, 1975).

The instrument was designed as a means of identifying favourable attitudes concerning the role of section leader. It is based on Thurstone's method of successive intervals for deriving an attitude scale (Thurstone, 1927a, 1927b; Edwards, 1957a). Eighty statements about the section leader role including managerial style, working relationships with controlling officers, peers and subordinates, job content and responsibility, were

collected. These statements were obtained from interviews with eight scientists from four divisions. Each statement was typed on a separate card. Another nine people from different levels and job categories within the organization, but all of whom were very familiar with the work of a section leader, were asked to read each statement and judge whether it would apply to a 'good' or 'poor' section leader or someone in between. They were asked to decide where on a nine-point scale the statement should come. For convenience nine cards were marked each with a letter A to I arranged in order so the judges could place the statement cards beneath them. "A" represented attitude statements applying to very good section leaders and "I" statements applicable to very poor section leaders (B, C, D, E, F, G, H were points in between).

To select the eighteen statements wanted for the attitude questionnaire, the mean and standard deviation and the frequency of responses for each of the scale points were calculated for each statement by means of assigning a score to each scale point (A = 9, I = 1). To select the two statements for the scale point A, we looked for statements with the highest mean (8+), lowest standard deviation, and the highest frequency of responses in the A category. For the scale point B, we looked at statements with means around 7, and again low standard deviations and a high number of responses in the B category. And so the process was repeated throughout. Statements were excluded from consideration if the standard deviation was very large as these tended to be ambiguous statements about which the individual judges showed little agreement. On the other hand, statements having exceptionally low standard deviations were scrutinized carefully before they were included as these tended to be either factual or somewhat trite statements for which only one socially desirable response was possible. If two statements seemed to be almost equal according to the three methods of selection above and one had to be excluded, the choice of statement was made on its verbal acceptability. In the end there were two statements for each scale point based on the greatest agreement amongst judges. The wording, also, of the statements was examined and modified slightly according to the usual considerations for questionnaire construction (see Edwards, 1957a, p.13-14, list of informal criteria for attitude statements).

It was decided that these eighteen statements should be listed randomly on a form and testees would be asked to read each one and decide whether they agreed or disagreed with it. There were two columns on the form marked 'agree' and 'disagree' and testees were asked to place a tick in the appropriate one according to their decision. If they really could not make up their minds about any particular statement they were asked to tick a third column marked "don't know".

The scoring was based on where each of the eighteen statements came on the nine-point scale, with each item receiving a positive weighting from 0-4. Thus a high overall score on this weighted scale was taken to indicate 'good' and low score 'poor' section leader attitude. For the eight statements which came in the A to D bracket, marks were given if the testee agreed with the statement. These statements were phrased in such a way that agreement with them indicated 'good' section leader attitude; A statements received 4 marks, B statements 3 marks, C statements 2 marks and D, 1 mark. The list of items selected with the weightings used in scoring and the final form of the attitude scale questionnaire is to be found in Appendix III, p. 279.

For the eight statements which came in the F to I bracket marks were given if the testee disagreed with the statement. All these statements were phrased in such a way that disagreement with them indicated 'good' section leader attitudes; F statements received 1 mark, G statements 2 marks, H, 3 marks and I, 4 marks.

The two E statements were regarded as being neutral and received 0 marks. A "don't know" response was also allocated 0 marks and this therefore had the effect of equating it with disagreement for the A to D statements and agreement with the F to I statements. This is unfortunate but somewhat unavoidable as it was thought necessary to have a "don't know" column to provide for those people who really could not decide. It is likely that they would have failed to record any answer at all for such statements. Thus the maximum score possible for the 18 item questionnaire was 40 and the minimum 0.

5.3.3 CONCLUDING COMMENTS

Unfortunately, there was no further time available to test and validate this measuring instrument but in view of its careful and

highly specialized construction and the fact that a similar one had been used successfully on previous occasions, it was decided to adopt it as one measure of attitude change in the main study. By doing so, results obtained could be compared with other measures of attitude change, in particular, responses to interview questions and behavioural changes observed in the work situation following training.

As noted earlier, these preliminary studies constituted a compromise between the need to operationalize and refine the instruments used to measure the dependent variables and the constraints of the practical situation. One of the constraints was the need to adapt to a tight time schedule. Other limitations included the availability and accessibility of members of the client organizations most of whom were scattered over a large geographical area and all of whom were highly involved in their own professional activities. Nevertheless, having set priorities for this research with the aim of maximizing possible gains, the present researcher was satisfied that the two most immediate aims, that of further criterion development and that of measuring instrument construction was sufficiently well advanced to proceed with the main evaluation.

6.1

INTRODUCTION

From the discussion of evaluation literature in preceding chapters, it can be seen that the measurement of training objectives or outcome evaluation may be directed at three levels,

- (1) immediate outcomes
- (2) intermediate outcomes
- (3) ultimate outcomes

(Warr, Bird and Rackham, 1976). However, due to the difficulties of establishing ultimate criteria for managerial performance (Campbell, Dunnette, Lawler and Weick, 1970) and on the assumption that changes occurring at the first two levels will ultimately help to achieve organizational goals (Hamblin, 1974) most of our efforts were concentrated on levels (1) and (2). At the same time, considerable effort was made to provide a monitoring function and to supply some feedback of information and results to the controllers of the management training programme during the time that the evaluation was in progress. This was implemented mainly by a series of written and verbal statements reporting findings at progressive stages of evaluation during the two year period. Additionally, attention was paid to the work environment of the trainees, the input or training methods used and the reactions of trainees to the course both immediately and from a longer term perspective.

It is significant that many of the evaluation techniques employed do not fall exclusively into one or other of the measurement categories but they overlap so that one technique may serve two purposes. An example of this is self-reported behaviour which was used to assess both trainee reaction and job performance within the one measure, following the suggestions of Dalziel, McWilliam, Strong, Haynes and Lunt (1972) and of Burgoyne (1973). A second example was the preliminary analysis of training needs which served to determine criteria for the evaluation of training outcomes as well as to establish pre-training levels of performance.

Immediate outcome evaluation consisted principally of measures of changes in knowledge, skills and attitudes by means of questionnaires, tests, interviews, self reports and the reports of controlling officers. Similar methods were used at the intermediate level but here interest shifted to the assessment of changes in work performance with emphasis on concrete examples of on-the-job behaviour. Such measures were taken at three, six and twelve months following training and a variety of different strategies were used to accomplish this part of the evaluation task. This use of mixed dependent variable measures, advocated by Suchman (1967) and Riecken (1977), was especially valuable in the present study where, as we shall see later, a very heterogeneous group of subjects made it necessary to adopt a flexible approach to measurement. A measure of organizational climate as well as a series of open-ended questions provided an evaluation of the work environment and formal session assessment forms together with questionnaire responses were used to assess the different training techniques employed. As we have seen in Chapter 5, the initial interviews conducted with trainees, their peers, subordinates and controlling officers provided data for the classification and description of the section leader role and facilitated the development of questionnaires and other measures and scales used during the study.

The evaluation progressed through a series of seven phases as outlined in Table 4.1, (p. 59). Within each of these phases there were a number of separate experiments which will be described in detail in subsequent chapters. Testing of subjects continued at intervals over a period of approximately 18 months from the administration of pre-training measures to the 12-month post-training follow-up and the evaluation of the following year's trainees.

6.2

SUBJECTS

The subjects in the present evaluation study were all members of government or semi-government institutions engaged in scientific and industrial research. The trainees themselves were senior scientists or technical officers who had reached or who would soon be promoted to positions of leadership as section or group leaders of teams of scientists and junior staff engaged in research projects, consultancy and

advisory work and the publication of material for a variety of audiences. The primary evaluation was conducted with the two intakes of trainees in the 1977 science management training course. The first of these, Group A, consisting of 24 participants who together with Group B (n = 23) constituted the principal set of subjects to be evaluated. Subsequently, two further intakes of trainees in 1978, Group C (n = 24) and Group D (n = 21) were similarly tested for comparison with the 1977 group results. The cooperation of two more samples of subjects was obtained for control purposes. Of these, a total of 11 ex-trainees from previous courses (Group E) and 19 future trainees (Group F) were selected (Table 6.1 & 6.2).

Table 6.1

Subjects involved in main evaluation study categorized
by group (rows) and type of employing
organization (columns)

Group	Government Department	Semi Government	Totals
1977 Trainees A - Experimental	20	4	24
1977 Trainees B - Experimental	20	3	23
1978 Trainees C - Replication	20	4	24
1978 Trainees D - Replication	17	4	21
Previously Trained E - Control	9	2	11
Future Trainees F - Control	18	1	19
Controlling Officers Superiors	20	3	23
Total			145

In addition to this, 23 controlling officers and directors of the above mentioned individuals were interviewed before and surveyed again after the 1977 training courses. Altogether there were 145 subjects involved and they were derived from a total of 29 separate divisions or associations of the organizations concerned.

Table 6.2

Biographical characteristics of subjects
in all groups

Group	No. of Subjects	No. of divisions or associations	No. of leaders ^a	No. of semi-government associations ^a	No. of technicians ^a
A	24	21	19(79)	4(17)	0(0)
B	23	19	17(74)	4(17)	1(9)
C	24	20	19(79)	4(17)	3(13)
D	21	20	12(57)	4(19)	2(1)
E	9/11 ^b	10	9(100)	2(22)	0(0)
F	16/19 ^c	10	9(56)	1(6)	1(6)

a - Number in parentheses indicates percentages.

b - Two subjects in this group did not supply biographical details.

c - Three subjects in this group did not supply biographical details.

In a field study of this kind, the researcher has little or no say in selecting matched groups for training and control purposes therefore she must abide by the organizational decisions in this matter. The question then arises of the similarity of the training and control groups prior to training. In this case the control Group F members were selected from a master list of possible candidates for future courses and therefore they tended to be slightly younger and less experienced within their organizations than the experimental groups. (Table 6.3)

Table 6.3

Average ages, tenure and leadership experience of
trainees and control group members

Group	AGE (Yrs)			TENURE (Yrs)			LEADERSHIP EXPERIENCE (Yrs)		
	Mean	S.D. ^a	Range	Mean	S.D. ^a	Range	Mean	S.D. ^a	Range
A	39.375	6.933	28-53	10.652	5.851	2-21	5.083	4.898	0-17
B	39.957	6.526	30-50	12.913	8.179	2-32	3.391	3.187	0-12
C	36.500	6.186	28-54	9.375	5.601	1-23	4.708	5.279	0-19
D	38.667	6.445	27-53	10.555	5.844	2-24	3.737	4.689	0-17
E	35.889	5.555	29-45	9.000	5.268	5-21	2.000	2.179	0-6
F	36.625	5.136	30-48	7.313	4.785	3-22	3.375	2.500	0-9

^a S.D. - stands for standard deviation

In order to examine the differences between Groups A and B (the main experimental groups for 1977), Groups C and D (the training groups involved in the 1978 replication study), Group E (trainees of previous courses) and Group F (untrained control group) an analysis of variance was performed over the six groups for each of the variables, age, tenure and leadership experience. The ANOVA tables are presented below.

ANOVA for Age

Source of Variance	df.	SS	MS	F
Between Group	5	265.230	53.047	1.339
Within Group	111	4395.914	39.603	(not significant)

ANOVA for Tenure

Between Group	5	338.669	67.734	1.778
Within Group	111	4229.142	38.100	(not significant)

ANOVA for Leadership Experience

Between Group	5	92.236	18.447	1.030
Within Group	111	92.236	17.907	(not significant)

Under the null hypothesis the differences within and between groups could be attributed to chance at the .05 level of significance for all three variables, age, tenure and leadership experience. Thus, we can assume that we have good comparability between all six groups, experimentals and controls on these important characteristics.

In general, a majority of the trainees had been employed by their organization for more than 5, but less than 20 years (72% of trainees) and had occupied a leadership position for less than 5 years (69% of trainees) and most were between the ages of 30 and 45 (77% of trainees). It is important to note that the people with whom we are dealing are a heterogeneous group in terms of age, length of service, scientific discipline and divisional affiliation (see Table 6.1 and 6.2). The disciplines represented ranged from physics and nuclear science to botany and entomology while the research institutions were associated with the wool, dairy, concrete, building and other industries. The divisions, while they tended to be organized along disciplinary lines, nevertheless, varied greatly in size, structure and status. Moreover, there were wide individual discrepancies in previous leadership experience and while some trainees would assume managerial positions immediately after training others would encounter a considerable time lapse before they could make use of their present training.

Lastly, the experimenter was not able to obtain an additional group to control for the effects of the experimental manipulation on subjects. It is always possible that the process of reading, considering and responding to items in a questionnaire and preparatory material for the courses serve as important stimuli upon each group member and moreover that the nature of these stimuli differs from pretraining members as opposed to the control group. These effects were controlled to some extent by the use of pre-course interviews and questionnaires which were administered to all subjects, experimental and control alike.

6.3

METHODS

A summary follows of the testing procedures used at different levels and stages of evaluation. Further details are given in later chapters where individual experiments are reported.

6.3.1 CONTEXT EVALUATION

Defined by Warr, Bird and Rackham (1976) as the process of obtaining and using information about the operational situation, context evaluation is concerned with the need to develop clear and objective criteria or goals of training and evaluation. This aspect has been repeatedly emphasized by Hamblin (1974), Warr, et al. (1976), Campbell, et al. (1970) and others engaged in evaluative research. They are also well aware of the grave difficulties inherent in the task especially when it is directed towards managerial behaviour. The writer has already referred to the considerable research effort which has gone into job analysis of the managerial role. In the present study we have concentrated on what Hamblin calls a problem - centred approach aimed at the identification and analysis of training needs. The method adopted here and described in detail in Chapter 5 of this report consisted of using a critical incident approach (Flanagan, 1954) during the preliminary interviews to collect a large number of behavioural statements. Additionally, the stated goals of training were discussed with training officers and other members of the organization to determine their opinions about those dimensions of organizational behaviour which required change. Further sorting techniques were used to eliminate redundant and irrelevant behavioural statements and to define performance dimensions (Smith and Kendall, 1963). Finally, critical job features were placed within appropriate dimensions by the experimenter while an independent judge was asked to make a separate judgement. The agreement between the two judge's classification of critical behaviours was assessed (See Section 5.2.6).

6.3.2 PRE-TEST MEASURES

Questionnaire I (Appendix III, p.277) was mailed out to all groups of subjects just before the first of the training courses in 1977 and again in 1978 (in the case of groups C and D). Questions referred to their reasons for attending courses, expected areas of personal and organizational change, preferred training techniques, trainees' pre-test understanding of topics and the perceived relevance of the topics. A biographical questionnaire (Appendix III, p.276) collected data on personal characteristics to permit identification of intergroup differences based on age, organizational and work experience variables. The atti-

tude scale developed during the preliminary study (Chapter 5) was also completed by the subjects at this time.

6.3.3 REACTIONS EVALUATION

The Leipas Scale (Appendix IV, p.280) was administered daily to Groups A and B to assess immediate reactions to individual sessions. A similar session assessment form had been used for the evaluation of a Rural Psychology Course at Flock House Training Institute in New Zealand in 1975. It is a particular example of a general type of questionnaire described by Hamblin (1974, pp.74-82). LEIPAS is a mnemonic which stands for Learning, Entertaining, Interest in topic, Participation opportunities, Application to job and Structure. Average scores on each of the scales were calculated and rank ordered to facilitate rapid overall appraisal. This procedure allowed data to be gathered leading to a comparison of the different sessions.

Questionnaire 2, (Appendix IV, p.281) administered on the last day of the course, asked trainees to indicate their current understanding of the topics covered, the relevance of the topics to their work and their opinions about length of sessions, methods of instruction and overall impressions of the course.

Questionnaire 3, (Appendix VI, p.284) which was completed at the end of the 3-4 months post-training period, similarly, was designed to assess reactions but, in this case, it was tapping longer-lasting and more mature reactions to the courses. As well as this it functioned as a self-report device for the measurement of intermediate outcomes. It can be seen that both of these questionnaires were linked with Questionnaire 1 (Appendix III, p.277) which had probed trainees' expectations concerning the course.

6.3.4 IMMEDIATE OUTCOME EVALUATION

(a) Changes in Knowledge: Unfortunately it was impossible to arrange the administration of pre- and post-training objective quizzes of content which would test the acquisition of factual information imparted during the training course. Such a measure would have required

advance knowledge of details of course content and this was not available prior to the course. As a compromise the evaluation had to rely on the self-reports of the trainees concerning their level of understanding of the topics covered, measured before and after training (immediately after and three months later).

(b) Changes in Skill: An attempt was made on a trial basis to monitor the syndicate group sessions through the entire course period for Group A. However, with three separate syndicate groups operating simultaneously and only one observer, the behavioural data obtained was too sketchy to attempt any type of statistical analysis and the practice was discontinued. There were indications, nonetheless, that such observational methods have good potential for providing valuable personal feedback on performance as Dalziel, et al. (1972) have suggested. It would also be a source of objective information under more favourable conditions for data collection and recording.

(c) Changes in attitudes: One of the objectives of training is the alteration in trainees' attitudes towards such topics as interpersonal and interdepartmental communications, the use of support services, project planning and resource management as well as changes in attitudes towards the organization as a whole and their particular role in it. The specially constructed attitude scale was administered for a second time at this stage for the purpose of detecting any post-test changes as well as differences between experimental and control groups.

(d) Changes in behaviour: As a means of obtaining base-line data and behavioural indices to facilitate subsequent measures of behavioural changes, subjects were asked to state changes they intended to make in their work behaviour as a result of things they had learned during the course. This follows the suggestions made by Dalziel, et al. (1972) and Morris (1972) of focussing the attention of trainees on personal goals which they hoped to achieve as the result of training.

(e) Work context evaluation: A standard Organization Climate Questionnaire (Appendix V, p.283) was administered to assess situational variables within the work setting which were likely to affect the train-

ees' ability to implement the new knowledge, skills and attitudes acquired during training.

6.3.5 INTERMEDIATE OUTCOME EVALUATION

As noted previously, all trainees were asked at the end of the training course to write down a brief description of one or more changes which they proposed to make in their work as a result of an idea gained during training. After an interval of 3-4 months subjects were requested by means of Questionnaire 3 (Appendix VI, p.284) to specify and explain the degree and type of any behavioural changes which had been attempted in the intervening period.

Fifteen of the subjects were personally interviewed, as well, at this time so that a more intensive investigation could be made of a smaller sub-sample. One of the problems of this particular programme is that not all trainees who complete the course, immediately move into managerial positions and for these people, it is difficult to obtain a realistic estimate of changes in skills and behaviour. Thus the interview subset consisted of those who had had more opportunity to test themselves out in the work situation. As with the first interviews, a critical incident approach was adopted to obtain the most detailed and objective data possible. This data was also used in the construction of a Kelly Repertory Grid (Kelly, 1955) employing work situations as elements of the grid and eliciting the personal constructs of individual subjects. This Repertory Grid was administered at the six month follow-up.

Questionnaire 3 assessed trainees' perceived changes in understanding and relevance of the five topics covered during the course, personal and organizational changes, books read on related topics and personal contacts maintained. These two latter questions served as indirect measures of the effectiveness of the training programme. Areas of organizational and personal change were assessed on a scale corresponding to the one on which precourse expectations were measured.

A six month follow-up provided a second opportunity to assess changes which had occurred during the intermediate outcome period. For this purpose Questionnaire 4 was administered to subjects (Appendix VII, p.287 and Appendix VIII, p.289) in order to assess:-

- (a) Changes in performance: Trainees, their controlling officers and controls all provided assessments of work behaviour by means of structured and open-ended responses to questionnaire items.
- (b) Work environment: An open-ended question gave subjects the opportunity to explain why they had not been able to make the desired changes in their work behaviour.
- (c) A measure of changes in work relationships: Changes in inter-personal relationships at work were assessed by the use of a Kelly Repertory Grid with a sub-sample of trainees and controls. This is an individually administered and ideographic measure which is particularly suitable for the individual case studies presented in chapter 9.

Finally, a twelve month follow-up was conducted in order to evaluate longer-term effects of the training programme. It consisted of:-

- (a) A questionnaire containing three brief scales in which subjects were asked to rate the effects of the previous year's training course on three important aspects of their work performance.
- (1) Effect on work relationships
 - (2) Effect on individual work performance
 - (3) Effect on the organization of their work groups or sections.
- (b) An open-ended statement (Appendix IX, p.293)

6.3.6 ULTIMATE OUTCOME EVALUATION

Although no direct attempt was made in this study to evaluate the management training course in terms of the ultimate objectives and values of the participating organizations, the fact that assessments of work behaviour were continued up until twelve months following training, increased the likelihood that any statements made about the effectiveness of the courses would have more relevance to the longer term goals of the participating organizations. Additionally, questions were put to the subjects at each stage, probing the reasons why they were unable to carry

out the changes which they had proposed. The free response answers to these questions together with responses to the Organization Climate Questionnaire offered some clues to the constraints operating within the organization context.

6.3.7 EXTERNAL VALIDITY AND RELIABILITY OF SCALES

External validity or the extent to which one can generalize the results obtained to other subjects, settings and times as well as the stability or reliability of the measuring instruments used in an important issue in this type of research. Some monitoring of these factors was made possible by the extension of the study to include a new set of trainees one year after the original evaluation was conducted. Groups C and D received a similar set of questionnaires and scales prior to their 1978 science management courses, immediately afterwards and at three months, following training. Results from these two groups were analysed and compared with those of the previous year (See Chapter 10).

6.4

CONCLUDING COMMENTS

It is hoped that this overview together with the more detailed records which follow of the experimental procedures used in the present study will illustrate how each successive experiment in the series provided additional information which eventually built up a much clearer picture of the effects of the training courses on the individuals and the organizations involved. The reader is referred again to the time chart (Table 4.1 p. 59-64) which summarizes phases in the testing sequence corresponding to the successive levels of evaluation outlined in the present chapter.

Regardless of the outcomes of evaluation the detailed nature of the evaluation study should provide sufficient pointers to important strengths and weaknesses in the programme to permit the identification of some of the reasons for success and failure of various aspects of the courses so that further action can be taken to correct mistakes and to modify and improve the existing programme along the lines suggested by the outcomes of evaluation. The writer has maintained from the start that both a formative and a summative evaluation is required and this study has been designed with the intention of fulfilling both of these critical requirements.

7.1

INTRODUCTION

In the pretest questionnaire, base-line data had been gathered on subjects' understanding and perceived relevance of course topics, their preferences concerning types of training techniques, their expectations and reasons for attending the course and their attitudes towards the role of section leader. These scales were administered again immediately at the conclusion of training together with the Organization Climate Questionnaire and another question to determine whether or not sufficient time had been allocated to each topic. These last two measures were continued in follow-up studies and results will be reported in Chapter 8. Thus, as well as the individual session assessment forms completed daily throughout the training course time was set aside on the last day of the course to allow trainees to fill out a questionnaire designed to assess any changes in attitude which may have occurred by the end of the course. This questionnaire covered:

- (1) Subjects' evaluation of course topics, in terms of their perceived understanding and relevance.
- (2) A second measure of preferred training techniques.
- (3) Opinions about the allocation of time to various topics.
- (4) Attitudes towards the section leader role.

The results of these immediate outcome measures together with responses to questions contained in the pre-test questionnaire and the session assessment forms are presented in this chapter. Finally, the trainees were asked, during this final session, to state any changes they intended to make in their work behaviour as a consequence of the course. The responses to this question provided base-line measures for the behavioural changes discussed in later sections.

7.2

REASONS FOR ATTENDING COURSESAim

To discover the reasons trainees give for deciding to attend the science management courses.

Procedure

Subjects: All 24 members of Group A and 23 members of Group B were asked to take part in this section but some declined to respond. The subject numbers are shown in Table 7.1.

Methods: As a pretest measure, shortly before the training courses began, Group A and Group B members were asked their reasons for attending the course. They were given six possible response categories as follows:

- (1) To make social/professional contacts.
- (2) To improve present job performance.
- (3) Personal interest/curiosity and general interest in the topics.
- (4) To enhance prospects for promotion.
- (5) Because I was given little choice in the matter.
- (6) Other - this last category provided the chance to offer a reason other than those suggested by the tester.

They were asked to place a number from 1 to 6 beside each response category thus ranking the six reasons in order of importance (See Appendix III, p.277).

Results

The modal ranking was ascertained for each category for Group A and Group B trainees, separately (See Table 7.1). Thus, the most frequently chosen ranking within each response category was used to form a priority list of reasons given by trainees for attending the course. Group A and B members both produced similar lists.

Table 7.1

Pre-test responses to question concerning reasons
for attending course

(from 1 = most applicable to me to
6 = least applicable to me)

Group A (course 4)

Response Scale	Social, Professional, Contacts	Improve Job Performance	Personal Interest	Promotion	Little Choice
1	0	11 ^a	0	1	8 ^a
2	3	5	11 ^a	2	0
3	8 ^a	5	8	4	1
4	4	1	3	6 ^a	2
5	5	1	1	5	4
6	3	0	0	4	4
	n=23	n=23	n=23	n=22	n=19
Group B (course 5)					
1	0	8 ^a	3	2	5 ^a
2	7 ^a	4	5	1	3
3	4	4	9 ^a	3	2
4	3	3	1	7 ^a	1
5	6	1	1	1	3
6	1	1	1	6	5 ^a
	n=21	n=21	n=20	n=20	n=19

^a Modal ranking on response scale (1 - 6)

The reasons for attending the course were given in the following order of preference:

Group A

- (1) Either, to improve present job performance or because they had little choice in the matter.
- (2) Personal interest/curiosity and general interest in the topics.
- (3) To make social/professional contacts.
- (4) To enhance prospects for promotion.

Group B people reversed reasons 2 and 3 but other than that the order was identical.

Given the opportunity to offer additional reasons of their own trainees mentioned the following:

- (a) Exchange experiences and ideas with others in similar jobs.
- (b) Gain formal management training (particularly science management).
- (c) To experience a group situation and have a chance to evaluate one's own effectiveness.

Discussion

According to these results, it seems that prospective trainees were undecided whether their main reason for attending was to improve their own work performance or because they were directed to do so. Of course, these two motives need not be mutually exclusive. Interest in the topics and the opportunity to meet their fellow scientists were secondary incentives, while they did not see attendance at the course as having much bearing on their chances of promotion. Other reasons considered to be of importance were to improve administration of their sections and their personal effectiveness as managers and to gain a better understanding of managerial practices in general and the functioning of their own organizations, in particular.

7.3 EXPECTED AREAS OF ORGANIZATIONAL AND PERSONAL CHANGE

Aim

To investigate subjects' expectations about the areas of organizational and personal change most likely to be affected by the training courses.

Procedure

As a pre-test measure, Group A and B members and controls (Group F) were asked to check areas of organizational and personal change which they thought would be most influenced by the training courses.

The question specified five organizational and five personal areas in which change could occur and subjects could check as many or as few categories as they chose. An additional open category was provided for subject's own comments. The alternatives suggested were as follows:

- (a) Organizational Change:
 - (1) Internal communication
 - (2) Staff turnover
 - (3) Increase in pool of potential section leaders
 - (4) Awareness of departmental policies
 - (5) Understanding between scientific staff and administrative staff
 - (6) Other

- (b) Personal Change:
 - (1) Job satisfaction
 - (2) Attitude to change
 - (3) Quality of managerial performance
 - (4) Human relations skills
 - (5) Social interaction between course members
 - (6) Other

Respondents were invited to indicate which of these categories were applicable to them.

Results

The results are presented for comparison of differences in expectations between the three groups before training (Tables 7.2 and 7.3). These results were followed up for the two training groups A and B three

months after training to see if there were any changes in their perception of effectiveness of the training courses following training (Section 8.2.2).

Table 7.2

Subjects' rank ordering of areas of organizational change expected

Areas of Organizational Change	Number of Responses			Rank Order		
	Gp A	Gp B	Gp F	Gp A	Gp B	Gp F
Awareness of policies	9	11	9	2	2=	2
Internal communication	7	13	12	3	1	1
Understanding between staff	13	11	8	1	2=	3
Pool of section heads	0	3	0	5	4	5
Staff turnover	1	2	1	4	5	4

Group A n = 24

Group B n = 22

Group F n = 15

As can be seen from the rank ordering of areas of organizational change, there is only a moderate degree ($r_{AB} = 0.65$, $r_{AF} = 0.60$, $r_{BF} = 0.85$) of similarity between subjects' expectations before they attended the training courses.

Table 7.3

Subjects' rank ordering of areas of personal change expected

Areas of Personal Change	Number of Responses			Rank Order		
	Gp A	Gp B	Gp F	Gp A	Gp B	Gp F
Social interaction	7	10	8	3	3	3
Human relations skills	16	13	10	2	2	2
Quality of managerial performance	17	16	15	1	1	1
Attitude to change	3	8	1	5	4	5
Job satisfaction	6	7	3	4	5	4

Group A	n = 24
Group B	n = 22
Group F	n = 15

The pre-training expectations about areas of personal change are quite similar between the three groups in this case ($r_{AB} = 0.90$, $r_{AF} = 1.00$, $r_{BF} = 0.90$).

Discussion

Although these results must be treated as tentative and the data does not lend itself to statistical tests, it appears that before training the three groups do not have a common set of expectations about how the courses will affect their work behaviour at an organizational level. On the other hand, when we consider areas of personal change, the three groups are in reasonable agreement concerning their expectations believing that they will firstly improve their managerial performance and human relations skills and provide a good opportunity for social interaction with their colleagues and to a lesser extent promote job satisfaction and affect their attitude to change at work.

7.4

TRAINEES' ASSESSMENT OF SESSIONS

Aim

To obtain participants' assessments of individual sessions held during the two week training course.

Procedure

Subjects: Twenty three members of Group B, the second group trained in 1977.

Method: Group B participants were asked to make an assessment of fifteen of the sessions held during the two week period. They were required to give a rating on a 6-point scale according to their impressions of the session on six criteria, namely,

- (1) Learning - did you learn a little or a lot of new information?
- (2) Entertainment - was it boring or entertaining (stimulating)?

- (3) Interest in topic - are you now only slightly or very interested in the topic?
- (4) Participation opportunities - was there little or plenty of opportunity to participate actively?
- (5) Application to job - will you be able to apply what you learned to your job?
- (6) Structure (logical development) - did you find the structure (pattern or presentation) logical and easy to follow or were you confused?

These ratings were made immediately after the close of each of the sessions and represent immediate reactions and feelings towards the content, presentation and effectiveness of the sessions as perceived by individual course members.

Only those sessions which formed part of the six major topics of the training course were included and in most cases the sessions were conducted by or in the presence of the topic supervisor. None of the lectures by guest speakers were included, nor were the syndicate sessions in which seven or eight course members got together to work on different aspects of the topics. Thus the fifteen sessions represented here are mainly lectures or other forms of presentation by the topic supervisors or else reporting sessions conducted in their presence.

Results

Participants' ratings between 0 (very low) and 5 (very high) were averaged on each of the six criteria as shown in Table Ia (Appendix I) and then the sessions were rank-ordered according to the averages obtained, from 1 (most favourable) to 5 (least favourable) in order to allow a quick comparison to be made between sessions (See Table 1b, Appendix I).

It is important to consider both average ratings *and* rank ordering in estimating the relative positions of the sessions on each of the six criteria. It was noted that some of the ratings were much more variable than others reflecting large differences in individual preferences

on some criteria but it was felt that the data and sample size did not warrant the use of standard deviations as a measure of variability, but rather, included in Table Ia (Appendix I) is the range, in each case, which is more suited to ordinal data.

Some of the 15 topics were not rated by all 23 subjects either because the particular session was not attended or else an assessment form was not completed at the end of the session. The number of subjects whose responses were averaged is indicated in each cell of Table Ia. Since 23 was the total number to take part in the course, this represents the maximum for any particular cell.

Discussion

It can be seen from the tables that topics like Personal and Group Relationships, Leadership, Motivation, Interviewing and Reporting and Marking were rated favourably on most of the criteria while ratings on Organization and Delegation and Communication were less favourable. It is not surprising that a session on Personal and Group Relationships lends itself to more participation and may be more entertaining, but it is likely to be less structured than other topics and this is reflected in the rankings in the tables. The three sessions on Reporting and Marking which received a favourable evaluation in terms of learning were also highly applicable to the job and well structured in presentation.

This is the type of information about specific sessions held during the course that course controllers find particularly useful in the detailed planning of future courses. When considering the overall effectiveness of topics it must be remembered that the guest speakers, whose sessions were not evaluated on these scales, also lectured on some of the topics and are likely to have had an impact on the final assessments.

7.5

EVALUATION OF COURSE TOPICS

Aim

To obtain subjects' ratings of topics included in the 1977 training courses, to measure changes in perceived *understanding* and *relevance* (usefulness) of topics following training and to compare ratings made by current trainees with those of two control groups.

Experimental subjects completed ratings on the same scales both before the training course began and immediately at the end of training. The two separate matched control groups used similar scales to rate their perceived understanding and relevance just prior to the training course. Thus it was possible to evaluate changes occurring in the experimental group following training using the sign test on the ordinal data obtained (Part a) and also to make comparisons between the experimental group and each of the two control groups (Parts b and c).

Results

Part a - Twenty four Group A members and 21 Group B members produced ratings both before and after training on the understanding scale, while there were 21 pairs of observations for each group on the relevance scale. A few trainees failed to complete ratings on both occasions and consequently were omitted from the sample. Changes in trainees' ratings from pre- to post-training were analysed for each topic by means of a two-tailed sign test omitting tied ratings. The sign test was chosen as the large number of ties prevented the use of a test designed specifically for ranked data.

The null hypothesis states that there would be no observed change in ratings for the group before and after training.

Decision Rule: Reject H_0 if $T < t$ or if $T > n - t$

where n = total number of pairs in the sample

T = computed statistic (number of positive changes)

t = critical value obtained from tables

and n' = total number of changes (less ties)

Results for both Groups A and B are shown in Table 7.4. These results indicate that in terms of changes in knowledge or understanding of topics from pre- to post-test, Resource Allocation and Reporting and Marking produced large and significant positive changes for both groups. Group A also recorded a positive change for the topic Organization and Delegation.

Table 7.4

Changes in perceived understanding and
relevance following training

Understanding					
Topic	T Number of positive changes		n' Number of changes (less ties)		
	Group		Group		
	A	B	A	B	
1. Organization & Delegation	16	10	18**	14	
2. Forecasting & Planning	7	10	15	17	
3. Resource Allocation	17	17	19**	17**	
4. Reporting & Marking	18 ^a	14	23*	16**	
5. Personnel Management	15	9	19	13	
	n=24	n=21			
Relevance					
1. Organization & Delegation	6	7	16	13	
2. Forecasting & Planning	3	6	15*	20	
3. Resource Allocation	6	13	14	15**	
4. Reporting & Marking	6	10	12	13	
5. Personnel Management	13	10	17*	17	
	n=21	n=21			

* Significant at $p \leq .05$

** Significant at $p \leq .01$

Others not significant at $p \leq .05$

^a Since $n > 20$ in this case, the binomial was approximated by the normal distribution for determining significance level i.e. $Z = 2.502$ which is significant at $p \leq .05$.

As far as relevance is concerned or the extent to which the topics were considered to be applicable to section leaders' day to day work, only Resource Allocation (Group B) and Personnel Management (Group A)

increased significantly while the relevance of Forecasting and Planning decreased for Group A subjects.

Part b

Ratings by the 14 members of the untrained control group (Group F) on the two scales were compared with the ratings produced by each of the experimental groups (A and B) before training. In the case of the "Understanding" scale, the number of subjects in Group F was reduced to 13 as one subject failed to complete the scale and the omission was not discovered until much later. Since none of these three groups A, B and F had received training at the time of testing, the hypothesis tested was one of no difference between the three groups before training. The ordinal data in each case was split into three blocks of similar proportions according to the overall median of the groups. That is, if the overall median was 3 then the data was divided into the number of ratings smaller than three, equal to three and greater than three. With such a small number of subjects per group this strategy was chosen in preference to the usual median test. A chi squared was then calculated on the three-by-three contingency table obtained. This was repeated for each course topic on both the understanding and the relevance scale. The first calculation only is given as an illustration. The overall median in this case was 3 over Groups A, B and F.

Understanding Scale (Pre-course)

	Group A	Group B	Group F
< 3	4	5	2
3	9	8	7
> 3	11	8	4
	n=24	n=21	n=13

$$\chi^2 = 1.55 \quad (4 \text{ degrees of freedom}).$$

This chi squared of 1.55 is not significant at $p \leq .05$. Thus, H_0 is not rejected and it cannot be claimed that the three groups are dissimilar. In fact, a χ^2 as small as 1.55 suggests that Group A, B and F are reasonably similar.

Table 7.5 shows overall medians and χ^2 values over these three groups for all topics on both scales.

Part C

Ratings were compared in a similar manner for Groups A and B, after training, with Group E results. These three groups had at this stage all received training. Again, overall medians were obtained and chi squares calculated on the 3 x 3 contingency table with scores split into blocks according to the new medians, as illustrated below. The median in this case was 4.

Understanding Scale (Post-course)

	Group A	Group B	Group E
< 4	4	10	1
4	9	5	5
> 4	11	8	3
	n=24	n=23	n=9

$$\chi^2 = 6.94 \quad (4 \text{ degrees of freedom})$$

This χ^2 is not significant at $p \leq .05$ and again, we cannot accept H_0 . Thus, the three groups are not dissimilar but the χ^2 of 6.94 suggests that they are not *highly* similar.

Out of the 10 tests performed in this part of the experiment, one of the chi squares did indicate a significant difference between Groups A, B and E. The "relevance" scale for the topic Resource Allocation was significant at $p \leq .05$. Results for part C are tabulated in Table 7.5.

Table 7.5

Medians and chi-squares of results obtained on the perceived understanding and relevance scales

Understanding				
Topic	Median Pre-	χ^2 Groups A, B & F	Median Post-	χ^2 Groups A, B & E
1. Organization & Delegation	3	1.55	4	6.94
2. Forecasting & Planning	3	2.02	3-4	4.27
3. Resource Allocation	2	2.62	4	1.59
4. Reporting & Marking	3	3.42	4	1.38
5. Personnel Management	3-4	0.36	4	1.91
Relevance				
Topic	Median Pre-	χ^2 Groups A, B & F	Median Post-	χ^2 Groups A, B & E
1. Organization & Delegation	4-5	2.54	3	3.52
2. Forecasting & Planning	3-4	4.74	3	7.45
3. Resource Allocation	3	5.03	4-5	11.52*
4. Reporting & Marking	3	4.65	4	4.01
5. Personnel Management	5	3.57	4-5	5.44

* Significant at $p \leq .05$

Apart from the chi squared value of 11.52 for Resource Allocation, there appears to be no significant differences between experimental groups following training and the previously trained control group. With the exception of the topics Organization and Delegation, Forecasting and Planning and Personnel Management on the relevance scale, the overall medians of the ratings given by subjects tended to increase from before to after training (see Table 7.5).

Discussion

Part a - The results indicate that on the understanding scale, Groups A and B trainees show some significant gains for the course topics Organization and Delegation, Resource Allocation and Reporting and

Marking while none of the other increases are significant. In fact, 47% of Group A trainees and 57% of Group B trainees show overall positive gains following training. It seems that many course members at least felt that they had acquired a better grasp of the topics studied by the end of the course than they had before. Resource Allocation was also judged to be much more relevant after training by Group B members while Forecasting and Planning was less relevant (Group A). It is worth noting that for both Resource Allocation and Reporting and Marking the topic supervisors were members of their own organizations and the syndicate group assignments were directly related to normal work activities. Consequently, the trainees were operating within a familiar framework and the new information presented could be applied directly to their own work situation. In addition to this there was much expertise available within the syndicate groups themselves and individuals were able to call upon their own and each other's prior knowledge, skill and experience as they worked together on the Resource Allocation and Reporting and Marking assignments. This is not to say that understanding will not continue to increase in *all* topic areas as the new knowledge is slowly assimilated and the new skills are incorporated into the behaviour of those who took part in the training programme. Further testing should reveal any longer term changes which may occur in any of the five topic areas.

As questionnaires asked for respondents' perceptions of their own understanding it is possible that those who reported a decrease had actually enlarged their concepts of the topics and consequently were more aware of their lack of understanding. On the relevancy scale, only Resource Allocation and Personnel Management increased (for one of the groups in each case) with trends towards increased relevancy for Organization and Delegation and Reporting and Marking and decreased relevancy for Forecasting and Planning. Additional comments indicated that respondents' opinions about the relevancy of topics were strongly influenced by the way in which topic supervisors presented their subject matter and Forecasting and Planning had earned a very unfavourable reputation in this respect.

These results were also checked against those obtained by the session assessment forms. This latter questionnaire was administered during the training course immediately at the end of each session involving a speaker such as the topic supervisor (See Section 7.4). Sessions

involving guest speakers and syndicate sessions were not evaluated in this way. Course members rated individual sessions separately on a 5-point scale according to their impressions of the session on 6 criteria, including "Learning" (Did you learn a little or a lot of new information?) and "Application to job" (Will you be able to apply what you have learned on your job?). These correspond roughly to the present two scales. On the session assessment scales, however, the effect of individual sessions rather than the whole topic was being rated and presumably the style of the speaker and the manner of presentation had a more direct influence on the trainees' ratings. When ratings for individual sessions within a particular topic were combined and topics rank ordered from 1 (most favourable) to 5 (least favourable) on the two scales, Reporting and Marking ranked first out of 5 on both "Learning" and "Application to job". On the "Application to job" scale Resource Allocation received the second highest rating. It was not possible to compare Resource Allocation on the "Learning" scale since session assessments were not carried out on the session where this topic was introduced. Thus there seems to be some consistency between results from the two forms of assessment.

Part b - It was observed that the differences between the first control group (matched sample, without training) and the two experimental groups was not significantly different when measured before training for any of the five topics on either the understanding or the relevancy scale (See Table 7.5). This means that the Group A and B trainees did not differ in their expectations about the training courses from other similar members of the organizations before training occurred. Thus any subsequent changes could be more readily attributed to the effects of training. It must be acknowledged that both trainees and controls felt that they had a moderate level of understanding even before training took place and this "ceiling" effect together with the diversity within the groups themselves makes it more difficult to detect differences when they are present.

Part c - When the second control group (previously trained subjects) was compared with experimental subjects after training, again there was no difference in most cases. The significant difference between groups on relevancy of the topic Resource Allocation has no obvious explanation. The results of Part c indicate that the attitudes of ex-trainees,

both the experimental subjects in the present study and people who had participated in previous courses were similar in terms of their understanding of course topics and perceived relevancy of topics.

General Discussion

When we consider the results of Parts a, b and c taken together we see that, not only is there a similarity between untrained experimental subjects and untraining controls *and* a similarity between trained experimentals and trained controls (Parts b and c) but there are also significant changes in experimental subjects from before to immediately after training (Part a).

An examination of the group medians in Table 7.5 reveals that there was a consistent increase on the understanding scale. These changes were significant for topics Resource Allocation, Reporting and Marking and Organization and Delegation (Table 7.4). The direction of the changes in relevancy was variable and Table 7.5 reveals signs of a decreasing trend on this scale for topics Organization and Delegation, Forecasting and Planning and Personnel Management. This result coincides with the general dissatisfaction expressed about how these topics, or at least parts of these topics, were presented.

The measures of learning and relevance used in this study were based on the participants' own perceptions concerning these variables. Several researchers have questioned the use of self ratings for judging the amount of learning achieved during training but suggest that they are useful for indicating where immediate changes are needed (Chapter 3, p.49). They have considerably less value if used as the sole means of evaluating training. Thus it is necessary to link the results of a study such as this with one aimed at discovering what specific things have been learned by the course members and what actual improvements have been made in the participants' work behaviour following the training course. Both of these types of questions were put to members of the science management course and were followed up over subsequent months.

The results of this section emphasize again that one of the great values of the course is the opportunity it provides during syndicate activities

for group members to share the knowledge and expertise which they possess. This is a readily available source of learning quite apart from the contributions made by invited speakers and topic supervisors, for it was in topic areas where the group could draw on a pool of relevant experience possessed by its members that the greatest gains in understanding were achieved.

7.6

TRAINING TECHNIQUES PREFERRED

Aim

To determine the methods of training preferred by trainees and to compare them with the preferences of other members of the organizations.

Procedure

Subjects: Twenty four members of Group A, twenty three members of Group B, thirteen members of Group F and nine members of Group E took part in this experiment.

Method

Members of Group A and B and both control groups were all asked to state their preferred methods of training before courses began and participants responded to the same question again immediately following training. The seven possibilities presented to all subjects were:

1. Lectures
2. Discussion with other course members (informal)
3. Syndicate group activities
4. Organized discussions
5. Practical activities
6. Discussions with course leaders (informal)
7. Other

Results

Respondents placed a tick beside any categories of their choice with the opportunity to offer their own suggestions for item 7. Responses

thus marked were tallied within each of the first six categories and training techniques were rank-ordered according to the frequency with which they were chosen by subjects (See Table 7.6).

Table 7.6

Number of subjects who indicated preference
for each of six types of training methods

	Group A		Group B		Group E	Group F
	Pre n=24	Post n=24	Pre n=22	Post n=23	n=9	n=13
Lectures	14	15	16	16	6	12
Discussion with other course members	10	18	14	16	5	8
Syndicate group activities	8	22	10	22	7	7
Organized discussions	14	7	16	11	4	7
Practical activities	5	9	9	11	3	11
Informal discussion with leaders	14	12	15	12	3	8

Pretest results showed that the 24 Group A members and 22 Group B members tested were very consistent in their rank ordering:

- 1st Lectures
- 2nd Organized discussions
- 3rd Discussions with course leaders (informal)
- 4th Discussion with other course members (informal)
- 5th Syndicate group activities
- 6th Practical activities

The similarity between Group A and B rank ordering, pretest, is demonstrated by listing the topics in the preferred order:

Group A (Pretest)

1. and 4. and 6.
 2.
 3.
 5.

Group B (Pretest)

1. and 4.
 6.
 2.
 3.
 5.

The 13 controls who had not yet attended a similar course produced a somewhat similar list but with organized discussion (4) and practical activities (5) transposed.

Group F (No training)

1.
 5.
 2. and 6.
 3. and 4.

These results suggest that there is considerable agreement among subjects prior to their training experience. Following the course, 24 Group A members and 23 Group B members produced a revised order of preferences. Again this was fairly consistent for both groups, as illustrated below:-

Group A (Post-test)

3.
 2.
 1.
 6.
 5.
 4.

Group B (Post-test)

3.
 1. and 2.
 6.
 4. and 5.

The revised list of preferences now became,

1st	Syndicate group activities
2nd	Discussions with other course members (informal)
3rd	Lectures
4th	Discussions with leaders (informal)
5th	Practical activities
6th	Organized discussions

Nine controls who had taken part in a previous science management course produced a list which showed the same major preferences but differed in details particularly in the lower ranks.

Group E (Previous trainees)

- 3.
- 1.
- 2.
- 4.
5. and 6.

Discussion

These results suggest that course members' attitudes changed and became highly favourable towards syndicate group activities because of their experience, whereas lectures and the more formal activities tended to lose favour. Other training techniques mentioned frequently by subjects included films and other audio-visual aids and reading lists of relevant literature. Case studies, role playing and other such practical activities were suggested as was the opportunity to hear directly about the functioning of head office and the work of other people who filled similar roles to themselves. Finally many requests were made for the chance to review the course as a whole at some stage towards the end of the second week.

7.7 CHANGES IN ATTITUDES TOWARDS THE SECTION LEADER ROLE

One of the aims of a training programme is to bring about changes and improvements in attitudes towards the trainee's role in the organization. In order to detect such changes, before-and-after measurements must be made. A questionnaire consisting of 18 opinion statements was constructed to measure attitude changes towards the role of section leader. Details concerning the construction of this scale have been given in Chapter 5 and the scale itself is reproduced in Appendix III.

The measurement of attitude change, especially verbally expressed attitudes, is a difficult task beset by uncertainty, the data being highly subjective in nature and strongly influenced by situational variables.

It is frequently obtained by using judgemental rating scales of one sort or another and, as a consequence, is subject to all the common biases for which such scales are notorious. In addition, such a measure is susceptible to social desirability effects, faking and distortion. Global rating scales have been found to be less satisfactory than attitude scales which focus on specific aspects of the job or work environment. This has been clearly demonstrated in the measurement of job satisfaction (Locke, 1976, p.1301). Therefore, the present writer decided to invest considerable effort in the construction of a specialized attitude scale, in the belief that the advantages of an attitude measure designed for the task in hand, would justify the time involved. Although a fully developed and well validated attitude scale could not be produced in the space of time available a good start could be made which would lay the groundwork for further refinement at a later date. As it turned out, the scale in its present form detected *no* significant attitude changes but demonstrated good potential for future development.

A questionnaire of this type tends to focus on the affective and cognitive aspects of the attitudes but does not tap the behavioural component at all. This latter aspect is more appropriately measured by subjects' stated behavioural intentions at the end of training and subsequent performance back on the job, both of which are documented in Chapter 8 of this report.

Aim

To measure verbally expressed attitudes towards the role of section leader before and after training in order to detect any changes which might occur in the short term.

Method

The experimental groups consisted of 22 Group A trainees and 19 Group B trainees who completed the questionnaire both before training commenced and on the last day of the science management courses. A further two control groups (Group E, n=9 and Group F, n=15) were administered the test on the first occasion only.

Results

Individual scores were calculated for all subjects and group means and standard deviations are shown in Table 7.7.

Table 7.7

Means and standard deviations of attitude scale scores obtained by experimental and control group subjects

Subjects	n	Mean	Standard Deviation
Group A - Precourse	22	30.5909	4.3715
Group A - Postcourse	22	32.6818	4.9414
Group B - Precourse	19	31.5455	5.6965
Group B - Postcourse	19	33.0000	4.3042
Group E (previous trainees)	9	32.8889	5.1058
Group F (future trainees)	15	29.2667	5.2978

The scores obtained by individuals ranged from 20 to 40 on the scale. As a preliminary analysis, t-tests were computed on the repeated measures for both experimental group and on the differences between experimentals and controls. Since all observed differences were found to be slight and non-significant (Table 7.8) further analysis was not attempted at this stage. Had there been any evidence of possible spuriously significant differences between means it would have been worthwhile to conduct an overall F test on the data but in the present case such analysis was redundant (Keppel, 1973, p.86-87). From the results obtained we must conclude that there was no differences between experimental and control groups initially on the components of the attitudes measured by the questionnaire, nor were there any statistically significant changes in the experimental group subjects following training.

Although the mean for Group A increased from 30.5909 to 32.6818 and for Group B increased from 31.5455 to 33.0000, the changes were not significant at $p < .05$. Likewise, members of previous courses (Group E) had a slightly higher mean than the untrained subjects and the untrained controls (Group F) had a lower mean score but all of these results could have occurred by chance in more than five cases out of one hundred.

Table 7.8

Results obtained when attitude scale scores
are compared for different groups

Group Comparisons	n	t value (2-tailed test)
Group A, Pre test - Post test	22	-1.873 (t-test for related samples)
Group A, Pre test and Group F (no training)	22 15	0.830
Group A, Post test and Group E (previous trainees)	22 9	-0.105
Group B, Pre test - Post test	19	-1.694 (t-test for related samples)
Group B, Pre test and Group F (no training)	19 15	1.194
Group B, Post test and Group E (previous training)	19 9	0.0601

Note: None of the t statistics are significant at the .05 level.

Discussion

From the results of the questionnaire, it appears that, on the affective/cognitive level, no measurable changes occurred immediately as a result of the training courses. However, two points need to be recognized concerning the measurement of attitude change in general and this scale in particular. Firstly, because of insufficient time to validate the questionnaire it is possible that the lack of significant results may be partly due to insensitivity of the measuring instrument itself. The narrow range of scores obtained, from 20 to 40 on a 40-point scale, suggests that the scores are seriously skewed towards the higher end. Other items need to be added to increase the spread of scores over the entire range. Clearly further refinement is necessary

to improve the accuracy of measurement. Secondly, as behavioural intentions and subsequent job related behaviours indicated, work performance contradicts the evidence of no change observed here. This result is in accordance with previous findings which generally reveal low correlations between the different components of attitude measurement. It is noted that attitude change is a highly complex process and subjects of the calibre of those who took part in the training course are likely to be particularly resistant to any attempts to bring about large and sudden attitude changes towards their work and work situations.

In addition to this, the orientation of the training course itself, consisting as it did of lectures and discussions for the most part, was designed to inform and educate rather than to persuade. It has been frequently observed that verbally expressed attitude change is more likely to follow rather than to precede the type of behavioural change observed in these subjects following the training courses.

CHAPTER 8FOLLOW-UP STUDIES

8.1

INTRODUCTION

As the discussion in earlier sections of this report has shown, the literature on evaluation confirms the present writer's opinion that the measurement of learning and attitudinal variables associated with immediate reactions of trainees must be complemented with further research into other outcome variables. In particular, we must consider work behaviours and attitudes which affect subsequent performance on-the-job. This long-term effect is, after all, the main concern of those who are conducting and financing the training programme. An answer must be given to the question, "Does the experience of training, enhance, inhibit or leave untouched the trainees' later performance at work?" It is only by means of longer term assessments of outcomes that an evaluation of the more durable effects of training can be ascertained. We have called this level of measurement intermediate outcome evaluation and the writer has endeavoured to distinguish between observed behavioural changes which can be attributed to training from those due to other environmental factors. Such measurements were taken at three, six and twelve months following the training course. Because of the serious difficulties, already discussed, of expecting busy women and men to respond to frequent questionnaires we were forced to limit our testing of group F members (untrained controls) and superior officers of the trainees to the 6 month follow-up and these results are included in Sections 8.2.6 and 8.2.7 of this report. At the three and twelve month follow-up periods trainees, only, were tested for changes in attitudes, knowledge and behaviour following training. All of these studies are described in the remainder of this chapter.

8.2

EVALUATION OF INTERMEDIATE OUTCOMES

8.2.1 EFFECTS ON WORK PERFORMANCE AFTER THREE MONTHS

Aim

To study the effects of the science management courses on work performance of section leaders after three months.

Procedure

A follow-up questionnaire was sent to 24 Group A and 23 Group B

trainees three months after their respective courses had ended. By this time one might expect the effects of the courses to have translated into some observable behaviour changes. Therefore, the questions concentrated on the current day-to-day practices of the ex-trainees within their normal work routines.

In a previous questionnaire completed immediately after the courses (See Appendix V, p. 281), 75% of Group A and 87% of Group B trainees had stated their intention of making some changes in their work behaviour based on ideas gained during training (Table 8.1). Additional comments suggested that the majority could name between 1 and 3 specific changes which they proposed to try out.

They included:

- (1) Better organization of work, particularly in the area of group projects, more emphasis on such matters as delegation, encouraging individual planning and goal-setting and group decision-making.
- (2) Improved communications within the section and in the wider sphere, for example, throughout the division, department or client community.
- (3) Better personnel management practices including interviewing, selection procedures, assessment and feedback to other section members.
- (4) More adequate human relations practices.

These same people were asked three months later if, in fact, their work performance had changed (even if only for a few days) as a result of the course (Part a). In Part b of this experiment, Group A and B members rated the amount of change in their work behaviour using the following scale:

(1)	(2)	(3)	(4)
A good deal	Some	A little	Don't know

In order to confirm this data and to obtain more information about the behaviour changes, a series of further questions were asked (See Appendix

VI, p. 286). In addition to this, 15 of the trainees were interviewed and given the opportunity to explain and expand on their brief questionnaire responses as fully as they wished. Part c consisted of a question which sought out reasons why some of these attempted changes had not been successful (Appendix VI, p.286).

Results

Part a - In response to the first question about whether or not their work behaviour had changed as a result of the course, 75% of Group A members and 86% of Group B members responded in the affirmative (Table 8.1).

Part b - Most of the 75% and 86% claimed either "Some" or "A little" degree of change. A further question asked if any of the changes made had actually survived by the end of the three month post-course period. In reply to this, 67% and 71% respectively of Group A and Group B members again responded in the affirmative (Table 8.1).

Table 8.1

Numbers of trainees who indicated changes in work behaviour

	Number who indicated change			Total
	Intended Change	Attempted Change	Change Survived	
Group A	18 (75%)	18 (75%)	16 (67%)	24
Group B	20 (87%)	^a 18/21 (86%)	^a 15/21 (71%)	23

^a Two members of Group B subsequently left their positions.

On closer examination of Group A responses, it was found that 14 out of 18 people could describe specific changes that they had attempted and of these, there were between 1 and 3 changes per person with a median of 1.5. For Group B trainees, 12 out of 18 could describe specific changes and one person had made as many as 5 or more changes the median being 2. (Table 8.2).

Table 8.2

Number of behaviour changes specified by individual course members after three months

	Number of Changes					Total Changes
	1	2	3	4	5 or more	
Group A	7	2	5	0	0	14
Group B	6	1	2	2	1	12

From the interviews it was clear that the majority of positive changes had occurred in the following areas:

- (1) The organization of group and project planning in co-operation with other section members had improved following training courses.
- (2) Improved communications were again placed second on the list of priorities, this being partially facilitated by more emphasis on regular and effective group or section meetings where matters concerning group activities were freely and openly discussed at some length.
- (3) As before, interviewing and assessment procedures and human relation skills were mentioned next.
- (4) Some positive gains had also been achieved in areas of receptiveness to change, improved personal planning of work and increased confidence in expressing one's ideas as well as being more prepared to follow them through.

Part c - The reasons why some of these attempted changes had not been successful were investigated next. In response to this item, the reason most frequently cited was the lack of support or unfavourable attitude of a person or persons in authority. This was associated with insufficient power on the part of the trainee to implement work changes of his/her own accord. Other reasons mentioned were, resistance from other staff, difficulty in "getting ideas across", lack of time and the pressure of other commitments, recent important changes in the department

itself and general economic conditions. Several people observed that any results of the changes initiated would take more time to become apparent. Because the successful implementation of changes in work behaviour depends upon the organizational climate of the work environment and because this varies so much from one division and one location to another, the variety of reasons given for failure to achieve change, must be taken into account when assessing the effectiveness of the training course. The greatest organizational differences exist between the government department and the research associations. For this reason, the comments of each category of subjects are recorded separately.

Examples given by members of research associations

- (1) The course is less relevant to research associations as it is not directly applicable to their needs which include problems associated with relationships with government departments, business and industry.
- (2) Some of the topics are not readily transferable to the work situation or to *science* management. More practical ideas are needed on running laboratories and doing cost-benefit analysis of scientific research projects.

Examples given by government department personnel

1. Reasons for lack of success in implementing change:
 - (1) The division is run on democratic lines and a consensus decision was taken not to initiate the suggested changes.
 - (2) There was insufficient support particularly from administration.
 - (3) There was resistance to change by superiors.
 - (4) Difficulty in tactfully establishing new procedures was encountered.
 - (5) The way in which the division is structured does not lend itself to changes of the type suggested.

- (6) Lack of authority to make changes was the main stumbling-block.
- (7) The trainee had insufficient authority to carry through ideas.
- (8) The organization has basic leadership problems which prevent change from taking place.
- (9) The current economic climate worked against initiating the desired changes.
- (10) Resistance from superiors prevented changes being made.

2. Criticisms of courses:

- (1) There was not enough feedback and summarization at the end.
- (2) There was nothing particularly new in the course; nothing directly applicable to our own situation.
- (3) The second half was a bit weak. We needed a more stimulating and competitive atmosphere.
- (4) Trainees needed more chance to discuss, informally, inter-divisional differences such as the different procedures used for the allocation of resources.
- (5) It was not a good idea to separate Forecasting and Planning topics. It created confusion and overload.
- (6) Better use should be made of time in the second week in summarizing and revising earlier learning.
- (7) More pre-planning is needed to prevent overlap and omissions; too much was covered in one course.
- (8) Follow-ups are needed on-the-job to improve the effectiveness of training.
- (9) Forecasting and Planning (as presented) was not relevant to the job.

- (10) Basic principles of management were neglected. Trainees would have been prepared to do some pre-course preparation in this area.
- (11) Syndicate group assignments missed the point. There was too much emphasis on content and not enough on how group decisions are made. More discussion and guidance on this aspect is necessary.
- (12) More profitable use could be made of the management forum. For example, questions should be submitted in advance to allow for better preparation.
- (13) Inviting more speakers from the private sector would improve the programme.

Discussion

These questions and discussions concerning behaviour changes were directed towards the trainees themselves and all the data was therefore based on self-report. Since this part of the evaluation was to be followed up on a broader basis after six months and would include both control group members and trainees' superiors, such a procedure was considered to be appropriate at this stage to give some indication of whether, in fact, the new knowledge and skills gained during the training courses were having any effect on subsequent performance. Indeed, only the trainees themselves could say whether or not they had attempted to incorporate any of the ideas gained, into their work behaviour. They could also cite their views on why those attempts had been unsuccessful in practice. As there are many reasons why on-the-job behaviour does not or cannot change despite the best intentions of the job occupants themselves, this intermediate step in the evaluation of course effectiveness needs to be taken. The fact that approximately 70% of all trainees in the two courses could claim some success in changing their work behaviour in line with new ideas gained during the training period suggests that the courses were effective up to a point. Moreover, because these changes could be specified and explained in such detail, it can be assumed that it was not merely a natural tendency to justify the time and effort invested in the training courses which had motivated them to claim some considerable success in introducing new ideas back on the job.

Summary of Findings

- 1) Between 75% and 86% of the science management course trainees claimed to have attempted some positive changes in their work behaviour as a result of the 1977 courses. Most of them were able to specify in detail what those changes were. Moreover, they considered that the changes were moderately successful.
- 2) Subsequently, the trainees noted some fall-off in this new behaviour so that by the end of three months 67% to 71% of trainees claimed to have maintained their new levels of performance. Again, particular items of behaviour could be described.
- 3) Four major areas of change were identified in order of importance, as follows:
 - (a) Section or group organization and joint planning of project work.
 - (b) Improved communications.
 - (c) Increased skill in personnel management.
 - (d) Increased confidence and personal satisfaction.
- 4) The chief problem encountered in introducing change and initiating new behaviour appeared to be that of overcoming varying amounts of rigidity in the structure of the organization in which they were working.
- 5) Research associates tended to find the courses less congruent with their needs than did government department section leaders. On the whole research associates were critical of the courses' applicability.
- 6) Criticisms of the course included lack of summarization and feedback to trainees, a general deterioration in interest and motivation during the second week, excessive diversity at the expense of depth and lack of relevance to the specific problems of science management.

8.2.2 AREAS OF ORGANIZATIONAL AND PERSONAL CHANGE

Aim

To investigate discrepancies between subjects' prior expectations

and post-test opinions about which areas of organizational and personal change were most affected by the training courses.

Procedure

Three months after training Group A and B trainees were asked a second time about areas of organizational and personal change most affected by the training courses. The pre-training assessment of expectations of both trainees and controls had shown that while subjects could agree on areas of personal change expected they were less certain about areas of organizational change (Section 7.3). In this section we will examine changes in experimental subjects' perceptions from before to after training as well as the extent of agreement between the two experimental groups A and B, after attending the courses.

The scale used here was similar to the one used before training to assess expected areas of organizational and personal change (see Section 7.3 and Appendix III).

Results

There was now perfect agreement ($r = 1.00$) between Groups A and B following training about areas of organizational change (Table 8.3).

Table 8.3

Subjects' rank ordering of areas of organizational change following training.

Areas of Organizational Change	Number of Responses			Rank Order		
	Group A	Group B	Group A + B	Group A	Group B	Group A + B
Awareness of policies	15	15	30	1	1	1
Internal communication	13	12	25	2	2	2
Understanding between staff	12	11	23	3	3	3
Pool of section leaders	1	8	9	4	4	4
Staff turnover	0	0	0	5	5	5

Group A n = 24, Group B n = 21, Total n = 45.

While there was only moderate similarity in trainees expectations pre-training about areas of organizational change most likely to be affected by the courses, the rank ordering produced by Groups A and B was identical after training. The overall rank order correlation of Groups A and B combined between pre- and post-test was moderate ($r = 0.53$). In other words, a considerable amount of change had taken place over that period and subjects were now much more certain about which areas of organizational change were affected. *Awareness of policies, Internal communication and Understanding between staff* took precedence over *Increase in pool of section leaders and Staff turnover*.

When we consider areas of personal change, we see that the post-training agreement between Groups A and B was only 0.55 and the pre-training expectations and post-training rank order coefficient was also moderate ($r = 0.50$). Table 8.4 shows Group A, Group B and combined responses.

Table 8.4

Subject's rank ordering of areas of personal change most affected by the training courses.

Areas of Personal Change	Number of Responses			Rank Order		
	Group A	Group B	Group A + B	Group A	Group B	Group A + B
Social interaction	12	14	26	1=	1	1
Human relations skills	12	10	22	1=	4	2
Quality of managerial performance	6	12	18	3	2=	3
Attitude to change	4	12	16	4	2=	4
Job satisfaction	3	2	5	5	5	5

Group A n = 24, Group B n = 21, Total n = 45.

Discussion

The results obtained indicate that in areas of organizational change, trainees pre-training uncertainty about which areas of organizational change would be most affected was resolved by the training courses and they now agreed that the order of priority was *Awareness of departmental policies, Internal Communication, Understanding between staff, Increase in pool of section leaders* and lastly, *Staff turnover*.

For personal change, the results suggest that although trainees could agree before-hand about their expectations, after training, their self-reported gains were much more variable. Again, neither *Job satisfaction* nor *Attitude change* was considered to be as important as *Quality of managerial performance*, *Social interaction* and *Human relations skills*. The only additional comment derived from the "other" category was to the effect that the contacts made between members of different divisions and the mutual understanding achieved during the training courses, led to a greater appreciation of their colleagues from other disciplines.

Conclusions

In retrospect, a 'no change' category should have been added to those used but since the emphasis was on changes in organizational and personal areas this was omitted at the time. Apart from this difficulty, the results indicate that while the courses enabled trainees to reach some consensus on organizational matters, when it came to personal changes, this is a much more complex matter. It is quite possible that individual trainees derive personal benefits from the course according to their own particular needs and the high degree of variability in responses to this part of the question suggests that this is, indeed, the case. The finding emphasizes the value of asking people specific questions about their behavioural changes, a policy which was followed in subsequent follow-up questionnaires. The type of global categories used here are sufficiently precise to classify organizational changes but inadequate when individuals are asked to pinpoint needs and changes at a personal level.

This phenomenon can be interpreted in terms of functional need theory which holds that the effect of any stimulus, including a training course or even a component of a training course, can only be understood within the context of the recipient's needs and personality (Wrightson, 1972, p.211). Thus, attitude change and associated behavioural changes take place following the training course, according to the individual's particular needs at the time of the course.

8.2.3 ALLOCATION OF TIME TO TOPICS

Aim

To determine whether or not the trainees are satisfied with the amount of training time allocated to the various topics.

Procedure

As soon as each training course ended and again after a time lapse of three months members were asked their views concerning the amount of time spent on each topic. The question posed was whether more time should be spent on a particular topic, less time, or whether they judged the time allocation to be about right (See Appendix V).

Results

Twenty three Group A members and twenty two Group B members responded and the percentages in Table 8.5 represent the proportion of subjects in Groups A and B who were satisfied or who indicated that they wanted more or less time devoted to different topics. The table shows responses at time t_2 (immediately after training) and t_3 (three months following training).

Table 8.5

Subjects' preferences for time allocation
to course topics

Topic	More Time		Time Correct				Less Time					
	Group A		Group B		Group A		Group B		Group A		Group B	
	t_2	t_3	t_2	t_3	t_2	t_3	t_2	t_3	t_2	t_3	t_2	t_3
Organization and Delegation	25	26	18	29	59	65	77	57	16	9	5	14
Forecasting and Planning	9	16	32	20	26	36	54	53	65	48	14	27
Resource Allocation	25	32	32	38	60	59	63	52	15	9	5	10
Reporting and Marking	15	18	0	14	71	68	95	76	14	14	5	10
Personnel Management	32	65	36	40	64	31	55	59	4	4	9	1

Note - All entries in table are expressed as the percentage of subjects

within their respective group, A or B.

Discussion

It can be seen from Table 8.5 that, in most cases, about half of the trainees were satisfied with the amount of time given to various topics. The exceptions to this were Reporting and Marking where considerably more than half were quite satisfied and the remainder were equally divided between wanting more time and less and Forecasting and Planning where less than half were satisfied. For most topics, subjects who were not satisfied indicated a distinct preference for either more or less time. Organization and Delegation, Resource Allocation and Personnel Management all required a larger time allocation while subjects felt that less time should be spent on Forecasting and Planning. This is to be expected considering the widespread dissatisfaction with this topic and the fact that trainees felt that the topic was not at all relevant to their work as section leaders.

Even after a time lapse of three months ex-trainees were reporting similar proportions in each category (see Table 8.5) and Group A and B are reasonably similar in their response patterns.

8.2.4 INDIRECT MEASURES OF EFFECTIVENESS

Aim

To study the effects of the training courses using indirect measures of behaviour at work.

Procedure

Twenty four members of Group A and twenty one members of Group B responded to three questions referring to books read on related topics, discussions engaged in and contacts maintained following the science management training courses. The questions posed were,

Question 1: Can you name any books, etc. that you have subsequently read on topics discussed during training?

Question 2: Following the science management course, have you discussed it with others in your work environment? With whom?

Question 3: Have you maintained contact with other course members? How often? By what means? (letter, telephone, face-to-face, etc.)

Results

In answer to Question 1, the following results show the percentage of course members who stated that they had undertaken subsequent reading:

	<u>Number of Subjects</u>	<u>%</u>	<u>n</u>
Group A	2	8.3	24
Group B	5	23.8	21

Responses to Question 2 indicate the percentage of course members who discussed the course with work colleagues:

	<u>Number of Subjects</u>	<u>%</u>	<u>n</u>
Group A	22	91.7	24
Group B	20	95.2	21

Question 3 shows the percentage of people who had subsequent contact with fellow course members:

	<u>Number of Subjects</u>	<u>%</u>	<u>n</u>
Group A	14	58.3	24
Group B	15	71.4	21

Discussion

As the results indicate, only a small proportion of trainees did follow up the course topics with subsequent self-directed reading in the three months following the training courses. However, the fact that at least some trainees were sufficiently motivated to pursue the topics of their own accord suggest that others may be encouraged to do so if topics leaders were to make a point of mentioning suitable literature or providing reading lists applicable to their particular areas of interest.

The books and journal articles sought out by the ex-trainees covered the following topics:

- a) A "rapid reading" course
- b) Management and executive effectiveness
- c) Human relations in industry and business
- d) Forecasting and planning
- e) Interviewing techniques

In general it was felt that the courses had facilitated comprehension of management literature. Again, most interest seems to be focused on the interpersonal areas of managerial behaviour and effectiveness with some emphasis on improving specific skills such as selection interviewing and assessment techniques.

From self-reported evidence, it seems that the majority of course members subsequently passed on information and impressions of their experience to their colleagues. This communication occurred both semi-formally, for example, in staff meetings and seminars and informally during casual conversations with fellow scientists, divisional directors, technicians, other research staff and administrators. Much of this discussion took place between members of Groups A and B and members of previous science management courses or those who hoped to participate in future courses, but also included other people who expressed interest in the programme. The figures do not include instances where a written report was submitted either spontaneously or in response to a request by the director or someone else in the division or section.

Results of Question 3 indicate that a fair proportion of people held, at least, brief interactions with their fellow course members following training. It is difficult to know to what extent these were interactions within the normal run of work (in which case they would probably have occurred regardless of the training courses). Indeed a few subjects indicated that they were more-or-less chance encounters. At least as many others suggested that they had purposely communicated with people they had

met during the courses, and that previous links with old colleagues had been strengthened. It was implied that many of the contacts were work-related rather than purely social. Most respondents indicated that these contacts were "infrequent" or "occasional" (about once per month) with people outside of their own location. Approximately 50% of all subsequent contacts had been of a face-to-face nature with the other 50% equally distributed between phone or letter communications.

Insofar as the participants' responses to the above questions can be taken as indicators of training course effectiveness, we may conclude that the training course experience did affect behaviour in the three areas mentioned. It is clear that participation in the courses leads to a considerable amount of discussion related to management practices and other topics presented during the courses as well as to overall impressions gained of the course itself. This cross-fertilization of ideas between different levels and between members of sections and divisions may be seen as a highly desirable outcome.

There is also evidence that the course afforded the opportunity to establish and strengthen links both within and between the divisions and that contacts made were maintained for some time following the courses. Perhaps more encouragement could be given to trainees to further their interests and understanding of relevant topics (both individually and within their work groups) by suggesting reading materials and by extending the opportunities to discuss the topics after they return to work.

8.2.5 PERCEIVED UNDERSTANDING AND RELEVANCE OF TOPICS

The experiment reported in Section 7.5 comparing subjects' perceived understanding and relevance of the five course topics showed that there were significant changes from before training to immediately afterwards in some areas of knowledge. Three months after finishing training when the subjects had returned to work they completed the same scales for a second time to test the durability of these effects.

Aim

To measure and compare trainees' perceived level of understanding

and relevance of topics studied in training courses before, immediately after and three months after training.

Subjects

A complete set of measures was obtained from all 24 Group A members and 20 out of 23 Group B members. The slight attrition in the sample was due to subsequent changes in course membership and staff turnover between measurement periods.

Method

On three separate occasions Group A and Group B members were asked to give ratings of their current understanding of the topics presented during training. The topics covered were:

1. Organization and Delegation
2. Forecasting and Planning
3. Resource Allocation
4. Reporting and Marking
5. Personnel Management

The individual ratings were made on a scale from 1 (vague understanding, only) to 6 (excellent understanding). The three occasions when measurements were taken were:

- (1) Immediately before training (t_1)
- (2) Immediately following training (t_2)
- (3) Three months after completion of training courses (t_3)

In addition to understanding of the topics, the trainees also rated perceived relevance of the five topics on the same three occasions.

Results

Each subject's set of ratings was ranked over the three occasions from highest (3) through medium (2) to lowest (1) and the Friedman's two

way analysis of variance by ranks (Friedman, 1937) was applied to the occasions by subjects matrix for each topic area. The Friedman Test is suitable to use with several related measurement sets of this type. The test statistic, T, is approximated by the Chi-square distribution and is a reasonably close approximation particularly as sample size increases (Conover, 1971, pp. 265-267). The results shown in Table 8.6 were obtained.

Table 8.6

Changes in trainees' perceptions of understanding
and relevance of topics studied during training

GROUP A					
Topic	<u>Understanding</u>				
	Organization and Delegation	Forecasting and Planning	Resource Allocation	Reporting and Marking	Personnel Management
T Value	11.70 ** (Increase)	0.02	17.15 ** (Increase)	11.52 ** (Increase)	10.19 ** (Increase)
<u>Relevance</u>					
T Value	0.81	7.65 * (Decrease)	1.58	3.67	0.77
GROUP B					
T Value	<u>Understanding</u>				
	4.08	1.43	20.58 ** (Increase)	12.70 ** (Increase)	2.43
T Value	<u>Relevance</u>				
	0.10	6.08 * (Decrease)	6.40 * (Increase)	3.70	1.08

* Significant at $p \leq .05$

** Significant at $p \leq .01$

Resource Allocation and Reporting and Marking show clearly significant increases for both Groups A and B on the "understanding" scale. These increases are significant at $p \leq .01$ (see Table 8.6). The results for Organization and Delegation and Personnel Management are less clear, but as Table 8.6 shows, the increase in understanding for Group A members for

both of these topics is highly significant ($p \leq .01$). However, for Group B, the increases are not significant. In the case of Forecasting and Planning no significant change in understanding is recorded by either Group A or Group B and the T statistics do not even approach significant levels.

Results on the relevance scale show that there is a significant decrease in perceived relevance for Forecasting and Planning for both Groups A and B. There is a significant increase for Group B, only, in Resource Allocation. Other than that changes are non-significant.

Discussion

When we take into account the results of the sign tests in Section 7.5, p.115, it becomes clear that for both Resource Allocation and Reporting and Marking the greatest increases in understanding occurred between the pre- and immediate post-test measurements, at times t_1 , t_2 respectively. The same is true of the topic Organization and Delegation for Group A members, with much smaller changes occurring between the immediate post-test and the three-month testing at time t_3 . Figure 8.1 illustrates graphically the changes in self-perceived understanding at times t_1 , t_2 and t_3 . It can be seen that Forecasting and Planning is virtually unaffected by the training experience for both groups. In the case of Group A, Organization and Delegation, Resource Allocation, Reporting and Marking and Personnel Management all show large and significant increases. Furthermore, Resource Allocation and Personnel Management continue a gradual upward trend between t_2 and t_3 , that is, up to three months following training. Organization and Delegation and Reporting and Marking decline slightly between t_2 and t_3 as often occurs in the three months post-training period (Bennis, 1963). Examining Group B data, it can be seen that the initial gains for Organization and Delegation and Personnel Management are not significant. Moreover, while Organization and Delegation continued to show a very gradual improvement on the understanding scale, Personnel Management declined more-or-less to pre-test level, within the first three months following the course. This effect does not appear to have any simple explanation but it is quite possible that negative post-training experiences in the work environment produced this rapid decline in newly acquired human relation skills.

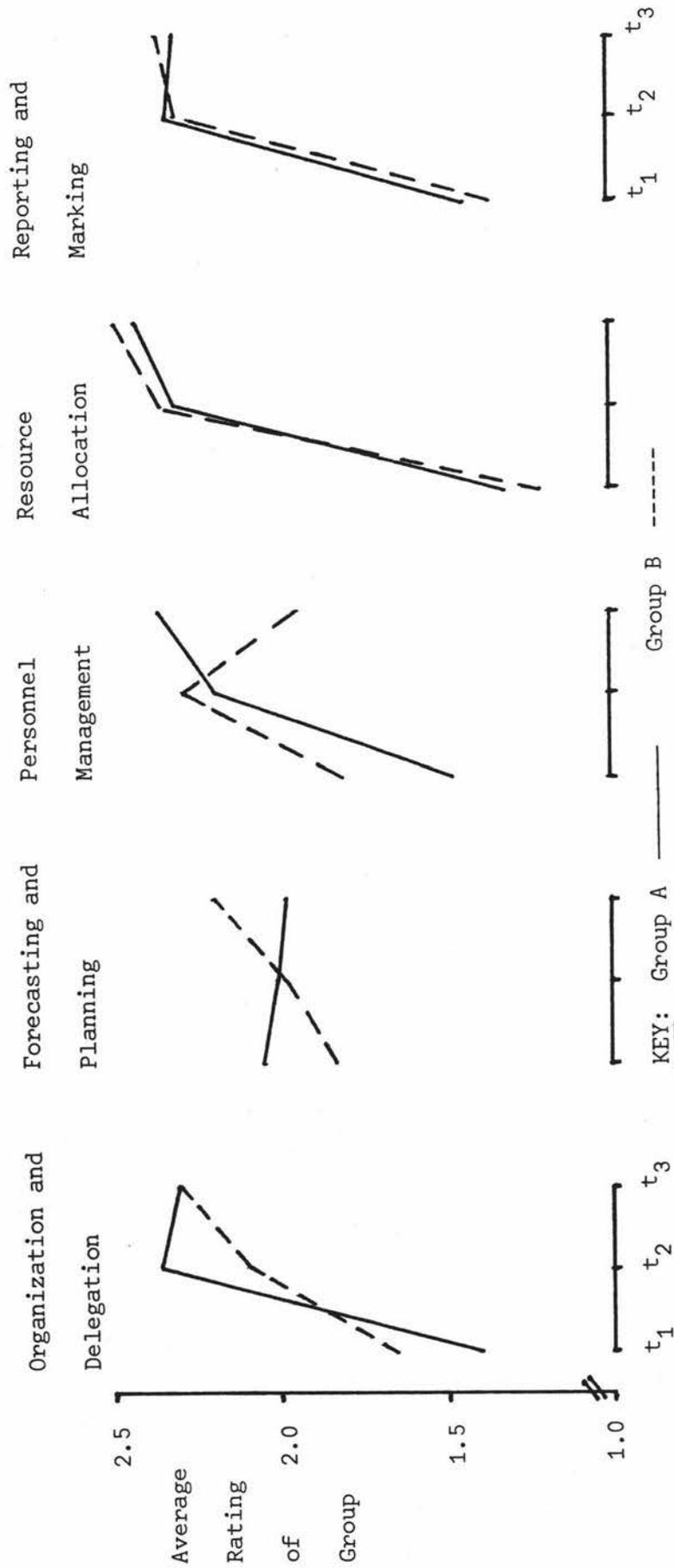


Figure 8.1 Changes in understanding of topics at three times, pre-test (t₁), immediate post-test (t₂) and three months (t₃).

Conclusions

The treatment of the two topics Resource Allocation and Reporting and Marking has clearly produced an increase in understanding as perceived by course members themselves. As noted previously (Section 7.5, p. 119) both of these topics were directly related to the trainees' normal work situations.

There is also some evidence of improved understanding in areas of Organization and Delegation and Personnel Management. The possibility that trainees may have perceived Forecasting and Planning as *less* relevant following the courses than before, goes some way towards explaining the failure of the subject matter to increase course participants' knowledge and understanding in these areas.

8.2.6 CHANGES IN BEHAVIOUR AFTER SIX MONTHS - TRAINEES' REPORT

Six months after completion of the 1977 training courses, questionnaires were used to elicit descriptions of behaviour changes adopted by trainees. Retrospective data was obtained directly from the section leaders themselves. That is to say, the leader was asked to state what he *thought* he did in his job. This may, in fact, differ from what he does when observed directly or when he is asked to keep an on-going record of his own behaviour over a continuous time period or on sample occasions. However, the latter methods are very time consuming and often yield an unwieldy amount of data. Furthermore, the presence of the observer is likely to unduly influence the people being observed.

Aim

To determine the extent and type of changes made in trainees' work behaviour six months after the training courses and to determine whether or not these changes are due to the courses, or to some other factors operating within the subjects' work environment.

Subjects

Responses to this questionnaire were received from 15 (63%) Group A members, 16 (70%) Group B members and 9 (48%) Group F members. This

represents a considerable attrition rate and naturally weakens the conclusions which can be drawn. It is quite possible that those trainees who were less satisfied with the courses were the ones who did not respond at this time and thus there was a tendency for the results to be biased in favour of more positive responses to questions concerning behavioural changes at work. However, over 50% of the control group members also failed to respond and a similar argument could be applied to that group in that those people who had made fewer changes were less likely to return the questionnaires.

Procedure

The measuring instrument used was a combination of highly structured items, for example,

'As a result of the course, did you try to make any changes in your work behaviour?'

(Yes/No)

and open-ended questions, such as,

'If "Yes", please specify what these changes were.'

(See Appendix VII, p.287) We were also interested in learning if section leaders met with any obstacles in trying to promote changes and asked the respondents to indicate which of their attempted changes had been unsuccessful. The open-ended questions were designed to permit respondents to use their own descriptive constructs that were both personally meaningful and organizationally relevant.

The verbal material generated in response to these questions and another one of similar form,

'If you did not succeed, please give the reason(s) in each case.',

required that it be organized and classified into one of several categories.

The classification scheme used was modelled on the original set of categories identified as training needs during interviews and questionnaires conducted with trainees, directors and other staff prior to the 1977 courses (Section 5.2).

These categories are listed in order of priority, below:

1. *Personal and Group Relationships*
 - (1) Internal relationships
 - i) General interpersonal skills
 - ii) Interviewing and selection
 - iii) Motivating staff (including encouraging participation)
 - iv) Communicating
 - v) Running of meeting, etc.
 - (2) Relationships with outside clients, etc.
2. *General Administration*
3. *Information about the Organizations*
4. *Forecasting and Planning*
5. *Organizational Skills* (including delegation of work and authority)
6. *Resource Management*
7. *Interaction between Divisions*
8. *Decision-making Skills*
9. *Reporting and Marking*
10. *Publishing Research*

Respondents were further required to rate the overall effect of the course on their work behaviour on a graduated scale consisting of scale points labelled 'A good deal', 'Some', 'A little' and 'Don't know'.

Results

The extent of changes in work behaviour

- a) From these questionnaires it was established that of those who responded, 67% of Group A trainees, 81% of Group B trainees and 67% of control group members claimed to have attempted some changes in their work

behaviour by the end of six months after the courses. (cf. 75% and 86% at 3 months - see Table 8.7). Thus it appears that people introduce changes in work behaviour whether or not they attend management courses.

Table 8.7

Number of subjects who indicated behavioural change following training courses

	GROUP A			GROUP B			GROUP F (Untrained Controls) 6-month
	Intention	3- month	6- month	Intention	3- month	6- month	
Ss	18 (75%)	18 (75%)	6 (67%)	20 (87%)	18 (86%)	13 (81%)	6 (67%)
n	24	24	15	23	21	16	9

b) A more important difference is the frequency with which the individual engages in innovative behaviour. Ten people from Group A, thirteen from Group B and six controls were able to specify particular changes which they had made.

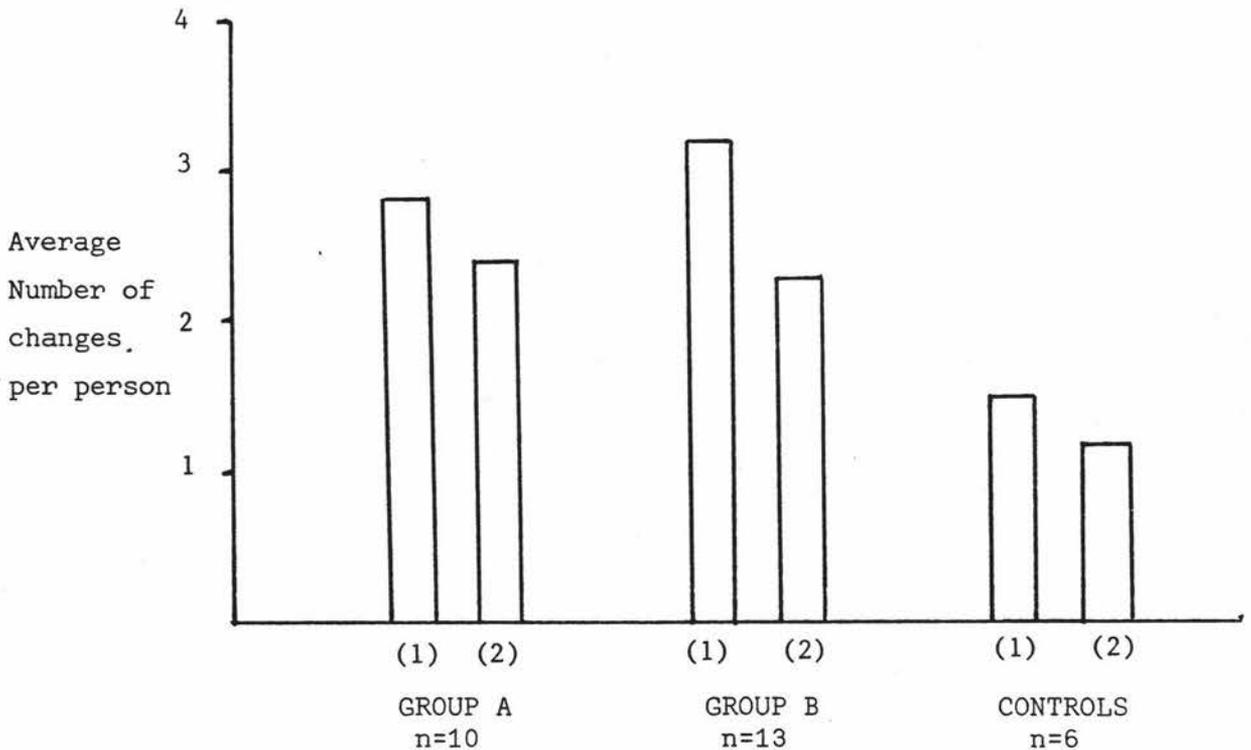


Figure 8.2 Average number of behavioural changes (1) attempted and (2) implemented by trainees and controls.

As the graphs in Figure 8.2 illustrate, of those who described changes made there is a trend for more changes to be introduced by trainees than by controls. In fact, from Table 8.8 experimental subjects appear to have initiated about two to three times as many changes per person as have controls. When these data were submitted to a median test there was a significant difference ($p \leq .05$) between Groups A and B (combined experimentals) and Group F (controls).

Median Test

Using 2 x 2 contingency table

	Groups		
	A + B	Group C	
≥ 3	15	1	16
< 3	8	5	13
	23	6	29

$n = 29$, $\chi^2 = 4.535$ (for 1 degree of freedom). This is significant at $p \leq .05$.

Table 8.8

Retrospective change behaviour reported by subjects
six months after training period

	Group A n = 10	Group B n = 13	Controls n = 6
Median attempted change score ^a	3	3	1
Median attempted change score ^a	2.9	3.2	1.5

^a Score derived by totalling discrete changes mentioned.

The above finding was obtained in response to a question which asked trainees to specify changes brought about 'as a result of the course' while controls were asked to name any changes at all which they had

attempted 'over the past year'. In line with the 3-months survey results, in all cases, subjects judged that either 'some' or 'a little' of the total work behaviour had been affected, rather than 'a good deal'. These results are not dissimilar to those obtained by Miles (1965) in studying a human relations training programme for 34 elementary school principals. He found that behaviour changes, reported retrospectively, occurred with three times as much frequency for experimental subjects as for controls.

c) At this stage it is appropriate to ask whether trainees' stated intentions at the end of the training course tend to be associated with the introduction of longer term changes in their work behaviour. Inspection of the data in Table 8.9 indicates that when subjects' immediate post-course intentions were categorised as low (0 or 1 changes), medium (2 or 3 changes) and high (4 and 5+ changes) and compared with the number of changes attempted at 6 months, those people who had every intention of making more adaptive behavioural adjustments immediately following the course also produced greater mean and median changes 6 months after returning to work. Correlation between the two was $r = 0.635$ significant at $p \leq .01$ ($n = 31$). Similarly, of the 45 trainees tested, those who stated their intentions of making work changes immediately following the courses were more likely to make changes at three months. There was a non-significant difference between the two groups using a chi squared test for related sample ($\chi^2 = 0.111$, for 1 degree of freedom). Change at three months was also related to change at six months ($\chi^2 = 2$, for 1 degree of freedom, $n = 32$). Back-home application of learning appears to be more probable for those whose attitudes are most favourably disposed to change at the end of the courses. Many parallel findings to this can be found in the literature.

Table 8.9

The relation between immediate measures of attitude change
and longer term change scores at six months

<u>Number of changes anticipated and attempted</u>		
<u>Behavioural Intentions</u>	<u>Mean changes attempted</u>	<u>Median changes attempted</u>
0 - 1 Low	1.3	1.5
2 - 3 Medium	3.1	3
4 - 5+ High	3.7	4

d) Work behaviour changes at 3 and 6 months as well as behavioural intentions were examined in relation to the personal characteristics of age, years of employment with the organization and number of years in a position of leadership. (See Tables 8.10 and 8.11). There was a slight, non-significant but fairly consistent tendency for younger people to initiate more changes and to perceive less obstacles to change back in the work setting. An implication of this is that the younger and earlier in his/her career a section leader attends a course, the more likely he/she is to register its full impact. The difference probably stems from a conservative tendency on the part of older and more experienced employees. However, the lack of statistical significance and the high degree of variability in the results suggest that neither age nor experience had a very large effect on the long-term adaptive changes made.

Table 8.10

Relationships between work behaviour and age, tenure and experience of Group A and B subjects

Relationship	Total n	Means		Standard Deviations		t Value ^a
		\bar{x}_1 (Yes)	\bar{x}_2 (No)	σ_1	σ_2	
1. Immediate post-course intentions/ Age	47	39.52 (n=39)	40.38 (n=8)	6.29	8.73	-0.3298
2. Age/Change at 3 months	45	39.22 (n=36)	41.11 (n=9)	6.50	7.62	-0.754
3. Tenure/Change at 3 months	44	11.67 (n=36)	11.10 (n=8)	13.94	4.54	0.113
4. Years leadership experience/Change at 3 months	45	3.62 (n=36)	3.47 (n=9)	3.61	2.94	0.1114
5. Age/Change at 6 months	31	39.71 (n=24)	42.29 (n=7)	6.48	7.52	-0.895
6. Age/Obstacles to change	31	41.21 (n=14)	39.53 (n=17)	5.83	7.40	0.692
7. Tenure/Obstacles encountered	30	13.64 (n=14)	11.26 (n=16)	9.10	5.77	1.261
8. Years leadership experience/Obstacles encountered	31	4.67 (n=14)	3.65 (n=17)	2.70	4.57	0.713

^a None of these t values were significant at $p \leq .05$.

Table 8.11

Correlation between frequency of change
and age, tenure and experience

Relationship between number of changes and age, etc.	n	Product-Moment Correlation Coefficient r^a
1. Number of intended changes/Age	47	+0.064
2. Number of changes at 6 months/Age	31	-0.138
3. Number of changes at 6 months/Tenure	30	0.0039
4. Number of changes at 6 months/Leadership experience	31	0.053

^a

None of these correlations are significant at $p \leq .05$.

The types of behavioural changes

a) Using the fifteen behavioural categories identified by content analysis of pre-course interview and questionnaire data together with the results obtained at 6 months, the average number of changes occurring in each category are presented graphically in Figure 8.3 for Groups A, B and control group members (only those areas of behaviour where some change occurred are shown). Once again, with the small samples available it can only be treated as trend data but it appears that the major changes occurred in organizational skills (including delegation), followed by attempts to motivate other staff and to improve communications within the work groups and divisions and lastly the improvement of interviewing techniques used in selection. None of the other behavioural categories indicate marked change. It seems that although reporting and marking and resource management produced the greatest learning in terms of information gained during the course neither of these topics affected trainees' on-the-job behaviour

as much as did those topics designed to increase sensitivity to others, skills of communication, leadership and group task and maintenance skills. This finding illustrates the importance of investigating long-term interpersonal changes which occur following management training courses.

Moreover, it highlights the fact that the results of immediate post-course reporting and evaluation are not necessarily a good indicator of the true effectiveness of such courses.

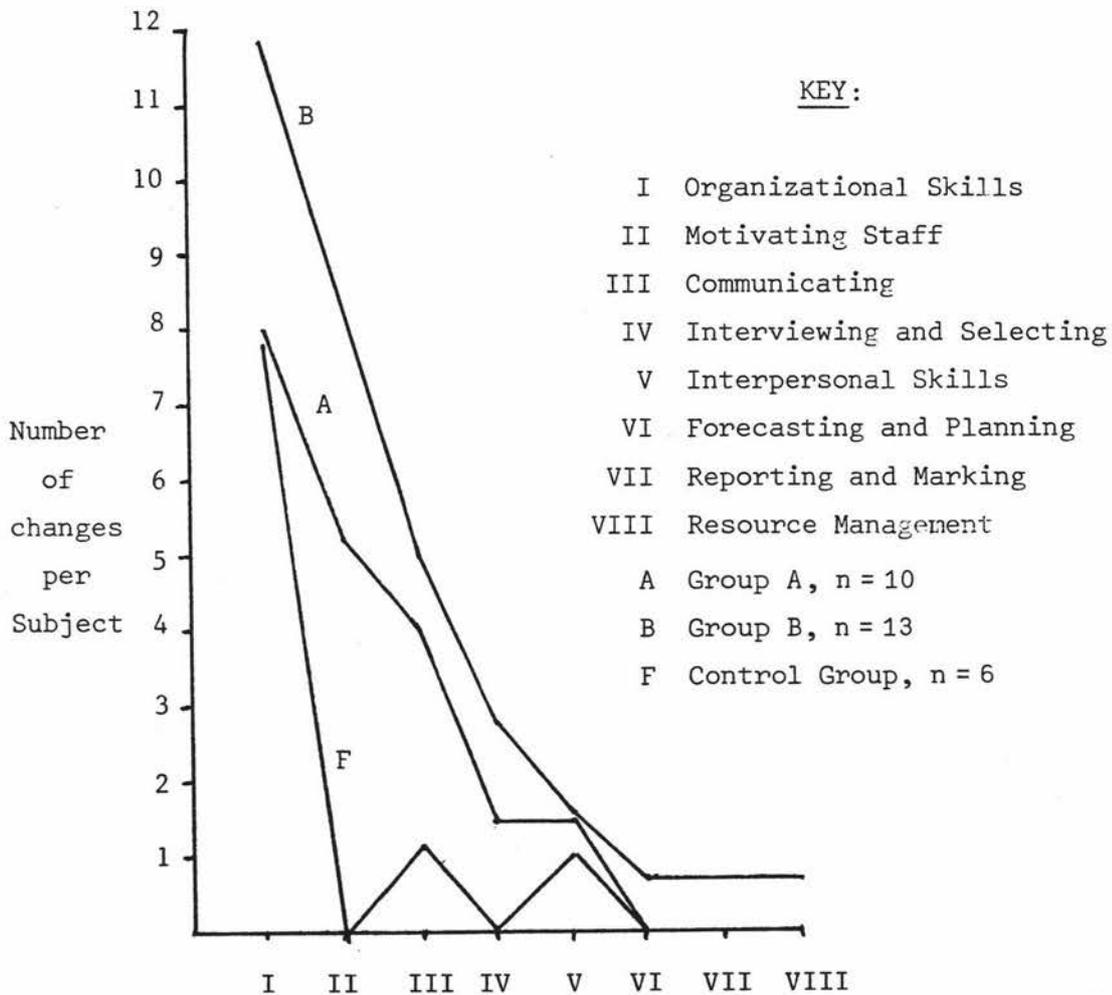


Figure 8.3 Number of behaviour changes made in each behavioural category expressed as a proportion of subjects in the group.

b) At the same time it has been pointed out that good leaders require a background knowledge of administrative procedures such as reporting and marking and resource allocation and appropriate technical/scientific know-

ledge and skills as well as managerial expertise. The ability to structure the work of others in an organization depends heavily on knowledge and experience as well as on the ability to use them (Smith & Wakeley, 1955, p.128).

c) In addition to these areas, other frequently mentioned gains included the opportunity to obtain information about the organizations and the unique opportunity offered by the courses to enjoy interaction between members of different divisions and the research associations. Neither of these latter experiences however, resulted in noticeable behaviour change at the end of six months. Response from the control group indicate less change in all areas (Figure 8.3).

d) A further interesting point to note was the actual description of the changes made by members of experimental and control groups. For example, whereas trainees commented that they had 'tried to ensure that oral and written messages among staff are clearly understood' and 'tried to improve communications', the comment of a non-trainee was to the effect that he had made 'a positive move to minimize time in discussion sessions and meetings'. Similarly, in the area of general interpersonal skills a trainee noted that he had become 'less negative with his controlling officer' while a non-trainee claimed to have 'put certain inter-personal relations on a more business-like footing'. Clearly there was a qualitative as well as a quantitative difference between members of the two groups. Trainees became more accepting of other people, more tolerant and more sensitive to the peculiarities of individual and group behaviour.

8.2.7 CHANGES IN BEHAVIOUR AFTER SIX MONTHS - SUPERIORS' REPORT

It is also possible to obtain retrospective descriptions from subordinates, peers and superiors and it is important to note that data collected from different sources do not always reveal the same picture (Fleishman, 1953a; Webber, 1970). It was for this reason that an attempt was made in this study to ask questions of both the leaders themselves and also their controlling officers in order that two different view points could be assessed.

Previous research has shown that self-recorded behavioural changes are generally more numerous and frequently qualitatively different from those observed by a work associate. (Miles, 1960). A separate assessment was made by subjects' immediate superior officers of the changes they had observed in the six months following the courses. It must be remembered that superiors are sometimes not in an ideal position to make such assessments because of geographical separation or other factors which mitigate against close contact between section leaders and their controlling officers. However, 14 controlling officers or their substitutes completed and returned a questionnaire referring to the work performance of their staff.

Aim

To ascertain the opinions of the controlling officers of both trainees and control group members regarding behaviour changes which they had observed during the six months following training courses.

Procedure

Controlling officers were presented with a set of 21 behavioural goals which were fully representative of those expressed as training needs prior to Courses 4 and 5. The items included:

- 1) Increased confidence and skill in handling and communicating with others.
- 2) Improved ability to communicate with support staff.
- 3) Improved planning and organization of work according to well defined objectives.
- 4) Increased concern with the development of technical staff.
- 5) More self confidence, generally, in coping with the work situation.
- 6) Increased appreciation of the need for good management and improved managerial skills.
- 7) Improved managerial skills.
- 8) Improved interpersonal relations within the work group.

- 9) Improved ability to organize the work of section or group so that it can be integrated with the overall aims of the department.
- 10) Greater understanding of the problems involved in the distribution and allocation of resources within the department.
- 11) Increased willingness to accept an administrative role.
- 12) A better knowledge of the goals and policies of the department.
- 13) A greater understanding of the part he/she plays (as an individual staff member) in the total organization.
- 14) Improved ability to plan the work of the section or group and to fully utilize the expertise of individual members.
- 15) Greater awareness of the relationship between individual research interests and the needs of the society.
- 16) Improved ability to motivate other members of staff.
- 17) Greater knowledge of other government departments.
- 18) Greater inclination and ability to delegate work.
- 19) Improved ability to encourage people to work together as a group.
- 20) Better able to make good decisions taking all relevant facts into account.
- 21) Improved skill in routine tasks e.g. letter-writing.

Concerning the trainees, the controlling officers were asked if, in their opinion, the courses had any effect on achieving each of the 21 objectives, to which they responded 'Yes', 'No', or 'Can't Say'. They were then asked to gauge on a 5-point scale the degree of change from very 'high' to very 'small'. An open-ended section following this, allowed the controlling officers to explain any other effects they had noticed. They were also given the opportunity to describe in detail changes in particular trainees and any changes in members of control group which they had observed in the year since the training courses. (See Appendix VIII, p.289).

Results

Table 8.12 contains the results obtained from the 14 controlling officers who responded to this questionnaire. Five of the 21 behavioural categories provided were found to be significantly ($p \leq .05$) affected.

They were:

- (1) Increased confidence and skill in handling and communicating with others.
- (2) Increased appreciation of the need for good management and improved managerial skills.
- (3) Improved managerial skills.
- (4) Greater understanding of the problems involved in the distribution and allocation of resources within the department.
- (5) A better knowledge of the goals and policies of the department.

In all cases, they felt that a 'moderate' degree of change had been implemented.

Table 8.12

Table showing number of "Yes", "No" and "Don't Know" responses made by fourteen controlling officers about their subordinates' work behaviour

Behavioural Categories	No. of "Yes" responses	No. of "No" responses	No. of "Don't Know" responses
1*	11	0	3
2	9	2	3
3	7	0	7
4	8	3	3
5	10	3	1
6*	13	1	0
7*	11	0	3
8	6	4	4
9	10	0	4
10*	13	0	1
11	9	3	2
12*	13	0	1
13	7	1	6
14	10	2	2
15	6	1	7
16	6	2	6
17	8	1	5
18	6	2	6
19	7	1	6
20	7	0	7
21	2	7	5

* Significant number of "Yes" responses ($p \leq .05$), using binomial test.

Discussion

The categories of work which were positively identified as having been affected by the training courses were general enough in nature to be readily visible to a controlling officer or were likely to be of particular interest and importance to him and therefore more likely to come to his attention. He is less likely to be in a position to judge whether his subordinate has 'improved inter-personal relations within the work group' or has 'greater inclination and ability to delegate work'. The latter two areas mentioned are similar to those emphasized by the trainees themselves

and in this respect the two points of view are indeed complementary. By surveying both groups of people a more complete picture of the changes that have occurred is obtained. Likewise, those categories to which a large proportion of controlling officers were unable to respond were either specifically concerned with sectional organization (categories 3 and 20) or referred to longer term outcomes which were probably not yet apparent, such as category 15, 'Greater awareness of the relationship between individual research interests and the needs of the society'.

The one category (21) to which more negative than positive responses were made was, 'Improved skill in routine tasks of letter-writing'. This is not surprising as the course made no attempt to provide training in routine administrative or clerical tasks. On the contrary, it placed emphasis on teaching leadership and non-routine tasks.

8.2.8 ORGANIZATIONAL CLIMATE

Although a careful study of behaviour is necessary, it is not sufficient if we are to make an accurate assessment of training effects. Litwin and Stringer (1968) point out that work behaviour is a function of the person *and* his environment. Our own study like that of Fleishman, Harris and Burt (1955) has illustrated that although human relations training produces immediate changes in self-perception this impact may soon begin to diminish once the trainee returns to his normal work setting. In fact, Mann (1962) makes the point that training which does not take the trainees' regular social environment into account will probably have little chance of modifying behaviour. Friedlander (1967) and Reid (1974) state that the newly acquired interpersonal skills can only be effective if the organizational setting supports the new behaviour, otherwise the result may be frustration and disenchantment. It is therefore clear that there is a need for consistency between the realities of an organization and the management development activities. Tagiuri and Litwin (1968) describe the global environment created by an organization (Organization Climate) as the relatively enduring quality of the internal environment of an organization that,

- (a) is experienced by its members
- (b) influences their behaviour

- (c) can be described in terms of the values of a particular set of characteristics (or attributes) of the organization.

It is interesting to note, in passing, that organizational climate refers to the perceived work environment and studies have shown that the different perceptions associated with different organizational roles are related to both the level of job satisfaction and attitudes towards one's work (Newman, 1975).

Unfortunately there are, as yet, no completely dependable, objective measures of organizational climate. One well-known measure is Likert's Profile of Organizational Characteristics (Likert, 1967). In the present study a variety of approaches were adopted to gauge the effects of climate on behaviour and to broaden the range of variables included in the investigations. The 'softer' exploratory techniques, such as interviews, open-ended questionnaire items and the practice of asking 'why' and 'why not' served to supplement the more structured types of questionnaire items. Another method employed was the Organization Climate Questionnaire completed by trainees and control group members using their own divisions or research associations as the frame of reference. Schneider (1975) has pointed out that climate measures may be aggregated at different levels of analysis. The justification for choosing the divisional level was based on the opinion of Katz and Kahn (1966) that the first step in research should always be to go to the next higher level of system organization when studying the dependence of the sub-system in question on the larger system. Furthermore, in discussions with employees it was found that the concept of the total organization was rather remote from the experience of most of the section leaders and conversations with all levels of personnel suggested that it was the *divisional* differences which had the most important effects on section or group performance. Conversely, the lower level system such as the group or section had reciprocal effects on the larger system. In other words the two were interdependent.

Aim

To study the organizational climate as perceived by the science management trainees in order to assess the effect of the prevailing climate on the impact of training.

Procedure

Twenty four Group A and 23 Group B trainees were administered the Organization Climate Questionnaire (Kolb, Rubin and McIntyre, 1974) at the end of the training courses. The information was used to complement answers to questions concerning "obstacles to change" obtained at the six month follow-up (Section 8.2.6). The Organization Climate Questionnaire used here (see Appendix V) is a 7-item questionnaire derived from Litwin and Stringer's (1966) 32-item version which provided scores on six dimensions based on the leadership style prevailing within the organization. The seven dimensions which comprise the present scale are:

- (1) Conformity: degree to which members feel they are regulated by set rules, procedures, policies and practices.
- (2) Responsibility: degree to which members feel they can make decisions and solve problems without checking with superiors.
- (3) Standards: organization's emphasis on quality performance and outstanding production.
- (4) Rewards: degree to which members feel they are recognised and rewarded for good work.
- (5) Organizational Clarity: degree to which members feel that things are well organized and goals clearly defined.
- (6) Warmth and Support: the feeling that members trust one another and offer mutual support.
- (7) Leadership: degree to which members are willing to accept and assume leadership based on expertise.

Campbell et al., (1970) have synthesized the dimensions of organizational climate proposed by various researchers like Litwin and Stringer (1966), Schneider and Bartlett (1968), Tagiuri (1966) and Kahn, Wolfe, Quinn, Snoeck and Rosenthal (1964) into just four dimensions:

- (1) The degree of autonomy people are given.
- (2) The degree of structure imposed on work positions.

- (3) The reward orientation, either in terms of individual satisfaction or company achievement.
- (4) The degree of consideration, warmth and support.

The present questionnaire also includes the task-related dimension of 'goal clarity'.

The importance of the work environment has been illustrated by many researchers including Drexler (1977) who found that a significant amount of variation in work behaviour could be accounted for by organizational effects. The main objections which have been raised to using this scalar type of measuring instrument are:

- (1) It is fundamentally an attitude test and as such is subject to all the defects of standardized attitude tests. For example, it lacks objectivity and it is evaluative rather than factual.
- (2) Questionnaire items which consist of closed-ended statements tend to be biased towards the views of the test constructor rather than the respondent.
- (3) The data obtained are ordinal at best.
- (4) 'Response set' is not fully controlled and many people tend to use a restricted range of scale values.

However, the practice of using more than one approach to the measurement of organizational climate, employed in this study served to provide a useful means of placing observed behavioural changes within a larger framework as has been recommended by numerous investigators of work performance.

Results

Responses to the Organization Climate Questionnaire were checked to see if the dominant attitude of the organization as perceived by employees could be determined. The groups' positions on each of the seven dimensions were plotted on profiles which are reproduced in Figures 8.4 and 8.5.

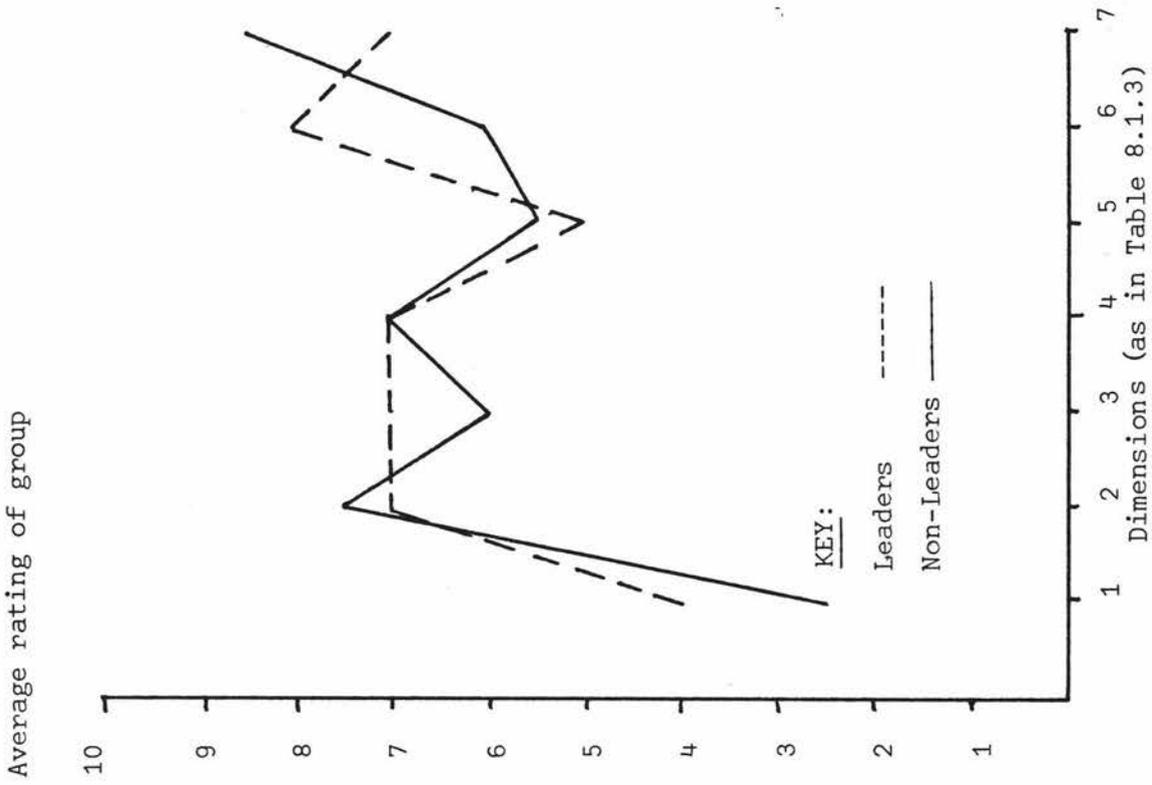


Figure 8.5 Summary profiles of organizational climate for leaders and non-leaders

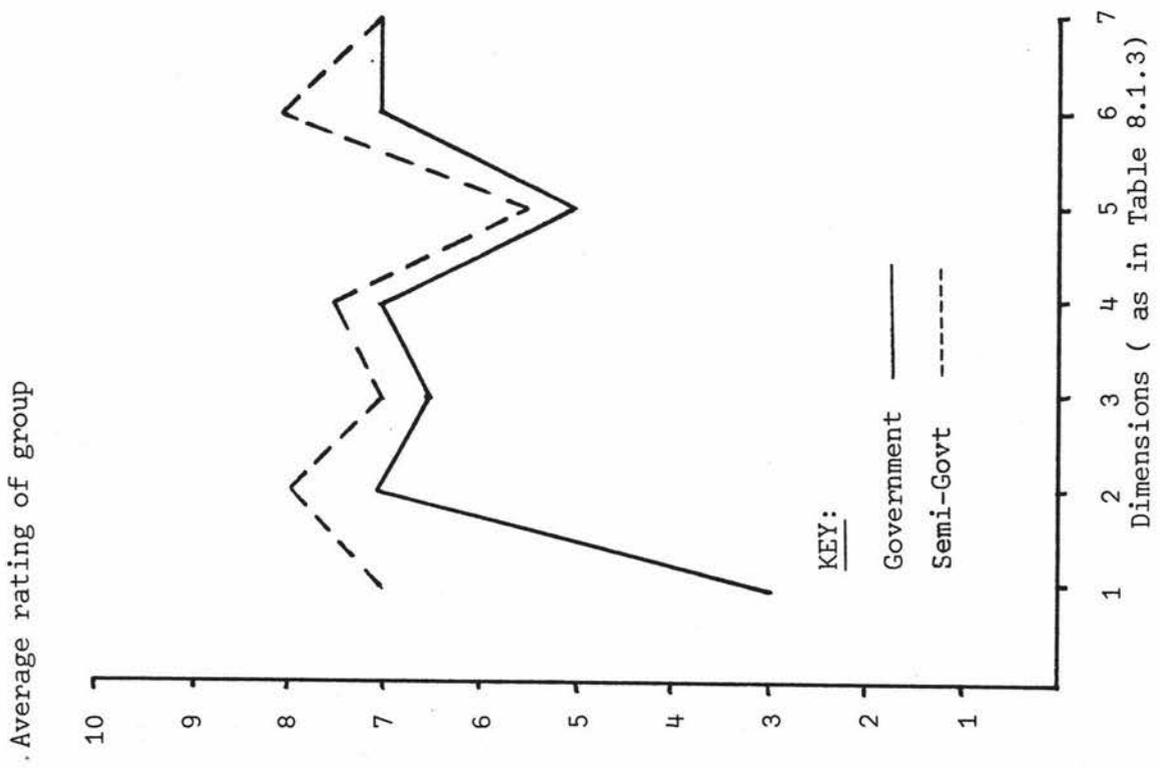


Figure 8.4 Summary profiles of organizational climate for government and semi-government organizations

All groups show somewhat similar perceptions but there are a number of differences to be noted:

- (1) Research associates tend to see their organizations as more 'conforming' than government department personnel. (Figure 8.4).
- (2) Most other characteristics (a sense of responsibility, high quality standards, reward structure, organizational clarity and warmth and support) with the exception of acceptance of leadership are seen in a slightly more positive light by research associates in relation to their own organizations (Figure 8.4).
- (3) Non-leaders are less aware of pressures to conform than are people in positions of leadership (Figure 8.5).
- (4) They are also less aware of performance standards and less conscious of a prevailing supportive climate possibly feeling less secure than their leaders (Figure 8.5).
- (5) They are more willing, however, to acknowledge and accept the leadership of experts (Figure 8.5).

From Table 8.13 which presents median ratings and ranges on all dimensions, it can be seen that the most remarkable feature is the high variability between subjects and within the groups (ranges of 5 to 8 on a 10 point scale). This variability together with the small sub-group sample sizes meant that it was not feasible to attempt to estimate the different perceptions of organizational climate between one division and another. Previous research has emphasized that variability in perceived work climate is a common feature of many organizations including government bureaucracies (Payne & Pugh, 1976). The employee's level in the hierarchy, the routine-ness of the work she performs and whether she is concerned with research and development or administration, finance and accounting all affect her attitudes and perceptions. Smith & Wakely (1955, p.192) suggest that since this is so and since modern organizations change so rapidly, perhaps 'adaptability' is, in fact, the most important trait of a leader. Therefore, the extent to which a management training course is judged to encourage flexibility and change may determine its ultimate success.

Table 8.13

Medians and ranges of ratings made by sub-groups
of subjects (training Groups A and B, combined)

	Total Gp. n=47		Government Dept n=39		Research Assoc. n=8		Leader n=36		Non-leader n=11	
	Med	Range	Med	Range	Med	Range	Med	Range	Med	Range
1. Conformity	4	8	3	8	7	6	4	7	2.5	8
2. Responsibility	7	7	7	7	8	5	7	6	7.5	4
3. Standards	7	7	6.5	6	7	7	7	7	6	5
4. Rewards	7	8	7	8	7.5	6	7	8	7	6
5. Organizational clarity	5	8	5	8	5.5	6	5	7	5.5	7
6. Warmth and support	7	8	7	7	8	7	8	7	6	7
7. Leadership	7	8	7	8	7	6	7	6	8.5	6

Obstacles to Change:

As another way of looking at training outcomes within the larger context of the work setting, each trainee and control group member was invited to describe the reasons why their attempts to bring about changes had failed. Six Group A members, eight Group B members and one control group member mentioned that they had met with obstacles when they tried to make a particular change in their work methods with the result that they had been unable to do so or only partially able to achieve the desired change.

The characteristics most commonly associated with the inability to make changes were:

- (1) Organization constraints, such as general resistance to change, lack of trust, staff restrictions or administratively enforced conformity.
- (2) Personal contingencies of time and opportunity such as pressure of work, rapid change within the section, unexpected, unplanned interruptions.

Discussion

In discussing the environmental context we have begun to move into the third level of evaluation, the ultimate outcomes where the interaction between the individual trainee and the total organization becomes more important. Psychologists began the study of organizations by stressing the importance of individuals and interpersonal relations and minimizing, by implication the salience of the larger organizational structure. However, later research has helped to modify this view (McCall, 1976). On the other hand both theory and research suggest that individual differences are equally important and the challenge of training is to match individuals with their environment so that the needs of the person, the organization and society at large are optimally satisfied. Vroom (1960) has shown that there are large individual differences in preference for the type of organizational climate, based on personality factors and other differences which arise from the attitudes and values a person acquires from his wider social environment. The individual section leader's own personal life history may be a significant intervening variable influencing reactions to the training experience. Korman (1971) says that work performance must be understood in terms of the social environment *and* the characteristics of the individuals and groups involved.

Conclusions

We see that individual differences play a large part in the perceptions of subjects in the present study. The individual variations tend to override any group differences between leaders and non-leaders or persons employed by government or semi-government organizations.

However, the results of the Organization Climate Questionnaire indicate that the trainees, as a whole, tend to perceive their organizations in a positive light. They see them as better than average in terms of the amount of responsibility delegated to them, professional standards, the reward structure, organizational clarity, supportiveness and strength of leadership. At the same time they perceive insistence on conformity to a set of rigid rules and regulations as being relatively low. This generally favourable attitude must have some impact on the willingness of section leaders to state their intentions and attempt to initiate changes in their

own and their group's work performance following training. Indeed, a high proportion (75-86%) of trainees did attempt such changes on their return.

Balanced against the result of the Organization Climate Questionnaire are the findings from the questions concerning obstacles to change encountered by ex-trainees. From these comments it is clear that when changes of the sort proposed were actually attempted, the trainees found more resistance to change than their subjective impressions of the amount of conformity had led them to suppose. Other pressures, as well, such as time constraints and financial and economic restrictions tended to increase the difficulties encountered in bringing about the desired changes.

Had we not followed up the Organization Climate Questionnaire administered on the last day of the training course with other questions aimed at eliciting more objective reasons why, in practice, ex-trainees had found themselves unable to make changes, this type of discrepancy in perceived climate could not have been detected. This illustrates the value of coupling subjective with more objective and longer-term measures of the variables of interest.

8.2.9 TWELVE MONTH FOLLOW-UP STUDY

It has been observed frequently by evaluators in the past that training programmes have little permanent impact on the attitudes and behaviour of the participants. This lack of durability of the developmental experiences means that a follow-up survey after a year or so is an important part of the evaluation procedure. Bennis (1963) uses the term 'fade-out' to describe this phenomenon and Miles (1965) found that some objective organizational factors actually mediated between short-run and longer term on-the-job changes. In our own study the possibility of this occurring was indicated by the numerous comments made by trainees concerning the difficulties which they experienced once they returned to their normal routines.

Aim

The 12-month follow-up was designed to evaluate the impact of the science management training courses, at least one year after the participants had resumed their usual roles within their organizations. It pro-

vides a longer term perspective on the effectiveness of the training courses against the background of the participants' normal work environment.

Subjects:

Of the original 47 subjects, 21 Group A members and 18 Group B members responded to the 12 month questionnaire. This represents quite a satisfactory 83% response rate and suggests that the training course had a strong "motivational" affect on those who attended over at least twelve months, since Groups A and B were still quite willing to respond to an evaluation questionnaire. This had not been true of the control group even at the six month follow-up when the attrition rate was found to be quite large (52%). For this reason and because many Group F members had by now taken part in the next year's training courses, no attempt was made to collect control group data in this phase. This is an unfortunate but unavoidable feature of a study such as this. Nevertheless, it was felt that it was still worthwhile to collect whatever information about longer term effects could be obtained on the understanding that other events may have occurred in the intervening twelve month period which could have influenced the behaviour of both the previous year's trainees and their peers who had received no training. Therefore although the following questions explicitly asked for the "effect of the training" we cannot be certain that the scales measure this effect, alone.

Procedure

Three simple 7-point scales were used to investigate the impact of training on three aspects of the trainees' roles as section leaders. They were:

1. Effect of the training on your interpersonal relationships at work.

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Did not affect work
relationships very much.

Affected work
relationships very much.

2. Effects of the training on your work performance.

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Had little effect on
my work performance

Affected my work per-
formance a lot.

3. Effect of training on the organization of your group or section.

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Not much effect on
my group or section

Affected my group or
section very much

Respondents were asked to mark their position on the scale from 1 representing minimal effect to 7 representing maximum effect.

Results

The group mean and standard deviation for each response is displayed in Table 8.14. Group A and Group B results are given separately.

Table 8.14

Group means and standard deviations of training effectiveness on three scales.

Scale	GROUP A			GROUP B		
	n	Mean	SD	n	Mean	SD
1. Interpersonal Relationships	21	3.62	1.32	18	2.89	1.23
2. Work Performance	21	3.48	1.29	18	2.89	1.53
3. Section Organization	20 ^a	3.30	1.45	18	3.28	1.90

^a One member of Group A omitted to respond to the third scale of the questionnaire.

Discussion

Mean responses were concentrated within the lower middle region of

the 7-point scale (2.89 - 3.62) suggesting that the perceived effectiveness of the courses after a time lapse of 12 months has been maintained at a moderate level. Estimates at earlier periods, namely, three months and six months, were likewise consistently within this middle range. Data collected at the end of three and six months showed that in the period between 3-6 months following training there was some slight fall-off in the percentage of ex-trainees attempting changes in their work behaviour but there has been less evidence of the major fade-out which might have been expected from the reports of previous research. This may be partly due to the fact many of the trainees on returning to their work not only discussed their experiences freely with their peers but also continued to have some personal contact with others who had taken part in similar training courses. Riecken (1952) found that those ex-trainees who had continuing contact with others from developmental experiences, were most likely to retain attitude changes.

As results from all previous follow-ups have shown the impact of the training courses may be seen as having only moderate effects on participants in areas of relationships, work performance and sectional organization. The most interesting and encouraging fact emerging from the present survey is that this effect may have more durability than might have been predicted on the basis of previous findings. However, one must recall that at the time of the twelve month evaluation we have no way of knowing for sure that the effects of work relationships and performance were not caused by other factors operating within the organizational setting despite the fact that the ex-trainees themselves attributed them to the training experience. The answer to this internal validity problem is one which only further study and extensive use of control groups can ascertain for certain.

8.3

SUMMARY OF FINDINGS

This study has indicated that participants of Courses 4 and 5 experienced moderate effects as a result of attending the science management courses. Criterion measures, for practical reasons, resolved themselves into simple judgements of improvement based as far as possible on observable, concrete behaviour back on the job. Back home organizational factors exerted an important influence on the training outcomes.

The improvements in effectiveness were observed mainly in areas which were of direct personal and organizational relevance to members and they endured for at least a period of twelve months beyond the training experience. As previous research has also shown, it is not easy to change managerial behaviour at work through training and successful training programmes have to be individually tailored to the work situation. Training programmes designed without regard for the work context are unlikely to produce favourable outcomes at the intermediate level. Thus these present courses, incorporating as they do some form of evaluation, have made a good beginning, but they must continue to be modified as additional evaluative information becomes available.

In general, the results have shown that people who have participated in these courses are:

- (1) More flexible and willing to initiate change.
- (2) More considerate and more oriented towards developing good human relations with other people in the work situation.

In particular, the following conclusions were drawn from follow-up studies of ex-trainees' subsequent performance in the work setting:

- (a) Between 67% and 81% of employees attempted to introduce changes in work behaviour following the six month post-course period. This applied whether or not they attended management courses.
- (b) However, participants in the science management courses attempted both quantitatively and qualitatively different changes to those implemented by non-trainees indicating that the frequency and type of innovative behaviour was affected by course attendance.
- (c) Those trainees who expressed their intentions of trying out new ideas at the end of the courses were more likely to implement more changes on returning to work.
- (d) Neither age nor work experience had a very large effect on the longer-term adaptive changes made but there was a slight tendency for older, more experienced employees to be more cautious in initiating change.

- (e) The consistent and discriminating changes between experimental and control groups were in areas of motivation, communication and interviewing skills. The extensive information imparted in the Reporting and Marking and Resource Allocation sessions had less obvious impact on subsequent behaviour.
- (f) There was a distinct qualitative difference between reports of changes made by trainees and non-trainees. It was clear that trainees became more accepting of other people, more tolerant and more sensitive to the peculiarities of individual and group behaviour.
- (g) Controlling officers of the course participants became aware of increased skill in handling and communicating with other people, improved managerial skills, greater understanding of the problems involved in resource allocation and a better knowledge of departmental goals and policies.
- (h) The greatest obstacles to change reported involved lack of trust and co-operation back in the work setting and lack of sufficient time and opportunity to implement the desired changes.

The evidence for these statements although somewhat tentative is derived from before and after measures of trainees as well as comparisons between trainees and non-trainees. Informal comments suggest that the inclusion of research association people as trainees, speakers and topic leaders is potentially of great value. Members of all the organizations benefit from contact with a wide range of people from diverse fields.

Since the section leader role is likely to become increasingly more varied and complex, direct help and training should be given in dealing with clients and contacts from outside of the organizations. This applies, in particular, to the research association members who tend to have more dealings with people from the larger community.

Given the importance of providing feedback to leaders to enable them to improve their performance, greater provision should be made in

this area both during the training course period and on an on-going basis within the normal work context.

In conclusion, it must be said that, notwithstanding inadequacies in the conceptual and practical applications of the evaluation methodology available as well as the inherent difficulties of evaluating management performance, the present evaluation has produced some consistent and discriminating effects on the experimental and control groups studied. Much of the data collected was of necessity based on subjective judgements by the trainees and their associates. This fact together with the limited number of subjects involved means that the conclusions reached must be treated with some caution. A management training programme designed from its inception with continuous evaluation in mind should provide more definitive answers to questions of cost-benefit and improved leadership performance. However, the present study has pointed clearly to many of the more immediate strengths and weaknesses of the existing programme and has paved the way for continuing criterion development, monitoring and evaluation of the effectiveness of the science management training courses.

CHAPTER 9

CASE STUDIES

9.1

INTRODUCTION

One of the ways in which a more intensive study can be made of the training process is to adopt a case study approach with a small number of trainees in order to gain a deeper insight into the effects of the intervention on a few individuals and to trace through the changes which occur in them through the training and post-training period. An ideal method of gathering such detailed information is by the use of the Kelly Repertory Grid which is designed to tap idiosyncratic and personalized information on individual subjects particularly in the area of their interpersonal relations. The interactions and relationships between subjects and their colleagues and associates at work is one such area of investigation and to this end the writer designed and employed specialized Repertory Grids which were administered during the twelve month follow-up evaluation.

9.2

AIM OF STUDIES

The purpose of this study was to augment and supplement the questionnaire, test and interview data already available with further information concerning a few individual subjects in order to obtain additional case study material for a more intensive study of the changes which did occur over the training and post-training period. In a study such as this it is most important to protect the anonymity of subjects and, if such data is reported, care must be taken to disguise the identity of individual subjects especially in circumstances where the information may pose a threat, either personally or professionally to the people concerned. On the other hand, Repertory Grid material is ideally suited for any type of counselling purposes aimed at individual and professional development and the writer commends its use in this way particularly if the emphasis of training is, as in the present case, concerned with managerial and leadership training. The small numbers involved in each course and the desire expressed by many trainees for more personal feedback and opportunity for self development makes methods such as the Repertory Grid a useful training aid as well as a means of evaluating the effectiveness of the programme.

The Repertory Grid as it was used here did provide a certain amount of information regarding the impact of the training courses as a whole and this aspect will be emphasized in the present report for reasons of confidentiality, while the depth of insight possible into individual trainee motivations and behaviours will be merely suggested in recognition of the fact that such information is more suited to the counselling situation, where it could be explored on a face-to-face basis with the clients concerned.

9.3

THE REPERTORY GRID TECHNIQUE

Kelly's Repertory Grid is the direct outcome of G.A. Kelly's (1955) Theory of Personal Constructs. Personal Construct theorists hold that each person comprehends his/her world (interpersonal or otherwise) by building up a system of highly individualized "personal constructs". These constructions of reality are constantly in the process of modification and change as new events and experiences impinge upon the individual either confirming or disconfirming previous perceptions. The person is thus motivated to reorganize his personal construct system in such a way that his interpretation of reality becomes more meaningful and comprehensible to him. Not only is the configuration of events which occur peculiar to the individual but also the personal interpretation is highly idiosyncratic. The system of constructs is hierarchical in nature with more important and overarching superordinate constructs having precedence over less pervasive and lower order subordinate constructs. This produces within the developing individual a highly complex cognitive pattern of parallel and interrelated minor construct systems within the grand scheme. The Repertory Grid was devised by Kelly to assess individual personal construct systems associated with any particular sphere of a person's life. It is administered by presenting him or her with a series of role titles frequently representing interpersonal relationships all of which are familiar to him. The role titles may refer to people, to situations, to objects or to a combination of these. He is then asked to identify for his own use, a corresponding list of names (persons, situations or objects) which are well-known to him and directly relevant to his own experience. In the present study we concentrated on the subject's interactions with particular people and situations within the work context (Appendix X, p.294). These become the elements of the Repertory Grid.

As the subject considers combinations of three of these elements at a time (triads) under instruction from the experimenter he begins to distinguish similarities and differences between the persons, situations or objects named on his list. These similarities and differences are expressed as bipolar descriptions and are called personal constructs. They represent the range of judgements commonly used by the subject when he expresses his ideas about this particular sphere of his life. After each bipolar construct is elicited, the subject makes a series of ratings or rankings on that construct with respect to all of the elements he has named on his list. The resulting matrix of responses is recorded in the form of a grid (Appendix X, p.294) and may be analyzed to detect the complex pattern of relationships between elements and personal constructs used by the subject. One consequence of this procedure is that the data obtained is highly personalized and is most suited to individual interpretation and counselling. However, depending on the type of role titles presented in the first place it can provide some useful information on perceived role relationships relevant to the work situation. By examining six such Repertory Grids in the present study and linking these with other data available about the subjects we can draw some tentative conclusions concerning more general aspects of the training course.

9.4

METHOD

During the six month interviews with a subsample of trainees and control group members fourteen people agreed to participate in the administration of the Repertory Grid. Careful instructions concerning the procedure to be followed were given and the instructions were repeated back to the experimenter to ensure that they had been understood but, due to pressure of time, the Repertory Grids were completed individually by the subjects in their own time and subsequently mailed back to the investigator. Consequently, it was found that two of the grids were unusable due to invalid use of constructs and a third individual had omitted a large number of cells in the grid. A fourth grid was found to be incorrectly coded and was also discarded from the sample. From the remaining ten grids six were chosen, four from experimental subjects and two from control group members in order to illustrate the use of the Repertory Grid and the type of data that can be obtained. The discussion is limited by the need to maintain confidentiality and the space available since Repertory Grids provide an almost endless supply of rich data for discussion.

In the present case, the elements elicited were derived from interpersonal roles closely associated with quite specific situations within the work environment. The ten role titles provided were:

- (1) A time when I delegated an important task to a co-worker.
- (2) The time I actively opposed the ideas of my controlling officer (or someone in authority).
- (3) A time I had to deal with a problem brought to me by a member of my staff.
- (4) A time I had to make an important decision concerning my research (or other work).
- (5) A time when I had a professional association with some outside organization (business, industry, etc.).
- (6) The occasion when I made (or proposed) changes in the running and conduct of section meetings or other procedures of a similar nature.
- (7) An occasion when I felt most satisfied with my work performance.
- (8) An occasion when I felt least satisfied with my work performance.
- (9) My professional self *now*.
- (10) My professional self *a year ago*.

Four of them (7, 8, 9 and 10) are generalizations which the subject can make about himself based on self observations of typical behaviour. They refer to perceptions of self, before training (self, a year ago), after training (self, now), least ideal self (least satisfied with work performance) and most ideal self (most satisfied with work performance). The remaining six roles were concerned with self in relation to specific work colleagues or work situations. As explained above, the Repertory Grid procedure was described individually to subjects during the interviews and then completed by them in their own time. The technique required, first of all, that the subject identify for his own private use during the test, particular persons and situations to replace

the ten role titles provided. His relationship with these persons and situations then became the elements of his grid and replaced the role titles for that subject. Next, he was asked to consider three of these elements at a time, the particular triads of role titles having been chosen in advance by the experimenter on the basis of whether they were likely to produce a wide range of useful and interesting comparisons for the purposes of the study. The ten triads of role titles chosen were:

1. (3), (8) and (7)
2. (10), (2) and (4)
3. (1), (9) and (6)
4. (5), (8) and (2)
5. (9), (10) and (7)
6. (3), (5) and (6)
7. (1), (4) and (10)
8. (7), (8) and (9)
9. (6), (2) and (4)
10. (5), (1) and (3)

Next the subject, using his own elements, decided which two elements of the triad were alike in some way and, by the same token, different from the third, in one important respect. He recorded the perceived similarity as the first pole of his construct (emergent pole) and the difference as the opposite pole (implicit pole). The next step was for the subject to rate each element on his list of ten on that particular personal construct. A seven-point rating scale was used for this task. Following this, the second triad of elements was carefully considered to produce a new personal construct by the subject. Elements were again rated one by one and the whole process repeated a total of ten times until ten separate and distinct personal constructs were elicited and used by each subject.

The responses resulting from this procedure were recorded on a grid and the results analyzed for each subject using the GENSTAT programme (Rothamsted Experimental Station, 1977) which produces a similarity matrix of the 10 x 10 variables employed. This procedure is a single

linkage hierachical cluster, principal coordinates analysis. The data from the similarity matrix can be displayed in the form of a dendrogram or tree diagram or as points plotted on a two-dimensional space, the distances between points representing the degree of similarity between constructs. The matrix was then transposed and a second dendrogram and graph displayed the similarities between the ten elements. This representation of data points, according to personal construct theory, at least to some degree, corresponds to the subject's cognitive structuring of events in relation to those particular elements and constructs.

While the original list of role titles and the composition of the triads were determined by the examiner the actual persons and events chosen for the elements, as well as the set of constructs used were selected entirely by the subjects themselves so that the final interpretation is best given in terms of individual case studies together with other information derived from tests, questionnaires and person interviews.

9.5

RESULTS AND DISCUSSION

The results reported for each case in question include the list of constructs elicited from the subject, the similarity matrix obtained from the cluster analysis, the dendograms and an example of the graphs of both constructs and elements. The matrices and graphs are located in Appendix II, p.270. The entries in the matrices correspond to the levels at which linkages occur between constructs or elements, measured on a scale of 0 to 100. An entry of 100 would represent perfect similarity while 0 on the scale represents zero similarity. These linkage levels are indicated on the dendograms in this chapter. On the graphs, these similarities are translated approximately as distances between adjacent points but the accuracy of these point distances is limited by the fact that they are projections of points onto a two-dimensional space.

The accompanying discussion is based on the Repertory Grid material detailed in the results and also the interview, questionnaire and test data collected during the evaluation together with supervisors' assessments made at the six month evaluation following training. While the writer has attempted to interpret and relate these different sources of information, she recognizes the tentative nature of such interpretation and she wishes

to emphasize that during the normal clinical or counselling process, the material would be discussed with the client, verified or disputed in a collaborative fashion and used as the basis for further exploration of the client's problems and personal developmental needs within and in relation to the needs and constraints of his organization.

While the Repertory Grid can serve both individual and organizational goals if used correctly in this consultative and cooperative manner, it can also pose an ethical dilemma to the researcher. Because the type of data made available by the Repertory Grid is of a highly personal and detailed nature it could provide a powerful tool in a conflict situation between organizational and individual interests. Indeed, the same may be said to a greater or lesser extent of any data collected during the course of evaluation and it is this aspect of applied research which may involve the experimenter in making value judgements. In any case the evaluator must come to a decision on the issue of the individual subject's privacy and she cannot make public information from Repertory Grids or other tests which might breach confidentiality any more than she can design an experiment to achieve scientific goals if, at the same time, it is likely to endanger or deprive the subjects taking part.

Since most evaluative research serves organizational or institutional aims, the use of information derived from evaluation studies tends to increase the control of the organization over the individual by increasing institutional interference with individual choice of action. On the other hand, the same detailed information may be valuable feedback to the individual trainee for his or her own personal development. Withholding information from subjects may disadvantage the very people for whom the training intervention was designed. This is an example of one of the problems that exist for the applied researcher in a situation where the main aim is educative or therapeutic. Sometimes, delaying such feedback temporarily to both subjects and training personnel may mean that an intermediate stage of evaluation can be completed together with summative evaluation before the information is used for its educative or therapeutic purposes. As Burgoyne and Cooper (1975) have said, action research of this type means that trainees are treated as active agents who are participants in the research rather than passive patients who are merely manipulated by the researcher.

In order to protect the identity of subjects and to ensure the confidentiality of the information provided by them during the course of this study, exact ages, periods of employment, leadership experience and group affiliation are not stated. Only enough detail is supplied to illustrate how the use of the Repertory Grid technique together with other test, questionnaire and interview data can complement each other and contribute towards the analysis of case study material. The first four case studies (A - D) presented are based on material obtained from experimental subjects, members of Groups A and B during the pre- and post-training evaluation period.

9.5.1 CASE STUDY A

Subject A is a male in his late thirties, employed by one of the organizations. He had nearly 15 years' experience as a scientist but less than a year in a section leader position at the time of training.

The dendrogram reproduced below displays the levels at which each of his ten elements are linked together by the clustering procedure. Elements one to ten correspond to the ten role titles listed on page 183. The distances between elements are derived from the similarity matrix in Appendix II, p. 270.

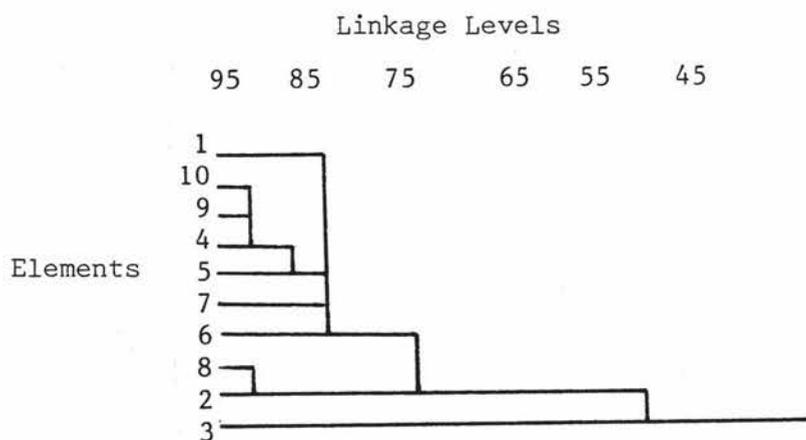


Figure 9.1 Dendrogram of elements for Case A.

Since element 9 (self, now) and element 10 (self, a year ago) are very similar = 91.7, it appears from his Repertory Grid that little change has taken place in the year since the training courses. Both, in fact, are closer to 7 (most satisfied) = 74.7 and 73.0 than to 8 (least satisfied) = 56.7 and 55.0.

7 (most satisfied) is related to element 5 (contact with outside organization) = 72.5. Both 9 and 10 are associated with 1 (delegated important task) = 75.0 and 83.3 and this in turn is related to 7 (most satisfied) = 83.0. It seems that delegating work is part of his normal and accepted work routine. Most elements, except 3 (problem by staff member) and to a lesser extent 1 (delegated important task) are associated with 5 (contact with outside organization). In other words, he is much involved with outside clients. This appears to be an important element of his work. Element 2 (opposed someone in authority) is related to both 5 (contact with outside organization) = 70.2 and to 8 (least satisfied) = 90.7. Element 4 (important decision) is close to 5 (contact with outside organization) = 85.5 to 6 (changes in work procedures) = 79.2, to 9 (self, now) = 91.7 and to 10 (self, a year ago) = 83.3.

Even a preliminary examination of the ten personal constructs elicited from this subject and listed below are informative. They indicate the subject's general attitudes and orientation and the basic dimensions which he tends to use in his judgements about others:

<u>Emergent Pole</u>	<u>Implicit Pole</u>
1. Scientific approach to problems	Emotional response to problems
2. Research oriented	Administration oriented
3. Not conservation minded	Conservation minded
4. Non laboratory worker	Laboratory worker
5. Non academic attitudes	Academic attitudes
6. Professional	Non professional
7. Flexible	Inflexible
8. Interested in science	More interested in other matters
9. Peers	Non peers
10. Intellectual	Non intellectual

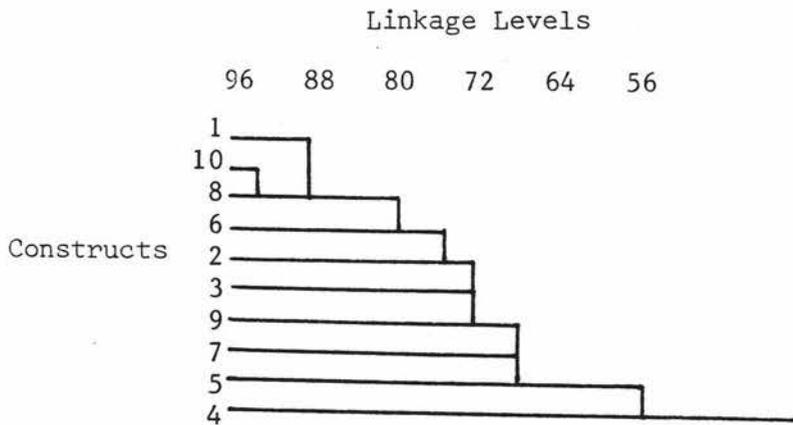


Figure 9.2 Dendrogram of constructs for Case A.

This subject's constructs include 1 (scientific approach to problems versus emotional response to problems) and this is closely associated with 6 (professional versus non-professional) = 82.7, to 8 (interested in science versus more interested in other matters) = 84.7, to 9 (peers versus non-peers) = 72.3 and to 10 (intellectual versus non-intellectual) = 91.7.

The theme of these constructs is continued in the associations between 2 (research oriented versus non-research oriented) and 7 (flexible versus inflexible) = 71.3, 8 (interested in science versus more interested in other matters) = 78.3 and 10 (intellectual versus non-intellectual) = 71.3.

There is an interesting change of theme in construct 3 (not conservation minded versus conservation minded) which is close to 7 (flexible versus inflexible) = 71.5, 8 (interested in science versus more interested in other matters) = 73.5 and to 10 (intellectual versus non-intellectual) = 70.5. It seems that the subject finds scientists somewhat inflexible in their thinking and not very responsive to conservationist attitudes. The next two constructs 4 and 5 appear to have a slightly different basis. Construct 4 (non laboratory worker versus laboratory worker) is not very close to any other construct used by this subject which suggests that his perceptions about laboratory workers are not highly elaborated and probably of lesser relevance to him. Interestingly, non-academic attitudes are related to flexibility while academic attitudes are related to inflexibility.

According to the other data available on this subject, before training his superior officer described him as a back-room researcher, not particularly outgoing and new to the job of section leader. The subject himself stated that he expected the course to be a waste of time, useful perhaps for learning a bit about how the organization works and how to get the most out of people. His choice of constructs seems to support this relatively narrow view of his work role. He rated himself as moderately knowledgeable in all topic areas included in the course and he expected to benefit most from lectures and discussions with course leaders rather than syndicate group or other activities. In fact, he changed his mind afterwards on this latter point and stated a preference for syndicate group activities as well as lectures.

Immediately after the course he rated himself higher in understanding on all topic areas covered. He commented that the course was a good opportunity to reflect on the operation of his group and that he had learnt a lot about organizational matters. He intended to improve communication with his colleagues and to build more unity into his work group.

As far as the organizational climate of the division was concerned he would prefer it to be higher on responsibility, clarity and leadership but he was satisfied with the other dimensions.

The three month follow-up showed that he still rated his understanding of course topics quite favourably, but he would have preferred to have spent more time on Forecasting and Planning. He felt that his work behaviour had changed (a little) and that the changes had endured up to the present time. The main areas of change mentioned by the subject were:

- 1) Interstaff communication
- 2) Assessment procedures
- 3) Increased interest in administration

His superior remarked that he had acquired greater confidence and acceptance of the managerial role. Changed behaviour (self-report) after six months related to:

- 1) Better communication with staff

- (2) More inclination to take into account his staff's research interests
- (3) Greater readiness to delegate responsibility

At the twelve month evaluation his self-ratings made on a 7-point scale (Section 8.2.9) over the three dimensions of behaviour affected by the course were:

- (1) Interpersonal relationships = 5
- (2) Work performance = 3
- (3) Organization of group = 3

Thus, he considers that while interpersonal relationships were affected to a considerable extent, the effect on both work performance and group organization was less pronounced.

His attitude score remained fairly stable and towards the high end of the scale. For this subject we must qualify the lack of change evidenced by the results of his Repertory Grid. On the basis of other follow-up questionnaire and interview data there are indications that he has shown more self confidence since the training course. In addition to this, he has displayed more interest in administrative matters and consideration for the people within his work group. The writer is of the opinion that it is always dangerous to accept the evidence of a single indicator in such cases, particularly when the data is largely qualitative and based on self-report or the perceptions of others. The value of multiple methods of measurement is amply illustrated throughout this research.

9.5.2 CASE STUDY B

This science management course trainee was a male scientist in his mid to late thirties with less than ten years' experience with the organization and less than five years as a section leader. A dendrogram of his Repertory Grid elements follows:

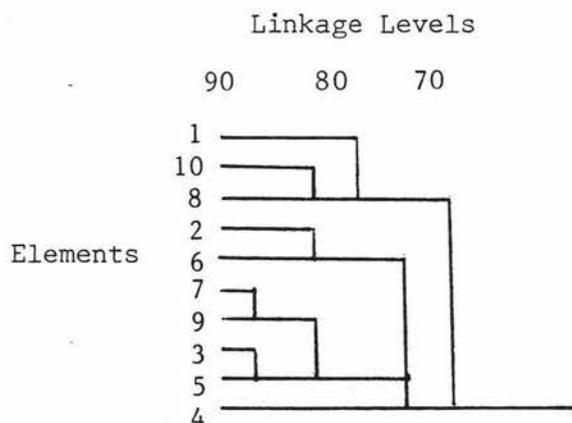


Figure 9.3 Dendrogram for elements for Case B

The elements 9 (self, now) and 10 (self, a year ago) were some distance apart, the association being only 43.3. Over the past year, the subject has moved much closer to "most satisfied" and further from "least satisfied".

A year ago, before training, his relationship with 8 (least satisfied) was 81.7 and with 7 (most satisfied) was 46.7. 9 (self, now) is now associated closely (86.7) with 7 (most satisfied) and only 41.7 with 8 (least satisfied). Both elements 9 (self, now) and 7 (most satisfied) are related to 2 (problem from staff) = 80.0, perhaps reflecting his present ability to deal with such problems.

9 (Self, now) is also related to 4 (important decision) = 73.3 and 5 (contact with outside organization) = 80.0 while 7 (most satisfied) is related to 5 (contact with outside organization) = 76.7 and to 6 (changes in work procedures) = 70.0. This cluster of related elements seems to indicate a change in his work emphasis towards an increased interest in working with external organizations. His constructs were:

<u>Emergent Pole</u>	<u>Implicit Pole</u>
1) Successful	Confusion
2) Disciplined	Undisciplined
3) Organizing/efficient	Cussed/awkward
4) Harmony	Conflict
5) Satisfied	Dissatisfied

	<u>Emergent Pole</u>	<u>Implicit Pole</u>
6)	Official interactions/ impersonal	Unofficial/personal
7)	Controlled	Uncontrolled
8)	Rational	Irrational
9)	Unselfish	Selfish
10)	Friendly	Unfriendly

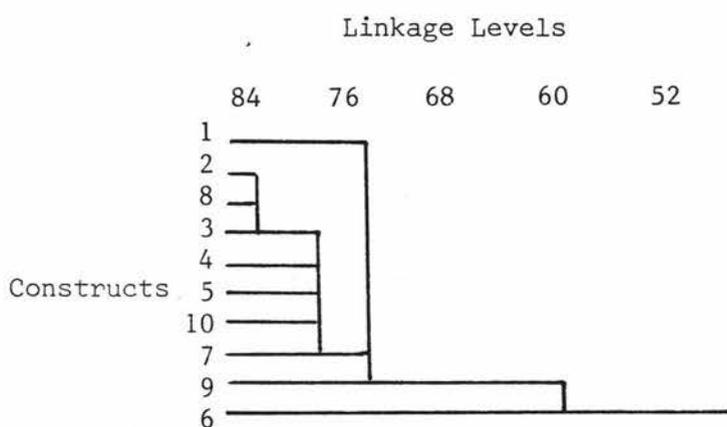


Figure 9.4 Dendrogram of constructs for Case B

Construct 1 (success versus confusion) which the subject has used as one scale of judgement is interesting, in that to most people it would seem to include two distinct ideas. However, to the present subject lack of success and confusion are so closely related that he has treated them as a single dimension. The success - confusion dimension is perceived as similar to construct 2 (disciplined versus undisciplined) with an association of 75.0 and to construct 8 (rational versus irrational) with an association of 70.0. Moreover the emergent poles, disciplined, rational and unselfish form a cluster as opposed to undisciplined, irrational and selfish. Not surprisingly, construct 3 (organized/efficient versus cussed/awkward) corresponds at the 70.0 level of association to 7 (controlled versus uncontrolled) and to 8 (rational versus irrational) at level 81.7. The subject also clusters harmony, satisfied and friendly together with controlled and rational as opposed to conflict, unsatisfied, unfriendly, uncontrolled and irrational. This provides additional insight into the subject's value system. The remaining

construct, 6 (official/impersonal versus unofficial/personal) stands, more or less alone, not bearing much similarity to other constructs suggesting that this dimension is less central to his thinking.

Prior to training this subject emphasized that he disliked administrative chores and avoided them as much as possible. He felt that the organizational climate of his division was a particularly difficult one for an administrator. In fact, on the Organization Climate Questionnaire he indicated that it was lacking on all dimensions with the exception of the amount of responsibility assigned to members. He clearly felt the need for more long-term planning and direction for project work. Moreover he felt a personal responsibility to provide some of that guidance which he believed could best be implemented by more team-work and discussion within the work group itself.

He rated himself in the middle region of the understanding scale on most topics prior to training and slightly higher on all topics except Forecasting and Planning after training. This ceiling effect which has been noted in other parts of this evaluation study, such as the measurement of attitudes (Section 7.7) is possibly the outcome of previous selection and training processes within the organization which, while imperfect in themselves, have resulted in the promotion of more competent employees to leadership positions and for selection into the management training scheme. Thus, in the development of scales to measure change in performance there is little room for movement at the top end of the scale and ratings tend to concentrate at the upper limits. He also perceived the topics as more relevant after training. Before training he expected that the courses would enhance trainees' prospects for promotion and he tended to favour discussion-type activities as training techniques. Following training he felt that more time should have been spent on Organization and Delegation and Communication and less on Forecasting and Planning which he noted was too long-winded and irrelevant. He commented, in fact, that the whole of the second week was heavy and monotonous.

His intentions included keeping his subordinates better informed, giving them more direction, passing on more information from the top down and providing more feedback on their performance. Altogether, his attitude towards his role as section leader increased on the scale from

30 to 37 from before to after training. As shown in Chapter 7 the means for the attitude scale were in the region of 31 out of a possible 40 points with a standard deviation of approximately five.

By the end of three months his understanding of the topics was still a little higher than before, he had changed his work behaviour to some extent but he did not feel that the changes had endured. On the other hand, his controlling officer after six months had noticed an effective change and indicated that the subject had developed well as a leader and manager. The subject did not return the twelve month evaluation questionnaire. One again it appears that the gains derived from these training courses are highly idiosyncratic and there is a tendency for trainees to glean from them the particular knowledge and skills which they believe they need to enhance their performance, according to their own set of work values.

9.5.3 CASE STUDY C

This subject was in his late thirties, a member of a research organization with nearly 15 years of leadership experience.

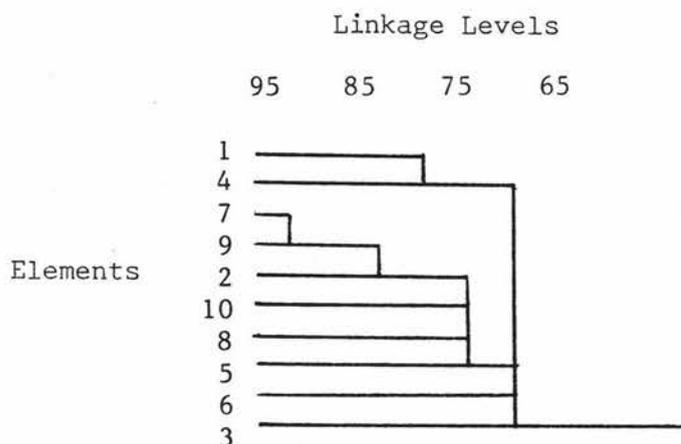


Figure 9.5 Dendrogram at elements for Case C

His Repertory Grid results indicated a fairly high level of similarity between 9 (self, now) and 10 (self, a year ago) = 72.8. Although relatively minor movement has occurred during the year, element 9 (self, now) is closer to 7 (most satisfied) = 90.5 than to 8 (least satisfied) = 47.7. Comparing this with his position one year previously he has moved closer to 7 (most satisfied) and further

from 8 (least satisfied). In fact, a year ago before training, he was almost equally distanced from most and least ideal self. Element 9 (self now) is also closer to 2 (opposed someone in authority) = 82.7 suggesting that he has become more assertive since training. Subject C produced the following set of bipolar constructs:

<u>Emergent Pole</u>	<u>Implicit Pole</u>
1) Ability to communicate	Lack of ability to communicate
2) Insecure	Confident
3) Clear and Certain	Vague
4) Assertive	Withdrawn
5) Calm, at ease	Worrying, anxious
6) Submissive	Demanding
7) Quiet, private	Talkative
8) Easy going	Agressive
9) Quick thinking, decisive	Phlegmatic
10) Solemn	Sense of humour

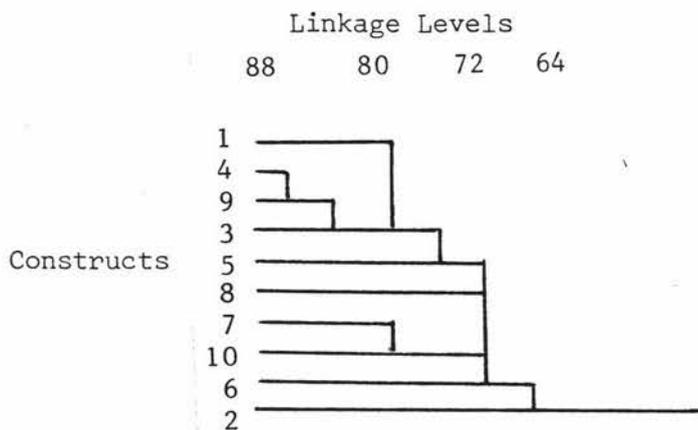


Figure 9.6 Dendrogram of constructs for Case C

An examination of his construct system reveals that he associates ability to communicate with being clear and certain, assertive, quick thinking and decisive as opposed to inability to communicate, vague, withdrawn and phlegmatic. During initial interviews the subject's immediate superior described him as a person who was very concerned about leadership

and human relations but lacking in confidence and somewhat awkward. The subject tended to see himself at that time as fulfilling a counselling type role with subordinates but he expressed a desire to improve his ability to communicate particularly with members of outside organizations and to become better organized. It seems that training has fulfilled some of these goals and he subsequently commented that the course had exceeded his expectations providing a "stimulating and profitable experience". At the end of training he expected to attempt five specific changes in his work behaviour related to his goals and at the end of three months he claimed that he had changed (a good deal) and that the attempted changes had survived. As he had been promoted to a senior position during that period, it is likely that the proposed changes could be more readily implemented. His self-rated understanding of course topics had all increased from before training.

By the six month follow-up he was able to confirm that he had made several changes as a result of training and he listed the following areas, specifically:

- (1) more delegation to subordinates
- (2) improved relationships and communication with staff
- (3) a more flexible approach to administration
- (4) greater perceptiveness in group situations, for example, improved communication and leadership skills.

He claimed only partial success in the changes that he had attempted. This he attributed to outside pressures caused by the new administration forcing a degree of conformity to old work patterns.

He stated that he had discussed the training course with numerous colleagues and had lectured to senior staff, introducing them to some of the ideas presented during the training course. He also acknowledged the possibility of a degree of interaction between the effect of training on his work behaviour and the fact that he had been promoted within the organization. By the twelve month evaluation, this subject rated the effectiveness of the courses quite high on the 7-point scales provided:

- (1) Interpersonal relationships = 5
- (2) Work performance = 5
- (3) Organization of group = 7

All in all, the results of the study suggest some positive gains from his experience with the training programme particularly in areas of self-confidence, ability to handle and communicate with subordinate staff and a greater willingness to assert himself in his new role of authority. His personal satisfaction with the course was apparent and his job satisfaction had increased considerably over the twelve month period. It is likely that much of this was due to his promotion and increased responsibility. This illustrates very well the complexity of the task of evaluating the effectiveness of training and the need to take situational and organization factors into account. Nevertheless, it is conceivable (indeed, it was the subject's opinion) that the impact of training immediately prior to these events in his career, enhanced his capacity to fulfil his new duties more adequately.

9.4.5 CASE STUDY D

This is an interesting case where the trainee, in his mid-forties with almost ten years' experience in the organization and nearly a year in a leadership role, experienced considerable changes including a severe decrease in job involvement during the year following the training course.

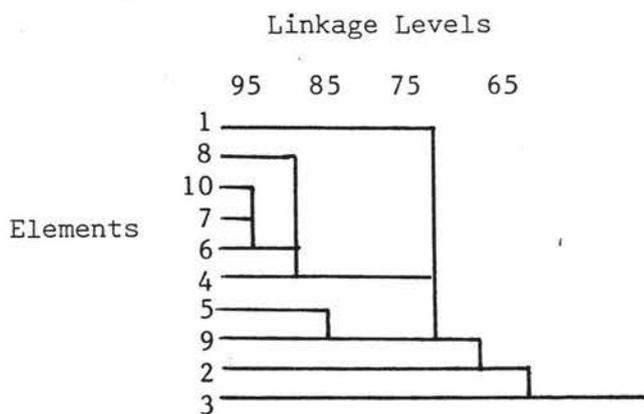


Figure 9.7 Dendrogram of elements for Case D

The distance between elements 9 (self, now) and 10 (self, a year ago) = 49.3 represents the largest change of all the subjects considered. Element 9 is not particularly close to either element 7 (most satisfied) = 56.0 or to element 8 (least satisfied) = 51.3. This suggests a degree of alienation from the work situation which is supported by other data. If we make similar comparisons with element 10 (self, a year ago) where there was a 93.3 association with element 7 (most satisfied) and an 87.3

association with 8 (least satisfied) we see that there was considerable ambivalence but much less of the detachment which is now present. Element 6 (changes in work procedures) is also close to 10 (self, a year ago) = 89.7 as well as to both 8 (least satisfied) = 84.3 and to 7 (most satisfied) = 93.0 illustrating his ambivalence about making changes at work. Element 9 (self, now) is quite close to 5 (contact with outside organization) = 80.7.

The subject's choice of constructs seem to reflect the confusion he is experiencing.

	<u>Emergent Pole</u>	<u>Implicit Pole</u>
1)	Conscientious	Irresponsible
2)	Intelligent	Stupid
3)	Detached	Lusting for power
4)	Amiable	Stringent
5)	Powerless	Powerful
6)	Straight forward	Devious
7)	Effective	Ineffective
8)	Relaxed	Tense
9)	Rational	Irrational
10)	Manly	Creepy

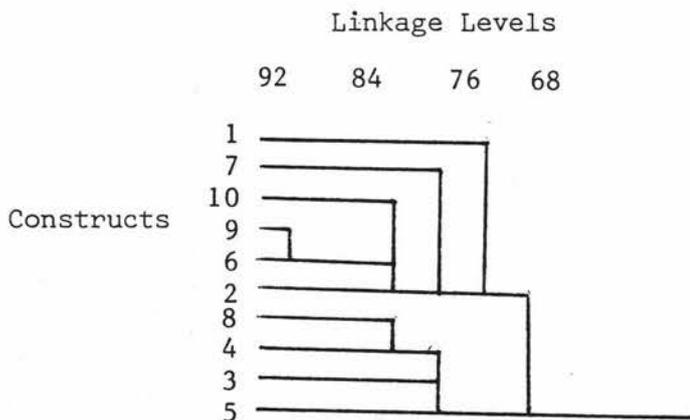


Figure 9.8 Dendrogram of constructs for Case D

The construct system displays a number of interesting items and interrelationships which would normally be explored with him if the technique were to be used for counselling purposes. Construct 7 (effective versus ineffective) is reasonably close to 1 (conscientious versus irresponsible) = 72.3 while intelligent, straight forward and rational form one cluster as opposed to stupid, devious and irrational. Similarly, detached and relaxed are related to each other and opposite to lusting for power and tense. Rational - irrational is also related to 10 (manly versus creepy) at the 80.8 level.

One notes the strength of the relationships between most of this subject's constructs, the vivid and sometimes extravagant descriptors used and the fact that he seems to evaluate power and "getting on in the system" in a particularly negative way while he sees detachment and a relaxed approach as being more rational, manly and intelligent.

This is all remarkably congruent with the comments made by his superior during the precourse interview when he described the subject as "tends to be anti-establishment but not particularly outspoken". The subject himself rated his prior knowledge of the topics studied as very low on the scale. He expected to improve internal communication and understanding between scientists and administrative staff as a result of the course and he hoped that it would lead to the development of common policies. Moreover, he expected to improve his managerial performance and human relations skills. He perceived the climate of his organization to be undesirably deficient in the amount of responsibility delegated to its members and in organizational clarity but high on warmth, support and leadership.

It was found at the intermediate post-training evaluation that he still rated his own understanding of the topics extremely low (even lower than before in some cases). He was particularly disappointed with the Forecasting and Planning topic. He felt that the course was "conscious-raising and motivating towards personal self-actualization and a stimulation to creativity" but he was cynical about the long-term effects.

He stated that he intended to be more outgoing, involved and positive in his work, for example, by giving more seminars and being more

optimistic about the results of his work. By the end of three months post-training he had given up his leadership and administrative position and was concentrating solely on research work. He indicated that he did not consider this to be the result of the training course and that it was partly related to other changes in his personal life. At this time he rated both his own understanding and the relevance of the course topics as extremely low. He had kept contact with one fellow course member but on only one occasion. He recorded no attempted changes whatsoever in his work behaviour. His superior officer declined to comment.

At six months, the subject responded to the evaluation by saying that the questionnaire was not relevant to him and at twelve months he gave each of the scales, interpersonal relationship, work performance and organization of work group a rating of 1, stating again that his personal life circumstances had greatly influenced his attitudes to his work. He added, however, that he thought the course stimulated creative thinking but was not applicable to scientists.

Altogether, there is evidence of considerable change over the year following training and although the movement may not be in the intended nor necessarily desired direction (from the point of view of the administration) it is possible that the training course experience had at least helped the subject to crystallize his thoughts and attitudes towards his work circumstances. The most remarkable finding is the consistency between all of the various measures used to tap attitudes towards work and the additional insight provided by the use of the Repertory Grid in this case.

9.5.5 CASE STUDY E

Case E is the study of one member of the untrained control group all of whom completed Repertory Grids at the same time as the experimental group. The role titles were exactly the same for all subjects, experimentals and controls. This subject was in his early thirties with less than ten years experience in the organization and less than five in a leadership position.

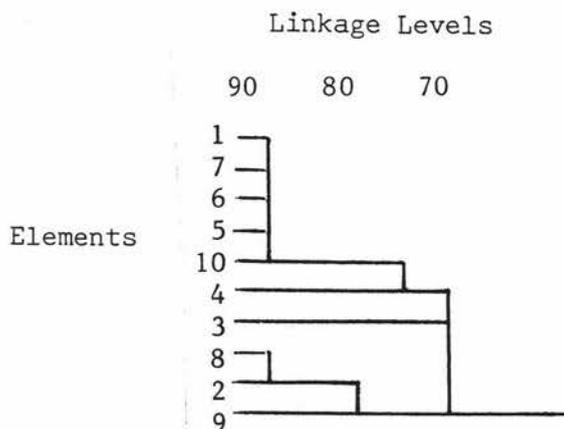


Figure 9.9 Dendrograms of elements for Case E

His elements 9 (self, now) and 10 (self, a year ago) are widely separated with a linkage level of 28.7. Here is a case where 9 (self, now) is much closer to 8 (least satisfied) = 77.0 and further from 7 (most satisfied). A year ago, the situation was reversed with 10 (self, a year ago) being much closer to 7 (most satisfied) = 79.7 and further from 8 (least satisfied) = 31.7.

Element 8 (least satisfied) is also close to 2 (opposed someone in authority) = 86.7 which may suggest some conflict in this area of his work. Moreover, it is interesting to note that 10 (self, a year ago) is fairly close to 1 (delegated an important task) = 74.7, to 4 (important decision) = 70.3, to 5 (contact with outside organization) = 81.3 and to 6 (changes in work procedures) = 88.0, whereas none of these elements are related to 1 (self, now). The element 3 (problems from staff) seems to stand in a, more or less, isolated position.

This subject's list of construct choices which follows similarly suggests a certain amount of recent conflict in his work situation.

<u>Emergent Pole</u>	<u>Implicit Pole</u>
1. Annoyed	Satisfied, Knowing that you have done a good job
2. Idealistic	Cynical
3. Beneficial	Destructive
4. Stimulating	Depressing
5. Disinterested	Interested
6. Disruptive	Supportive
7. Decisive	Indecisive
8. Relieved	Discontented
9. Pleased	Disappointed
10. Cooperation	Disagreement

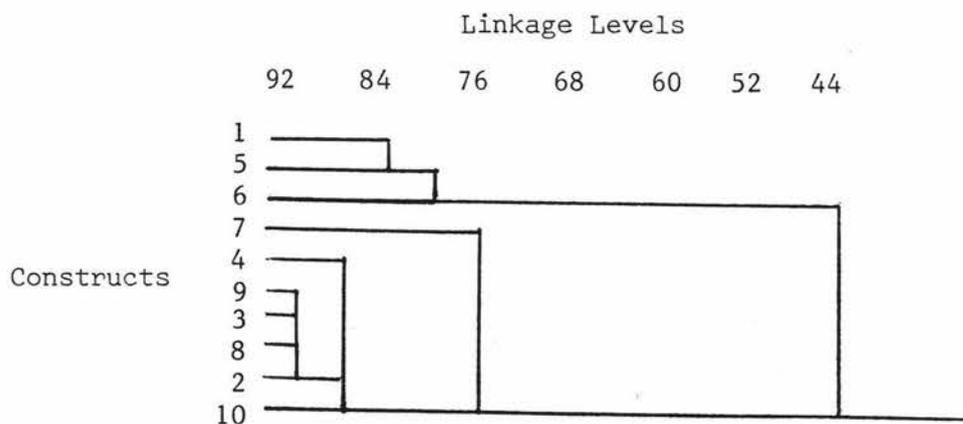


Figure 9.10 Dendrogram of constructs for Case E

His construct number 1 (annoyed versus satisfied, knowing you have done a good job) is related to the constructs 5 (disinterested versus interested) = 83.3 and to 6 (disruptive versus supportive) = 75.2 while construct number 2 (idealistic versus cynical) is related to 10 (cooperation versus disagreement) = 77.2, to 9 (pleased versus disappointed) = 88.5, to 8 (relieved versus discontented) = 82.3, to 4 (stimulating versus depressing) = 73.5 and to 3 (beneficial versus destructive) = 82.0.

At that time the subject's self report indicated that he would like people to be able to approach him but that he was uncertain about others' expectations of him. The superior's report after 6 months indicated no significant change. From the subject's self-report at 6 months we learn that he had tried to make changes, for example, he has delegated more (this was necessary because he had been away for three months during the year). He estimated (a little) change in his work behaviour. Taken altogether, the evidence from this case suggests substantial deterioration in the subject's work performance over the past year. There have been a number of disruptions which appear to have been detrimental to his job satisfaction and his work relationships have not improved despite his hopes to the contrary. It is possible that if he were to take part in the management development programme he would have the opportunity to crystallize some of the performance goals which he appears to be seeking.

9.5.6 CASE STUDY F

This male section leader in his late thirties had less than ten years experience with the organization most of which was in a position of leadership. He was a member of the control group who had not taken part in the training programme.

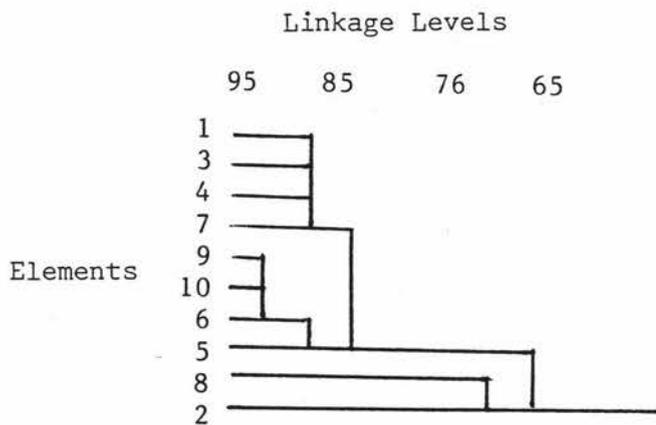


Figure 9.11 Dendrogram of elements for Case F

Repertory Grid results show, element 9 (self, now) and 10 (self, a year ago) are very similar = 90.8. 9 (self, now) is close to 7 (most satisfied) = 84.0 and further from 8 (least satisfied) = 64.7. 10 (self,

a year ago) was also fairly close to 7 (most satisfied) = 79.8 and a little further from 8 (least satisfied) = 60.5. In other words, the amount of change that has taken place is quite small. 9 (self, now) and 10 (self, a year ago) are both near 5 (contact with outside organization) = 79.8 and 84.0 as well as to 6 (change in work procedures) = 90.7 and 88.2 respectively. Element 1 (delegated an important task) is related to 3 (problem from staff member) = 88.8 and to 4 (important decision) = 80.5. Element 2 (opposed someone in authority) is not particularly close to any other element which suggests that this is not an important feature of his work life. 4 (important decision) is fairly close to 6 (changes in work procedures) = 76.8 as well as to 9 (self, now) = 78.2.

He produced the following list of personal constructs:

	<u>Emergent Pole</u>	<u>Implicit Pole</u>
1.	Satisfaction with job	Dissatisfaction with job
2.	Self confidence	Diffidence
3.	Efficiency	Inefficiency
4.	Discontentment	Contentment
5.	Fluctuation in job satisfaction	Stability in job satisfaction
6.	Effectiveness with people	Ineffectiveness with people
7.	Effectiveness for getting job done	Ineffectiveness for getting job done
8.	Current satisfaction	Current dissatisfaction
9.	Assertiveness	Reticence
10.	Ability to communicate	Lack of ability to communicate

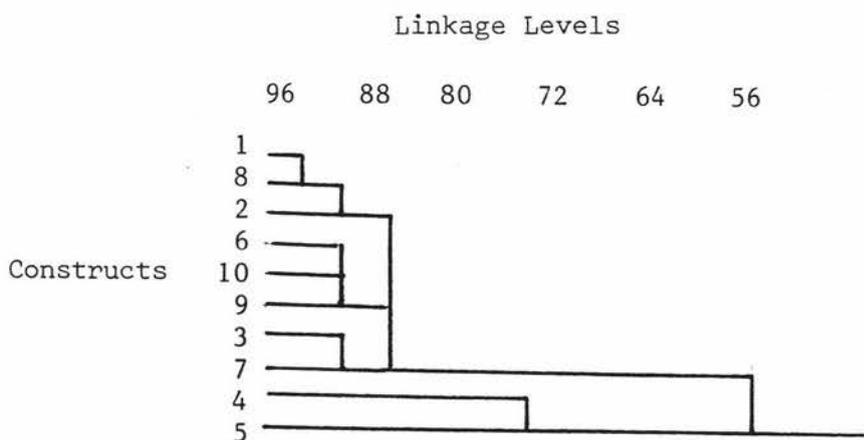


Figure 9.12 Dendrogram of constructs for Case F

The most remarkable feature of this subject's construct system is his relatively "tight construing". That is, there is a very close similarity between all of his constructs, most of which are linked together by level 84.7. This suggests a degree of cognitive simplicity. He appears to be working on a relatively few separate and distinct personal constructs in relation to his work environment. Construct 1 (satisfaction with job versus dissatisfaction with job) is closely related to every other construct with the exception of constructs 4 (discontentment versus contentment) and 5 (fluctuation in job satisfaction versus stability in job satisfaction) which continue the same theme of job satisfaction but with the positive and negative poles reversed. Thus we conclude that this subject is operating very much on the two closely related criteria of job satisfaction and effectiveness.

The initial report from his controlling officers described him as a typical scientist with very narrow objectives who found difficulty in relating his work to the needs of society. He regarded him as somewhat immature and protected but a hard worker and genuine. The subject's own comments indicated that he felt that the lines of communication within his organization were not very clear. He noted that if he wanted to know something about management practices he would rather go to the library and read a book on it, then to attend a management training course. To him the prospect of the courses seemed irrelevant and a waste of time.

His attitude towards the job of section leader rated a score of 20, much lower than average. When asked at the end of six months if he had tried to make any changes in his work behaviour over the past year he responded in the negative but thought that there had been (a little) change in his behaviour over that time.

Altogether, the various sources of information on this subject tend to confirm the Repertory Grid evidence of little or no change as well as indications of a relatively restricted attitude towards his work environment.

9.6

CONCLUSIONS

Possibly the most outstanding feature of the data presented in this chapter is the highly specific and variable results obtained from both experimental and control group subjects. For this reason it certainly does not offer conclusive evidence for the overall success or otherwise of the training courses. The case studies do underline the fact that the situations with which we are dealing are extremely variable and the movement of individuals is idiosyncratic, certain people benefiting or not from the training experience depending upon their previous levels of functioning and the events peculiar to their own experience. Factors which affect the impact of training upon individuals include relatively stable tendencies and previous life history as well as subsequent events such as promotions, the opportunity to try out new ideas and the flexibility and supportiveness of the organizational climate in which they work.

Generally speaking, the four trainees discussed in this chapter moved closer to ideal self but in some cases the movement was very small. Certainly other changes did occur which would strongly suggest that most trainees and their controlling officers were more satisfied with their work performance after training than before. Compared with the controls, the experimental subjects tended to become more perceptive about their interpersonal relationships at work and were more willing to try out new ideas. They possibly also become more self critical.

The stability of the relationship between element 9 (self, now) and element 10 (self, a year ago) perhaps should not come as a surprise

since this type of self consistency is a common feature underlying all human functioning.

In some instances and in one case in particular (Case D) the results of training were unanticipated and certainly unintended. In others the outcomes of training were clearly beneficial both from the point of view of the trainee and of the organization, but in still others the impact was negligible. This may well be an indication that there is a need to be more selective about the candidates for the training programme. It would be possible to develop some mechanism for choosing those who are most likely to benefit, instead of attempting to send every section and group leader who can be persuaded to attend. Another alternative or supplementary approach which the present researcher tends to favour is to conduct modules of training, on site, within the familiar context of the work environment and natural work group. In this way close individual and peer monitoring and feedback can occur.

Inevitably the cost factor involved is an important one and the organization must decide whether it can afford the luxury of either type of training programme, particularly in the present economic climate. However, balanced against this is the question of whether the organization can afford not to continue such a programme of personal development for its managerial staff. The weight of evidence from overseas suggests that this type of activity is continuing at ever-increasing pace despite severe financial restraints (Department of Employment, Work Research Unit, 1981).

Apart from the intrinsic interest of these case studies we must ask ourselves what, if anything, they have shown about the effectiveness of the training programme as a whole. The Repertory Grid on which these case studies are based has obvious uses in certain contexts particularly as a means of monitoring and giving feedback to individual subjects. There is no doubt in this writer's mind about their value in that context. As far as the Repertory Grid as a means of evaluating the general effectiveness of the training courses is concerned, there are probably easier and more straight forward ways of obtaining data which although it provides less depth and insight is still sufficiently detailed for most purposes. Moreover, one cannot expect to observe overall effectiveness under the conditions of high variability which are well illustrated by the present case studies but the additional insight made available by

Repertory Grid technique argues strongly in favour of, at least, the limited use of such means as a supplementary measure of general effectiveness and certainly as an aid to a deeper understanding of individual changes than is possible from the more superficial information gathered by other techniques.

The results discussed so far were obtained by means of a variety of scales and measures selected or developed specifically for the purpose of evaluating the science management training courses. In previous chapters we have been concerned with internal validity, or the degree to which they fulfilled their intended purpose of measuring those variables associated with the effectiveness of the training programme. To this end the evaluation study utilized as far as possible multiple dependent variable measures both objective and subjective, matched control groups, representative samples and monitored follow-up periods.

In this chapter we are more concerned with the generalizability of results and the use of replication as a means of assessing external validity. Williamson, Prost and George (1978) and Wortman (1975) regard it as the fifth feedback procedure in the evaluative process. Replication can also give some indication of the reliability or consistency of measurement which is a prerequisite of internal validity. This reliability refers to the effect of random, unsystematic error which is present in all evaluative research. Suchman (1967) who reviewed much of the pioneering work in evaluative research has enumerated four different types of reliability, all of which may be applied to the present study. These are:-

- (1) The congruence of several indicators or the extent to which several indicators measure the same thing: This type of consistency is illustrated by the similarity in results obtained between the various measures used including behavioural measures, questionnaire responses, verbal scales, interview data and Repertory Grid results. For example, according to the questionnaire data, approximately 75% of trainees made positive changes in their work behaviour following training as a result of the courses and this percentage was quite stable over the three measurement periods (immediately, three months and six months, after training). Other assessments of the effects of

training on work performance, e.g. indirect measures of performance, superiors' reports and Repertory Grid data, tended to confirm this evidence of change in the way in which trainees performed their work.

Similarly, the trainees' reactions to the course were explored by a number of means, direct questions, training techniques preferred, session assessment forms and amount of time devoted to various topics. Again, there was some congruence between the findings from all of these sources.

Attitudes towards the section leader role was approached by both verbal and behavioural means but here as has been found in the past, the congruence was less apparent, the changes in behavioural tendencies being more pronounced than the verbally expressed attitudes. Thus there is an element of systematic variance here which is found to be consistent over much attitude research.

Finally, the work environment of these section leaders was assessed by means of both a formal Organization Climate Questionnaire and by inquiring why the intended changes in trainees' behaviour could not be carried out when they returned to the job and again the findings from these two sources tended to be confirmatory.

- (2) Secondly, the precision of an instrument or the extent to which the same indicator is consistent for a single observer: This type of reliability could not be tackled directly in the present study because of time restrictions and the nature of the study itself. One would expect change to occur following training and for this reason before and after measures were used with individual subjects. Such change did occur and was detected by the measuring instruments used thus eliminating the possibility of assessing this type of reliability within the main study. However, as a compromise we were able to administer a similar set of scales and measures to a corresponding group of trainees in the following year. This is the aspect of reliability with which the replication study is concerned.

- (3) Thirdly, the objectivity of an instrument or the extent to which the same indicator is consistent for two or more observers: Some of the questionnaire items required closed-ended responses which provided them with a degree of objectivity but this type of reliability is much more difficult to assess with self-report, questionnaire and interview measures. The degree of objectivity obtainable when we are dealing with human beings' perceptions of other people and events is always open to some question. The present researcher attempted to deal with this problem by developing scales and collecting data which was as closely tied to concrete, observable behaviours as possible so that for both the subjects and the evaluator the questions asked and the responses made were open to a minimum of individual interpretation. Wherever possible, the responses were cross-checked with other observers so that both the subjects and their immediate controlling officers were interviewed and asked to complete questionnaires. A great deal more could be done in this direction but as previously discussed the experimenter found that field research imposes severe limits of time and opportunity on such activities.
- (4) The consistency of the object measured or the extent to which the object being measured does not fluctuate is the fourth type of reliability. The initial analysis of training needs together with the pre-course questionnaires enabled the evaluator to define some training goals. These, in turn, became the criteria of judgement against which the outcomes of training were evaluated. Since this was a sequential process it is difficult to say positively whether the interpretation of these criteria changed over the twelve month period but there is no reason to believe from the comments and interview data that was being gathered almost continuously over that time period that there was any shift in the perceived goals of training by either the trainees or their controlling officers. Certainly, there was a change in the trainees' priorities as a result of training with more emphasis being placed on organizational and communication skills and less on purely routine and factual exchange but those types of changes were monitored and included in reports of the evaluation.

As the last three points have illustrated, a complicating factor in a treatment programme of this sort is that actual changes do and should occur. Suchman has observed in his 1967 review that it becomes particularly difficult to estimate the stability of instruments and the criteria of measurement within a single study. In this chapter we turn to the immediate task of replicating several of the tests and measures used in the original study as an indirect test of the second type of reliability and as a measure of external validity.

10.2

AIM

To compare the 1977 evaluation results with the 1978 results by replicating some of the experiments carried out with Group A and B trainees in 1977 and their controls, using a new set of subjects who participated in similar courses the following year. The two specific aims of the replication study were:-

- (1) to determine whether results of the evaluation could be generalized to other groups of trainees at other times.
- (2) to test the precision of the measuring instruments when used with other subjects at later training courses. This refers to the second type of reliability mentioned by Suchman.

10.3

SUBJECTS

The replication was conducted with the following years' trainees. They comprised 24 members of course 6 (Group C) and 21 members of course 7 (Group D). These subjects were chosen on the same basis as Groups A and B and were trained under similar conditions exactly one year later.

10.4

PROCEDURE

Members of the sixth and seventh courses were tested by the evaluator pre-, post- and three months after training. It was not possible to administer the complete set of scales and measures of reactions, attitudes, knowledge and behaviour to this new intake of trainees but those that were, occurred under strictly similar

conditions to the previous year. They included measures of:-

(1)	Reasons for attending the course	See Section 7.2
(2)	Training techniques preferred	7.6
(3)	Attitudes to the section leader role	7.7
(4)	Work performance after three months	8.2.1
(5)	Allocation of time to topics	8.2.3
(6)	Indirect measures of effectiveness	8.2.4
(7)	Trainees' understanding of topics	7.5 & 8.2.5
(8)	Measurement of organizational climate	8.2.7

Relationships between age and other biographical variables and work performance were also examined within this replication study.

Because of the type of data available, some of these comparisons can only be attempted by examining the descriptive statistics for each of the two years' trainees. The size of the samples and the level of data does not warrant in these cases further statistical analysis and it would be inappropriate to attempt anything more precise. Nevertheless, it is useful to examine the degree of consistency of the findings from one year to the next and to try to determine any trends which might identify those scales which are sufficiently reliable for future trials and those that need further refinement to develop them into adequate tools of evaluation since one of the purposes of the study was to suggest the means of ongoing monitoring of the courses.

10.5

RESULTS

10.5.1 REASONS FOR ATTENDING COURSES

Using the same five reasons presented to Groups A and B the previous year (Section 7.2), Groups C and D subjects provided ratings on a six point scale (1, most applicable to me, to 6, least applicable to me) for each of the 5 possible categories. Results for Groups C (n=23) and D (n=21) were combined giving a sample size of n = 44. Modal rankings were again calculated for each category of responses.

Scale 1: To make social/professional contacts

Modal Ranking = 4 (frequency = $\frac{18}{44}$)

- Scale 2: To Improve Job Performance
 Modal Ranking = 1 (frequency = $\frac{18}{44}$)
- Scale 3: Personal interest/curiosity and general interest in topics
 Modal Ranking = 2 and 4 (frequency = $\frac{12}{44}$)
- Scale 4: To enhance prospects for promotion
 Modal Ranking = 5 (frequency = $\frac{19}{44}$)
- Scale 5: I was given little choice in the matter
 Modal Ranking = 1 (frequency = $\frac{17}{44}$)

Table 10.1 sets out comparative rank orders for Groups A, B and C + D over the two years.

Table 10.1

Modal rankings of reasons for attendance given by training group members

Group	Reason				
	Social/ Professional Contact	Job Performance	Interest/ Curiosity	Promotion	Little Choice
A	3 (n=23)	1 (n=23)	2 (n=23)	4 (n=22)	1 (n=19)
B	2 (n=21)	1 (n=21)	3 (n=20)	4 (n=20)	1 & 6 (n=19)
C + D	4 (n=44)	1 (n=44)	2 & 4 (n=44)	5 (n=44)	1 (n=44)

Note: Rating Scale: 1 (most applicable to me) - 6 (least applicable to me)

10.5.2 TRAINING TECHNIQUES PREFERRED

It will be recalled from Section 7.6 that Groups A and B consistently altered their preferences for training techniques after training in favour of syndicate group activities instead of more formal and organized types of discussions and lectures. The results from

Groups C and D suggest that there is now a preference for syndicate group activities right from the start with organized discussions and practical activities last. The pre-training preference for the 1978 trainees were:-

- 1st Syndicate group activities
- 2nd & 4th Lectures, informal discussions with course members,
informal discussions with leaders
- 5th & 6th Organized discussions, practical activities

Following training the preferred order was:-

- 1st Syndicate group activities
- 2nd Informal discussions with course members
- 3rd Practical activities
- 4th Lectures
- 5th Informal discussions with leaders
- 6th Organized discussions

10.5.3 ATTITUDE SCALE

The attitude of the section leaders towards their work role was assessed for Group C and D subjects before training courses began (t_1) and immediately afterwards (t_2) and individual subjects' change score calculated. This was compared with change scores for Groups A and B. The change score is calculated by subtracting attitude score at time t_1 from score at time t_2 .

It was found that, in contrast to the previous year (Groups A and B) the mean score for Groups C and D actually *decreased* from before to immediately after training (see Table 10.2).

The four groups were then compared by performing a one way analysis of variance between groups on subjects' individual change scores over the two time periods. The ANOVA was calculated as follows:

Source of Variance	df	SS	MS	F
Between Groups	3	142.770	47.590	
Within Groups	81	1549.505	19.130	3.732
Totals	84	1692.275	66.720	

Table 10.2

Attitude scale means and standard deviations for groups A, B, C and D at times, before (t_1) and immediately after (t_2) training

GROUP	n	Before Training (t_1)		After Training (t_2)	
		Mean	S.D.	Mean	S.D.
A	22	30.5909	4.3715	32.6818	4.9414
B	19	31.5455	5.6965	33.0000	4.3042
C	24	30.8333	4.9490	30.3333	4.0397
D	20	31.8000	6.6301	30.7500	5.3594

This gives an F statistic of 3.732, significant at $p \leq .05$

This significant difference between groups occurs largely between the 1977 trainees (Groups A+B) and the 1978 trainees (Groups C+D) as suggested by the reversal in the sign of the changes in score.

In order to obtain a conservative estimate about which members of the four groups A, B C and D contributed to this overall significance, three independent t-tests were performed on the three orthogonal contrasts, A with B, C with D and A + B with C + D, using the formula,

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2(1/n_1 + 1/n_2)}} \quad , \text{ for 81 degrees of freedom}$$

Comparing Group A with Group B,

$$t = \frac{0.6172}{1.3698} = 0.451 \text{ (not significant)}$$

Comparing Group C with Group D

$$t = \frac{0.5500}{1.324} = 0.415 \text{ (not significant)}$$

Comparing Groups (A+B) with Groups (C+D),

$$t = 2.684 \text{ (significant at } p \leq .01).$$

Thus we see that the significant difference in change scores occurs only between the 1977 and 1978 training groups. This difference is highly significant ($p \leq .01$) and occurs because 1978 trainees showed a decreasing rather than an increasing trend in attitude score from before to after training (Table 10.2).

10.5.4 PERFORMANCE AFTER THREE MONTHS

Groups C and D subjects were sent questionnaires three months after training. These were similar to the ones sent to Group A and B trainees (see Section 8.2.1). Since there was some degree of non-response particularly at these later follow-ups the results are presented for comparison in Table 10.3 together with sample size for intended changes (immediately following training) and attempted and successful changes at the end of three months.

Table 10.3

Number of trainees who indicated change

	Intended	Attempted	Successful Change
Group A	75% (18/24)	75% (18/24)	67% (16/24)
Group B	87% (20/23)	86% (18/21)	71% (15/21)
Group C	83% (20/23)	68% (13/19)	42% (8/19)
Group D	95% (20/21)	76% (16/21)	71% (15/21)

In response to whether their work behaviour had changed (even if only for a few days) as a result of the course,

(1)	(2)	(3)	(4)
A good deal	Some	A little	Don't know

the majority of trainees again stated that they had changed their behaviour, (2) Some.

There were no interviews conducted with the 1978 trainees so that the data is less detailed than for the trainees of the previous year. However, Group C and D subjects were asked to specify

particular changes they made in their work behaviour and the number of such changes recorded is tabulated for comparison (Table 10.4). The comparison is based on those subjects who could specify changes made and therefore the n's in the four groups were small.

Table 10.4

Number of changes specified by individual
course members after three months

Groups	Number of Changes					Number of Subjects
	1	2	3	4	5+	
Group A	7	2	5	0	0	14
Group B	6	1	2	2	1	12
Group C	1	4	1	0	0	6
Group D	4	4	1	1	0	10

A median test was calculated to determine whether there were any differences between the four groups A, B, C and D using the overall median of 2 changes per subject. The number of responses in the eight cells of the contingency table are based on ≤ 2 changes and ≥ 2 changes with subjects giving the median response of 2 being split between the upper and lower half of the table, thus providing a more conservative estimate of the null hypothesis.

Median Test:

Groups	A	B	C	D
≥ 2	8	6.5	3	6
≤ 2	6	5.5	3	4

$\chi^2 = 0.0907$ for 3 degrees of freedom indicating that there is no significant difference between the four groups.

10.5.5 ALLOCATION OF TIME TO TOPICS

The aim of this scale as reported in Section 8.2.3 was to determine the reactions of trainees to the amount of time devoted to the different topics and the question was asked immediately (t_2) and

three months (t_3) after training. The percentage of subjects who stipulated "more time needed", "time correct" and "less time needed" was recorded in Table 8.5, page 141 for Group A and Group B. Results for Groups C and D are shown in Table 10.5 below. There were 43 respondents to the scale at t_2 and 40 at t_3 (Groups C and D combined).

Table 10.5

Subjects' preference for amount of time allocated
to course topics

Topics	More Time Needed		Time Correct		Less Time Needed	
	Group C + D		Group C + D		Group C + D	
	t_2 n=43	t_3 n=40	t_2 n=43	t_3 n=40	t_2 n=43	t_3 n=40
Organization and Delegation	2	10	67	63	30	28
Forecasting and Planning	29	13	36	33	36	27
Resource Allocation	33	28	54	43	14	5
Reporting and Marking	42	38	51	50	7	13
Personnel ^a Management	33	38	45	50	21	13

Note: All entries in table are expressed as a percentage of subjects within the combined group C + D.

^a Personnel Management included Personal and Group Relationships, Communication and Leadership.

10.5.6 INDIRECT MEASURES OF EFFECTIVENESS

The experiment reported in Chapter 8.2.4 was replicated with Group C and D subjects. Responses to the three questions are summarized in Tables 10.6, 10.7 and 10.8.

Question 1: Can you name any books, etc. that you have subsequently read on topics discussed during training?

Table 10.6

Subjects who read books related to
course topics

	Number of Subjects who gave a positive response	Percentage of Subjects	Total Subjects
Group A	2	8.3	24
Group B	5	23.8	21
Group C	2	10.5	19
Group D	3	14.3	21

Question 2: Following the science management course have you discussed it with others in your work environment? With whom?

Table 10.7

Subjects who discussed course with
work colleagues

	Number of Subjects who discussed course with colleagues	Percentage of Subjects	Total Subjects
Group A	22	91.7	24
Group B	20	95.2	21
Group C	19	100.0	19
Group D	20	95.2	21

Question 3: Have you maintained contact with other course members? By what means? (letter, phone, face-to-face, etc).

Table 10.8

Subjects who maintained contact with
other course members

	Number of Subjects who maintained contact with other course members	Percentage of Subjects	Total Subjects
Group A	14	58.3	24
Group B	15	71.4	21
Group C	10	52.7	19
Group D	14	66.7	21

10.5.7 UNDERSTANDING OF TOPICS

In a similar manner to the analyses performed on Group A and B results (see Sections 7.5 and 8.2.5), the 1978 data from Group C and D trainees was subjected to a sign test to determine whether there were significant differences in understanding from before (t_1) to immediately after (t_2) the training courses on any of the five topics studied. Additionally, Friedman's rank order ANOVA was applied to the ordered data over the three measurement periods, before training (t_1), immediately after training (t_2) and at three months following training (t_3). Graphs illustrating the average ratings given by each group of trainees on all five topics are presented in Figure 10.1. These graphs permit a visual comparison to be made between groups and it can be seen that this visual representation accords with the results of the statistical tests of changes over time.

Results of the pre- and immediate post-test level of understanding of the four training groups were analyzed by means of a sign test (similar to the test applied to Groups A and B trainees in Section 7.5) to determine changes in understanding occurring over the training course period. Table 10.9 contains comparative data for Groups A, B, C and D. Results of the Friedman test are shown in Table 10.10.

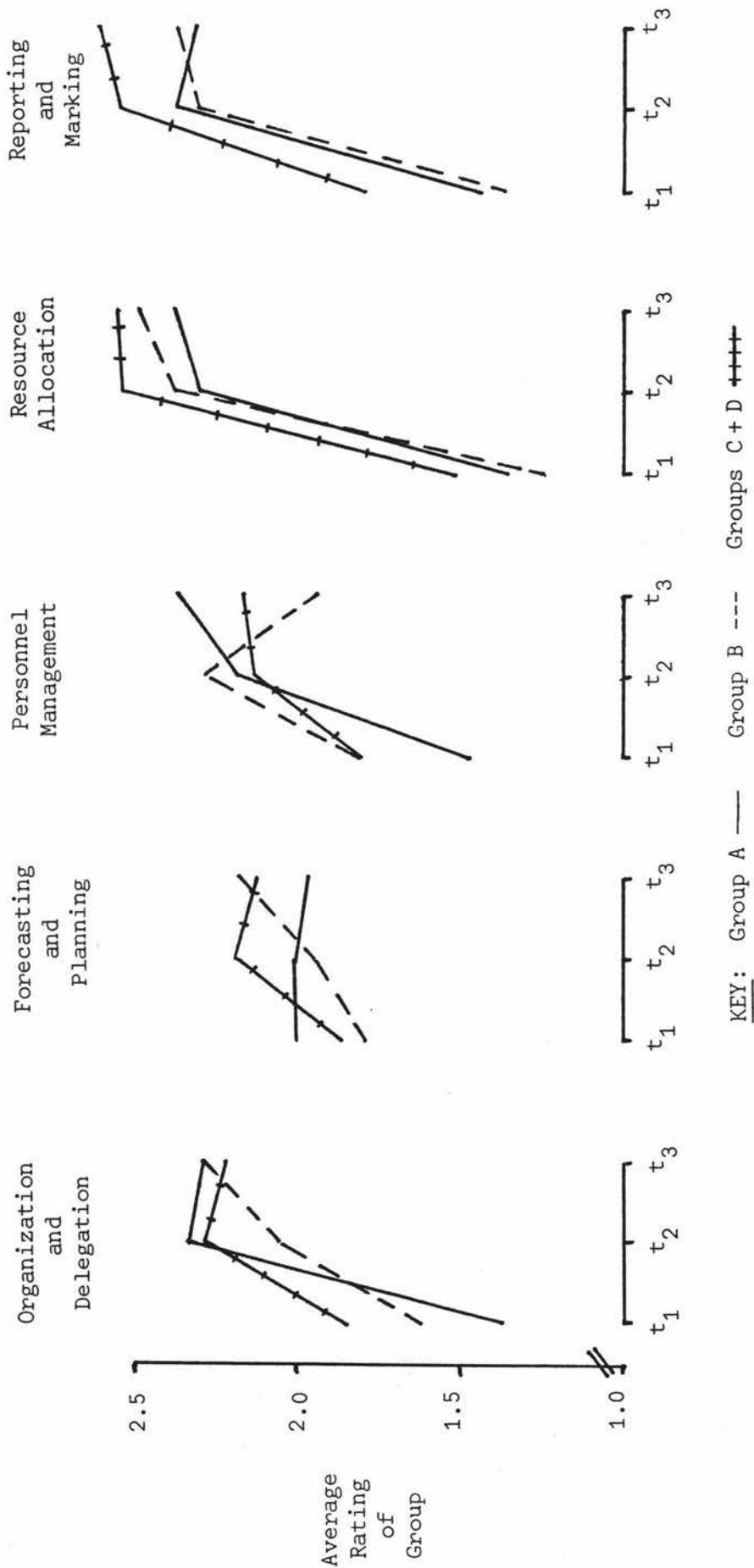


Figure 10.1 Changes in understanding of topics at three times, pre-test (t_1), immediate post-test (t_2) and three months (t_3).

Table 10.9

Number of positive changes in understanding of
course topics for Groups A, B, C and D

Group	n	Organization and Delegation	Forecasting and Planning	Resource Allocation	Reporting and Marking	Personnel Management
A	24 ^a	16/18** ^b	7/15	17/19**	18/23*	15/19
B	21	10/14	10/17	17/17**	14/16**	9/13
C	23	14/15**	13/19	18/20**	16/17**	13/15**
D	19	11/13**	12/15**	16/16** ^c	13/14** ^c	11/16

* Significant at $p \leq .05$

** Significant at $p \leq .01$

Notes: ^a Z score used to test for significance when number of changes (less ties) is greater than 20 since the binomial distribution is approximated with the normal distribution when $n > 20$.

^b second number is n' = total number of changes (less ties)

^c $n = 18$, as one subject failed to respond to this scale

Table 10.10

T statistics obtained using Friedman's test on subjects' rank ordering of level of understanding of topics at times t_1 , t_2 and t_3

Group	n	T value				
		Organization and Delegation	Forecasting and Planning	Resource Allocation	Reporting and Marking	Personnel Management
A	24	11.70**	0.020	17.150**	11.520**	10.190**
B	20	4.075	1.425	20.575**	12.700**	2.425
C	18	6.861*	5.028	15.194**	18.778**	4.694
D	18	14.083**	6.028*	20.361**	8.583*	3.694

* Significant at $p \leq .05$

** Significant at $p \leq .01$

Thus we see that Resource Allocation and Reporting and Marking are consistently significant over the four training courses. The results for Organization and Delegation are slightly less consistent but for these three topics there is evidence of increased understanding from before to immediately after training. On the other hand, results for the topics Forecasting and Planning and Personnel Management do not suggest that consistent positive gains in understanding have been made.

In order to compare changes in understanding over the three testing periods t_1 , t_2 and t_3 (see Section 8.5), the results of the two-way analysis of variance by ranks (Friedman, 1937) is presented in Table 10.10.

10.5.8 ORGANIZATION CLIMATE QUESTIONNAIRE

For the Organization Climate Questionnaire which is reported in Section 8.2.8 the results of the replication are illustrated by presenting all four profiles obtained from each group for visual comparison, together with the medians and ranges of each of the seven dimensions. Once again, the results were subdivided into government department/semi-government groupings and leader/non-leader groupings (see Table 10.11).

Table 10.11

Medians and ranges of ratings made by Group C and D trainees on the climates within their organizations

Dimensions	Total Group n=45		Govt Dept n=37		Semi-Govt Research n=8		Leader n=31		Non-Leader n=14	
	M^a	R^b	M	R	M	R	M	R	M	R
	1. Conformity	4	8	4	7	4	8	4	8	5.5
2. Responsibility	7	8	7	8	7	7	8	8	7	5
3. Standards	6	5	6	5	7	4	7	5	6	2
4. Rewards	7	6	7	6	7	4	7	6	6	4
5. Organizational Clarity	6	8	6	6	6	8	6	8	5	6
6. Warmth and Support	7	9	9	8	7	7	8	9	7	6
7. Leadership	7	8	7	8	7	6	7	8	6.5	4

a_M = Median b_R = Range

All groups produced similar profiles when average ratings were plotted over the seven dimensions of organizational climate (Figures 10.2 and 10.3).

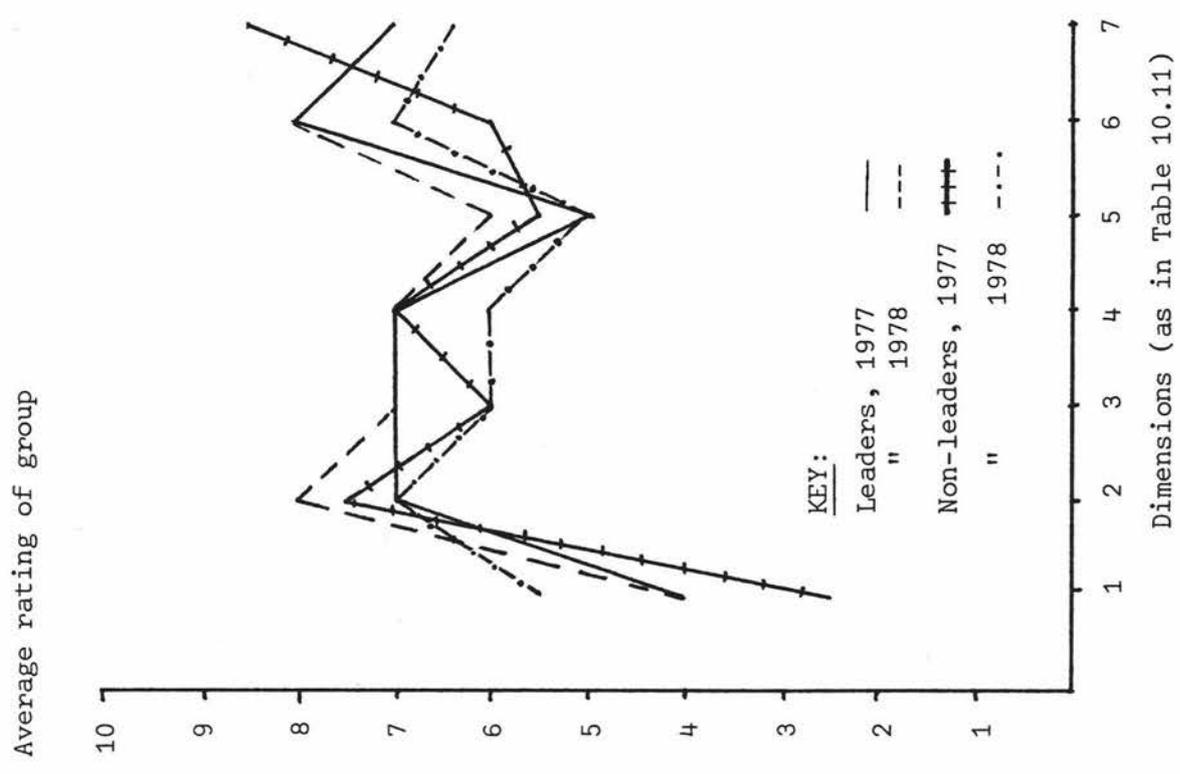


Figure 10.3 Summary profiles of organizational climate for leaders and non-leaders

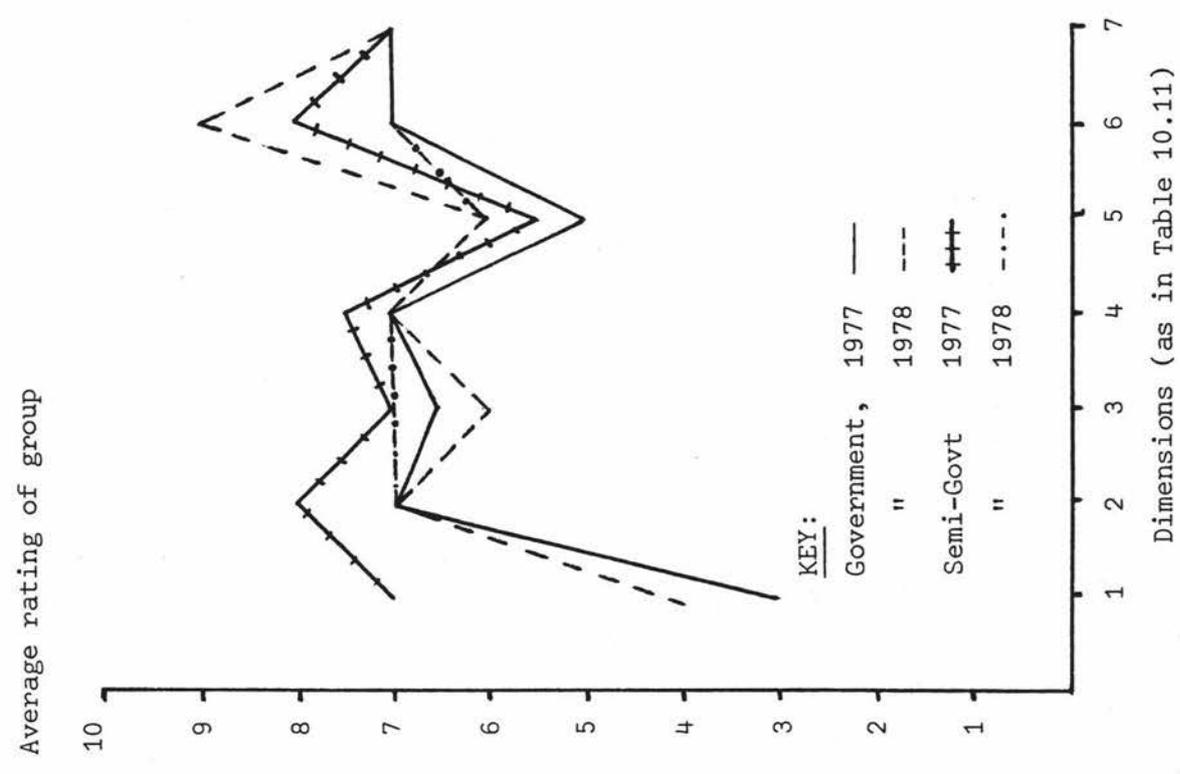


Figure 10.2 Summary profiles of organizational climate for government and semi-government organizations

10.5.9 EFFECTS OF AGE, TENURE AND LEADERSHIP EXPERIENCE

During the six month follow-up evaluation (Section 8.2.6) the relationship between the variables age, tenure and leadership experience and changes in work behaviour for Groups A and B were examined. Several of these analyses were repeated for Groups C and D in 1978. These results are tabulated (see Table 10.12) and may be compared with those in Table 8.10, page 157. Since the 1978 trainees were followed up for only three months after training, results were compared for:-

1. Immediate post course intentions and age
2. Age and change at three months
3. Tenure and change at three months
4. Years of leadership experience and change at three months.

Table 10.12

Relationship between work behaviour and age, tenure and
experience of Group C and D subjects

Relationship	Total n	Means		Standard Deviations		t ^a Value
		\bar{X}_1 (Yes)	\bar{X}_2 (No)	σ_1^2	σ_2^2	
1. Immediate post- test intentions/ Age	43	36.87 (n=39)	39.75 (n=4)	5.27	10.90	-0.9345
2. Age/Change at 3 months	39	37.59 (n=29)	37.30 (n=10)	5.997	7.56	0.1217
3. Tenure/Change at 3 months	38	10.07 (n=29)	9.00 (n=9)	6.36	5.52	0.4528
4. Years of Leadership experience/Change at 3 months	37	4.07 (n=29)	3.63 (n=8)	4.85	6.63	0.2116

^a None of these t values were significant at $p \leq .05$

10.6

DISCUSSION

A brief discussion follows of the results of each replication experiment.

10.6.1 REASONS FOR ATTENDING COURSES

From the results obtained we see that the ordering of priorities for the 1978 trainees is quite similar to that of Group A and Group B. For both years the preferences were:

- (1) To improve present job performance or because trainees had little choice in the matter
- (2) To make social/professional contacts or from general interest in the topics

and finally,

- (3) To enhance prospects for promotion.

The consistency between groups of subjects over the five possible reasons suggests that this scale produces similar results across different groups of trainees and that their priorities have not altered very much over two consecutive years.

10.6.2 TRAINING TECHNIQUES PREFERRED

Again, following training, syndicate group activities occurred in first position and organized discussions last as was the case for Groups A and B. There was considerable variation in the remaining preferences. However, results from the replication study suggest that syndicate group activities are now preferred to more formal methods even before training begins. The writer suspects that, as a result of the extensive discussion that was occurring between past and future trainees following the 1977 courses (as shown in Chapter 8.2.4) the word was spreading within the organizations about the value of syndicate group activities and trainees were now beginning to attend the courses more prepared to accept such informal group experiences.

10.6.3 ATTITUDE SCALE

From our previous discussion in Section 7.7 we saw that there were no significant changes for Groups A and B from before training to after training. Nor were there significant differences between experimental and control groups (Groups E and F) on the attitude scale. When we compare the mean scale score for all four training groups, A, B, C and

D over times t_1 and t_2 , before and after training, we see that while these changes continue to be small and non-significant, there is actually a decrease in mean attitude scale score in 1978 (Groups C and D), as shown in Table 10.2, page 217.

A comparison of the change scores obtained by subtracting score at t_2 minus score at t_1 for subjects in each of the four groups gave an F statistic of 3.7316 which is significant at $p \leq .05$. Furthermore, the independent t-tests showed that most of this difference occurs from years 1977 to 1978 where there is a decrease rather than an increase in mean attitude scale score. This additional information supports the previous statement that the attitude scale in its present form has failed to detect any significant and consistent changes in verbally expressed attitudes.

10.6.4 PERFORMANCE AFTER THREE MONTHS

From Table 10.3 we note the similarity in proportion of respondents who planned, attempted and were successful in implementing work changes following the courses. The results of the median test between the four groups A, B, C and D strongly support the null hypothesis so that we must assume that there is no difference between Groups A and B trainees (in 1977) and Groups C and D trainees (in 1978) in the number of changes made in work performance after three months. In this case the number of subjects who chose to report details of their behaviour changes was quite small but so far as we can tell from these results, this particular question has produced stable results over the four groups of trainees.

Additional comments by the 1978 trainees revealed that the types of changes made were similar to those of the previous year's trainees with the addition of:

- (1) Maintaining useful work contacts established with other departmental scientists and with administrative staff
- (2) Improved understanding of departmental functioning
- (3) Coming to a realization and acknowledgement of the fact that a particular trainee was not prepared to accept administrative duties.

When asked about reasons for lack of success, members of this group specified:-

- (1) Lack of support and/or resistance by other staff or by management
- (2) Shortage of time, insufficient staff or heavy workload
- (3) Inability to "get through" with new ideas.

On the whole, both the extent of changes made by trainees following the courses, the types of changes in behaviour and the reasons for lack of success were similar for both years' subjects.

10.6.5 ALLOCATION OF TIME TO TOPICS

From an examination of the proportion of subjects who responded in each time category, the following information is available about each of the course topics:-

Organization and Delegation: In 1978 about 2/3 of the trainees were satisfied while almost 1/3 would have preferred less time on this topic. Only a small fraction wanted more time. The previous year, again, approximately 2/3 were satisfied but on that occasion the majority of the remaining 1/3 felt that they needed more time.

Forecasting and Planning: In both years the tendency was for trainees to be divided between wanting less time for this topic and being content with the existing state of affairs. On neither occasion did many trainees require more time on these topics.

Resource Allocation: The 1978 trainees tended to be equally divided on whether they wanted more, less or the same amount of time for Resource Allocation. In 1977, while approximately 2/3 were satisfied, 1/3 wanted more time.

Reporting and Marking: Half of the trainees were satisfied in 1978 and a further 1/3 would have preferred more time whereas in 1977 well over 2/3 of the course members were satisfied with the allocated time.

Personnel Management: In 1978 approximately 1/2 the trainees said that the time was correct and another 1/3 would have liked more time. The rest wanted less time. In 1977 subjects were divided almost 50-50 between being satisfied and wanting more time, with a very small percentage requiring less time.

This appears to be a useful scale as it allows the reaction of the trainees to individual topics to be closely monitored. We would expect fluctuations in the results to occur because while approximately the same total amount of time was allocated to topics for all four courses, both the content of the topics and the speakers and course supervisors changed from one year to the next. Taking this into account, the opinions of the trainees in the years 1977 and 1978 remained remarkably stable.

10.6.6 INDIRECT MEASURES OF EFFECTIVENESS

The responses to the first question in this section were consistent in so far as none of the groups of trainees followed up the training courses with much additional reading in the management area. Positive responses were counted only when particular books or articles were specified. For question two, the answers show that, once again, as for the previous year, Group C and D members held discussions with a wide range of people including their controlling officers, fellow scientists, technical staff and many other staff members. The writer believes that this widespread discussion could be an important and significant feature of the success and lasting effectiveness of the training courses.

The amount of contact maintained between ex-trainees from the courses is encouraging. Again, it seems that more than 50% of trainees have some subsequent interaction with others from the course and that the majority of these contacts were work related (Question 3). The reported interactions were "occasional" or "infrequent" except in the case of people within their immediate work environment. Approximately 63% were face-to-face for the 1978 trainees compared with 50% for the 1977 groups. The remainder were by letter or telephone.

10.6.7 UNDERSTANDING OF TOPICS

If we consider the results from the sign test between t_1 and t_2 with the results from Friedman's test for ordered data over times t_1 , t_2 and t_3 for all four groups, A, B, C and D we see that where there were overall gains in understanding, most of the improvement occurred during the training courses themselves with a gradual

fall-off in the rate of change in the three months following training. The slope of the graphs in Figure 10.1 illustrates this point.

The consistency of this scale over the four groups of trainees is well supported by these results. In 1978, as in the previous year, when considering overall changes in understanding of topics from before training to immediately after training to three months after training, the two topics Resource Allocation and Reporting and Marking show positive significant results while the remainder of topics are more variable, Organization and Delegation indicating some degree of consistent improvement.

10.6.8 ORGANIZATION CLIMATE QUESTIONNAIRE

Figures 10.2 and 10.3 illustrate the similarity in profiles obtained from the trainees on the seven dimensions of organizational climate over the four groups A, B, C and D. Tentative distinctions between leaders and non-leaders and between government and semi-government employees were not supported by the replication, thus we cannot draw any conclusions at all about these sub-group differences in perception. However, an overall similarity between the four groups of trainees is clearly discernable.

10.6.9 EFFECTS OF AGE, TENURE AND LEADERSHIP EXPERIENCE

As we had found in the previous year, there was no significant relationship between age, tenure or leadership experience and either the intention to make work changes or the implementation of such changes at three months. In both 1977 and 1978 there was a positive relationship between intentions and attempted changes by the end of the three month period ($\text{Chi}^2 = 7$, significant at $p \leq .01$). These results are consistent over the years 1977 and 1978.

10.7

CONCLUSIONS

The first aim of these replication studies was concerned with the external validity of the evaluation results obtained. The question to be considered here is whether or not the evaluation results can be generalized to other groups of trainees at other times. The similarity of the results between the four training groups, A, B, C and D over the

two years 1977 and 1978 suggest strongly that for a population of section leader trainees from organizations similar to the ones described here, there is reasonable consistency of results from group to group and from one year to the next. This would suggest that the scales used are suitable for the measurement of changes following training in science management trainees of the type described. However, we cannot reasonably extrapolate to other types of management trainees in other types of training schemes for the study is concerned with a specialized group of subjects undergoing a particular type of management training for a specific purpose. We may conclude that, within these population parameters, the results have external validity. It is the writer's opinion that this is the essence of management training. It must be designed with a high degree of specialization and the conclusions to be drawn about its effectiveness must be similarly limited (Williams and Berger, 1972).

The second aim of replication was to test the reliability of the scales and measures used with different but similar groups of subjects and at different times. As the results in this chapter demonstrate, there is a certain amount of stability of results over the four groups of trainees, A, B, C and D. The complicating factor is that the courses themselves tend to change from year to year due to the availability of speakers and topic supervisors. Moreover, courses will continue to change as they are further developed in response to the evaluation feedback. Thus reliability of measures is not an easy or straightforward concept to assess under these circumstances. In spite of this difficulty, Burgoyne and Cooper (1975) stress the need to develop the means of continuous monitoring of changes occurring in a developing programme of training. This is particularly important when the evaluator must supply up-to-date information on the immediate situation within an action-research context (Burgoyne, 1973).

We must expect the training programme to change from year to year and this may mean discarding certain of the scales and measures as they are found to be inadequate or redundant. The goal of continuous monitoring must extend to the means of evaluating training as much as to the training itself and one way of ensuring that the evaluation techniques are fulfilling their function is to continue to make comparisons at frequent points throughout the process so that objective and informed judgements can be made about the entire

training-evaluation cycle. It is for the purpose of monitoring the measuring instruments used that some form of reliability studies are needed.

In the next chapter, the writer will attempt to gather together the various strands of information obtained throughout the entire series of studies, so that some general conclusions about the effectiveness of the training programme can be drawn, remembering that such conclusions are valid at a particular point in time. As Wortman (1975) has shown by his training-evaluation model, both the training courses themselves and the means used to evaluate them are not static but are subject to the effects of the feedback loops within the total system of programme operation and evaluation.

CHAPTER 11 THE EFFECTIVENESS OF THE TRAINING COURSES

11.1

INTRODUCTION

If we accept the broad definition of evaluation as the measurement of valued consequences of an action (Chapter 1, p. 4) we may now ask ourselves what the results have shown about the effects of training and begin to assess the value of the training programme in the light of this information. Working within the practical frameworks of evaluation suggested by Hamblin (1974) and Warr, Bird and Rackman (1976) there has been a deliberate effort made to examine the *context* of training and provide an answer to the question of what needs to be changed. We have also examined the *input* of training and the techniques used to determine which procedures are most likely to bring about change. We have assessed the trainees' *reactions* and opinions about the worth of the courses. Finally, we have investigated the *outcomes* of training to discover what evidence there is that change has occurred.

Using the Wortman model as an ideal theoretical basis for evaluation research the writer can now go a step further and begin to make explicit some of the additional variables which were found to have moderating effects on the observed outcomes of training. Figure 11.1 presents an adaptation of Hamblin's cycle of evaluation including some of these moderator variables which emerged from the empirical study.

11.2

CONTEXT OF TRAINING

Initially, the context of training was investigated by means of a careful analysis of training needs in which members of the participating organizations from all levels were interviewed and questioned to help tease out the critical requirements of the job of section or group leader. Then, on the basis of this information, the problems of measurement of the relevant variables were tackled, the aim being to develop a set of scales which demonstrated a close correspondence between the concepts which had been identified as important criteria of success and the actual measures of the dependent variables. This task is an ongoing one in the training-evaluation process and one would not expect to complete it within a single study.

MODERATING VARIABLES

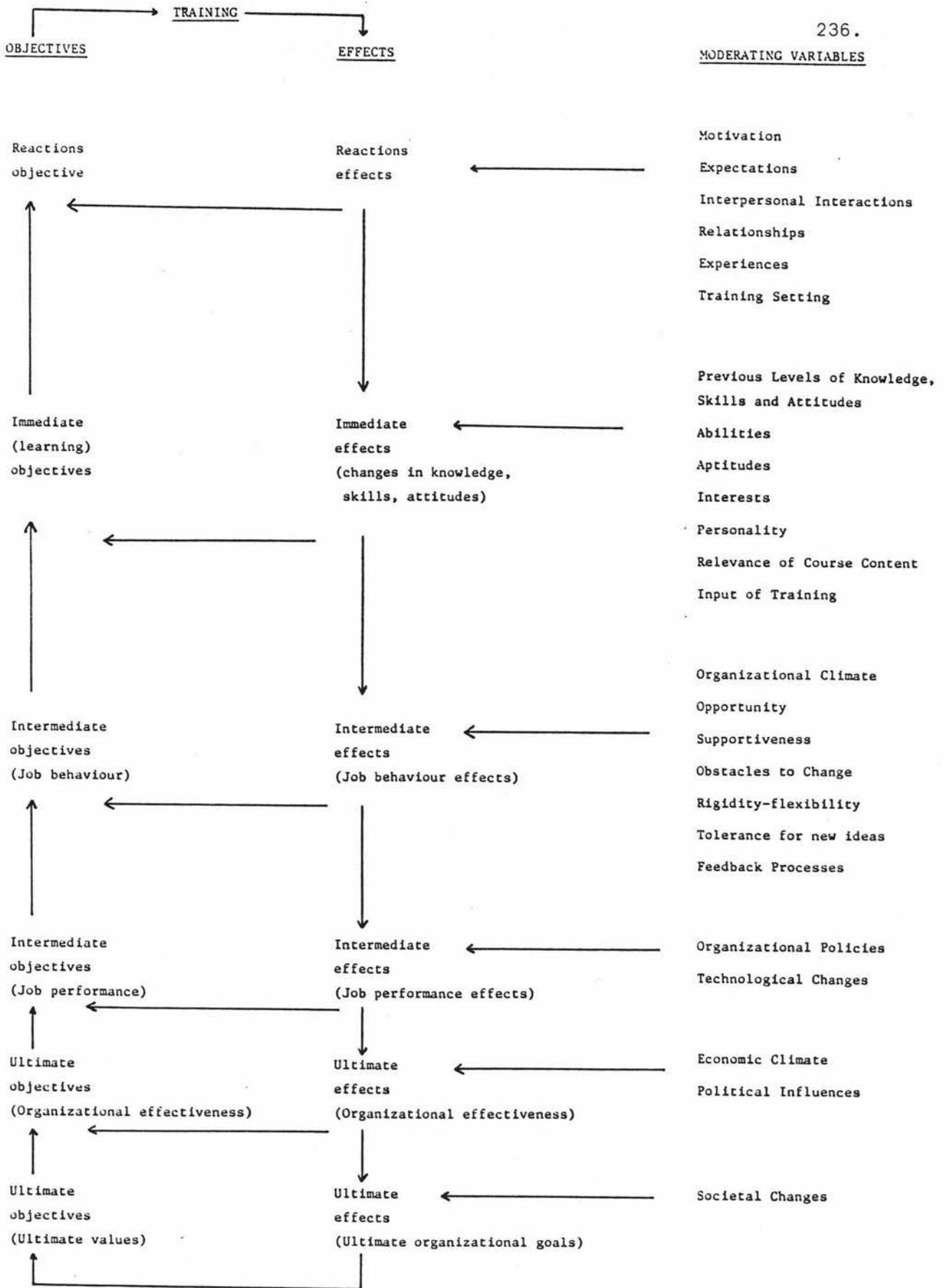


Figure 11.1 The cycle of evaluation including moderator variables
(Adapted from Hamblin, 1974)

The critical requirements of success in the managerial and leadership role are too variable to assume that measures chosen will be applicable in the long term. Rather, the evaluator is faced with the prospect of a dynamic set of criteria requiring a flexible and multi-dimensional approach which changes and develops over time.

11.3

ULTIMATE OBJECTIVES OF TRAINING

There are several distinct levels of objective setting and outcome evaluation for training, the ultimate level, which includes the evaluation of organizational effectiveness being the final and most distal one. This level of objective setting and evaluation involves the broader areas of organizational change including employee satisfaction, morale and productivity. If this phase is to be attempted at all, it requires longitudinal research and extensive follow-up procedures incorporating criteria of relevance to the long-term goals of the organization. Unfortunately, these same distal criteria are particularly susceptible to other environmental influences which often cannot be predicted or measured easily. Most evaluators suggest that such long-term assessments of effectiveness are rarely practicable. The present study did not attempt any direct measurement of this type of organizational change but rather concentrated on the behaviour changes observed in its members over a twelve month period following training so that not only immediate behaviour change was assessed but also the degree of relevance and transfer of training to the work situation. This is a partial solution, only, to the problem because, ultimately, evaluation is most concerned with the valid prediction of organizational goals and this implies ongoing and continuous research. In the present case this could possibly proceed along the lines of a study by Hinricks (1978) who evaluated the effectiveness of a management assessment centre designed to predict the potential of members of a marketing organization. In that study, personal characteristics of employees identified at the assessment centre were found to be strongly related to the organizational level attained by the managers eight years later. In a similar manner, characteristics of the section leaders measured at the end of training could be related to objective measures of organizational variables some time in the future. Some possible distal criterion variables are staff turnover, number of scientific papers published, measures of job satisfaction, salary level attained, group

or sectional output, number of projects completed, and so on. However, all of the outcomes mentioned are subject to many external, uncontrolled influences and may not, in the long run, add a great deal of understanding to the immediate problems of the training courses themselves.

On the other hand, a concern with long-term goals does force one to consider what might be called the construct validity of the total programme (Williamson et al., 1978) and calls for an examination of the whole rationale on which the programme is based. One piece of information of organizational relevance often sought by an organization involved in personnel development is an estimation of cost-effectiveness of its managerial training programme. In order to comply with this demand two conditions must be met. Firstly, the effects of training at the organizational level must be precisely identified and, secondly, there must be evaluation techniques available which adequately measure in financial terms, the kinds of benefits which may result from training. This latter requirement is particularly difficult to achieve when the goals of training are largely educative and developmental. How do we know when people like managers and section leaders are making a full contribution to their organization? It is much easier when we are evaluating a competitive profit oriented organization because an entirely different approach to evaluation can be adopted. In the latter case, if we wish to assess the benefits of training in economic terms, the relevant dependent measures may include such variables as reduction in spoilage, error rates, overtime, staff turnover, tardiness and so on. The resulting cost savings may then be calculated in objective, financial terms and a cost-benefit analysis conducted.

However, organizations of the type we are discussing do not have objectives which are exclusively or even largely financial. They are aimed partly at saving money but also at promoting societal benefits and economic gains of a very long term and global nature. To these ends the effects of management training cannot always be precisely identified. The organizational objectives are often necessarily vague and include such things as flexibility, adaptability and the willingness of managerial and research staff to grasp unforeseen opportunities when they arise. These may take a long time to be reflected as economic profitability and they are extremely susceptible to the effects of other intervening and

moderator variables.

Besides the complexity of the aims in the present study, it was not possible to obtain the necessary information for conversion of these aims into financial terms. Consequently, a cost-benefit analysis was not attempted.

Williamson et al.(1978) maintain that one can have reasonable confidence in the programme rationale if a positive evaluation is obtained of the general effectiveness of the training, the training techniques used, the reliability and validity of measurement and the congruence between training needs and training outcomes. However, the rationale and theoretical basis on which a training programme is based should be constantly monitored and if cause-and-effect relationships between training procedures and training outcomes appear to be lacking then new rationales, goals and training procedures must be considered.

11.4

TRAINING INPUT

In the present case there were indications that changes in course content or the input of training were required in certain areas. For example, although improved internal communication was one of the major goals attributable to the training courses, it was clear that communication between the organizations and various outside groups and clients from government departments and industry had not been significantly improved. Comments from course participants indicated that except at the very general level of enhanced self-confidence and interpersonal skills, existing courses did not provide training in implementing the flow of information to and from the outside community. Yet, clearly, the organizations concerned recognized a need to consciously train people for the important task of effectively controlling and promoting such interchange. Thus, it was suggested that training organizers should emphasize the broader issues of communication in future courses.

Another important variable in improving managerial courses and one which is often neglected in leadership research as well as practice is the feedback process - the mechanism whereby a leader finds out how well he or she is doing. Evidently, this type of consistent and precise

feedback on performance had been lacking in the past for these section leaders. It was shown that this could be provided during the training course itself, particularly if the course was designed to create situations reflecting the daily demands of the leadership role. Then, by means of immediate feedback from topic supervisors and fellow trainees they could study at first hand the impact of their behaviour on others. Ideally, of course, the concept of feedback should not be limited just to the training period but should form part of a continuous, supportive, in-service training orientation. Although it may not replace the existing formal management training courses, longitudinal work with section leaders could certainly supplement and enhance their effectiveness. Moreover, feedback on work performance is probably essential to consolidate the learning achieved during a brief two-week course. Indeed, training research has increasingly placed heavy emphasis on such techniques. According to Blake and Mouton (1964), training must be followed by the development of work teams within the normal work context.

Expanding into more experiential and situational training methods would naturally reduce the time available for imparting purely factual and theoretical information but the evidence from measures such as the session assessment forms and direct questions relating to training techniques suggested that these changes in procedure would enhance the training outcomes, while much of the content contained in such topics as Reporting and Marking and Resource Allocation could be imparted equally well in written form by means of a departmental handbook. Clearly, the more factual topics produced the greatest immediate learning effects (see Chapter 7) but, by the same token, they had less impact on long-term performance outcomes (Chapter 8). The need for section leaders to have a thorough knowledge of their organizations should not be minimized and they do need a good understanding of topics like the two mentioned above. This factual information is needed in addition to their technical and scientific skills because, as this evaluation has shown, the section leaders' ability to structure and organize their work groups depends not only on technical expertise but also on ability to use such knowledge effectively to achieve organizational goals. By producing a handbook which could be readily updated some of the informational material which had hitherto taken up a considerable proportion of the time at the annual courses could be more effectively presented in a form suitable for easy

reference, thus allowing for more efficient use of the limited training time available. The desirability of these types of changes only became apparent when some of the underlying organizational goals were revealed during context level evaluation. Such changes represent some basic shifts in the philosophy governing the training programme.

11.5

IMMEDIATE OUTCOMES OF TRAINING

At this level of objective setting and outcome evaluation we must consider the immediate outcomes of training, including changes in knowledge, attitudes and skills. In both Chapters 7 and 8 we have examined in some detail the learning (knowledge) achieved both immediately following and three months after the training courses ended. The changes were estimated from a baseline of pretest levels of understanding of course topics and, as previously explained, all of these measures were made on the basis of self-reported understanding rather than on objective measures of knowledge of course content which would have been the preferred method of data collection. Over the three measurement periods some significant gains were observed for all topics with the exception of Forecasting and Planning. Analysis of the following year's courses revealed a similar pattern of results (see Figure 10.1). Further, it was interesting to note that it was in the topics of Resource Allocation and Reporting and Marking that a significant amount of learning occurred during the training period itself. For other topics there was a tendency for the increases in understanding to occur more gradually from pre-course to at least three months after training. In the case of Personnel Management (Group B) there was an initial gain during training but by the end of three months, trainees' self-reported understanding had returned almost to precourse levels. It is possible that this latter effect may be a function of a work environment which was unfavourable to the fostering of good interpersonal skills but this suggestion would need more intensive investigation. When compared with the control groups (untrained and previously trained subjects) before and immediately after training there is some evidence that all of these changes are due to the subjects' participation in the training courses (Section 7.5). These results indicate that training has produced some positive effects on subjects' knowledge of course topics.

The evidence for changed attitudes towards the section leader role has been discussed in Chapter 7. Although there was a trend towards a positive change in verbally expressed attitudes immediately after training, none of these results were significant and this trend was not replicated by the following year's trainees. At the behavioural level, however, the outcome was much more promising. At the end of three and six months a high proportion of trainees recorded attempted and successful efforts to change their managerial behaviour (Section 8.2.6). These attempts were more frequent and of a different type to those made by untrained controls, the former showing more concern for developing good interpersonal relations and improving intra-departmental communication than the latter. The following year's trainees again displayed no significant increases on the verbal attitude scale but measures of the behavioural component were of similar magnitude to those of the previous year (89% intended change, 73% attempted change and 58% introduced change with moderate success). Thus, some positive attitude change may be inferred but this was certainly not observed at the self-expressed verbal level in response to the questionnaire used here. The writer considers that further development of this scale is needed to improve its sensitivity to the relatively small changes which are likely to occur in subjects' verbal attitudes.

11.6

INTERMEDIATE OUTCOMES OF TRAINING

At the next level of objective setting one must specify intermediate objectives of training, such as changes in work performance which can then be evaluated by the measurement of actual changes in work attitudes, skills and behaviours and by estimating the amount of transfer of training to the work situation. Thus, the setting of intermediate objectives is related to the measurement of intermediate outcomes which were assessed by observing both general and specific changes in behaviour from before to immediately, three, six and twelve months after training. The types of behaviour changes noted in Chapter 8 provide some very clear indications of those aspects of the existing courses which met with the specific requirements of the section leader role. These were based on the critical requirements of the job as defined in the initial survey and included:

- (1) Administrative skills
- (2) Information about the organizations
- (3) Interactions between divisions
- (4) Forecasting and planning
- (5) Organizational skills
- (6) Resource management
- (7) Publishing
- (8) Conducting meetings
- (9) Decision-making skills
- (10) Reporting and marking
- (11) General interpersonal skills
- (12) Communicating
- (13) Interviewing and selection
- (14) Motivating staff
- (15) External relationships

By adopting a critical incidence approach when conducting the interviews, it was relatively straightforward to elicit the specific behaviours associated with each of these performance dimensions. It was found that categories (4), (5), (6), (10), (11), (12), (13) and (14) were all affected to some extent by the training courses (Section 8.2.6, Figure 8.3). This method conveniently provides a direct link between behaviour, performance and organizational effectiveness, the three levels of outcomes in the Campbell, Dunnette, Lawler and Weick model of organizational effectiveness (see Figure 12.1).

In the present study, evaluation of trainees' behaviour, an intermediate level outcome, has shown first of all that moderate improvements occurred particularly in areas which had direct personal and organizational relevance to the people who took part in the training courses. Moreover, these areas of change could be predicted fairly accurately by trainees prior to training (Section 7.3). However, while there was reasonable agreement on broad areas of organizational change,

when it came to personal change, this was much more variable from one individual to another and dependent on the particular needs of the trainee (see Section 8.2.2 and case studies, Chapter 9). This observed variability in behaviour suggests that there is no standard or stereotyped outcome towards which the training should conform. Bunker's (1965) study similarly supports this argument. He concludes that the long-term outcomes of human relations training may be, "an increased capacity for adaptive orientation to the individual trainee's particular situation and an increased personal freedom to act on the basis of information acquired and processed during the training experience."

The results of the present study seem to indicate that training affects different people in different ways. The outcomes are, in fact, partly a function of each person's specific needs based on his past experiences and partly the result of his mode of interaction with the training experience itself. It is relevant to this training program that the writings of Bennis (1966a) and other researchers stress that modern organizations need above all people equipped with skills to cope with rapid and unexpected social and technological growth and change, flexibility and adaptability being of primary importance. Because of these indications early on in the evaluation study, it was decided to allow the subjects themselves to name and specify the particular behavioural changes they had made at subsequent three and six month follow-ups. This meant that a large body of qualitative material was collected as well as the quantitative data. The writer maintains that this was the only way in which a full and accurate assessment could be made of the actual changes which did occur.

At three months, self-reported behavioural changes were recorded by means of questionnaires and individual and group interviews. Additionally, a number of indirect indices of behaviour were employed, namely, the number of books read on course related topics, discussions held and contacts maintained after the course had ended. Similar methods were employed with the 1978 trainees. The proportion of course participants who successfully implemented changes was approximately 70% (Section 8.2.1). Behavioural changes were verified and substantiated by specific incidents related during interview sessions. Four major areas

of change were identified. They were:

- (1) Section or group organization including joint planning of project work.
- (2) Improved communications.
- (3) Improved personnel management skills, such as selection interviewing.
- (4) Increased confidence and personal satisfaction.

On the indirect measures, only 24% of the experimental group (13% in 1978) could name relevant books they had read but 95% had discussed courses with work colleagues (98% of 1978 trainees) and 71% had maintained contact with their course companions (60% of 1978 trainees).

For the six month follow-up, experimental and control group members plus their controlling officers were asked to describe any changes observed in work performance since the time of the courses. As stated earlier, 67% and 81% of ex-trainees and 67% of non-trainees could specify such changes (see Section 8.2.6). However, the important difference seems to be in the frequency with which individuals engage in innovative behaviour. In fact, ex-trainees had initiated two to three times as many changes per person as had the comparative group of non-trainees and there was a qualitative difference between the two as well (Section 8.2.6). The above judgements were confined to trainees' self-reports and since previous research has shown that self-recorded behavioural changes may be both quantitatively and qualitatively different from those observed by superiors, fourteen directors also completed a questionnaire referring to the work performance of their staff. From Section 8.2.7 we see that five of the 21 behavioural categories provided were significantly ($p \leq .05$) affected by the training courses and that these tended to complement the trainees' self-reports.

The twelve month follow-up was a simple three scale questionnaire designed to assess briefly trainees' opinions, in retrospect, of the impact of the courses on their subsequent behaviour on three job related dimensions (Section 8.2.9). These dimensions were:

- (1) Work relationships
- (2) Personal work performance
- (3) Group or sectional performance

They recorded their ratings for each dimensions on a seven point scale. Mean responses were concentrated slightly below the middle on all three scales, indicating that the perceived effectiveness of the courses after a time lapse of twelve months had been maintained at a moderate level. Thus, although there is evidence of some gradual decline in the effectiveness of the courses over a period of twelve months following training, the impact of the science management training courses appears to have been more lasting at job performance level than previous research might suggest. At this point the focus of interest has shifted from specific behaviours to what Campbell et al. (1970) describe as "performance level outcomes".

As previously observed (Section 7.5) some of the topics studied were perceived by trainees to be more relevant to their work than others. Notably, Forecasting and Planning and Motivation were considered to be less relevant than other topics. The manner in which the topics were handled and the degree to which a deliberate effort was made to relate topic content to the trainees' own work context seem to be important factors here. However, it is interesting to observe that while most of the topics studied increased trainees' knowledge significantly, for many of them, the gains were accomplished over a longer period of time, up to at least three months following training. Reporting and Marking and Resource Allocation, both of which were considered to be most relevant and applicable initially, also achieved the greatest learning during the training period. In other words, their effects were more immediate, possibly *because* it was easier for trainees to see their direct applicability.

Assuming that the extent of the behavioural changes produced in the trainees is a reflection of the effectiveness of training, we can conclude that the course was moderately effective at least for a period of twelve months beyond training. Of the 15 training objectives, communication, improved interpersonal skills and the ability to organize sectional activities and projects were satisfactorily met while others

such as outside communication and report writing were relatively unaffected. In the case of interpersonal skills the type and extent of change was quite variable with large improvement occurring in particular instances as illustrated by the results of the individual Repertory Grids (Chapter 9).

Moreover, those topics which appeared to be better understood at the time were not necessarily those which effected the most significant longer term changes in behaviour. Motivational factors such as the trainees' stated intentions were strongly related to the subsequent implementation of such changes. These findings also have important implications for future training courses. Apart from an on-going need to re-evaluate and set new priorities for training goals as organizational and societal conditions change, the organizers of the training programme must pay attention to personal goal-setting which should promote both self and peer monitoring of performance. Personal goal-setting was further recommended because similar managerial-type goals may be achieved by a variety of different types of behaviour and conversely apparently similar behaviours may have different organizational effects depending on other factors operating within the situation. A modification of the training programme along these lines would promote what Williamson, et al. (1978) call goal-outcome congruence, since specific training goals could be more closely related to training effects. If a reasonable proportion of the training time during the second week of the course is to be set aside for personal goal setting, then it becomes clear that the total course content would have to be substantially reduced. It is the evaluator's opinion that far too much was attempted during the two week period. If more intensive training is to be undertaken in areas of interpersonal skills and leadership techniques as well as incorporating a goal-setting phase in the final stages, then the use of a departmental handbook to convey factual information becomes even more important. A further bonus to be derived from the introduction of a summary and goal-setting phase is the increased likelihood that training effects will transfer to the work setting.

Finally, at the intermediate level of evaluation one should examine the external validity of measurement to ensure that the results can be generalized to other groups of trainees, employees of

government and non-government departments, people with prior leadership experience and those with little or none. In this way, if certain groups appear to be under-served then those aspects of the programme that contribute to this deficiency can be examined. In the present case, the small sub-scale sizes frequently limited this activity, but, as the replication study (Chapter 10) has shown, some consistency of results was observed with other groups of trainees from different locations and at other times. At several points in this study it became clear that more thought should be given to the place of members of the non-government institutions (Section 8.2.1). The senior officers, in particular, from these organizations who attended the courses encountered little that was new to them. It is likely that a training programme aimed at higher levels of management would be more suitable for such people. On the other hand, their presence at the courses contributed to the overall success of the programme and helped to forge useful links between the different research organizations. It is worth noting that this type of detailed information was obtained mainly from discussions and interviews rather than from analysis of questionnaire responses. This illustrates the value of combining objective measures with a more subjective approach to evaluation.

11.7 ORGANIZATIONAL ENVIRONMENT

All of the changes in behaviour, knowledge and attitudes discussed so far must be considered against the background of the work environment which tends to constrain the type and extent of change which can occur. These environmental factors are the moderating variables operating within the total context and the relationship between them and the other dependent and independent variables must be studied if a complete theoretical model of training and evaluation is to be developed and tested. This is a long-term project and a simple study, even one of longitudinal design can only begin to move towards this goal. The writer believes, however, that this should be one of the aims of evaluative research and it can only be accomplished by longitudinal studies and the use of multiple measures of outcome and moderator variables including measures of organizational climate.

The global environment created by an organization is known as the organizational climate. In the present study, two separate approaches were adopted to test the effects of climate on behaviour and to broaden the range of variables on which the study was based (see Section 8.7).

A seven dimensional questionnaire derived from Litwin and Stringer's (1966) 32-item test was administered and used to check the dominant attitudes of the organizations as perceived by leaders, non-leaders, government and semi-government personnel. There was a high degree of variability between subjects within these groups but the profiles obtained from all four were relatively consistent (see Figures 8.4 and 8.5) with very high mean ratings on the dimensions of:

- (1) Responsibility - degree to which members feel that they can make decisions and solve problems without checking with supervisors
- (2) Rewards - degree to which members feel they are recognized and rewarded for good work
- (3) Warmth and support - the feeling that members trust one another and offer mutual support.

Slightly above average mean ratings were obtained for:

- (4) Standards - organization's emphasis on quality performance and outstanding production
- (5) Organizational clarity - degree to which members feel that things are well organized and goals are clearly defined
- (6) Leadership - degree to which members are willing to accept and assume leadership based on expertise.

A below average mean rating was recorded on:

- (7) Conformity - degree to which members feel they are regulated by set rules, procedures, policies and practices.

The same test applied to the following year's trainees corresponded closely to the original results (see Chapter 10).

Rundquist (1967) suggests that it is important to include not only such relatively stable factors as organizational climate but also situational factors like the effects of current economic conditions on the particular section or work group concerned. Although, in this instance, the high intra-group variability of the subjects tended to mask such group differences, further investigation may show that variations between different sections argue in favour of training at sectional or work group level rather than at cross-divisional and organizational level.

As a secondary means of looking at training outcomes within the broader context of the work setting, each trainee and control group member was invited to describe the reasons why his/her attempts to bring about work changes had failed. Eight of the 23 experimental group members and one of the 11 controls mentioned that they had met with obstacles which prevented them from achieving a desired change. The barriers most commonly cited were:

- (1) Organizational constraints, such as general resistance to change, lack of trust, staff restrictions, administratively enforced conformity
- (2) Personal contingencies of time and opportunity such as pressure of work, rapid changes within the section, unexpected, unplanned interruptions.

Therefore, it seems that organizational conformity, although perceived as relatively low from the results of the Organization Climate questionnaire, may, nevertheless, place certain limitations on innovative behaviour. Additionally, for section leaders, with substantial job-related responsibilities and work load, (characteristic of the unpredictability and stresses involved in any leadership position) the opportunity to implement new ideas is limited.

Environmental pressures of this nature may well account for the failure of a topic like Personnel Management to have an enduring impact on trainees for as Fineman and Warr (1971, p280) have insisted, training in human relations skills at immediate outcome level reveals either continued increase or decrease to original levels of functioning depending on leadership climate of the work group. This is another

reason why continued support and training once the trainees return home was recommended by the evaluator.

11.8

REACTIONS TO TRAINING

Since it can be argued on a theoretical basis that a well motivated trainee is more likely to be an apt learner, the present evaluator was interested in testing the reactions of trainees towards different aspects of the courses. This included both the immediate and the longer term reactions and an attempt was made to link these responses to precourse expectations about the training programme.

Reactions of trainees were explored by a series of measures both formal and informal at various stages throughout the programme. These data were used in the formative and corrective part of the evaluation but they also provide a good indication of how acceptable the programme was to participants. The general consensus of opinion among trainees was clearly a favourable one although specific criticism about certain aspects of the courses was not uncommon. The vast majority of trainees responded more favourably than either they or course organizers had anticipated. Results and conclusions drawn from the pertinent measures are summarized below.

1. Reactions to individual sessions (Section 7.4)

The only subjects tested on the Leipas Scale were members of the 1977 courses. Therefore, it is not known to what extent reactions were similar in 1978. However, 1977 results indicated that both Reporting and Marking and Resource Allocation were interesting and relevant while the material on Motivation was more difficult to relate to their work. They felt that the Forecasting and Planning topic was not at all relevant to them and Organization and Delegation as presented was rated rather poorly on all scales. While Personal and Group Relations was not highly structured, it was judged to be entertaining, informative, moderately interesting and relevant to their work. The Communication topic was criticised for its lack of structure and its failure to produce learning results. The Leipas scale ratings referred only to sessions conducted by topic supervisors

while the overall effects of the topics was also influenced by the quality of the syndicate group activities associated with them.

2. Preferred Training Techniques (Sections 7.6 and 10)

Results from 1977 were borne out by those obtained in 1978, (1977 and 1978 trainees' rank ordering of techniques correlated 0.83, significant at .05 level) and supported by analysis of control group results. They indicated clearly that the less formal methods, namely, syndicate group activities and informal discussions increased in popularity as a result of in-training experiences while the formal types of presentation such as instructional sessions and organized discussions which, prior to the courses were the most preferred methods, decreased in popularity.

3. Perceived Relevance of Topics (Sections 7.4, 7.5 and 10)

It was important to consider the reaction of trainees to the topics studied during the course; whether or not they seemed applicable to their everyday work. Quantitative measures were taken before and immediately after the courses and it was found that there was a tendency to perceive Resource Allocation and Reporting and Marking (Personnel Assessment) as more relevant following training than before. This was confirmed for both the previous and the 1978 course members. On the other hand, Forecasting and Planning, in both year 1977 and 1978 was seen as less relevant following training. When results were compared with the "Applicability" rating on the Leipas scale, it appears that a critical factor in determining this reaction was the success of the presentation of the particular topic either by the supervisor during formal instructional sessions or by the quality of the syndicate group activities when trainees engaged in intensive projects related to these topics.

4. Trainees' Reactions to Amount of Time Allowed for Different Topics (Sections 8.2.3 and 10)

About 50% of both years' trainees were satisfied with time allotment to various topics. The topics which produced the

greatest ambivalence were Forecasting and Planning (thought by many to need less time) and Personal and Group Relations (Personnel Management in 1978) believed by some to require more attention.

5. Expectations about Courses (Sections 7.2, 7.3, 8.2.2 and 10)

Participants were quizzed before attending the courses about their expectations. They gave their reasons for attending and what they expected to gain from the courses. It was found that trainees in 1977 and 1978 agreed that the two main reasons for participating were a desire to improve their personal work performance and because they were asked to attend by a controlling officer.

At organizational level, they expected courses to produce such results as better understanding between scientific and administrative staff, greater awareness of departmental policies and improved internal communication, rather than increases in the pool of section leaders or reduction of staff turnover. Thus they expected that the course would affect understanding and relationships within the organization rather than their conditions of employment.

As for their expectations concerning personal change, they anticipated improved managerial performance and human relations skills rather than changes in their attitudes or general job-satisfaction.

At the end of the course, trainees indicated that their expectations had been met but in addition they now rated social interactions between course members as an important outcome. These results were quite consistent over the years 1977 and 1978.

6. Behavioural Intentions (Sections 8.2.1 and 10)

A practical indication of the reaction to the courses might be gauged from the number of people who indicated that they intended to make positive changes in their work behaviour following training. For courses 4, 5, 6 and 7 these percentages were 75%, 87%, 83% and 95% respectively. In each case their

statement of intentions was included in the data only if they were able to specify what these changes were to be.

These evaluations of trainees reactions to the courses together with information gained about numbers of books read, contacts maintained and discussions held as well as internal comments made by trainees assist in the monitoring of the training course and indicate how they might be modified in the future to increase their effectiveness. For example, the discrepancies revealed between trainees' expectations and later reactions pointed to some possible weaknesses and omissions which could be rectified in future courses. In this course, there was no instruction given on routine administrative tasks such as writing official letters, chairing meetings and acting on committees nor about decision-making and problem-solving or communication with outside clients. By studying the data from reactions evaluations like this, it is possible to identify those sub-groups of trainees who need this sort of instruction. Moreover, it was clear that some adjustments should be made to the allocation of time to certain topics covered in the course as well as to the emphasis placed on topics.

The session assessment forms could be used more fully to provide the necessary feedback to trainees during the course, by presenting course members with average ratings together with their own individual ratings for purposes of discussion and comparison. Reactions to syndicate activities could be monitored by this method and allow the trainees to study the mechanisms of group processes and their own behaviour in a group situation. This method combined with a detailed behavioural analysis of these syndicate activities would help to fill an important gap in trainees' learning experiences. Where there are several sessions within a particular topic, the scores for separate sessions could be combined to give an overall assessment and rank ordering of the topics. Finally, as suggested earlier, particular sub-groups of trainees (non-leaders, certain organizational divisions or age groupings) who have special needs could be identified and given specialized treatments.

Much of this reaction data is also applicable to the level of evaluation designated as "input evaluation" which is concerned with the procedures, methods and techniques used to bring about the desired changes. Williamson, et al. (1978) refer to this as "mean-ends" analysis and it is concerned with the careful study of the elements of the training programme itself. This includes the training techniques employed and the content of the topics studied to ensure that they match closely with the identified programme goals for as Rundquist (1967) has pointed out, it is necessary to make the training content and the job content as nearly identical as possible.

The foregoing analysis at all levels of evaluation has led to the following conclusions about the particular methods and techniques used in this programme. It is recommended that course organizers should, in future, pay careful attention to the planning and integration of topics included in the programme. More emphasis should be placed on prior preparation of trainees and their orientation towards the goals of training. Increased efforts should be made to maintain interest and motivation particularly in the latter stages of the training courses by means of introducing new material and devoting time to summarization and personal goal-setting. Continued use should be made of the more informal and less structured methods of training with possibly the choice and composition of syndicate groups being delayed or changed during the training period to provide more interesting and varied peer feedback. Since the work environment determines to a large extent whether trainees carry out the desired changes in performance, ample opportunity should be given to discuss important aspects of the work setting and ways in which the individual can cope with it. Emphasis on personal and group relations should continue. Follow-up activities in the form of seminars, short courses and conferences should be arranged to consolidate newly acquired skills, attitudes and knowledge and to promote the transfer of these skills to the work setting. Follow-up work should be carried out with intact work teams such as complete sections, including members from different hierarchical levels, to enhance communication.

11.10

CONCLUSION

One of the clear messages which has emerged from this study has been the need for an ongoing endeavour to establish well-defined but flexible and dynamic goals of training. The writer believes that the present study has made a substantial beginning but much development still remains to be done and for this to be accomplished, more time and continued evaluation effort is required. Another is the need to provide for the consolidation of the learning acquired during training and the means of continuous self-monitoring of performance during the training period and beyond, for individual section leaders. McCall (1976) has shown that both follow-up training and feedback on performance are seriously lacking in the experience of most individuals who find themselves in a leadership position.

The reader will recall that the Wortman model was offered in Chapter 2 as an ideal model of evaluative research. The present study has attempted to apply this paradigm to a practical situation. It forms the basis of the experimental design, measurement of variables and testing of hypotheses. At some points, for practical and ethical reasons, the model has had to be very loosely interpreted and to be modified to meet the needs of the 'real life' situation. In this regard it is relevant to quote another recent researcher, C.T. Schreiber (1979) who comments:

"Perhaps to approximate the physical sciences, we have come to equate scientific control with control of the research process and the research environment. Certainly such control is more possible in a laboratory setting, for both the physical and the behavioural sciences. In a field or naturalistic setting control is a scarce commodity. There are many sources of potential noncontrol in field research; so many facets of the research procedure are subject to influences beyond the control of the researcher. Sometimes these environmental elements are labelled error variance and accepted. But often they seem evidence of nonrigor in our work".

(Schreiber, 1979, p148-149)

She goes on to say that the strong temptation is to brush over or fail to report noncontrol fearing that it will be equated with incompetence. Of late, however, senior behavioural scientists like Cronbach (1975) have encouraged researchers to observe and report variables in both controlled and uncontrolled contexts, variables including personal characteristics and spontaneous events which occur throughout the whole treatment and measurement process. Many of the factors contributing to noncontrol here have been discussed at some length in previous chapters and the writer has indicated where the weaknesses are due to factors beyond the control of the evaluator, a problem shared with other researchers in applied areas. Moreover, it is relevant that the position of student research associate which the

writer occupied did not carry with it a large measure of power in the sense discussed by Bonoma (1977) and there were times when she was acutely aware of this disparity within the power context. This state of affairs, while it does not excuse a failure to implement the best compromises available, nevertheless must influence the ability of the researcher to negotiate optimal conditions in a field experiment. For example, under other circumstances it may have been possible to call a moratorium on the programme for a period long enough to fully investigate the goals of training. A full scale analysis of training needs would have required a much more substantial time commitment on the part of the organizations prior to beginning the main evaluation study. Secondly, a great deal more could have been accomplished in terms of randomization of subjects to experimental and control groups and alternative treatments had the researcher not been obliged to abide by earlier decisions made at organizational level. To resolve matters of this nature requires a high level of collaboration and a more complete commitment to an "action research" type of approach by the participating organizations, but it would have allowed some formative evaluation to be accomplished as a preliminary step. This could have been followed up by a more tightly controlled summative evaluation than was possible in this instance.

What the study does show, however, is the advantage of having a model of evaluation and a research design which permits one to identify causal relationships between variables whether they be changes in personal attributes, environmental conditions or events. The present study has endeavoured to include some of these variables and by asking probing questions of the subjects to establish the linkages between them over an extended period of time. Much of this activity could be classified as construct validation of the interdependent variables in accordance with the integrated criterion model proposed by James (1973). James' model is based on the general criterion model of managerial effectiveness described by Campbell, et al. (1970). In this model the training and development experiences depicted in the middle portion of the diagram (Fig 12.1) are part of the organizational situation which intervenes between personal attributes of the individual and organizational outcomes. The job behaviour, job performance and organizational outcome measures represent the three

levels of criterion measurement. The integration occurs between the general and the multiple criterion models for James stipulates that rather than assuming a single, underlying general factor for all facets of job performance, multiple criteria of performance must be obtained on the assumption that many job performance measures are factorially independent of one another. Thus it is necessary to collect multiple measures of job performance including behaviourally based job performance ratings from different raters, objective measures of job performance, situational data and global measures of performance, to determine the independent dimensions underlying the criteria.

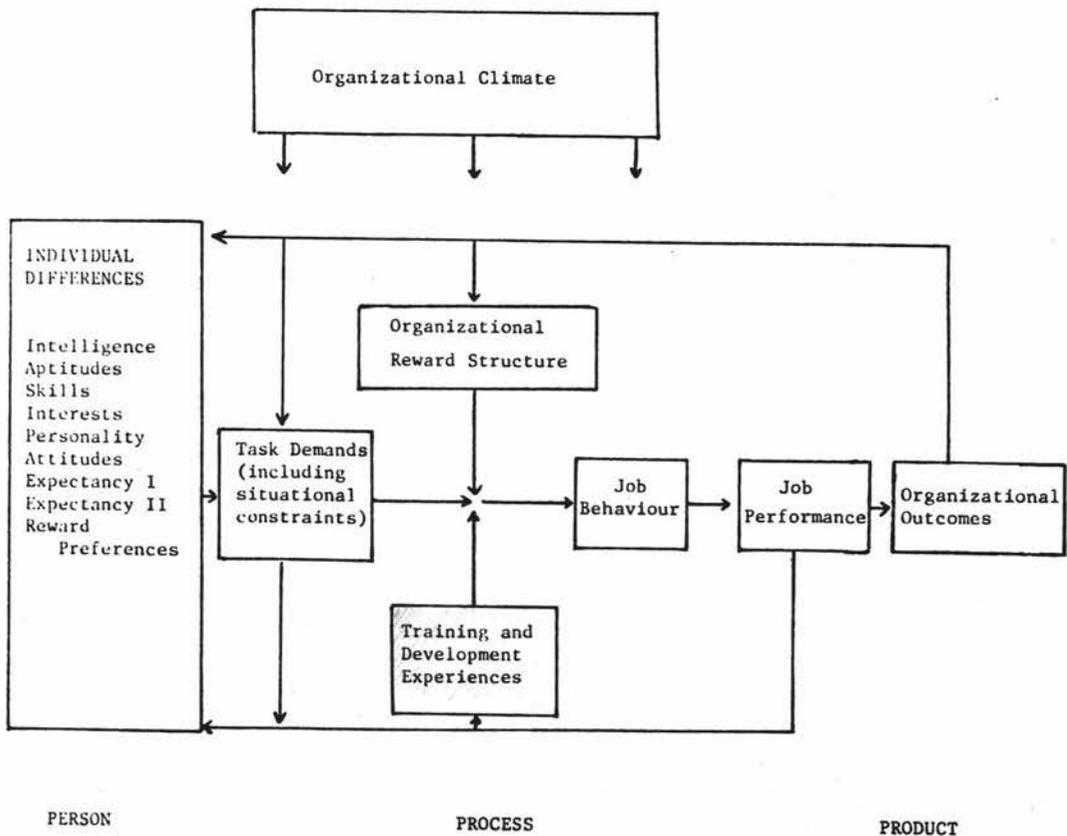


Figure 12.1 General criterion model of managerial effectiveness (Campbell, Dunnette, Lawler and Weick, 1970).

It is to be noted that the methods used in the present study to gather contextual information were based on Smith and Kendall's (1963) behavioural expectation scaling technique (Section 5.2.3) which seeks to determine both the underlying performance dimensions and the identifiable job behaviours that are related to organizational outcomes. Thus there is a degree of correspondence between levels of evaluation and data collection methods.

The integrated criterion model emphasizes the dynamic nature of managerial criteria. As James says, one would expect that task demands, expectancies, interests and behaviours would change over time as a function of both the feedback loop in the model and the changing environment (Fig 12.1). This provides a strong argument for longitudinal investigations of change rather than static studies.

A successful training evaluation study should also establish an effective feedback loop between ultimate goals as they are defined by the organization and the various sources of training input; trainers, trainees, training techniques and whatever resources and facilities are involved. Here, the onus is on the researcher to establish multiple measures of the dependent and independent variables for, in order to ensure construct validity, one must know that the relationships assumed between the theoretical constructs underlying the training-evaluation model are valid. For example, it is not sufficient to show that certain topics like Resource Allocation and Reporting and Marking produce large and significant immediate learning effects and further assume that they will produce ultimately the greatest behavioural change. It is also necessary to understand the subjects' motivation to learn, the immediate acceptability or relevance of those particular topics, the methods used to present them to the trainees and even the type of interaction that has taken place between trainees and trainers. Moreover, in considering the motivational effects of the trainees, one must be aware of the bias produced within the experimental subjects by their own expectations as well as those generated by the course organizers and by the evaluator. Riecken (1977) refers to this as ecological validity. These in themselves may produce differences between the experimental group and the controls quite apart from any direct effects of the training programme itself. This type of validity is closely

associated with what Campbell and Stanley (1966) have labelled internal validity. Thus, in this writer's opinion, it is not possible to make definitive statements about the effectiveness of a training programme unless at least one control group is included in the experimental design. The proper choice and use of control groups is probably the major difficulty encountered in applied, naturalistic research and the conclusions drawn are inevitably weakened to the extent that such control is improperly implemented. This is, indeed, the reason why people like Riecken and Boruch argue so strongly for randomization of both subjects and treatments but the fact remains that true randomization is an ideal which is rarely achieved in the practical situation. The best compromise that could be achieved in the present study was the use of both future and previous trainees who were matched on a number of important variables and who provided some valid basis for comparison with the main experimental groups.

Attempts were made to eliminate motivational discrepancies between experimentals and controls by providing similar pre-test treatments for both groups and by choosing control group members from a master list of previous or future trainees. It was reasonable to assume that such controls would have a similar orientation towards the management training programme.

To ensure both good construct validity and "conclusion" validity which is defined by Wortman as a subset of internal validity, the experimenter has emphasized the need to refine and validate measures of the dependent and independent variables.

The validation of dependent variable measures using a construct validation approach (Cronbach and Meehl, 1955) requires that formal rules of correspondence, preferably mathematical, be established between perceptual or observable variables (measures of behaviour, performance and organizational effectiveness), between observables and theoretical constructs and between constructs and other constructs. James (1973) claims that it is only by determining the complex relationships between different levels of variables that the constructs became meaningful and have useful explanatory

power which will lead to an empirically and scientifically substantiated theory of managerial effectiveness.

It is for this reason that the present writer believes that the ongoing development, refinement and validation of the tools of measurement is the second major requirement of good evaluative research. This involves both multiple operationalization and measurement of theoretical constructs and greater sophistication of the measuring techniques.

Conclusion validity is more concerned with the effects of small sample size, differences in the administration of training techniques, unreliable measuring instruments and inappropriate tests. Methods of observing and recording psychological data lag far behind the availability of the techniques of statistical analysis. On the other hand, it is not reasonable to suggest that we confine ourselves to totally objective, easily quantifiable data which is amenable to statistical analysis because the concepts involved are complex and inevitably subjective. This gap between the two, imperfect largely subjective data, on the one hand, and sophisticated statistical procedures on the other, may never be completely bridged, but more attention must be paid to the degree of correspondence at this interface. There is evidence that subjective data can be utilized and more flexible statistical techniques are now available which can cope with psychological data. Meanwhile, one must take care not to misuse relatively weak data. The present writer has favoured non-parametric approaches to analysis and even confined herself to descriptive statistics on a number of occasions in recognition of the small changes likely to occur in a training session of this nature and of the relatively few subjects which she had to work with. It is incorrect to use statistical models which require very strong assumptions about the underlying distribution of the data when this would mean treating subjective judgements as quantitative data. This is so whether we choose to employ decision-oriented or null hypothesis testing approaches. It does not, however, prevent the observation of important trends which can lead to some tentative conclusions which will, in turn, prompt further testing and verification. The writer is of the opinion that evaluation must be a multi-stage procedure. Wortman (1975) observes that statistical inference is a necessary prerequisite for causal inference and there are new developments afoot

within statistics which are providing means of studying multivariate problems of the type which we have encountered here. The psychologist for his or her part must continue to perfect judgemental and observational techniques for the measurement of both input and outcome (criterion) variables in the training situation.

The fourth type of validity included in the Wortman model is external validity and to this end the present study has utilized two separate experimental groups in the first year and two more in the second so that all four results could be compared for reliability and replicability. Once again, it was not possible to adopt random sampling methods. Therefore, we must be cautious in drawing conclusions from these comparisons, but the degree of similarity in the findings does suggest that the results obtained in the original experiment could be extended to other similar populations. The writer is further convinced that such replicability is another important feature of evaluation research and unless such evidence of consistency in results can be demonstrated then the conclusions to be drawn from any evaluation study may be seriously questioned.

In order to conduct a fully satisfactory summative evaluation, it is necessary to extend the study over a much longer period than the present one and to use distal criteria of training effectiveness which are closer to organizational goals both in terms of time and quality. Examples of such goals include quality and quantity of scientific research produced by the section leaders and their work groups, but as we have already discussed, this is by no means an easy task and, indeed, may not be feasible in this case. For one thing, long-term research output is, generally speaking nowadays, a group activity rather than an individual one and the measurement of such criteria may be an impossible task. Secondly, research output is largely under the control of economic and political factors such as the availability of funds and bureaucratic priorities. There are other intermediate goals of training whose achievement may be quite closely related to organizational goals such as staff turnover and job satisfaction but, once again, it is very difficult to obtain valid indicators of some of these factors. Moreover at the time of the present evaluation study there appeared to be little consensus within

and between different organizational levels about the type of job performance that was related to long-term organizational goals. This may indeed, be the usual state of affairs in research organizations of this kind where in principle, at least, freedom and creativity are highly valued and the imposition of external goals is strongly resisted.

This means that a major emphasis of the evaluation must focus on what Wortman terms formative evaluation. The present evaluator has taken some trouble to ensure that a large measure of formative evaluation has occurred in this study. By the use of session assessment forms and by tapping the reactions of the trainees at various points over the evaluation period, she has endeavoured to monitor the effectiveness of the topics and of the training techniques used. This information has been fed back to course organizers to aid in the ongoing development of the training courses. Besides this, short term assessments of goal oriented effectiveness were passed on to the organizers (but not to the subjects) by means of written reports and discussions between the evaluator and training organizers at regular intervals over the two year period. This did provide considerable evaluative feed-back in the formative sense and allowed decisions to be made as they were needed along the way. The writer is aware that this does not really constitute a complete 'action research' approach for the important ingredients of continuous change and monitoring of such changes was not possible on a long term basis. However, this method was a reasonable compromise to make in the circumstances. Moreover, it did help to fulfil the political and ethical requirements concerning the need for open discussion and collaboration between all parties involved in the evaluation, by providing a forum where doubts and uncertainties concerning the evaluation study could be raised and explanations and assurances given. It would be unrealistic to suggest that this always worked out as smoothly as the evaluator would have wished but for the most part it served the purpose well, despite the difficulties and reservations mentioned at the beginning of this chapter. It is the writer's firm belief, based on this experience, that such cooperation is vital to the success of any evaluation study and, indeed, probably to most applied research.

Having argued in favour of Wortman's ideal model of evaluative research, the writer acknowledges that evaluation is ultimately a

decision-making exercise and as such takes place within a particular organizational context. Writers such as Wortman and Guttentag would agree that finally a decision must be made (even if it be the decision to ignore the evaluation results) and this is usually the prerogative of people other than the evaluator. However, if the applied researcher is to have a constructive role in this decision-making and implementation phase, she is required to meet not only the demands of science and ethics but also the political demands and economic realities of the situation. Wortman, for his part, sees the feedback loops of his process model providing the mechanisms whereby the necessary information is passed between evaluator, training administrators and policy-makers and leaves it at that. Guttentag approaches the issue more directly and conceptualizes evaluation as primarily a decision-making problem.

A third option is to adopt a model which attempts to combine scientific hypothesis-testing with a decision-making approach. Super (1976) has offered a career decision-making model which seems to be applicable to evaluation research. In his life stage model of career decision making, he combines the concept of individual vocational development and maturation with the need to make not one but a whole series of career decisions in the course of a lifetime. The intermediate decision points form the basis of the new, updated information on which subsequent decisions are made. This process involving sequential decisions of an exploratory nature forms what Super calls a developmental model of emergent career decision-making. As an individual approaches an impending career decision, he formulates the question, reviews his premises, identifies the facts of the situation, seeks new data, evaluates and weighs the old against the new data and identifies lines of action. He then considers their possible outcomes and their respective utilities and weighs the alternatives in terms of his values and objectives. Finally, he selects the preferred plan of action, stores the alternatives for future reference, and pursues his plan on either an explanatory basis or with a more definite but still tentative commitment. In any case, more data collection must take place by means of the evaluation of outcomes, with modification of plans or with recycling of decisions.

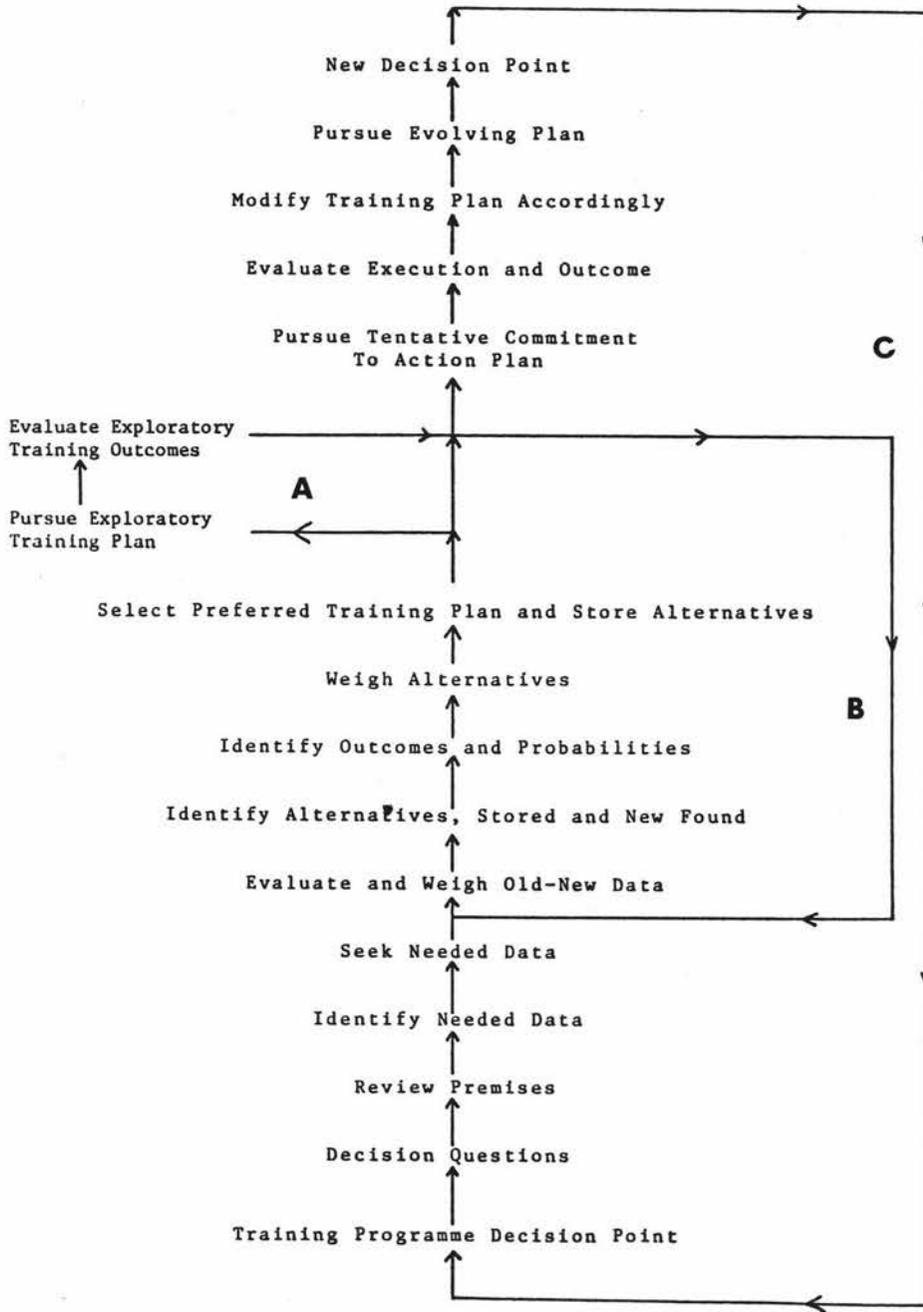


Figure 12.2 Decision-making model of evaluation
(Adapted from Super, 1976)

This model of evaluation which is illustrated in Figure 12.2 appears to parallel the process of combined formative and summative evaluation because it requires on-going monitoring of the existing programme as well as fulfilling the need to reach intermediate decisions and make appropriate changes in the training programme. It provides in Super's terms, a "mini-loop" in which explanatory adjustments and evaluations can occur within the total "maxi-cycle" of major and more permanent changes needed.

In the centre of the diagram (Fig 12.2) the explanatory mini-cycle (A) on the left and the recycled decision (B) on the right correspond to formative evaluation while the maxi-cycle (C) leading to a new decision point at the top of the diagram represents the outcome of summative type evaluation. The point is, that both are proceeding simultaneously. New information is constantly being collected and tested scientifically and used for either intermediate or terminal decisions. Thus, this adaptation of Super's model provides the basis for a training-evaluation model, regardless of whether the evaluator has a large or a small part to play in organizational problem-solving and decision making. The degree of involvement depends on the extent of the evaluator's participation in the political processes of the organization and the commitment that has been made to an action-research type of approach to the evaluation of training.

Finally, what can be said about the future of evaluation research in the light of the present investigation? The key to questions concerning the direction that evaluation research is likely to take lies in the concept which we, as psychologists, have of scientific investigation. Particularly in the area of evaluation research where understanding is equally as important as prediction and control, we must be prepared to acknowledge the complexity of the laws which govern human behaviour. In the process, we must avoid a very narrow and rigid view of applied science. At the same time, as social scientists, we recognize that the basic purpose of evaluation is to "measure" and to discover "causes". For this latter reason we will probably continue to adopt a strategy that Glass and Ellett (1980) attribute to Cronbach, namely,

"tend towards the rigorous end of the naturalistic-experimental continuum".

That must not prevent us from seeing evaluation from a much broader perspective than we have in the past. Methodologically this means adopting greater flexibility of experimental design and statistical analysis. Conceptually evaluation must be seen in relation to its practical purpose, that of assisting in the decision-making processes of our clients. However, we must recognize that evaluation and decision-making are conceptually distinct even though it may be appropriate for the psychologist, as evaluator, to participate in both phases. Indeed, whether we like it or not, both are likely to be occurring simultaneously. It is up to the psychologist to maintain the logical separation between the two. Thus we need to continue to develop our ideas concerning concurrent formative and summative evaluation. One way of achieving this, as suggested by the present research, is to learn to work within an action research framework which can form the rapprochement between the rigorous but inflexible classical experimental approach and a newer, more adaptable but equally scientific mode of investigation. In this way the advantages of the scientific method can be brought to bear upon the field of applied psychology known as evaluation research.

APPENDIX I

RESULTS OF SESSION ASSESSMENT FORMS OF SECTION 7.4

Table Ia

Number of responses, mean score and range on each criterion

Title of Session	Learning			Entertainment			Interest			Participation			Application			Structure		
	n	Mean	Range	n	Mean	Range	n	Mean	Range	n	Mean	Range	n	Mean	Range	n	Mean	Range
Organization and Delegation - Session 1	23	2.87	1-4	23	2.67	1-4	23	3.43	2-4	23	3.09	1-5	23	3.09	0-5	23	3.43	1-4
Organization and Delegation - Session 2	19	1.26	0-3	19	1.74	0-4	18	2.89	0-4	19	2.63	0-5	19	2.47	0-4	19	1.58	0-4
Forecasting and Planning - Session 1	23	3.17	1-4	23	3.87	2-5	23	3.13	1-5	23	3.17	1-5	23	2.74	1-5	23	2.96	1-4
Forecasting and Planning - Session 2	21	3.43	2-5	21	2.90	2-5	21	3.81	3-5	21	3.10	1-5	21	3.38	1-5	21	3.43	2-5
Forecasting and Planning - Session 3	22	2.91	1-4	22	3.64	2-5	21	3.67	2-5	22	3.73	2-5	22	3.09	2-5	22	3.14	2-4
Leadership	22	3.05	1-5	22	4.09	2-5	22	4.00	2-5	22	3.27	0-5	22	3.77	3-5	22	3.50	1-5
Motivation	19	3.32	1-5	19	3.95	3-5	19	3.89	2-5	19	3.26	1-5	19	3.37	2-5	19	3.53	2-5
Interviewing	23	2.87	0-5	23	4.00	0-5	23	3.74	2-5	23	3.48	1-5	23	3.96	3-5	23	3.52	1-5
Personal and Group Relations	22	3.27	0-4	22	4.41	3-5	22	3.77	2-5	22	4.68	3-5	22	3.45	2-5	22	2.77	1-4
Resource Allocation	20	2.55	0-4	20	3.35	2-4	20	3.45	1-5	20	3.95	2-5	20	3.50	0-5	20	2.40	1-4
Reporting and Marking - Session 1	23	3.17	1-4	23	2.74	1-4	23	3.61	2-5	23	2.70	1-5	23	3.61	1-5	23	3.57	2-5
Reporting and Marking - Session 2	21	3.00	1-5	21	2.86	1-4	21	3.43	1-5	21	3.38	1-5	21	3.67	1-5	21	3.38	2-5
Reporting and Marking - Session 3	22	4.09	2-5	22	3.18	1-5	22	3.82	1-5	22	3.36	2-5	22	3.91	2-5	22	3.82	2-5
Communication - Session 1	22	2.68	0-4	22	3.05	1-5	22	3.45	2-5	22	3.00	2-5	22	3.45	1-5	22	2.82	1-4
Communication - Session 2	21	2.67	1-4	21	3.43	2-5	21	2.43	1-5	21	4.14	3-5	21	3.29	1-5	21	2.86	1-4

Notes (a) Best possible rating = 5, worst possible rating = 0

(b) Maximum possible number of subjects in any cell, n = 23.

Appendix I continued

Table Ib

Rank-ordering of session assessments on each criterion.

Title of Session	Learning	Entertainment	Interest	Participation	Application	Structure
Organization and Delegation - Session 1	10=	14	11=	12=	12=	6=
Organization and Delegation - Session 2	15	15	14	15	15	15
Forecasting and Planning - Session 1	5=	5	13	10	14	10
Forecasting and Planning - Session 2	2	11	4	11	9	6=
Forecasting and Planning - Session 3	9	6	7	4	12=	9
Leadership	7	2	1	8	3	5
Motivation	3	4	2	9	10	3
Interviewing	10=	3	6	5	1	4
Personal and Group Relationships	4	1	5	1	7=	13
Resource Allocation	14	8	9=	3	6	14
Reporting and Marking - Session 1	5=	13	8	14	5	2
Reporting and Marking - Session 2	8	12	11=	6	4	8
Reporting and Marking - Session 3	1	9	3	7	2	1
Communication - Session 1	12	10	9=	13	7=	12
Communication - Session 2	13	7	15	2	11	11

Note: 1 is most favourable, 15 is least favourable

APPENDIX II

ANALYSIS OF REPERTORY GRID DATA OF
CHAPTER 9

CASE STUDY A

Similarity Matrices

1	----										
2	49.0	----									
3	30.0	34.3	----								
4	66.7	59.0	33.3	----							
5	62.2	70.2	47.8	85.5	----						
6	62.5	68.5	44.2	79.2	80.3	----					
7	83.0	62.7	27.0	69.7	72.5	69.5	----				
8	55.0	90.7	31.7	65.0	72.2	62.5	64.7	----			
9	75.0	50.7	35.0	91.7	77.2	70.8	74.7	56.7	----		
10	83.3	49.0	36.7	83.3	75.5	69.2	73.0	55.0	91.7	----	
	1	2	3	4	5	6	7	8	9	10	

Elements

1	----										
2	69.7	----									
3	62.2	68.5	----								
4	41.0	54.0	55.5	----							
5	50.0	63.0	50.5	57.7	----						
6	82.7	52.3	55.5	45.0	32.7	----					
7	54.3	71.3	71.5	52.0	71.0	37.0	----				
8	84.7	78.3	73.5	52.3	44.7	74.0	63.0	----			
9	72.3	69.3	61.2	30.0	39.0	58.3	68.0	71.0	----		
10	91.7	71.3	70.5	49.3	51.7	81.0	56.0	93.0	64.0	----	
	1	2	3	4	5	6	7	8	9	10	

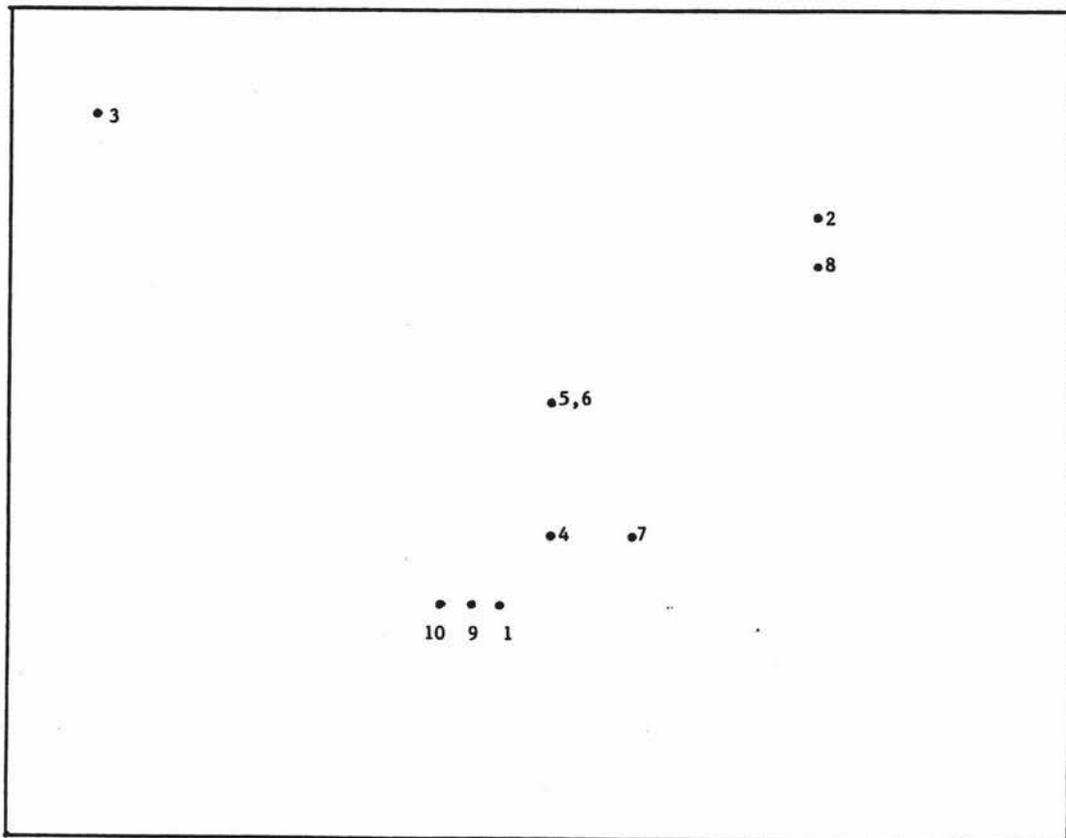
Constructs

Appendix II continued

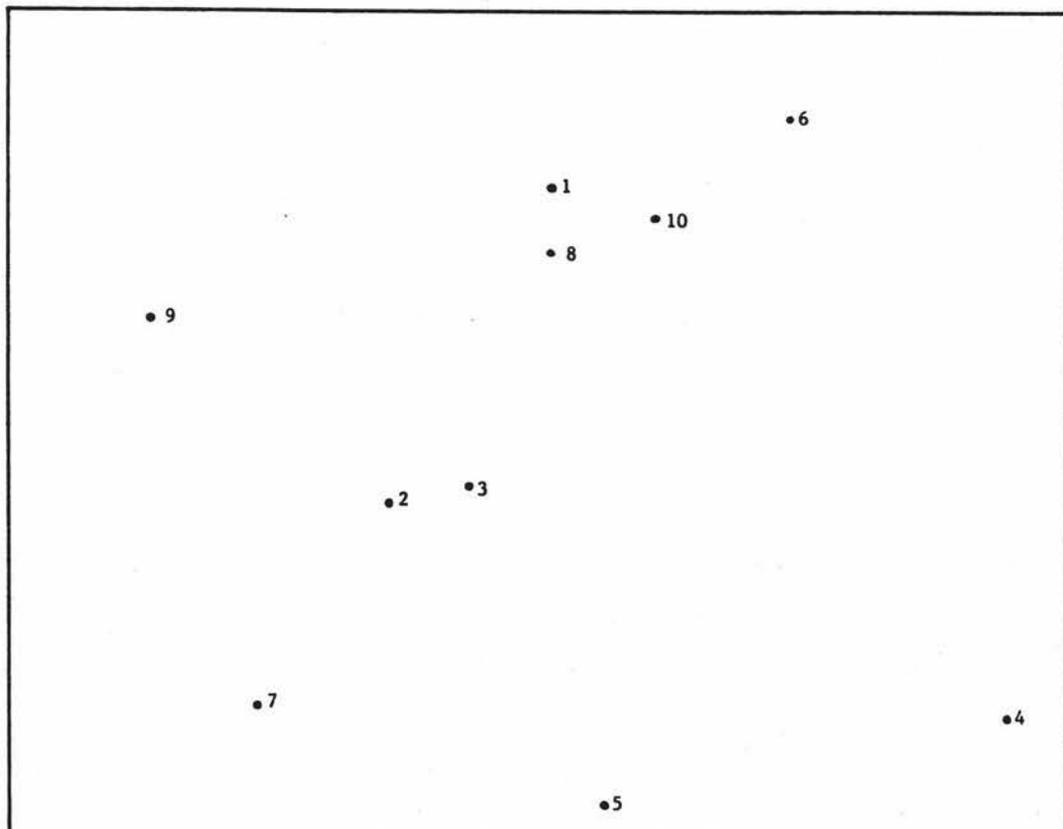
CASE STUDY A (continued)

Map of points from Repertory Grid

ELEMENTS



CONSTRUCTS



Appendix II continued

CASE STUDY D

Similarity Matrices

1	----										
2	65.2	----									
3	63.3	50.2	----								
4	71.7	43.5	56.3	----							
5	47.7	16.2	40.3	72.7	----						
6	70.0	51.8	50.7	87.7	64.4	----					
7	63.0	48.2	43.7	87.3	68.0	93.0	----				
8	72.3	60.8	53.0	79.3	55.3	84.3	84.0	----			
9	47.0	17.2	53.0	68.7	80.7	56.3	56.0	51.3	----		
10	66.3	54.8	47.0	80.7	61.3	89.7	93.3	87.3	49.3	----	
1	----										
2	48.2	----									
3	25.7	59.5	----								
4	36.3	54.8	69.3	----							
5	15.2	53.7	66.5	78.8	----						
6	58.2	82.0	63.5	56.8	49.0	----					
7	72.3	70.8	49.3	54.0	36.2	70.2	----				
8	29.8	64.3	79.8	81.5	78.7	59.7	57.5	----			
9	58.5	81.7	67.2	61.2	53.3	88.3	73.8	67.3	----		
10	57.3	69.8	68.3	65.0	47.2	77.2	77.0	68.5	80.8	----	

Elements

Constructs

CASE STUDY E

Similarity Matrices

1	----										
2	33.0	----									
3	51.7	64.7	----								
4	58.3	48.7	66.7	----							
5	80.0	33.0	58.3	58.3	----						
6	80.0	29.7	55.0	61.7	80.0	----					
7	88.3	21.3	46.7	50.0	88.3	88.3	----				
8	23.0	86.7	68.0	45.3	26.3	23.0	14.7	----			
9	23.3	73.7	55.0	58.3	23.3	20.0	11.7	77.0	----		
10	74.7	38.3	59.7	70.3	81.3	88.0	79.7	31.7	28.7	----	
1	----										
2	25.7	----									
3	29.7	82.0	----								
4	33.8	73.5	76.8	----							
5	83.3	25.7	29.7	33.8	----						
6	75.2	33.8	33.2	41.3	76.8	----					
7	43.5	55.8	70.5	74.0	33.5	42.0	----				
8	36.7	82.3	89.7	80.5	36.7	40.2	63.5	----			
9	22.2	88.5	88.5	85.0	22.2	29.7	62.3	85.5	----		
10	30.2	77.2	76.5	83.0	30.2	41.7	57.0	76.8	84.7	----	
	1	2	3	4	5	6	7	8	9	10	

Appendix II continued

CASE STUDY F

Similarity Matrices

1	----										
2	23.7	----									
3	88.8	30.8	----								
4	80.5	43.2	87.7	----							
5	47.5	52.8	54.7	67.0	----						
6	57.3	59.7	64.5	76.8	85.2	----					
7	69.7	50.0	80.8	85.2	68.8	78.7	----				
8	34.0	69.7	41.2	53.5	52.5	63.3	53.7	----			
9	58.7	61.0	69.8	78.2	79.8	90.7	84.0	64.7	----		
10	49.5	56.8	60.7	69.0	84.0	88.2	79.8	60.5	90.8	----	
	1	2	3	4	5	6	7	8	9	10	

Elements

1	----										
2	88.8	----									
3	83.5	84.7	----								
4	28.2	35.3	26.7	----							
5	31.8	34.0	30.3	73.0	----						
6	80.0	87.2	78.5	48.2	36.8	----					
7	75.2	76.3	91.7	23.3	27.0	75.2	----				
8	95.8	89.7	84.3	29.0	32.7	80.8	76.0	----			
9	68.8	80.0	71.3	55.3	39.0	88.8	68.0	69.7	----		
10	71.3	82.5	73.8	52.8	41.5	91.3	70.5	72.2	90.8	----	
	1	2	3	4	5	6	7	8	9	10	

Constructs

APPENDIX III

PRECOURSE INTERVIEWS AND QUESTIONNAIRES

A Interview schedule for trainees and controls

Name of scientist

Group

Division

Position title

Job description :

What would you like to be able to do in a managerial sense as a result of this course? (adapted for groups 2 and 3)

B Interview schedule for controlling officers

Division

Director

What changes would you like to see in a managerial sense as a result of these courses?

What changes would you like to see in specific staff members?
(adapted for groups 2 and 3)

Appendix III continued

C Biographical questionnaire

Name:

Age:

Sex: (M or F)

Title: (Section leader, etc.)

Division: (or Research Assoc.)

Experience in D.S.I.R. years months

Experience in Leadership role (above) years months

No. of Staff in Section (or Group) responsible to you:

Scientists

Technicians

Clerical

Appendix III continued

D Precourse questionnaire

As participants in the Science Management training program you may have certain expectations about the course which you will attend

(a) What are your reasons for attending the course? Place the appropriate number beside each of the following reasons given, from 1 Most Applicable to me to 6 Least applicable to me.

- To make social/professional contacts
- To improve present job performance
- Personal interest/curiosity and general interest in the topics
- To enhance prospects for promotion
- I was given little choice in the matter
- Other:

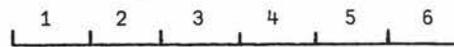
(b) What do you expect to get out of the subject matter discussed? Rate each topic according to the following scales:

How useful (relevant) do you think this topic will be for your work?



Minimally useful in practice Highly relevant to daily work

How do you rate your present understanding of this topic?



Vague understanding only Excellent understanding

Topic 1	<input type="checkbox"/>	Topic 1	<input type="checkbox"/>
Topic 2	<input type="checkbox"/>	Topic 2	<input type="checkbox"/>
Topic 3	<input type="checkbox"/>	Topic 3	<input type="checkbox"/>
Topic 4	<input type="checkbox"/>	Topic 4	<input type="checkbox"/>
Topic 5	<input type="checkbox"/>	Topic 5	<input type="checkbox"/>
Topic 6	<input type="checkbox"/>	Topic 6	<input type="checkbox"/>

Appendix III continued

D Precourse questionnaire (continued)

- (c) What do you expect to get out of the course in each of the two areas below.
Please tick none or more items.

Areas of organisational change:

- Internal communication
- Staff turnover
- Increase in pool of potential
Section Heads
- Awareness and acceptance of
departmental policies
- Understanding between scientific
staff and administrative staff
- Other:

Areas of personal change:

- Job satisfaction
- Attitude to change
- Quality of managerial performance
- Human relations skills
- Social interaction between
course members
- Other:

- (d) Which methods or instruction do you prefer for a course in management?

Please tick none or more items.

- Lectures
- Discussions with other course members (informal)
- Syndicate Group activities
- Organised discussions
- Practical activities
- Discussions with course leaders (informal)
- Other:

Here are some statements about which we would like your opinion. Please read each one and decide whether you agree or disagree with it. Place a tick in the appropriate column according to your decision. If you really cannot make up your mind about any particular statement, please tick in the third column marked 'don't know'.

* (scores assigned to responses, giving a total of 40 points).

	Agree	Disagree	Don't Know
A group or section leader should be able to discuss any problems that arise, freely, with his controlling officers	4*		
Part of the leader's job is to take every opportunity to keep other group members informed of any Departmental policy changes ..	3		
Individual scientists should have complete freedom to choose and direct their own research activities		1	
Communication from the top down to section or group level is very poor		3	
It is essential for the group to devise a standard method for submitting requests to their leader for new equipment and for deciding upon priorities		3	
The leader's main job is the maintenance of scientific programmes within the section including generating ideas and assessing equipment needs to expedite this end	1		
There are times when the lack of structure in groups leads to unnecessary confusion and disorganization	0	0	
There is little that can be done about unfinished projects and unused data		4	
Planning ahead is made particularly difficult by lack of prompt response from higher authority		1	
The leader must act as the representative of his work group to higher authority (e.g. the Director) and vice versa	2		
Group members should feel free to approach the leader (or each other) to discuss various aspects of their work	4		
A section leader (or equivalent) is required to spend too much time on the routine aspects of the job		2	
A work group should be structured quite loosely so that individuals may pursue their scientific activities unhampered ..	0	0	
Regular meetings with other group members are necessary for the smooth running of the section (or group)	1		
The leader has the opportunity of providing the sort of environment which will lead to increased interest and work satisfaction among his staff	3		
The leader must be able to anticipate new demands and needs in the community and be constantly on the alert for information and ideas which may lead to new research topics for his group.	2		
Since the management of people takes up only a small part of a leader's time and interest there is little point in devoting much time to learning management principles		2	
One of the most time-consuming and frustrating aspects of the job of section or group leader is feeling obliged to help people with personal problems which are affecting their on-		4	

APPENDIX IV
SESSION ASSESSMENT FORM

This scale is to be used to evaluate individual sessions throughout the day.

Session Number:

Please mark YOUR personal rating on a scale of

0 = very low to 5 = very high

Your rating should indicate how you feel about the individual criteria, i.e., did you learn a little or a lot of NEW INFORMATION, was it boring or entertaining (stimulating), are you now only slightly or very interested in the topic(s), was there little or plenty of OPPORTUNITY TO PARTICIPATE ACTIVELY, will you be able to APPLY what you learned in your job and did you find the structure (pattern of presentation) logical and easily followed - or were you confused?

CRITERIA

Learning

Entertaining

Interest in topic(s)

Participation opportunities

Application to job

Structure (logical development)

	0	1	2	3	4	5
Learning						
Entertaining						
Interest in topic(s)						
Participation opportunities						
Application to job						
Structure (logical development)						

Mnemonic - LEIPAS

WRD 7/74

APPENDIX V

IMMEDIATE POST-COURSE QUESTIONNAIRES

We would like to know your immediate reactions to some aspects of the course in which you have participated. Please read the following items and express your opinions in the manner indicated.

(a) Which of the methods of instruction used during this course did you prefer?

Please tick none or more items.

- Lectures
- Discussions with other course members (informal)
- Syndicate Group activities
- Organised discussions
- Practical activities
- Discussions with course leaders (informal)

(b) What additional methods would you like to see used?

.....

.....

.....

(c) What did you get out of the subject matter discussed? Rate each topic according to the following scales:

How useful (relevant) was this topic to you in your work?

1	2	3	4	5	6
Minimally useful in practice			Highly relevant to daily work		

How do you rate your present understanding of this topic?

1	2	3	4	5	6
Vague understanding only			Excellent understanding		

Topic 1	<input type="checkbox"/>
Topic 2	<input type="checkbox"/>
Topic 3	<input type="checkbox"/>
Topic 4	<input type="checkbox"/>
Topic 5	<input type="checkbox"/>
Topic 6	<input type="checkbox"/>

Topic 1	<input type="checkbox"/>
Topic 2	<input type="checkbox"/>
Topic 3	<input type="checkbox"/>
Topic 4	<input type="checkbox"/>
Topic 5	<input type="checkbox"/>
Topic 6	<input type="checkbox"/>

Appendix V continued

(d) What is your opinion of the time allocation to various topics?

Tick one for each topic.

	More time should be spent on this topic	Time spent on this topic was about right	Less time should be spent on this topic
Topic 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Topic 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Topic 3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Topic 4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Topic 5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Topic 6	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

(e) Write a paragraph or two describing your feelings and impressions about the course at this moment.

(f) Do you plan to make any changes as a result of ideas gained from this course?

Yes/No (Cross out one)

If "Yes", what are the changes and from what ideas.

(g) Attitude scale (as in Appendix III).

Appendix V continued

Organization Climate Questionnaire

Introduction

For each of the seven organization climate dimensions described below place an (A) above the number that indicates your assessment of the organization's current position on that dimension and an (I) above the number that indicates your choice of where the organization should ideally be on this dimension.

1. *Conformity*. The feeling that there are many externally imposed constraints in the organization; the degree to which members feel that there are many rules, procedures, policies, and practices to which they have to conform rather than being able to do their work as they see fit.

Conformity is not characteristic of this organization 1 2 3 4 5 6 7 8 9 10 Conformity is very characteristic of this organization

2. *Responsibility*. Members of the organization are given personal responsibility to achieve their part of the organization's goals, the degree to which members feel that they can make decisions and solve problems without checking with superiors each step of the way.

No responsibility is given in the organization 1 2 3 4 5 6 7 8 9 10 There is a great emphasis on personal responsibility in the organization

3. *Standards*. The emphasis the organization places on quality performance and outstanding production including the degree to which the member feels the organization is setting challenging goals for itself and communicating these goal commitments to members.

Standards are very low or nonexistent in the organization 1 2 3 4 5 6 7 8 9 10 High challenging standards are set in the organization

4. *Rewards*. The degree to which members feel that they are being recognized and rewarded for good work rather than being ignored, criticized, or punished when something goes wrong

Members are ignored, punished, or criticized 1 2 3 4 5 6 7 8 9 10 Members are recognized and rewarded positively

5. *Organizational clarity*. The feeling among members that things are well organized and goals are clearly defined rather than being disorderly, confused, or chaotic.

The organization is disorderly, confused, and chaotic 1 2 3 4 5 6 7 8 9 10 The organization is well organized with clearly defined goals

6. *Warmth and support*. The feeling that friendliness is a valued norm in the organization; that members trust one another and offer support to one another. The feeling that good relationships prevail in the work environment.

There is no warmth and support in the organization 1 2 3 4 5 6 7 8 9 10 Warmth and support are very characteristic of the organization

7. *Leadership*. The willingness of organization members to accept leadership and direction from qualified others. As needs for leadership arise members feel free to take leadership roles and are rewarded for successful leadership. Leadership is based on expertise. The organization is not dominated by, or dependent on, one or two individuals.

Leadership is not rewarded, members are dominated or dependent and resist leadership attempts 1 2 3 4 5 6 7 8 9 10 Members accept and reward leadership based on expertise

APPENDIX VI

THREE MONTH FOLLOW-UP QUESTIONNAIRE

PART A

1. What did you get out of the subject matter discussed? Rate each topic according to the following scales:

- 1) How useful (relevant) were the following topics to you in your work?
- 2) How do you rate your present understanding of this topic?



- 1) Minimally useful in practice 1) Highly relevant to daily work
- 2) Vague understanding only 2) Excellent understanding only

	1) Relevance	2) Understanding
Topic 1: Organisation & Delegation	<input type="checkbox"/>	<input type="checkbox"/>
Topic 2: Forecasting	<input type="checkbox"/>	<input type="checkbox"/>
Topic 3: Planning	<input type="checkbox"/>	<input type="checkbox"/>
Topic 4: Resource Allocation	<input type="checkbox"/>	<input type="checkbox"/>
Topic 5: Reporting and Marking	<input type="checkbox"/>	<input type="checkbox"/>
Topic 6: Personnel Management	<input type="checkbox"/>	<input type="checkbox"/>

3) What is your opinion of the time allocation to various topics? Tick one for each topic.

	More time should be spent on this topic	Time spent on this topic was about right	Less time should be spent on this topic
1. Organisation & Delegation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Forecasting	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Planning	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Resource Allocation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Reporting and Marking	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Personnel Management	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Appendix VI continued

PART B

1. What types of changes, if any, do you consider were brought about by the course? Please tick none or more items.

1) Areas of organisational change:

- Internal communication
- Staff turnover
- Increase in pool of potential Section Leaders
- Awareness and acceptance of departmental policies
- Understanding between scientific staff and administrative staff.
- Other.....

2) Areas of personal change:

- Job satisfaction
- Attitude to change
- Quality of managerial performance
- Human relations skills
- Social interaction between course members
- Other.....

2. Can you name any books etc. that you have subsequently read on the topics discussed during training?

3. Following the science management course, have you discussed it with others in your work environment?

Yes No

With whom?

- \$. Have you maintained contact with other course members?

Yes No

How often?

By what means? (letter, phone, face-to-face, etc.)

Appendix VI continued

PART C

1. Do you think that your behaviour changed - even if it is only for a few days immediately after the course - as a result of the course?

Yes

No

2. Never mind for how short a time, how much was your behaviour affected?

A good deal

Some

A little

Don't know

3. Have any of the behaviour changes survived until now?

Yes

No

4. At the end of the course you were asked to write down one or more changes you proposed to make in your work as a result of something you learnt on the course:

- 1) Were you able to produce the proposed changes?

- 2) Did you try to introduce any other changes?

- 3) How successful were these changes? If not successful, why were they not able to be accomplished?

APPENDIX VII

SIX MONTH FOLLOW-UP QUESTIONNAIRE FOR
TRAINEES AND CONTROL SUBJECTSA Trainees

As a result of the course,

1. Did you try to make any changes in your work behaviour?

YES

NO

If 'YES', please specify what these changes were:

- 1) _____
- 2) _____
- 3) _____
- 4) _____
- 5) _____
- 6) _____

2. Did you succeed? Place a tick in the box opposite to the changes which were successful.

3. If you did not succeed, please give the reason(s) in each case:

- 1) _____
- 2) _____
- 3) _____
- 4) _____
- 5) _____
- 6) _____

4. Overall, how much was your work behaviour affected as a result of the course?

A GOOD DEAL

SOME

A LITTLE

DON'T KNOW

5. If you discussed the training course with co-workers after your return, what was the content of these discussions? e.g. your impressions of the course, subject material covered by the course topics, certain topic areas in particular, other.

Appendix VII continued

B Control subjects

1. Have you tried to make any changes in your work behaviour over the past year?

YES

NO

If 'YES', please specify what these changes were:

- 1) _____
- 2) _____
- 3) _____
- 4) _____
- 5) _____
- 6) _____

2. Did you succeed? Place a tick in the box opposite to the changes which were successful.

3. If you did not succeed, please give the reason(s) in each case:

- 1) _____
- 2) _____
- 3) _____
- 4) _____
- 5) _____
- 6) _____

4. Overall, how much has your work behaviour (in a managerial sense) changed over the past year?

A GOOD DEAL

SOME

A LITTLE

DON'T KNOW

APPENDIX VIII

SIX MONTH FOLLOW-UP QUESTIONNAIRE
FOR CONTROLLING OFFICERS

Have the Courses, in your opinion, had any effect on achieving the following goals?

SECTION A:

	YES	NO	CAN'T SAY						
1. Increased confidence and skill in handling and communicating with others.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	If <input type="checkbox"/> YES please indicate degree of change.	VERY HIGH DEGREE	HIGH DEGREE	MOD-ERATE DEGREE	SMALL DEGREE	VERY SMALL DEGREE
					<input type="checkbox"/>				
2. Improved ability to communicate with support staff.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>				
3. Improved planning and organisation of work according to well defined objectives.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>				
4. Increased concern with the development of technical staff.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>				
5. More self confidence, generally, in coping with the work situation.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>				
6. Increased appreciation of the need for good management and improved managerial skills.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>				
7. Improved managerial skills.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>				

Appendix VIII continued

- | | | | | | | |
|--|--|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 8. Improved interpersonal relations within the work group. | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> |
| 9. Improved ability to organise the work of section or group so that it can be integrated with the overall aims of the department. | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> |
| 10. Greater understanding of the problems involved in the distribution and allocation of resources within the department. | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> |
| 11. Increased willingness to accept an administrative role. | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> |
| 12. A better knowledge of the goals and policies of the department. | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> |
| 13. A greater understanding of the part he/she plays (as an individual staff member) in the total organisation. | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> |
| 14. Improved ability to plan the work of the section or group and to fully utilise the expertise of individual members | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> |
| 15. Greater awareness of the relationship between individual research interests and the needs of the society. | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> |

Appendix VIII continued

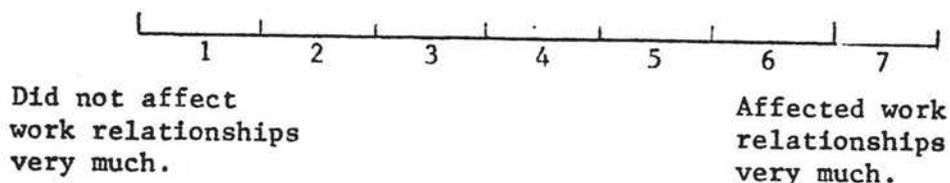
- | | | |
|--|--|--|
| 16. Improved ability to motivate other members of staff. | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 17. Greater knowledge of other Government departments. | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 18. Greater inclination and ability to delegate work. | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 19. Improved ability to encourage people to work together as a group. | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 20. Better able to make good decisions taking all relevant facts into account. | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |
| 21. Improved skill in routine tasks e.g. letter-writing. | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> |

APPENDIX IX

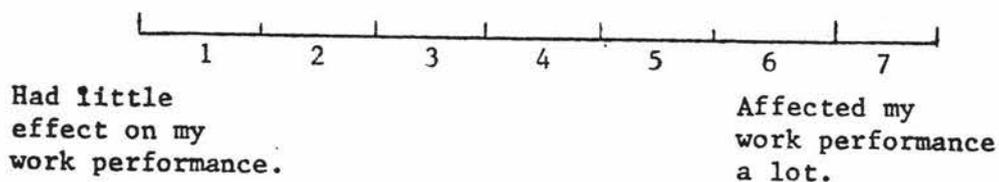
TWELVE MONTH FOLLOW-UP QUESTIONNAIRE

On the following three scales, indicate your opinions by placing a cross (x) in the interval from one to seven along each of the scales.

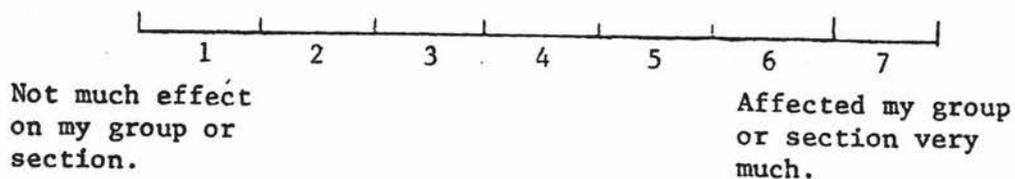
1. Effect of the training on your interpersonal relationships at work.



2. Effect of the training on your own work performance.



3. Effect of the training on the organisation of your group or section.



OPEN STATEMENT

At this considerable distance from the course what are the main things about the course that stand out most clearly in your mind?

ANSWER:

APPENDIX X

THE KELLY REPERTORY GRID

Role titles for elementsINTERPERSONAL SITUATIONS

1. A time when I delegated an important task to a co-worker.
2. The time I actively opposed the ideas of my controlling officer (or someone in authority).
3. A time I had to deal with a problem brought to me by a member of my staff.
4. A time I had to make an important decision concerning my research (or other work).
5. A time when I had a professional association with some outside organisation (business, industry, etc.).
6. The occasion when I made (or proposed) changes in the running and conduct of section meetings or other procedures of a similar nature.
7. An occasion when I felt most satisfied with my work performance.
8. An occasion when I felt least satisfied with my work performance.
9. My professional self now.
10. My professional self a year ago.

Personal constructsBI-POLAR DESCRIPTIONS

A ----

(Similarly for constructs B through J)

Repertory grid matrix

(Subjects were presented with a blank matrix having columns 1 through 10 as elements and rows A through J as constructs)

BIBLIOGRAPHY

- Apsler, R. In defense of the experimental paradigm as a tool for evaluation research. *Evaluation*, 1977, 4, 14-20.
- Argyris, C. Problems and new directions for industrial psychology. In M.D. Dunnette (Ed.) *Handbook of Industrial and Organizational Psychology*. Chicago: Rand McNally, 1976, 151-184.
- Ashton, D. and Gibbon, B. Evaluating a project-based management development programme. *Industrial Training International*, 1974, 9, (8), 254-256.
- Baynes, M. An examination of the major variables in the management training process. *Personnel Review*, 1975, 4, (1), 33-45.
- Bennis, W. G. A new role for the behavioural sciences: effecting organizational change. *Administrative Science Quarterly*, 1963, VIII, 125-165.
- Bennis, W. G. *Changing Organizations*. New York: McGraw-Hill, 1966, (a).
- Berger, P. Participant learning from group dynamics courses: an exploratory study. *North Western Polytechnic Working Paper*, 2, 1968.
- Berger, M. and Berger, P. An evaluation study of a communication course of managers. *North Western Polytechnic Working Paper*, 7, 1969.
- Bernstein, I. N. Validity issues in evaluative research: an overview. *Sociological Methods and Research*, 1975, 4 (1), 3-12.
- Blake, R. R. and Mouton, J. S. *The Managerial Grid*. Houston: Gulf, 1964.
- Blood, M. R. Spin-offs from behavioural expectation scale procedures. *Journal of Applied Psychology*, 1974, 59 (4), 513-515.
- Blum, M. L. and Naylor, J. C. *Industrial psychology: its theoretical and social foundations*. New York: Harper & Row, 1968.

- Bonoma, T. V. Social psychology and social evaluation, *Representative Research in Social Psychology*, 1976, 7, 147-159.
- Bonoma, T. V. Overcoming resistance to changes recommended for operating programs. *Professional Psychology*, 1977, 451-463.
- Borman, W. C. *The ratings of individuals in organizations: an alternative approach*. Unpublished doctoral dissertation, University of California, Berkeley, 1972.
- Boruch, R. F. Coupling randomized experiments and approximations to experiments in social program evaluation. *Sociological Methods and Research*, 1975, 4, 31-53.
- Boruch, R. F. and Gomez, H. Sensitivity bias, and theory in impact evaluations. *Professional Psychology*, 1977, 411-434.
- Boruch, R. F. and Rindskopt, D. On randomized experiments, approximations to experiments and data analysis. In Rutman (Ed.), *Evaluation Research Methods: A basic Guide*. Beverley Hills, California: Sage, 1977, 141-176.
- Bouchard, T. J. (Jr.). Field research methods: interviewing, questionnaires, participants observations, unobtrusive measures. In M. D. Dunnette (Ed.), *Handbook of Industrial and Organizational Psychology*. Chicago: Rand McNally, 1976, 363-413.
- Brennan, R. L. and Light, R. J. Measuring agreement when two observers classify people into categories not defined in advance. *British Journal of Mathematical and Statistical Psychology*, 1974, 27, 154-163.
- Brown, N. R. Job analysis by multidimensional scaling. *Journal of Applied Psychology*, 1967, 51, (6), 469-475.
- Bunker, D. R. Individual applications of laboratory training. *Journal of Applied Behavioural Science*, 1965, 1, 131-147.
- Burgoyne, J. G. An action research experiment in the evaluation of a management development course. *Journal of Management Studies*, 1973, 10 (1), 8-14.
- Burgoyne, J. G. and Cooper, C. Evaluation methodology. *Journal of Occupational Psychology*, 1975, 48, 53-62.

- Campbell, D. T. and Fiske, D. W. Convergent and discriminant validation by the multitrait - multimethod matrix. *Psychological Bulletin*, 1959, 56, 81-105.
- Campbell, D. T. and Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1966.
- Campbell, J. P., Dunnette, M. D., Arvey, R. D., and Hellervik, L. W. The development and evaluation of behaviourally based rating scales. *Journal of Applied Psychology*, 1973, 57 (1), 15-22.
- Campbell, J. P., Dunnette, M. D., Lawler, E. R. and Weick, K. E. *Managerial behaviour performance and effectiveness*. New York: McGraw-Hill, 1970.
- Cannell, C. F. and Kahn, R. L. The collection of data by interviewing. In L. Festinger and D. Katz (Eds.), *Research Methods in the Behavioural Sciences*, New York: Holt, Rinehart and Winston, 1953.
- Cantril, H. *Gauging public opinion*. Princeton: Princeton University Press, 1944.
- Carlson, R. E., Thayer, P. W., Mayfield, E. C. and Petersen, D. A. Improvements in the selection interview. *Personnel Journal*, 1971. 50, 268-275.
- Cherns, A. B. Social research and its diffusion. *Human Relations*, 1969, 22 (3), 210-218.
- Conover, J. C. *Practical Nonparametric Statistics*. New York: Wiley, 1971.
- Cook, T. D. and Campbell, D. T. Experiments in field settings. In M. D. Dunnette (Ed.), *Handbook of Industrial and Organizational Psychology*, Chicago: Rand McNally, 1976, 223-326.
- Cronbach, E. Beyond the two disciplines of scientific psychology. *American Psychologist*, 1975, 30, (2), 116-127.
- Cronbach, L. J. and Meehl, P. E. Construct validity in psychological tests. *Psychological Bulletin*, 1955, 52, 281-302.

- Dalziel, P., McWilliam, S., Strong, R., Hayes, B., and Lunt, R.
The evaluation of training. *Industrial Training International*,
1972, 7, (3), 88-91.
- Department of Employment Work Research Unit, *1980 Report of the
Tripartite Steering Group on Job Satisfaction*. London, 1981.
- Dillehay, R. C., Bruvold, W. H. and Seigel, J. P. Attitude, object
label and stimulus factors in responses to an attitude object.
Journal of Personality and Social Psychology, 1969, 11, 220-223.
- Drexler, J. A. Organizational climate: its homogeneity within
organizations. *Journal of Applied Psychology*, 1977, 62, (1),
38-42.
- Dubin, R. Theory building in applied areas. In M. D. Dunnette (Ed.),
Handbook of Industrial and Organizational Psychology. Chicago:
Rand McNally, 1976, 17-39.
- Edwards, A. L. *Techniques of attitude scale construction*. New York:
Appleton-Century Crofts, 1957(a).
- Edwards, A. L. and Kilpatrick, F. P. A technique for the
construction of attitude scales. *Journal of Applied Psychology*,
1948, 32, 374-384.
- Edwards, W. Social utilities. *The Engineering Economist*, Summer
Symposium Series, VI, 1971.
- File, Q. Q. and Remmers, H. H. *How Supervise? Manual 1948 revision*.
New York: Psychologist Co-operation, 1948.
- Fineman, S. & Warr, P. B. *Managers: Their effectiveness and
training*. In P. B. Warr (Ed.), *Psychology at Work*. Middlesex:
Penguin, 1971, 259-282.
- Flanagan J. C. The critical incident technique. *Psychological
Bulletin*, 1954, 51, (4), 327-358.
- Fleishman, E. A. The description of supervisory behaviour.
Journal of Applied Psychology, 1953, 37, 1-6 (a).

- Fleishman, E. A. A leader behaviour description for industry. In R. M. Stogdill and A. E. Coons (Eds.), *Leader Behaviour: Its Description and Measurement*. Columbus: Ohio State University Bureau of Business Research, Research Monograph No 88, 1957, 103-119 (a).
- Fleishman E. A. The leadership opinion questionnaire. In R. N. Stogdill and A. E. Coons (Eds.), *Leader behaviour: Its Description and Measurement*. Columbus: Ohio State University Bureau of Business Research, Research Monograph No. 88, 1957, 120-133. (b)
- Fleishman, E. A. Development of a behaviour taxonomy for describing human tasks: a correlation experimental approach. *Journal of Applied Psychology*, 1967, 41, 1-10.
- Fleishman, E. A., Harris, E. E. and Burtt, H. E. Leadership and supervision in industry: an evaluation of a supervisory training programme. *Bureau of Educational Research Monograph, No. 33*, Columbus, Ohio: Ohio State University Press, 1955.
- Frederiksen, N., Saunders, D. R. and Ward, B. The In-basket Test. *Psychological Monographs*, 1957, 71 (9, Whole No. 438).
- Friedlander, R. The impact of organizational training laboratories upon the effectiveness and interaction of ongoing work groups. *Personnel Psychology*, 1967, 20 (3) 286-308.
- Freidman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 1937, 32, 675-701.
- Glaser, E. M. and Backer, T. E. A clinical approach to program evaluation. *Evaluation*, 1972, 54-60.
- Glaser, E. M. and Backer, T. E. A look at participant evaluation. *Evaluation*, 1 (3), 1973, 46-49.
- Glass, G. V. and Ellett, F. S. Jr. Evaluation research. *Annual Review of Psychology*, 1980, 31, 211-228.
- Grant, D. L. and Anderson, S. B. Issues in the evaluation of training. *Professional Psychology*, 1977, 659-673.

- Guttentag, M. Subjectivity and its use in evaluation research. *Evaluation*, 1 (2), 1973, 60-65.
- Guttman, L. A basis for scaling quantitative data. *American Sociological Review*, 1944, 9, 139-150.
- Hamblin, A. C. *Evaluation and control of training*. New York: McGraw-Hill, 1974.
- Hand, H. H. and Slocum, J. W. A longitudinal study of the effects of a human relations training program on managerial effectiveness. *Journal of Applied Psychology*, 1972, 56, (5), 412-417.
- Hemphill, J. K. Dimensions of executive positions. Ohio Studies in Personnel. *Research Monographs*, No. 98, Ohio State University Bureau of Business Research, 1960.
- Hinrichs, J. R. An eight year follow-up of a management assessment centre. *Journal of Applied Psychology*, 1978, 63, (5), 596-601.
- House, R. J. and Rizzo, J. R. Toward the measurement of organizational practices: scale development and validation. *Journal of Applied Psychology*, 1972, 56, (5), 388-396.
- James, L. R. Criterion models and construct validity for criteria. *Psychological Bulletin*, 1973, 80, (1), 75-83.
- Jones, D. and Huczynski, A. Research note: evaluation of management training: some aspects of methodology. *Personnel Review*, 4 (1), 1975, 59-61.
- Kahn, R. L., Wolfe, D. M., Quinn, R. P., Snoek, J. D. and Rosenthal, R.A. *Organizational stress: studies in role conflict and ambiguity*, New York: Wiley, 1964.
- Katz, D. and Kahn, R. L. *The social psychology of organizations*. New York: Wiley, 1966.
- Kelly, G. A. *The psychology of personal constructs*. New York: Norton, 1955.
- Keppel, G. *Design and analysis: a researcher's handbook*. Englewood Cliffs, New Jersey: Prentice-Hall, 1973.

- Kimber, D. The evaluation of training. *Industrial Training International*, 5, (6), 1970, 276-278.
- Kirkpatrick, D. L. Evaluation of training. In R. L. Craig and L.R. Bittel (Eds.), *Training and Development Handbook*. American Society for Training and Development. New York: McGraw-Hill, 1967.
- Kolb, D. A., Rubin, I.N. and McIntyre, J.M. *Organizational Psychology*, (2nd Ed.). Englewood Cliffs, New Jersey: Prentice-Hall, 1974.
- Korman, A. K. *Industrial and organizational psychology*. New Jersey: Prentice-Hall, 1971.
- Lawler, E. E. Management and performance as seen from above, below and within. *Journal of Applied Psychology*, 1967, 51, 247-253.
- Lawler, E. E. The multitrait-multirater approach to managerial job performance. *Journal of Applied Psychology*, 1967, 51, (5), 369-381.
- Likert, R. A technique for the measurement of attitudes. *Archives of Psychology*, 1932, 140, 44-53.
- Likert, R. *The human organization: its management and value*. New York: McGraw-Hill, 1967.
- Litwin, G. H. and Stringer, R.A. *The influence of organizational climate on human motivation*. Paper presented at a conference on organizational climate, Foundation for Research on Human Behaviour, Ann Arbor, Michigan, 1966.
- Litwin, G. H. and Stringer, R. A. *Motivation and organizational climate*. Boston: Division of Research, Harvard Business School, 1968.
- Locke, E. A. Nature and causes of job satisfaction. In M.D. Dunnette (Ed.), *Handbook of Industrial and Organizational Psychology*. Chicago: Rand McNally, 1976, 1297-1349.
- Mann, F. C. Studying and creating change. In W. G. Bennis, K. D. Benne and R. Chin (Eds.), *The Planning of change*. New York: Holt-Rinehart and Winston, 1962.

- Matheson, D. W., Bruce, R. L. and Beauchamp, K. L. *Introduction to experimental psychology*, (2nd Edition). New York: Holt-Rinehart and Winston, 1974.
- Miles, M. B. Human relations training: processes and outcomes. *Journal of Counselling Psychology*, 1960, 7 (4), 301-306.
- Miles, M. B. Changes during and following laboratory training: a clinical experimental study. *Journal of Applied Behavioural Science*, 1965, 1, 215-242.
- Morris, G. Organizational development in a multinational company - a case study. *European Training*, 1 (1), 1972, 78-88.
- McCall, M. W. Leadership research: choosing gods and devils on the run. *Journal of Occupational Psychology*, 1976, 1 (3), 139-153.
- McCormick, E. J., Jeanneret, P. R. and Mechan, R. C. Study of job characteristics and job dimensions as based on the Position Analysis Questionnaire. *Journal of Applied Psychology*, 1972, 56, 347-368.
- Newman, J. E. Understanding the organizational structure: job attitude relationship through perceptions of the work environment. *Organizational Behaviour and Human Performance*, 1975, 14, 371-397.
- Payne, R. and Pugh, D. S. Organizational structure and climate. In M.D. Dunnette (Ed.), *Handbook of Industrial and Organizational Psychology*. Chicago: Rand McNally, 1976, 1125-1173.
- Rackham, N., Honey, P. and Colbert, M. (Eds.) *Developing interactive skills*. London: Wellens Publishing, 1971.
- Reid, T. J. The context of management development. *Personnel Journal*. 1974, 53 (4), 280-287.
- Riecken, H. W. *The volunteer work camp: a psychological evaluation*. Cambridge, Mass: Addison-Wesley, 1952.
- Riecken, H.W. Principal components of the evaluation process. *Professional Psychology*, 1977, 392-410.
- Riecken, H.W. and Boruch, R. F. (Eds.). *Social experimentation: a method for planning and evaluating social intervention*. New York: Academic Press, 1974.

- Riecken, H. W. and Boruch, R. F. Social experiments. *Annual Review of Sociology*, 1978, 4, 511-532.
- Rothamsted Experimental Station Staff. *Genstat: a general statistical program*. The Statistics Department, Rothamsted Hertfordshire, 1977.
- Rosenthal, R. *Experimenter effects in behavioural research*. New York: Appleton-Century-Crofts, 1966.
- Rundquist, E. A. *Course design and redesign: manual for job training courses, (1st ed.)*. San Diego: U.S. Naval Personnel Research Activity (Technical Bulletin STB 66-17, Revised), 1967.
- Schein, E. H. *Organizational psychology*. New York: Prentice-Hall, 1965.
- Schneider, B. Organizational climates: an essay. *Personnel Psychology*, 1975, 28, 447-497.
- Schneider, B. and Bartlett, C. J. Individual differences and organizational climate 1. The research plan and questionnaire development. *Personnel Psychology*, 1968, 21, 323-334.
- Schreiber, C. T. *Changing Places*. Cambridge, Mass.: The MIT Press, 1979.
- Scriven, M. The methodology of evaluation. In C. H. Weiss (Ed.), *Evaluating action programs: readings in social action and education*. Boston: Allyn and Bacon, 1972.
- Shouksmith, G. *Assessment through interviewing (2nd ed.)*. Oxford: Pergamon Press, 1978.
- Smith P. C. and Kendall, L. M. Retranslation of expectations: an approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 1963, 47, (2), 149-155.
- Smith, R. J. and Siegel, A. I. A multi-dimensional scaling analysis of the job of civil defence director. *Journal of Applied Psychology*, 1967, 51, (6), 476-480.

- Smith, C. M. and Smith, M. C. Who is versatile? *Industrial and Commercial Training*, 8, 1975, 389-381.
- Smith, H. C. and Wakeley, J. H. *Psychology of industrial behaviour*. New York: McGraw-Hill, 1955.
- Soloman, R. L. An extension of control group design. *Psychological Bulletin*, 1949, 46, 137-150.
- Stewart, R. Classifying different types of managerial jobs. *Personnel Review*, 1975, 4, (2), 25-30.
- Stroud, P. V. Evaluating a human relations training program. *Personnel*, 1959, 36, 52-60.
- Suchman, E. A. *Evaluation research*. New York: Russell Sage Foundation, 1967.
- Super, D. E. Vocational guidance: emergent decision-making in a changing society. *Bulletin: Association Internationale d'Orientation Scolaire et Professionnelle*, 1976.
- Tagiuri, R. *Comments on organizational climate*. Paper presented at a conference on organizational climate. Foundation for Research on Human Behaviour, Ann Arbor, Mich, 1966.
- Tagiuri, R. and Litwin, G. H. *Organizational climate: exploration of a concept*. Harvard Business School, 1968.
- Thurstone, L. L. A law of comparative judgement. *Psychological Review*, 1927, 34, 273-286(a).
- Thurstone, L. L. Psychophysical analysis. *American Journal of Psychology*, 1927, 38, 368-389(b).
- Thurstone, L. L. Attitudes can be measured. *American Journal of Sociology*, 1928, 33, 529-554.
- Thurstone, L. L. The measurement of social attitudes. *Journal of Abnormal and Social Psychology*, 1931, 26, 249-269.
- Tiffin, J. and McCormick, E. J. *Industrial Psychology*. Englewood Cliffs, New Jersey: Prentice-Hall, 1965.

- Tornow, W. W. and Pinto, P. R. The development of a managerial job taxonomy: a system for describing, classifying and evaluating executive positions. *Journal of Applied Psychology*, 1976, 61(4) 410-418.
- Triandis, H. C. *Attitude and attitude change*. New York: Wiley, 1971.
- Vroom, V. H. *Some personality determinants of the effects of participation*. Englewood Cliffs, N.J.: Prentice-Hall, 1960.
- Vroom, V. H. Leadership. In M.D. Dunnette (Ed.), *Handbook of Industrial and Organizational Psychology*. Chicago: Rand McNally, 1976, 1527-1551.
- Warr, P., Bird, M. and Rackham, N. *Evaluation of management training*. London: Gower Press, 1976.
- Webber, R. A. Perception of interactions between superiors and subordinates. *Human Relations*, 1970, 23, 235-248.
- Weiss, C. H. *Evaluation research*. Englewood, Cliffs, New Jersey: Prentice-Hall, 1972.
- Williams, R. and Berger, M. The evaluation of management development: its possible relevance to a levy/grant system. *Management Education and Development*, 1972, 3, 37-48.
- Williamson, G. A., Prost, J. and George, C. E. The professional psychologist as evaluator. *Professional Psychology*, 1978, 308-314.
- Wortman, P. M. Evaluation research: a psychological perspective. *American Psychologist*, 1975, 562-575.
- Wortman, P. M. and Muirhead, S. Toward the proper conduct of social program evaluations. *Evaluation*, 1977, 4, 189-192.
- Wrightsmann, L. S. *Social psychology in the seventies*. Monterey, California: Brooks-Cole, 1972.