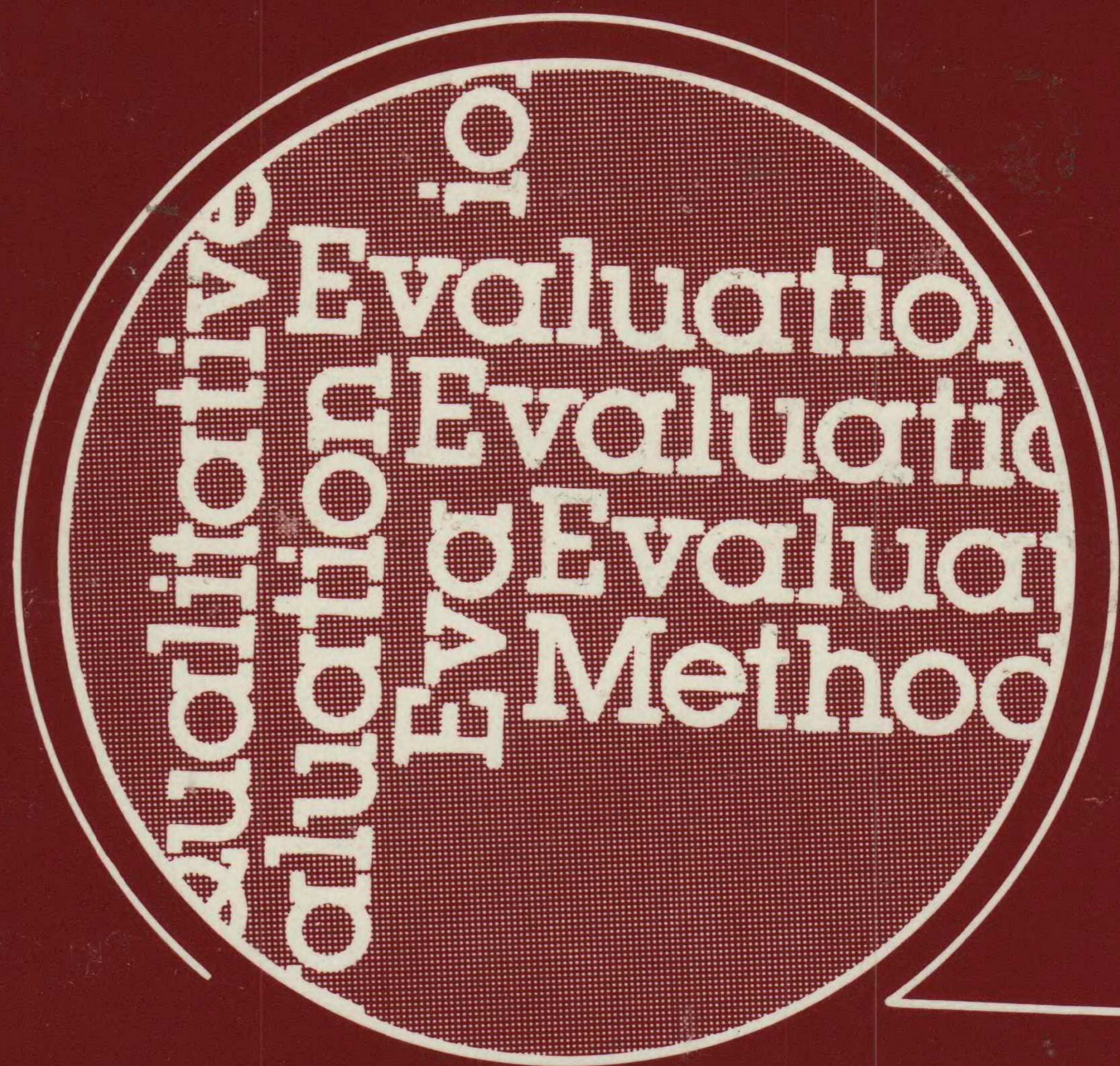


MICHAEL QUINN PATTON



# Qualitative Evaluation Methods

---



The emphasis throughout *Qualitative Evaluation Methods* is on strategies for generating valid, useful, and credible qualitative information for decision-making. Different approaches to field observations, in-depth interviewing, and qualitative data analysis are explained; and examples of different research strategies are provided, including samples of qualitative instruments and excerpts from naturalistic inquiry reports.

This book provides evaluators with the first comprehensive framework for deciding when it is appropriate to use qualitative evaluation methods. Patton's concrete approach represents a unique and much-needed contribution to the evaluation research literature.

**ISBN 0-8039-1395-8**



Michael Quinn Patton is Director of the Minnesota Center for Social Research and Adjunct Professor in the Hubert H. Humphrey Institute of Public Affairs, both at the University of Minnesota. He is the author of numerous articles, reports, and conference papers in the field of evaluation research, and regularly conducts evaluation workshops. Dr. Patton has also served as an evaluation consultant to many educational and human services projects in the United States and abroad. He is the author of *Utilization-Focused Evaluation* (Sage, 1978).



*By the same author . . .*

# UTILIZATION- FOCUSED EVALUATION

**Michael Quinn Patton**

Combining the theoretical with the practical, Patton examines both how and why to conduct evaluations. He reviews the history of the development of evaluation research and presents a new, comprehensive model for evaluation—"the utilization-focused approach." This model is derived from many sources, including the study of the utilization of federal evaluation research conducted through the University of Minnesota's NIMH training program in evaluation; the author's own experience conducting over thirty evaluations during the past five years; recent developments in decision-making theory and political analysis; and organizational theory and the rapidly growing body of evaluation research literature.

"This book probably represents the best discussion of the art of program evaluation which I have ever read. It maintains a high level of scholarship while being extremely pertinent to the practice of program evaluation."

—*The Applied Sociologist*

1978

hardcover ISBN 0-8039-0982-9

softcover 0-8039-0981-0



**SAGE** PUBLICATIONS

The Publishers of Professional Social Science

Beverly Hills London



MICHAEL QUINN PATTON

**Qualitative**  
Evaluation  
Methods



SAGE Publications Beverly Hills London



All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

*For information address:*

SAGE Publications, Inc.  
275 South Beverly Drive  
Beverly Hills, California 90212



SAGE Publications Ltd  
28 Banner Street  
London EC1Y 8QE, England

Printed in the United States of America

**Library of Congress Cataloging in Publication Data**

Main entry under title:

Patton, Michael Quinn.

Qualitative evaluation methods.

Includes index.

1. Social sciences—Methodology. 2. Evaluation research

(Social action programs) I. Title.

H62.P3218 300'.7'23 79-24392

ISBN 0-8039-1395-8

**NINTH PRINTING, 1988**

Permission to use the Sufi stories on pages 227, 326, and 333 was granted by Curtis Brown Ltd., Copyright 1972, 1973 by Mulla Nasrudin Enterprises Ltd.



TO

Hal and Malcolm  
Clyde and Harris  
Gray/Lenke Associates  
Learninghouse  
and the  
Southwest Field Training Project



# CONTENTS

Preface	9
---------	---

## PART I

### CONCEPTUAL ISSUES IN THE USE OF QUALITATIVE METHODS FOR EVALUATION RESEARCH

1	A Paradigm of Choices	17
2	The Nature of Qualitative Data	21
3	The Strategy of Qualitative Methods	39
4	Evaluation Research Strategies Using Qualitative Methods	49
	Evaluation Models Compatible with Qualitative Methods	53
	Beyond Models: The Practice of Evaluation	58
	Checklist of Evaluation Situations for which Qualitative Methods are Appropriate	88
5	Evaluation Designs	91
	Critical Trade-Offs in Evaluation Design	95
	Methodological Mixes	108
	The Case of Operation Reach-Out: Variations in Program Evaluation Design	111

## PART II

### COLLECTING QUALITATIVE DATA

6	Evaluation Through Observation	121
	Variations in Observational Methods	126
	What to Observe	137
	Doing Fieldwork: The Data-Gathering Process	160
	The Stages of Fieldwork	169
	The Observer and What Is Observed: Unity and Separation	189
7	Qualitative Interviewing	195
	Variations in Qualitative Interviewing	197
	The Content of Interviews: What Questions to Ask	207
	The Wording of Questions	211
	Recording the Data	246
	Personal Reflections on Interviewing	252
	Appendix 7.1	255
	Appendix 7.2	258



PART III  
DATA ANALYSIS

8	The Purpose of Qualitative Analysis: A Perspective	267
	Truth	268
	Theory	274
	Generalizations	279
	A Documentor's Perspective	283
9	Qualitative Analysis and Interpretation	295
	Qualitative Description	302
	Case Studies	302
	Inductive Analysis	306
	Logical Analysis	313
	Validation and Verification	326
	Presentation of Findings	340
	Appendix 9.1	347
	Appendix 9.2	351
	Appendix 9.3	359
	References	369
	Index	375
	About the Author	381



# PREFACE

Question: How does a sociologist whose graduate training was entirely in the hypothesis-testing approach to research based on quantitative measurement and statistical analysis come to write a book on qualitative evaluation methods?

Answer (select one):

- (a) rebellion against authority
- (b) flawed socialization
- (c) must be genetic
- (d) mixture of b and c
- (e) idle hands are the devil's playthings
- (f) a microlevel function of being caught in the historical intersection of the ownership of the means of production and the changing forces of production (in other words, lost a bad bet)
- (g) all of the above
- (h) none of the above.

For readers who, like myself, may experience some initial separation anxiety in moving away from standardized psychological tests and multiple-choice questionnaires to the open-ended, nonstandardized world of qualitative methods, I thought it might ease the transition to begin with a familiar format. You should know, however, that from this point forward the closed categories and standardized response sets are out.

The question, however, is a reasonable one. One of the cardinal principles of qualitative methods is the importance of background and context to the processes of understanding and interpreting data. It therefore seems appropriate in this preface to establish the background of and a context for understanding and interpreting this book. My graduate training did not offer courses in either qualitative methods or evaluation research. In pursuing an interest in applying social science at the program level I quickly discovered that my graduate school repertoire of research methods was too limited to serve me well in the variety of situations I was encountering. I was



forced to learn about alternative methods strategies in order to be of service to the staff members of innovative education programs who refused to subject their students, themselves, or their programs to “impersonal numbers” and “neat social science categories.” During two years of fieldwork at the Center for Teaching and Learning, University of North Dakota, working with Vito Perrone, Chuck Nielson, Nancy Miller, Lowell Thompson, and Tom Pace, I had the opportunity to explore ways of using numbers to understand programs while also learning about other ways of studying the world. Discussions with Patricia Carini, Ted Chittenden, Ann Cook, Brenda Engel, Sharon Feiman, George Hein, Herb Mack, Debbie Meier, Vito Perrone, Susan Stoldosky, and Lillian Weber through the North Dakota Study Group on Evaluation were critical in my coming to understand and use qualitative evaluation methods. It was the work with and support from this study group that led to my first writings on qualitative methods.

Participation as a postdoctoral fellow in the NIMH-supported Evaluation Methodology Training Program at the University of Minnesota gave me both time and the opportunity to think about and explore alternative evaluation research methods. I am indebted to Harold Finestone, University of Minnesota, who introduced me to the sociological literature on qualitative methods during this period. The faculty advisers in the program, John Brandl of the Hubert H. Humphrey Institute for Public Affairs and Ray Collier of the Department of Statistics, were particularly tolerant and supportive of trainees’ explorations into a variety of methodological approaches. The Minnesota training program supported our research on the utilization of evaluation research that led to the writing of *Utilization-Focused Evaluation* (Patton, 1978). In part of that book (Chapter 10) I advocated a flexible approach to selection of evaluation methods and argued that qualitative methods should be considered an appropriate, useful, and legitimate evaluative paradigm for certain evaluation situations. What I did *not* do in that book was explain precisely what those situations were or how to actually use qualitative methods. The purpose of this book is to make up for those omissions and respond to questions about qualitative evaluation methods that have been raised by colleagues and students since the publication of *Utilization-Focused Evaluation*.

That this book was written at this point in time is due entirely to the persuasive powers of Sara Miller McCune, President and Publisher of Sage Publications. Over a period of several months she provided me with reasons why the book should be written *now*, finally



convincing me to make a commitment during a trip to Minnesota in the fall of 1978. It has been an enormous pleasure to work with a publisher who is so thoroughly knowledgeable about evaluation research and the issues of concern to evaluators. That explains at least part of the reason why Sage Publications has become the leading publisher of evaluation research literature.

Several people were invaluable to me in the actual writing of the book. Greg Stark, a staff member of the Minnesota Center for Social Research, played the critical role of posing the questions to which I respond throughout the book; he was a constant source of inspiration and direction. Greg's own personal inquiry into qualitative evaluation methods kept me focused on the reason the book was being written and helped me know what I needed to cover. His careful reading of early drafts of the manuscript led to major revisions, deletions, and additions. Greg's own fieldwork furnished examples of qualitative analysis that are featured in the book.

Malcolm Gray and Hal Lenke, though they were not aware of it at the time, played critical roles in the writing of the book. They directed the Southwest Field Training Project, the evaluation of which gave me the opportunity to develop and test, test and develop my understandings about qualitative evaluation methods. Their commitment to that evaluation process, their trust in me, and their ongoing struggle to make sense of the program and to use the evaluation observations to that end convinced me that a book like this needed to be written, that it could be written, and that I ought to be the one to do it because of what I had to share from my experience with them. Because they do not believe in making a distinction between "personal" and "professional," they will, I hope, forgive me if I use this space to publicly acknowledge my thanks to them for the opportunities they provided me for personal and professional learning. None of that, however, explains the dedication of the book to them. The dedication is to our friendship.

I also wish to thank Ron Geizer and Chris Hauptert who kept the Minnesota Center for Social Research running smoothly while I was preoccupied with this book. Joyce Keller and Beth Alberty contributed original pieces of writing that illuminate important aspects of qualitative methods that have not been part of my own experience. Janet Donicht and Leslie Bernstein worked diligently and caringly in typing (and editing) the manuscript. (Leslie is in no way responsible if the word "impact" appears in these pages as a verb; she did her best to delete it every time it appeared in my raw verbiage.)



As other authors know, there is no way to really recognize the contribution of one's family to a book like this, the writing of which was a struggle and matter of endurance for both family and author. While Sara McCune was persuading me that the book should be written, Jeanne was persuading me that we could nurture together both a new book and a newborn child. (Having been left out of that decision, the newborn child subsequently made it clear he didn't always agree.) The contribution of Jeanne to the book exemplifies the kind of situation where the personal and professional cannot be and ought not be separated. Jeanne's reflections on her own evaluation fieldwork and interviewing experiences helped me clarify and break through some particularly difficult sections of the book. Her editorial advice was invaluable (it's embarrassing to think about some of the things that would have been left in without the cutting edge of her editorial axe). Those are her tangible contributions; the intangibles she contributed are the things that made the book happen.

Some prefatory comments on style and substance seem to be in order. This is not a recipe book. The nature of both evaluation research and qualitative methods makes it inappropriate to follow recipes or formulae. It is therefore not a "how to" book. At one time I had begun to tell people I planned to write a "how to" book. At the North Dakota Study Group on Evaluation Meetings in February 1979, Lillian Weber heard me say that and commented gently, "Michael, why don't you just write some suggestions about how people might proceed sometimes. We don't need a 'how to' book. Those are the kinds of books that got evaluation into the trouble it's already in."

The book is also not simply about methods. Qualitative modes of understanding the world are rooted in philosophical and epistemological traditions that require explication in order to establish a context for making decisions about the usefulness, credibility, validity, and appropriateness of various qualitative evaluation strategies. Thus, "suggestions" about methods are integrated with "suggestions" about some ways to think about methods and evaluation.

This last point means that this is not a neutral textbook which carefully balances alternatives and gives equal weight to all points of view. A definite point of view about evaluation is offered. The perspective presented here is grounded in the observation that much of what passes for evaluation research is not very useful; that *evaluation research ought to be useful*; that one factor affecting



usefulness is the appropriateness of the methods used; and that, therefore, some different things must be done if appropriate methods are to be employed to produce useful information for decision-making. This book tries to suggest some of the different things that can be done.

Saint Paul, Minnesota

— *Michael Quinn Patton*



## *Conceptual Issues in the Use of Qualitative Methods for Evaluation Research*

- Conceptualizing an evaluation depends on understanding self-interest: yours and theirs. Useful evaluations put theirs first. Then there are the others. . . .
- Hennes' Lament: Evaluators do IT under difficult circumstances.
- The scientific status of a methodological approach has nothing to do with its appropriateness. And vice versa.
- Evaluation is too serious a matter to be done by someone who has never been a client in a program.
- Gigo's Law of Deduction: Garbage in, garbage out.  
Halcolm's Law of Induction: No new experience, no new insight.
- An evaluation not worth doing is not worth doing well.

*From Halcolm's Evaluation Laws*



## A PARADIGM OF CHOICES

And it is recorded that the students came unto Halcolm, the Wise. “Teach us, Master, the right methods to use when we evaluate.” And he said:

“Issues of evaluation methodology are issues of strategy, not of morals. Purity of method is no virtue. That strategy is best which matches research methods to the evaluation questions being asked. The challenge is to decide which methods are most appropriate in a given situation. The science of making methods decisions is no less highly developed than the technology for making other simple decisions, for example, how to choose a spouse, career, city of residence, or which toothpaste to use.

“Blessed are the poor in choices, for they will have no trouble making up their minds.”

*From Halcolm's Evaluation Beatitudes*

UTILIZATION-FOCUSED EVALUATION:  
A LARGE REPERTOIRE OF METHODS

Evaluation research is coming of age. It has its own scholarly professional societies, a large and growing literature, and a multitude of practitioners. It also has growing pains, many of them occasioned by the increasing consciousness that the world is complex and not always manipulable according to one's wishes. Like a child losing its innocence, evaluation research has grown beyond the simple days when the answer to every evaluation problem was the administration of a standardized test to experimental and control groups.

Today's evaluator must be sophisticated about matching research methods to the nuances of particular evaluation questions and the idiosyncracies of specific decision maker needs. In *Utilization-*



*Focused Evaluation* (Patton, 1978) I described the evaluator's role as "active-reactive-adaptive" in working with decision makers and information users to focus evaluation questions and make methods decisions (Patton, 1978). In order to be active-reactive-adaptive the evaluator must have a large repertoire of research methods and techniques available to use on a variety of problems. Thus, today's evaluator may be called on to use any and all social science research methods, including analyses of quantitative data, questionnaire results, secondary data analysis, cost benefit and cost effectiveness analyses, standardized tests, experimental designs, unobtrusive measures, participant observation, and in-depth interviewing.

The active-reactive-adaptive evaluator works with decision makers to design an evaluation that includes any and all data that will help shed light on evaluation questions, given constraints of resources and time. Such an evaluator is committed to research designs that are relevant, rigorous, understandable, and able to produce useful results that are valid, reliable, and believable. On many occasions—indeed, for most evaluation problems—a variety of data collection techniques and design approaches will be used. Multiple methods and triangulation of observations contributes to methodological rigor. This style of evaluation is in the tradition of social science research aimed at "*sophisticated rigor*." Norman K. Denzin introduced this phrase to describe a commitment "to making data and explanatory schemes as public and replicable as possible." It is a phrase intended to describe "any and all [evaluators] who employ multiple methods, seek out diverse data sources, and attempt to develop behaviorally grounded theories" (Denzin, 1978:167).

Of course, the ideal of the active-reactive-adaptive evaluator being methodologically flexible, sophisticated, and able to use a variety of methods to study any particular evaluation question runs headlong into the realities of the evaluation world. Those realities include limited resources, political considerations, and the narrowness of disciplinary training available to most evaluators. These constraints mean that the imagery of being active-reactive-adaptive includes the evaluator as a negotiator who strives to obtain the best possible design and the most useful answers within the real world of politics, people, and methodological prejudice.

Concern about methodological prejudice led me to compare two alternative paradigms of evaluation measurement and design in *Utilization-Focused Evaluation*. That comparison included a lament about the dominance of one paradigm over the other.



Evaluation research is dominated by the largely unquestioned, natural science paradigm of hypothetico-deductive methodology. This dominant paradigm assumes quantitative measurement, experimental design, and multivariate, parametric statistical analysis to be the epitome of “good” science. This basic model for conducting evaluation research comes from the tradition of experimentation in agriculture, which gave us many of the basic statistical and experimental techniques most widely used in evaluation research. . . .

By way of contrast, the alternative to the dominant hypothetico-deductive paradigm is derived from the tradition of anthropological field studies. Using the techniques of in-depth, openended interviewing and personal observation, the alternative paradigm relies on qualitative data, holistic analysis, and detailed description derived from close contact with the targets of study. The hypothetico-deductive, natural science paradigm aims at prediction of social phenomena; the holistic-inductive, anthropological paradigm aims at understanding of social phenomena. From a utilization-focused perspective on evaluation research, neither of these paradigms is intrinsically better than the other. They represent alternatives from which the active-reactive-adaptive evaluator can choose; both contain options for identified decision-makers and information-users. . . . *The problem from a utilization-focused approach to evaluation is that the very dominance of the hypothetico-deductive paradigm with its quantitative, experimental emphasis appears to have cut off the great majority of its practitioners from serious consideration of any alternative evaluation research paradigm or methods.* The label “research” has come to mean the equivalent of employing the “scientific method,” of working within the dominant paradigm. There is, however, an alternative [Patton, 1978:203-204, 207].

The dominant paradigm no longer seems so ominous. The 1978 meeting of the Evaluation Research Society devoted substantial program time to consideration of qualitative methods. Donald Campbell and Lee Cronbach, considered major spokesmen for the dominant paradigm in the past, have recently advocated the appropriateness and usefulness of qualitative methods (Cronbach, 1975; Campbell, 1974). Ernest R. House, in describing the role of “qualitative argument” in evaluation research, notes that “when two of the leading scholars of measurement and experimental design, Cronbach and Campbell, strongly support qualitative studies, that is strong endorsement indeed” (1977:18).

The issues of selecting methods is no longer one of the dominant paradigm versus the alternative paradigm, of experimental designs with quantitative measurement versus holistic-inductive designs



based on qualitative measurement. The debate and competition between paradigms is being replaced by a new paradigm—a *paradigm of choices*. The paradigm of choices recognizes that different methods are appropriate for different situations.

This book is aimed at increasing and expanding the evaluation repertoire of applied social scientists. No attempt will be made here to argue that qualitative methods are “better” than quantitative methods. The purpose of this book is to assist those evaluators who want to employ qualitative methods so that they will know when it is appropriate to do so, and so that they can use qualitative methods in ways that will produce useful and valid data. The emphasis throughout is on *strategies* for using qualitative evaluation methods.



## THE NATURE OF QUALITATIVE DATA

### THE FRUIT OF QUALITATIVE METHODS

There was once a man who lived in a country which had no fruit trees. This man was a scholar and spent a great deal of time reading. In his readings he often came across references to fruit. The descriptions of fruit were so enticing that he decided to undertake a journey so that he could experience fruit for himself.

He went to the market and asked everyone he met if they knew where he could find fruit. After much searching he located a man who knew the directions to the country and place where he could find fruit. The man drew out elaborate directions for the scholar to follow.

With his map in hand, the scholar carefully followed all of the directions. He was very careful to make all of the right turns and to check out all of the landmarks that he was supposed to observe. Finally, he came to the end of the directions and found himself at the entrance to a large apple orchard. It was springtime and the apple trees were in blossom.

The scholar entered the orchard and proceeded immediately to take one of the blossoms and taste it. He liked neither the texture of the flower nor the taste. He went to another tree and sampled another blossom, and then another blossom, and another.

Each blossom, though quite beautiful, was distasteful to him. He left the orchard and returned to his home country, reporting to his fellow villagers that fruit was a much over-rated food. Being unable to recognize the difference between the spring blossom and the summer fruit, the scholar never realized that he had not experienced what he was looking for.

*From Halcolm's Evaluation Parables*



## RECOGNIZING QUALITATIVE DATA

The later chapters in this book discuss how to collect qualitative data and how to apply qualitative methods. First, it may be helpful to discuss the fruit of qualitative methods; it is important to know what qualitative data look like so that you will know what you are looking for. In so doing it will also be possible to develop some criteria concerning the quality of qualitative data.

Qualitative measurement has to do with the kinds of data or information that are collected. Qualitative data consist of *detailed descriptions* of situations, events, people, interactions, and observed behaviors; *direct quotations* from people about their experiences, attitudes, beliefs, and thoughts; and excerpts or entire passages from documents, correspondence, records, and case histories. The detailed descriptions, direct quotations, and case documentation of qualitative measurement are raw data from the empirical world. The data are collected as open-ended narrative *without* attempting to fit program activities or peoples' experiences into predetermined, standardized categories such as the response choices that comprise typical questionnaires or tests.

Quantitative measurement relies upon the use of instruments that provide a standardized framework in order to limit data collection to certain predetermined response or analysis categories. The experiences of people in programs and the important variables that describe program settings are fit into these standardized categories to which numerical values are then attached. By contrast, the evaluator using a qualitative approach to measurement seeks to capture what people have to say in their own words. Qualitative measures describe the experiences of people in depth. The data are open-ended in order to find out what people's lives, experiences, and interactions mean to them in their own terms and in their natural settings. Qualitative measures permit the evaluation researcher to record and understand people in their own terms.

Qualitative data provide *depth* and *detail*. Depth and detail emerge through direct quotation and careful description. The extent of depth and detail will vary depending upon the nature and purpose of a particular study. At the simplest level, depth and detail may emerge from responses to open-ended questions on a questionnaire. A good example of the difference in depth and detail obtained from open-ended questions versus standardized questionnaire items can be observed by comparing the two kinds of data from a single study.



### *Comparing Two Kinds of Data: A Case Example*

In the early 1970s the school system of Kalamazoo, Michigan, implemented a new accountability system. It was a complex system that included using standardized achievement tests administered in both fall and spring; criterion reference tests developed by teachers; performance objectives; teacher peer ratings; student ratings of teachers; parent ratings of teachers; principal ratings of teachers; and teacher self-ratings.

The Kalamazoo accountability system began to attract national attention. For example, the *American School Board Journal* reported in April 1974 that “Kalamazoo schools probably will have one of the most comprehensive computerized systems of personnel evaluation and accountability yet devised” (p. 40). In the first of a three-part series on Kalamazoo the *American School Board Journal* had asserted:

Take it from Kalamazoo: *a comprehensive, performance-based system of evaluation and accountability can work* [ASBJ, 1974:32; italics in the original].

Not everyone agreed with that positive assessment, however. The Kalamazoo Education Association charged that teachers were being demoralized by the accountability system. Some school officials, on the other hand, argued that teachers did not want to be accountable. In the spring of 1976 the Kalamazoo Education Association, with assistance from the Michigan Education Association and the National Education Association, sponsored a survey of teachers to find out the teachers’ perspective on the accountability program (Perrone et al., 1976).

The education association officials were interested primarily in a questionnaire consisting of standardized items. One part of the questionnaire provided teachers with a set of statements with which they could agree or disagree. A representative sampling of results from those items are listed in Table 2.1.

It is clear from the most cursory look at the questionnaire results that teachers felt the accountability system was largely ineffective and inadequate. Ninety percent of the teachers *disagreed* with the school administration’s published statement, “The Kalamazoo accountability system is designed to personalize and individualize education.”<sup>1</sup> Eighty-eight percent reported that the system does *not* assist teachers to become more effective. Ninety percent responded that the accountability system has not improved educational planning



Table 2.1 Quantitative Measures of Teacher Reaction to the Kalamazoo Accountability System

	<i>Strongly Agree</i>	<i>Agree</i>	<i>Disagree</i>	<i>Strongly Disagree</i>	<i>Number of Respondents</i>
1. The Kalamazoo accountability system is designed to personalize and individualize education.	0%	10%	36%	54%	512
2. The accountability system makes teachers more conscious of their responsibility to attend to basic academic skills.	4%	32%	35%	29%	506
3. The accountability system assists teachers to become more effective in their teaching.	1%	11%	44%	44%	511
4. Accountability as practiced in Kalamazoo creates an undesirable atmosphere of anxiety among teachers.	67%	26%	4%	3%	519
5. The accountability system has greatly improved educational planning in Kalamazoo.	1%	9%	42%	48%	507
6. The accountability system is mostly a public relations effort.	50%	40%	7%	3%	496
7. The accountability system enables citizens to determine what kind of return they are receiving on their investment in the local educational process.	1%	5%	45%	49%	520
8. Teachers who criticize the accountability system just don't want to be accountable for their teaching.	2%	3%	33%	62%	524
9. Accountability as practiced in Kalamazoo places too much emphasis on things that can be quantified so that it misses the results of teaching that are not easily measured.	61%	36%	3%	0%	522
10. Accountability data are used by administrators in a positive, constructive manner.	2%	8%	45%	45%	491



in Kalamazoo. Ninety-three percent believed, “Accountability as practiced in Kalamazoo creates an undesirable atmosphere of anxiety among teachers.” Ninety percent asserted, “The accountability system is mostly a public relations effort.” Nor did teachers feel that the accountability system fairly reflected what they did as teachers. Ninety-seven percent of the teachers *agreed* on the following:

Accountability as practiced in Kalamazoo places too much emphasis on things that can be quantified so that it misses the results of teaching that are not easily measured.

Teachers felt that the school administration had failed to develop a useful, positive approach to educational accountability. Ninety percent reported that administrators do *not* use accountability data in a positive, constructive manner.

It is relatively clear from these statements that most teachers who responded to the questionnaire were negative about the accountability system. When school officials and school board members reviewed the questionnaire results, however, many of them immediately dismissed those results by arguing that they had never expected teachers to like the system, teachers didn't really want to be accountable, and the teacher unions had told their teachers to respond negatively anyway. In short, many school officials and school board members dismissed the questionnaire results as biased, inaccurate, and the results of a teacher union campaign wherein teachers were told how to respond to the questionnaire in order to discredit the school authorities.

The same questionnaire included two open-ended questions. The first was placed midway through the questionnaire, and the second came at the end of the questionnaire.

1. Please use this space to make any further comments or recommendations concerning any component of the accountability system.
2. Finally, we'd like you to use this space to add any additional comments you'd like to make about any part of the Kalamazoo accountability system.

Three hundred seventy-three teachers (70 percent of those who responded to the questionnaire) took the time to respond to one of these questions. All of the comments made by teachers were typed verbatim and included in the report. These open-ended data filled



101 pages of single-spaced comments. When the school officials and school board members rejected the questionnaire data, rather than argue with them about the meaningfulness of teacher responses to the standardized items, we asked them to turn to the pages of open-ended teacher comments and to simply read at random what teachers said. The kind of comments they read, and could read on virtually any page in the report, are reproduced below. These are six representative responses from the middle pages of the report.

*Teacher Response #284:*

I don't feel that fear is necessary in an accountability situation. The person at the head of a school system has to be human, not a machine. You just don't treat people like they are machines!

The superintendent used fear in this system to get what he wanted. That's very hard to explain in a short space. It's something you have to live through to appreciate. He lied on many occasions and was very deceitful. Teachers need a situation where they feel comfortable. I'm not saying that accountability is not good. I am saying the one we have is *lousy*. It's hurting the students—the very ones we're supposed to be working for.

*Teacher Response #257:*

This system is creating an atmosphere of fear and intimidation. I can only speak for the school I am in, but people are tense, hostile and losing their humanity. Gone is the good will and team spirit of administration and staff and I believe this all begins at the top. One can work in these conditions but why, if it is to "shape up" a few poor teachers. Instead, it's having disastrous results on the whole faculty community.

*Teacher Response #244:*

In order to fully understand the oppressive, stifling atmosphere in Kalamazoo you have to "be in the trenches"—the classrooms.

In 10 years of teaching I have never ended a school year as depressed about "education" as I have this year. If things do not improve in the next 2 years I will leave education. The Kalamazoo accountability system must be viewed in its totality and not just the individual component parts of it. In toto it is oppressive and stifling.

In teaching government and history, students often asked what it was like to live in a dictatorship—I now know firsthand. The superintendent with his accountability model and his abrasive condescending manner has managed in three short years to destroy teacher morale and effective creative classroom teaching.



Last evening my wife and I went to an end of the school year party. The atmosphere there was strange—little exuberance, laughter or release. People who in previous years laughed, sang and danced were unnaturally quiet and somber. Most people went home early. The key topic was the superintendent, the school board election, and a millage campaign. People are still tense and uncertain.

While the school board does not “pay us to be happy” it certainly must recognize that emotional stability is necessary for effective teaching to take place. The involuntary transfers, intimidation, coercion and top to bottom “channelized” communication in Kalamazoo must qualify this school system for the list of “least desirable” school systems in the nation.

*Teacher Response #233:*

I have taught in Kalamazoo for 15 years and under 5 Superintendents. Until the present superintendent I found working conditions to be enjoyable and teachers and administration and the Board of Education all had a *good* working relationship. In the past 4 years—under the *present* superintendent—I find the atmosphere deteriorating to the point where teachers distrust each other and teachers do not trust administrators *at all!* We understand the position the administrators have been forced into and feel compassion for them—however—we still have *no trust!* Going to school each morning is no longer an enjoyable experience.

*Teacher Response #261:*

A teacher needs some checks and balances to function effectively; it would be ridiculous to think otherwise—if you are a concerned teacher. But in teaching you are not turning out neatly packaged little mechanical products all alike and endowed with the same qualities. This nonsensical accountability program we have here makes the superintendent look good to the community. But someone who is in the classroom dealing with all types of kids, some who cannot read, some who hardly even come to school, some who are in and out of jail, this teacher can see that and the rigid accountability model that neglects the above mentioned problems is pure “BULLSHIT.”

*Teacher Response #251:*

“Fear” is the word for “accountability” as applied in our system.

My teaching before “Accountability” is the same as now. “Accountability” is a political ploy to maintain power. Whatever good there may have been in it in the beginning has been destroyed by the awareness that each new educational “system” has at its base a political motive. Students get screwed. . . .



The bitterness and hatred in our system is incredible. What began as “noble” has been destroyed. You wouldn’t believe the new layers of administration that have been created just to keep this monster going. Our finest compliment around our state is that the other school systems know what is going on and are having none of it. Lucky people. Come down and visit in hell—sometime.

## THE PURPOSE OF OPEN-ENDED RESPONSES

The questionnaire results sampled above illustrate the difference between qualitative measurement based on responses to open-ended questions and quantitative measurement based on scales composed of standardized questionnaire items. Quantitative measures are succinct, parsimonious, and easily aggregated for analysis; quantitative data are systematic, standardized, and easily presented in a short space. By contrast, the qualitative measures are longer, more detailed, and variable in content; analysis is difficult because responses are neither systematic nor standardized. Yet the open-ended responses permit one to understand the world as seen by the respondents. The purpose of gathering responses to open-ended questions is to enable the researcher to understand and capture the points of view of other people without predetermining those points of view through prior selection of questionnaire categories. As Lofland explains:

In order to capture participants “in their own terms” one must learn *their* categories for rendering explicable and coherent the flux of raw reality. That, indeed, is the first principle of qualitative analysis [Lofland, 1971:7; italics added].

Direct quotations are a basic source of raw data in qualitative measurement, revealing respondents’ level of emotion, the way in which they have organized their world, their thoughts about what is happening, their experiences, and their basic perceptions. The task for the qualitative methodologist is to provide a framework within which people can respond in a way that represents accurately and thoroughly their points of view about the world, or that part of the world about which they are talking—for example, their experience with a particular program being evaluated. As Denzin (1978:10) has noted, too often social scientists “enter the field with preconceptions that prevent them from allowing those studied to ‘tell it as they see it.’”



I purposefully selected questionnaire data to illustrate the differences between quantitative and qualitative measurement because open-ended responses on questionnaires represent the most elementary form of qualitative data. There are severe limitations to open-ended data collected in writing on questionnaires; limitations related to the writing skills of respondents, the impossibility of probing or extending responses, and the effort required of the person completing the questionnaire. Yet, even at this elementary level of measurement, the depth and detail of feelings revealed in the open-ended comments of the Kalamazoo teachers illustrate the fruit of qualitative methods.

While the comparison above is based on qualitative data collected from open-ended questionnaire items, the major way in which the qualitative methodologist seeks to understand the perceptions, feelings, and knowledge of people is through in-depth, intensive interviewing. The chapter on interviewing will discuss ways of gathering high quality information from people—data which reveal experiences with program activities and perspectives on program impacts from the points of view of participants, staff, and others involved in and knowledgeable about the program being evaluated.

And what was the impact of the qualitative evaluation data collected from teachers in Kalamazoo? You will recall that many of the school board members initially dismissed the standardized questionnaire responses as biased, rigged, and the predictable result of the union's campaign to discredit school officials. However, after reading through a few pages of teachers' own personal comments, after hearing about teachers' experiences with the accountability system *in their own words*, the tenor of the discussion about the evaluation report changed. School board members could easily reject what they perceived as a "loaded" questionnaire. They could not so easily dismiss the anguish, fear, and depth of concern revealed in the teachers' own reflections. Discussion of the evaluation results shifted from an attack on the measures used to the question, "What do you think we should do?"

During the summer of 1976, following discussion of the evaluation report, the superintendent "resigned." The new superintendent and school board in 1976-1977 used the evaluation report as a basis for starting fresh with teachers. A year later the teacher association officials reported a new environment of teacher-administration cooperation in developing a mutually acceptable accountability system. The evaluation report did not directly cause these changes. Many other factors were involved in Kalamazoo at that time. However, the qualitative information in the evaluation report revealed



the full scope and nature of teachers' feelings about what it was like to work in the atmosphere created by the accountability system. The depth of those feelings as expressed in the teachers' own words became part of the impetus for change in Kalamazoo.

## MEASUREMENT BY OBSERVATION

What people say is a major source of qualitative data, whether what they say is obtained verbally through an interview or in written form through document analysis or survey responses. There are limitations, however, to how much can be learned from what people say. To understand fully the complexities of many program situations, direct participation in and observation of the program may be the best evaluation method. Howard S. Becker, one of the leading practitioners of qualitative methods in the conduct of social science research, argues that participant observation is the most comprehensive of all types of research strategies.

The most complete form of the sociological datum, after all, is the form in which the participant observer gathers it: an observation of some social event, the events which precede, and follow it, and explanations of its meaning by participants and spectators, before, during, and after its occurrence. Such a datum gives us more information about the event under study than data gathered by any other sociological method [Becker and Geer, 1970:133].

Observational data, especially participant observation, permits the evaluator to understand a program to an extent not entirely possible using only the insights of others obtained through interviews. Of course, not everything can be directly observed or experienced—and participant observation is a highly labor-intensive and therefore relatively expensive research strategy. In a later chapter strategies for using observational methods, including both participant and nonparticipant approaches, will be discussed at length. My purpose at this point is simply to give the reader another taste of the fruits of qualitative methods. Before discussing how to collect observational evaluation data, it is helpful to know what such data should look like.

The purpose of observational analysis is to take the reader into the setting that was observed. This means that observational data must have depth and detail. The data must be descriptive, sufficiently descriptive that the reader can understand what occurred and how it occurred. The observer's notes become the eyes, ears, and perceptual senses for the reader. The descriptions must be factual, accurate, and



thorough without being cluttered by irrelevant minutiae and trivia. The basic criterion to apply to a recorded observation is whether or not that observation permits the reader to enter the situation.

The observation which follows is meant to illustrate what such a descriptive account would look like. This evaluation excerpt describes a two-hour observation of a parent discussion in a parent education program. The purpose of the program, one of twenty-two such state-supported programs, is to increase the skills, knowledge, and confidence of parents. The program is also aimed at providing a support group for parents. In funding the program, legislators emphasized that they did not want parents to be told how to rear their children. Rather, the purpose of parent education sessions was to increase the options available to parents so that they could make conscious choices about their own parenting styles and increase their confidence about the choices they make. Parents were also to be treated with respect and to be recognized as the primary educators of their children—in other words, the early childhood educators were not to impose their expertise upon parents, but rather to make clear that parents are the real experts about their own children.

Site visits were made to all programs, and parenting discussions were observed on each site visit. Descriptions of these sessions then became the primary data of the evaluation. In short, *the evaluators were to be the eyes and ears of the legislature and the state program staff, permitting them to understand what was happening in the parent sessions.* Descriptive data about the sessions also provided a mirror for the staff who conducted those sessions, a way of looking at what they were doing to see if that was what they wanted to be doing. What follows is a description from one such session. The criterion that should be applied in reading this description is whether or not sufficient data are provided to take readers into the setting and permit them to make a judgment about the nature and quality of parent education being provided.

#### **OBSERVATION DATA ILLUSTRATED: A DISCUSSION FOR MOTHERS OF TWO-YEAR-OLDS**

The discussion component of this parent education program operates out of a small classroom in the basement of a church. The toddler center is directly overhead on the first floor so that noises made by the children these mothers have left upstairs can be heard during the discussion. The room is just large enough for the twelve mothers, one staff person, and myself to sit along three sides of the



room. The fourth side is used for a movie screen. (The staff person told me afterward that smoking had been negotiated and agreed on among the mothers.) The seats are comfortable, including two couches. A few colorful posters with pictures decorate the walls. Small tables are available for holding coffee cups during the discussion. The back wall is lined with brochures on child care and child development, and a metal cabinet in the room holds additional program materials.

The session begins with mothers watching a twenty-minute film about nursery school children. The film forms the basis for getting discussion started about what two-year-olds do. Mothers are asked by the staff person to begin by picking out from the film things that their own children do and talking about the way that some of the problems with children were handled in the film. For the most part, mothers share positive experiences about their two-year-olds. The focus of the discussion turns quickly to what happens as children grow older, how they change and develop.

Staff person comments: "Don't worry about what kids do at a particular age. Like don't worry that your kid has to do a certain thing at age two or else he's behind in development or ahead of development. There's just a lot of variation in the ages at which kids do things."

The discussion is free-flowing and, once begun, is not directed by the staff person. Mothers talk back and forth to each other, sharing experiences about their children. A mother will bring up a particular point and other mothers will talk about their own experiences as they want to. For example, one of the topics is the problem a mother is having with her child urinating in the bathtub. Other mothers share their experiences with this problem, ways of handling it, and whether or not to be concerned about it. The crux of that discussion seems to be that it is not a big deal and not something that the mother ought to be terribly concerned about. It is important not to make it a big deal for the child; the child would outgrow it.

The discussion turns to things that two-year-olds can do around the house to help their mothers. . . . This is followed by some discussion of the things that two-year-olds can't do and some of their frustrations in doing things. There is a good deal of laughing, sharing of funny stories about children, and sharing of frustrations about children. The atmosphere is informal and there is a good deal of intensity in listening. Mothers seem especially to pick up on things that they share in common about the problems they have with their children.



Another issue from another mother is the problem of her child pouring out her milk. She asks, "What does it mean?" This question elicits some suggestions about using water aprons and cups that don't spill and other mothers' problems with this kind of issue, but the discussion is not focused and does not really come to very much closure. The water apron suggestion brings up a question about whether or not a plastic bag is okay. The discussion turns to the safety problems with different kinds of plastic bags. About twenty minutes of discussion have now taken place. (At this point one mother leaves because she hears her child crying upstairs.)

The discussion returns to the questions around giving children baths. Staff member interjects: "Two-year-olds should not be left alone in the bathtub." With reference to the earlier discussion about urinating in the bathtub, a mother interjects that urine in the bathwater is probably better than lake water that kids swim in. The mother with the problem of urination in the bathtub says again, "It really bugs me when he urinates in the bathtub." Staff member responds, "It really is your problem, not his. If you can calm yourself down, he'll be okay."

At a lull in the discussion, the staff member asks: "Did you agree with everything in the movie?" The mothers talk a bit about this and the question focuses upon an incident in the movie where one child bit another. Mothers share stories about problems they've had with their children biting. Staff member interjects: "Biting can be dangerous. It is important to do something about biting." The discussion turns to what to do. One mother suggests biting the child back. Another mother suggests that kids will work it out themselves by biting each other back. Mothers get very agitated, more than one mother talking at a time. Staff person asks people to "cool it," so that only one person talks at a time. (Mother who had left returns.)

The discussion about biting leads to a discussion about child conflict and fighting in general, for example, the problem of children hitting each other or hitting their mothers. Again, the question arises about what to do. One mother suggests that when her child hits her she hits him back, or when her child bites her she bites him back. Staff person interjects: "Don't model behavior you don't like," and goes on to explain that her philosophy is that you should not do things as a model for children that you don't want them to do. She says that that works best for her; however, other mothers may find other things that work better for them. Staff member comments that hitting back or biting back is a technique suggested by Dreikurs. She says she dis-



agrees with that technique, "but you all have to decide what works for you."

(About forty minutes have now passed since the film, and seven of the eleven mothers have participated, most of them actively. Four mothers have not participated.)

Another mother brings up a new problem. Her child is destroying her plants, dumping plants out, and tearing them up. "I really get mad." She says that the technique she has used for punishment is to isolate the child. Then she asks a question: "How long do you have to punish a two-year-old before it starts working?"

This comment is followed by intense discussion with several mothers making comments: (This discussion is reproduced relatively in full to show the type of discussion that occurred.)

Mother No. 2: "Maybe he needs his own plant. Sometimes it helps to let a child have his own plant to take care of and then he comes to appreciate plants."

Mother No. 3: "Maybe he likes to play in the dirt. Does he have his own sand or dirt to play in around the house?"

Mother No. 4: "... oatmeal is another good thing to play in ..."

Staff: "Rice is another thing that children like to play in and that is clean, good to use indoors."

Mother No. 5: "Some things to play in would be bad or dangerous. For example, powdered soap isn't a good thing to let kids play in."

Mother No. 2: "Can you put the plants where he can't get at them?"

Mother with problem: "I have too many plants, I can't put them all out of the way."

Staff: "How old is your child?"

Mother with problem: "Two."

Staff: "Can you put the plants somewhere else or provide a place to play with dirt or rice?"

Mother with problem kind of shakes her head no. Staff person goes on: "Another thing is to tell the kid the plants are alive, to help them learn respect for living things. Tell him that those plants are alive and that it hurts them. Give him his own plant that he can get an investment in."

Mother with problem: "I'll try it."

Mother No. 2: "You've got to be fair about a two-year-old. You can't expect them not to touch things. It's not fair. I try hanging all my plants."

Staff interjection: "Sometimes just moving a child bodily away from the thing you don't want him to do is the best technique."



Mother No. 4: "They'll outgrow it anyway."

Mother with problem: "Now he really dumps them and I get angry."

Staff: "Maybe he feels a rivalry with the plants if you have so many. Maybe he's trying to compete."

Mother No. 3: "Let him help with the plants. Do you ever let him help you take care of the plants?"

Mother No. 6: "Some plants are dangerous to help with."

Staff: "Some dangerous house plants are poison."

She reaches up and pulls down a brochure on plants that are dangerous and says that she has brochures for everyone. Several people say that they want brochures and she goes to the cabinet to make them available. One mother who has not participated verbally up to this point specifically requests a brochure.

This is followed by a discussion of child-proofing a house as a method of child rearing versus training the child not to touch things, but with less emphasis on child-proofing, that is, removing temptation versus teaching children to resist temptation. One parent suggests, in this context, that children be taught one valuable thing at a time. Several mothers give their points of view.

Staff: "The person who owns the house sets the rules. Two-year-olds can learn to be careful. But don't go around all day long saying, "no, no."

The time had come for the discussion to end. The mothers stayed around for about 15 minutes talking to each other informally, going up and getting their children and getting them dressed. Some brought them back down. They seemed to have enjoyed themselves and continued the discussions informally. One mother with whom the staff person had disagreed about the issue of whether or not it was all right to bite or hit children back as a way to teach them not to do something stopped to continue the discussion. The staff person said she hoped the mother understood that she respected her right to have her own views and that she wasn't trying to tell her what to do, but she disagreed and felt everybody had a right to their own opinion; that part of the purpose of the group was for everyone to be able to come together to appreciate their own points of view and understand what worked for them. The mother said that she certainly didn't feel badly about the disagreement and she knew that some things that worked for other people didn't work for her and that she had her own ways, but that she really enjoyed the group. The staff person cleaned up the room, and the session ended.



## THE RAW DATA OF QUALITATIVE MEASUREMENT

The description of this parenting session is aimed at permitting the reader to understand what occurred in the session. These data are *descriptive*. The purpose of the description is to take the reader into the setting. The data do not include judgments about whether or not what occurred was good or bad, appropriate or inappropriate, or any other interpretive judgments. The data simply describe what occurred. State legislators, program staff, parents, and others used this description, and descriptions like this from other programs, to discuss what they wanted the programs to be and do. The descriptions helped them make explicit their own judgmental criteria. In a later chapter the issue of interpreting qualitative data will be addressed at some length.

This chapter has been aimed at describing what qualitative data look like. Basically, qualitative data consist of quotations from people and descriptions of situations, events, interactions, and activities. The purpose of these data is to understand the point of view and experiences of other persons.

Sociologist John Lofland has suggested that there are four elements in collecting qualitative data. First, the qualitative methodologist must *get close enough* to the people and situation being studied to be able to understand the depth and details of what goes on. Second, the qualitative methodologist must aim at capturing what actually takes place and what people actually say: *the perceived facts*. Third, qualitative data consist of a great deal of *pure description* of people, activities, and interactions. Fourth, qualitative data consist of direct quotations from people, both what they speak and what they write down.

The commitment to get close, to be factual, descriptive and quotive, constitutes a significant commitment to represent the participants *in their own terms*. This does not mean that one becomes an apologist for them, but rather that one faithfully depicts what goes on in their lives and what life is like for them, in such a way that one's audience is at least partially able to project themselves into the point of view of the people depicted. They can "take the role of the other" because the reporter has given them a living sense of day-to-day talk, day-to-day activities, day-to-day concerns and problems. . . .

A major methodological consequence of these commitments is that the qualitative study of people *in situ* is a *process of discovery*. It is of necessity a process of learning what is happening. Since a major part of what is happening is provided by people in their own terms, one must



find out about those terms rather than impose upon them a preconceived or outsider's scheme of what they are about. It is the observer's task to find out what is fundamental or central to the people or world under observation [Lofland, 1971:4].

## THE FRUIT OF QUALITATIVE METHOD REVISITED

This chapter began with the parable of the man who traveled far in search of a widely proclaimed food called "fruit." When finally directed to a fruit tree, he confused the spring blossom of the tree with the fruit of the tree. Finding the blossom to be tasteless, he dismissed all he had heard about fruit as a hoax—and went on his way. This chapter has been aimed at illustrating what qualitative data look like so that the person in search of the fruits of qualitative methods will know what to look for—and to know when the real thing has been obtained. To close this chapter it may be instructive to consider two other short parables about the search for fruit.

While the first seeker after fruit arrived too early to experience the ripened delicacy, and tasted only the blossom, a second seeker after fruit arrived at a tree that had been improperly cultivated, so that its fruit was shriveled and bitter. This bad fruit had been left to rot. Not knowing what good fruit looked like, he sampled the bad. "Well, I've seen and tasted fruit," he said, "and I can tell you for sure that it's terrible. I've had it with fruit. Forget it. That stuff is awful." He went on his way and his journey was wasted.

One can hope that such a foolish mistake is less likely today, since early in school students are taught the danger of generalizing from limited cases. Yet rumors persist that some people continue to reject all qualitative data as worthless (and "rotten"), having experienced only bad samples produced with poor methods.

A third seeker after fruit arrived at the same tree which produced the shriveled and bitter fruit. He picked some of the rotting fruit and examined it. He took the fruit to a farmer who cultivated fruit trees with great success. The farmer peeled away the rotten exterior and exposed the stone inside. The farmer told him how to plant the stone, cultivate the resulting tree, and harvest the desired delicacy. The farmer also gave him a plump, ripe sample to taste. Once the seeker after fruit knew what fruit really was, and once he knew that the stone he held in his hand was a seed, all he had to do was to plant and tend properly the tree's growth and work for the eventual harvest—the fruit. Though there was much work to be done and many things to be learned, the resulting high quality fruit was worth the effort.



## NOTE

1. The results reported here are for combined “strongly agree” and “agree” responses *or* combined “strongly disagree” and “disagree” responses.



## THE STRATEGY OF QUALITATIVE METHODS

“Tell us again, Master, how it was in the beginning.”

“In the beginning special gifts were given to different groups of people. The caregivers were endowed with compassion for the less fortunate. The engineers were given the ability to see what was not yet there. The carpenters were given patience to set straight lines and perfect angles. The technicians were provided with diligence so that they might conscientiously follow the blueprints and detailed directions of others. The experimental scientists were given the certain belief that the world could be manipulated according to their vision of it. . . . And finally there remained one last group and one last gift. These were the explorers. To them was given the gift of curiosity that they might forever observe the world as it is and seek to understand the many wonders of the world and the special gifts given to others.”

“But what of the evaluators? You have not mentioned their special gift.”

Halcolm smiled. “The evaluators, dear children, were spread throughout all the other groups, each endowed with the special gift of their own group, and each using that gift in a special way.”

“But does that not make for much arguing among evaluators about who has the most special gift of all?”

Halcolm smiled.

*From Halcolm's Evaluation Oracles*

### THE NEED FOR A STRATEGIC FRAMEWORK

A strategy is a plan of action. A strategy provides basic direction. It permits seemingly isolated activities to fit together; it moves separate efforts toward a common, integrated purpose. An evaluation



research strategy, then, provides basic direction for the evaluator; it provides guidance in selecting particular techniques or methodological practices for specific settings.

The term “strategy” is used purposefully. Methods decisions represent strategic choices. The term “strategy” is meant to make clear that there are various levels of decision-making involved in designing an evaluation study. The evaluator moves back and forth between general methods strategies and specific evaluation questions to establish the relevance and meaningfulness of particular operational procedures to the program under study. The strategy of a research approach must be translated into concrete data-gathering techniques, instruments, and operations. How that translation occurs will be the subject of later chapters. The purpose of this chapter is to present the strategy of qualitative methods, including fundamental assumptions and epistemological ideals.

### *A Holistic View*

Researchers using qualitative methods strive to understand phenomena and situations as a whole; evaluators using qualitative methods attempt to understand programs as wholes. The researcher strives to understand the gestalt, the totality, and the unifying nature of particular settings. This holistic approach assumes that the whole is greater than the sum of its parts; it also assumes that a description and understanding of a program’s context is essential for understanding the program. Thus, it is insufficient simply to study and measure the parts of a situation by gathering data about isolated variables, scales, or dimensions. In contrast to experimental designs which manipulate and measure the relationships among a few carefully selected and narrowly defined variables, the holistic approach to research design is open to gathering data on any number of aspects of the setting under study in order to put together a complete picture of the social dynamic of a particular situation or program. This means that at the time of data collection each case, event, or setting being studied is treated as a unique entity with its own particular meaning and constellation of relationships emerging from and related to the context within which it exists.

### *An Inductive Approach*

A qualitative research strategy is inductive in that the researcher attempts to make sense of the situation without imposing preexisting expectations on the research setting. Qualitative designs begin with specific observations and build toward general patterns. Categories or dimensions of analysis emerge from open-ended observations as



the researcher comes to understand organizing patterns that exist in the empirical world under study. This contrasts with the hypothetico-deductive approach of experimental designs which require the specification of main variables and the statement of specific research hypotheses *before* data collection. A specification of research hypotheses based on an explicit theoretical framework means that general principles provide the framework for understanding specific observations or cases. The researcher must then decide in advance what variables are important and what relationships among those variables are expected. The strategy in qualitative designs is to allow the important dimensions to emerge from analysis of the cases under study *without presupposing in advance what those important dimensions will be*. The qualitative methodologist attempts to understand the multiple interrelationships among dimensions which emerge from the data without making prior assumptions about the linear or correlative relationships among narrowly defined, operationalized variables. In short, an inductive approach to evaluation research means that an understanding of program activities and outcomes emerges from experience with the program. Theories about what is happening in a program are grounded in this program experience, rather than imposed on the program a priori based on hypothetico-deductive constructions.

Program evaluations may be inductive in two ways. *Within* programs an inductive approach begins with the individual experiences of program participants without pigeonholing or delimiting what those experiences will be in advance of fieldwork. *Between* programs the inductive approach looks for unique program characteristics that make each program a case unto itself. At either level generalizations may later emerge as case materials are content analyzed, but the initial focus is on fully understanding individual cases *before* combining or aggregating those unique cases.

### *Naturalistic Inquiry*

Qualitative designs are naturalistic in that the researcher does not attempt to manipulate the research setting. The research setting is a naturally occurring event, program, relationship, or interaction that has no predetermined course established by and for the researcher. Rather, the point of using qualitative methods is to understand naturally occurring phenomena in their naturally occurring states.

Willems and Raush (1969:3) define naturalistic inquiry as "the investigation of phenomena within and in relation to their naturally occurring context." Egon Guba (1978), in his extensive review of naturalistic inquiry in educational evaluation, identifies two dimensions



along which types of scientific inquiry can be described: (1) the extent to which the scientist manipulates some phenomenon in advance in order to study it, and (2) the extent to which constraints are placed on output measures; that is, the extent to which predetermined categories or variables are used to describe the phenomenon under study. He then defines "naturalistic inquiry" as a "discovery-oriented" approach which minimizes investigator manipulation of the study setting and places no prior constraints on what the outcomes of the research will be (in other words, an inductive approach). Naturalistic inquiry is thus contrasted to experimental research where ideally the investigator attempts to completely control conditions of the study by manipulating, changing, or holding constant external influences and where a very limited set of outcome variables are measured.

The experimentalist evaluator enters the program, in its simplest form, at two points in time, pretest and posttest, and compares the treatment group to some control group on a limited set of standardized measures. Such designs assume a single, identifiable, isolated, and measurable treatment. What is more, such designs assume that, once introduced, the treatment remains relatively constant and unchanging.

While there are some narrow, largely technical treatments that fit this description, in practice program efforts are often quite comprehensive and anything but static treatments. Programs frequently change as practitioners learn what works and what does not, as they experiment and grow and change their priorities. This, of course, creates considerable difficulty for experimentalist evaluators who need specifiable, unchanging treatments to relate to specifiable, predetermined outcomes. Experimental evaluation designs work best when it is possible to limit program adaptation and improvement so as not to interfere with the rigor of the research design (Parlett and Hamilton, 1976).

By contrast, under field conditions where programs are subject to change and redirection, naturalistic inquiry replaces the fixed treatment/outcome emphasis of the controlled experiment with a dynamic, process orientation. A dynamic evaluation is not tied to a single treatment and predetermined goals or outcomes, but focuses on the actual operations and impacts of a program over a period of time. The evaluator sets out to understand and document the day-to-day reality of the setting or settings under study. The evaluator makes no attempt to manipulate, control, or eliminate situational variables or program developments, but accepts the complexity of a



changing program reality. The data of the evaluation include outcomes, changes in treatments, and patterns of action, reaction, and interaction.

## GETTING CLOSE TO THE PHENOMENON UNDER STUDY

The strategic mandate to be holistic, inductive, and naturalistic means getting close to the phenomenon under study. The evaluator using qualitative methods attempts to understand the setting under study through direct personal contact and experience with the program. Engaging in holistic-inductive research through naturalistic inquiry represents a comprehensive strategy for describing and understanding human service and education programs that includes specification of the role of the researcher in conducting the evaluation.

Qualitative research designs require that the evaluator get close to the people and situations being studied in order to understand the minutiae of program life. The evaluator gets close to the program through physical proximity for a period of time, as well as through development of closeness in the social sense of intimacy and confidentiality. That many quantitative methodologists fail to ground their findings in qualitative understanding poses what Lofland calls a major contradiction between their public insistence on the adequacy of statistical portrayals of other humans and their personal, everyday dealings with and judgments about other human beings.

In everyday life, statistical sociologists, like everyone else, assume that they do not know or understand very well people they do not see or associate with very much. They assume that knowing and understanding other people require that one see them reasonably often and in a variety of situations relative to a variety of issues. Moreover, statistical sociologists, like other people, assume that in order to know or understand others one is well advised to give some conscious attention to that effort in face-to-face contacts. They assume, too, that the internal world of sociology—or any other social world—is not understandable unless one has been part of it in a face-to-face fashion for quite a period of time. How utterly paradoxical, then, for these same persons to turn around and make, by implication, precisely the opposite claim about people they have never encountered face-to-face—those people appearing as numbers in their tables and as correlations in their matrices! [Lofland, 1971:3]

The desire to get close to the situation in order to increase understanding, to generate a holistic description of the situation, to



proceed inductively, and to study programs in their naturally occurring complexity “involves the studied commitment to actively enter the worlds of interacting individuals” (Denzin, 1978:8-9). This makes possible description and understanding of *both* externally observable behaviors and internal states (world view, opinions, values, attitudes, symbolic constructs, and the like). Attention to inner perspectives does not mean administering attitude surveys. “The inner perspective assumes that understanding can only be achieved by actively participating in the life of the observed and gaining insight by means of introspection” (Bruyn, 1963:226). Actively participating in the life of the observed means, at a minimum, being willing to get close to the sources of data.

## THE ROOTS OF A QUALITATIVE RESEARCH STRATEGY

This comprehensive strategy of qualitative methods is derived from a variety of philosophical, epistemological, and methodological traditions. Qualitative methods are derived most directly from the ethnographic and field study traditions in anthropology (Pelto and Pelto, 1978) and sociology (Bruyn, 1966). More generally, the holistic-inductive paradigm of naturalistic inquiry is based on perspectives developed in phenomenology (Bussis et al., 1973; Carini, 1975), symbolic interactionism and naturalistic behaviorism (Denzin, 1978), ethnomethodology (Garfinkel, 1967), and ecological psychology (Barker, 1968). An integrating theme running through these traditions is the fundamental notion or doctrine of *verstehen*.

The basic dispute clustering around the notion of *verstehen* has typically sounded something like the following: The advocate of some version of the *verstehen* doctrine will claim that human beings can be understood in a manner that other objects of study cannot. Men have purposes and emotions, they make plans, construct cultures, and hold certain values, and their behavior is influenced by such values, plans, and purposes. In short, a human being lives in a world which has “meaning” to him, and, because his behavior has meaning, human actions are intelligible in ways that the behavior of nonhuman objects is not. The opponents of this view, on the other hand, will maintain that human behavior is to be explained in the same manner as is the behavior of other objects of nature. There are laws governing human behavior. An action is explained when it can be subsumed under some such law, and, of course, such laws are confirmed by empirical evidence [Strike, 1972:28].



The *verstehen* approach assumes that the social sciences need methods different from those used in agricultural experimentation and natural science because human beings are different from plants and nuclear particles. The *verstehen* tradition stresses understanding that focuses on the *meaning* of human behavior, the context of social interaction, an *empathetic* understanding based on subjective experience, and the connections between subjective states and behavior. The tradition of *verstehen* or understanding places emphasis on the human capacity to know and understand others through sympathetic introspection and reflection from detailed description and observation.

Bogdan and Taylor (1975) contrast the *verstehen* tradition rooted in phenomenology to logical positivism, the dominant social science perspective in the twentieth century:

Two major theoretical perspectives have dominated the social science scene. One, *positivism*, traces its origins to the great social theorists of the nineteenth and early twentieth centuries and especially to August Comte and Emile Durkheim. The positivist seeks the *facts* or *causes* of social phenomena with little regard for the subjective states of individuals. Durkheim advises the social scientist to consider “social facts,” or social phenomena, as “things” that exercise an external and coercive force on human behavior.

The second theoretical perspective, which, following the lead of Irwin Deutscher, we will describe as *phenomenological*, stems most prominently from Max Weber. The phenomenologist is concerned with *understanding* human behavior from the actor’s own frame of reference. . . . The phenomenologist examines how the world is experienced. For him or her the important reality is what people imagine it to be.

*Since the positivists and the phenomenologists approach different problems and seek different answers, their research will typically demand different methodologies* [Bogdan and Taylor, 1975:2; italics in the original].

The phenomenological tradition in qualitative methods proposes an active, involved role for the social scientist/evaluation researcher.

Hence, insight may be regarded as the core of social knowledge. It is arrived at by being on the inside of the phenomena to be observed. . . . It is participation in an activity that generates interest, purpose, point of view, value, meaning, and intelligibility, as well as bias [Wirth, 1949:xxii].



This is a quite different scientific process from that envisioned by the classical, experimental approach to science.

This in no way suggests that the researcher lacks the ability to be scientific while collecting the data. On the contrary, it merely specifies that it is crucial for validity—and, consequently, for reliability—to try to picture the empirical social world as it actually exists to those under investigation, rather than as the researcher imagines it to be [Filstead, 1970:4].

Thus, the importance of such field techniques as participant observation, in-depth interviewing, detailed description, and qualitative field notes.

### FROM STRATEGIC IDEALS TO PRACTICAL TACTICS

Holistic-inductive research through naturalistic inquiry is a strategic ideal. In conceptualization a pure qualitative methods strategy emphasizes a holistic approach where the researcher neither manipulates the setting under study nor predetermines what variables or categories are worth measuring. In practice, however, it is important to recognize that *holistic-inductive analysis and naturalistic inquiry are always a matter of degree*. In making this point Guba (1978) has depicted the practice of naturalistic inquiry as a wave on which the investigator moves from varying degrees of a “discovery mode” to varying emphasis on a “verification mode” in attempting to understand the real world. As the research begins the investigator is open to whatever emerges from the data, a discovery or inductive approach. Then, as the inquiry reveals patterns and major dimensions of interest, the investigator will begin to focus on verifying and elucidating what appears to be emerging—a more deductive approach to data collection and analysis. In program evaluation in particular the researcher, through feedback of descriptions and analysis to program participants and staff, may also begin to affect directly and help manipulate the program, thus moving away from a purely naturalistic inquiry. As evaluative feedback is utilized the researcher may then move back into a more naturalistic approach to observe how the changes in the program unfold.

In the same vein, the attempt to understand the program as a whole does not mean that the investigator never becomes involved in component analysis and in looking at particular variables, dimensions, and parts of a program as separate entities. Rather, it means that the



qualitative methodologist consciously works back and forth between parts and wholes, separate variables, and complex, interwoven constellations of variables in a sorting-out then putting-back-together process. Guided by the strategy that mandates the importance of striving to present a holistic picture of the program, the qualitative evaluator recognizes that certain periods during data collection and analysis may focus on component, variable, and less-than-the-whole kinds of analysis.

The practice and practicalities of fieldwork also mean that the strategic mandate to “get close” to the program or setting under study is not an absolute and fixed approach. Closeness to and involvement with the program under study are most usefully viewed as variable dimensions. The personal styles and capabilities of evaluators will permit and necessitate variance along these dimensions. Variations in types of programs and evaluation purposes will affect the extent to which the evaluator can or ought to get close to the program staff and participants. Moreover, closeness is likely to vary over the course of the evaluation. At times the evaluator may become totally immersed in the program experience. These periods of immersion may be followed by times of withdrawal and distance (for personal as well as methodological reasons), to be followed still later by new experiences of immersion in and intimacy with the program.

Qualitative methods can be used both to *discover* what is happening and then to *verify* what has been discovered. What is discovered must be verified by going back to the empirical world under study and examining the extent to which the emergent analysis fits the phenomenon and works to explain what has been observed. Glaser and Strauss (1967:3) describe what it means for results to fit and work. “By ‘fit’ we mean that the categories must be readily (not forcibly) applicable to and indicated by the data under study; by ‘work’ we mean that they must be meaningfully relevant to and be able to explain the behavior under study.” Discovery and verification mean moving back and forth between induction and deduction, between experience and reflection on experience, and between greater degrees and lesser degrees of naturalistic inquiry.

The practical tactics of qualitative methods do not undermine the strategic ideals of qualitative research. Those strategic ideals provide the basic framework out of which the practical tactics are developed and in which actual field procedures are grounded. Holistic-inductive analysis based on naturalistic inquiry constitutes the strategy of qualitative methods which provide a framework for and guidance in making practical, tactical decisions about the evaluation. A Sufi



story about the wise fool Mulla Nasrudin illustrates the importance of understanding the connections between strategic ideals and immediately practical tactics in real-world situations.

In his youth Nasrudin received training in a small monastery noted for its excellence in the teaching of martial arts. Nasrudin became highly skilled in self defense and after two years of training his superior abilities were recognized by both his peers and his teachers.

Each day it was the responsibility of one of the students to go to the village market to beg for alms and food. It happened that a small band of three thieves moved into the area and began laying in wait each night for the student from the monastery to return laden with food and alms. They would then rob the student.

After three days of such losses the monastery's few supplies were exhausted. It was Nasrudin's turn to go to the village market. His elders and peers were confident that Nasrudin's skills were more than sufficient to overcome the small band of thieves.

At the end of the day Nasrudin returned ragged, beaten, and empty-handed. Everyone was amazed. Nasrudin was immediately taken before the Master. "Nasrudin," he asked, "how is it that with all your skill in our ancient arts of defense you were overcome?"

"But I did not use the ancient arts," replied Nasrudin.

All present were dumbfounded. An explanation was demanded.

"All of our competitions are preceded by great and courteous ceremony. We have learned that the opening prayers, the ceremonial cleansing, the bow to the East—these are essential to the ancient ways. The ruffians seemed not to understand the necessity for these things. I didn't find the situation ideal enough to use the methods you have taught us, Master."

On more than one occasion evaluators have told me of their belief in the potential usefulness of qualitative methods but "I just haven't found the ideal situation in which to use them."

By way of expanding evaluators' methodological repertoire and enlarging the capacity for making real choices, it is important to consider under what conditions qualitative strategies, tempered by recognition of practical tactics, are particularly useful and appropriate. In the next chapter those evaluation situations which particularly lend themselves to the use of qualitative methods will be described and discussed. Chapter five will then discuss in more detail some of the methodological trade-offs involved in adapting the strategic ideals of qualitative methods to the practical realities of conducting program evaluations.



## EVALUATION RESEARCH STRATEGIES USING QUALITATIVE METHODS

e . val . u . a . tion (i-val ū-a sh n), n.

*Fr.* evaluator e-(L. *ex* -), out + *value*; see VALUE , e.val.u.ate, v.t.

1. to find the value or amount of; determine the worth of; appraise. 2. to determine the extent to which goals have been attained. 3. to judge effectiveness. 4. types of: outcomes, process, formative, summative, internal, external, goal-based, goal-free, systems analysis, effort, efficiency, discrepancy, expensive, inexpensive, long-and-arduous, quick-and-dirty, good, bad, indifferent, ones-you-like, ones-you-don't-like, . . . 5. synonyms: information, investigation, judgment, research, propaganda, inquisition, inveiglement, conjecture, . . . 6. family of: planning, needs assessment, quality assurance, management information systems, monitoring, politics, public relations, professional confusion, prostitution, . . .

And they said unto Halcolm, "Tell us, Aged One, what is evaluation?"

"It is many things to many people, my children," he replied solemnly, "many different things."

"So many things to so many people, Master. We are still unclear. Is there not some way to tell us more simply so we can understand what it means to evaluate?"

"It is written simply in the Great Book. To evaluate is to do unto others as you would *not* have them do unto you."

*From Halcolm's Universal Encyclopedic  
Dictionary of Evaluation*

### EVALUATION MODELS AND QUALITATIVE STRATEGIES

The purpose of this book is to review qualitative research models or strategies. The specific aim of this chapter is to



suggest when it may be particularly appropriate to use qualitative methods in evaluation research.

Certain models of evaluation research are more consonant with qualitative methods than others. Among recent efforts, House (1978) has developed perhaps the most comprehensive taxonomy of major evaluation models. He identifies eight separate models distinguishable by the audiences they address, what they assume consensus on, the outcomes they examine, the typical questions they ask, and the methods they employ. The eight models are: (1) systems analysis, (2) behavioral objectives, (3) decision-making, (4) goal-free, (5) art criticism, (6) accreditation, (7) adversary, and (8) transaction. In the first part of this chapter I shall review the compatibility of qualitative methods with these models given the epistemological assumptions of each.

## INCOMPATIBLE MODELS

Systems analysis and behavioral objectives are the traditional and dominant evaluation models concerned with efficiency, productivity, and quantitative changes in specified effects or outcomes. Variables and outcomes are specified in advance, measured in operational quantities, and, where possible, subjected to experimental manipulation. Qualitative methods derive from a different and conflicting paradigm (Patton, 1978); there is little compatibility between comprehensive qualitative methods strategies and evaluation approaches based on systems analysis or behavioral objectives.

Systems analysis requires (1) identifying the important input and output variables for programs, (2) quantifying those input and output measures, and (3) statistically analyzing the relationships between program inputs and program outcomes. Inputs are quantities such as program budget, staff size, staff-client ratios, client characteristics, baseline performance levels of clients (pretest scores), and the like. Output measures include client performance levels after the program (posttest scores), placement rates, monetary value of new client skills, new client wage levels or equivalents, and so on. Systems analysis has dominated evaluation research thinking in the federal government. It is an approach that depends on component analysis (breaking the program system into its separate parts), in contrast to the holistic perspective of qualitative methods. Important variables are identified deductively and logically before data collection begins in contrast to the inductive analytical strategy of qualitative research where important dimensions emerge from the work in the field; and program variables are manipulated or controlled where possible



(either programmatically or statistically) to test hypothesized relationships in contrast to naturalistic inquiry where the investigator studies what occurs in the program without trying to control, manipulate, or artificially interrelate what happens. The systems analyst's view of the world is incompatible with a phenomenological perspective.

The second model in the House taxonomy, the behavioral objectives model, has dominated educational evaluation since its inception. Educational outcomes are specified in terms of specific student performance criteria that are measured by standardized tests. These tests are statistically analyzed to compare individual students and groups of students. The criteria of achievement are the same for all students—not only all students in a particular school but, with nationally used norm-referenced tests, all students in the country. This standardized approach to assessing the effects of educational programs is incompatible with the assumption in qualitative research that student performance can only be understood in a specific context and in relation to the particular meanings an individual student attaches to the outcomes of his or her personal school experience. This point will be developed in greater depth later in this chapter. For the moment it is sufficient to establish the incompatibility of a behavioral objectives approach with a naturalistic inquiry approach to evaluation.

To say that qualitative methods are essentially incompatible with systems analysis and behavioral objectives is *not* to say that both kinds of data-gathering—quantitative and qualitative—cannot go on at the same time. At the 1978 meeting of the Evaluation Research Society, William Filstead presented several arguments in favor of collecting both kinds of data. Nor would I disagree that one can learn a great deal through data triangulation. But to the extent that qualitative data are collected and used in an evaluation, that evaluation is no longer an exemplar of a pure systems analysis or behavior objectives model. By definition, by epistemology, and by the very nature of the basic assumptions upon which the two models are based, qualitative methods do not fit. To make them fit is to change the models.

## SOMETIME-COMPATIBLE MODELS

Art criticism, accreditation, and the adversary approach are professional review models. Experts examine the program and make judgments based on their subjective perceptions about the standards



that are appropriate. The potential use of qualitative methods is different, however, in each case.

Accreditation teams typically rely heavily on analysis of program documents, informal interviews, and site visit observations. However, their data collection procedures are seldom systematic and rigorous; they are more likely to report judgments than data; and their methods are usually deductive—applying general criteria (either explicitly or implicitly) to a particular program to see if it meets minimal standards. Qualitative data may be collected as part of accreditation evaluations, and a comprehensive qualitative methods strategy could be used to make the accreditation approach more systematic and rigorous, but to do so would considerably change the model by diminishing the near total reliance on the prior expertise and knowledge of the selected experts.

The art criticism model is a “connoisseurship” approach articulated most fully by Elliot W. Eisner through the Stanford Evaluation Consortium. The imagery of evaluators as “connoisseurs” making critical appraisals of programs is analogous to the traditional way in which literary and artistic connoisseurs and critics work, Eisner having considerable experience as an art educator.

Guba (1978) argues that the connoisseur approach has many elements which relate it to naturalistic inquiry. My own view is that such comparisons do a disservice to both approaches. The examples of evaluation connoisseurship thus far produced by Eisner seem to indicate that there are more differences than similarities between the two approaches. Where the qualitative methodologist seeks to understand the program in its own terms, the connoisseur searches for a metaphor to portray the program. Where naturalistic inquiry is inductive, the evaluator as critic is predisposed to look at a program according to standards of quality based on experience, reference group preferences, and comparisons with some Platonic ideal. The connoisseur has no commitment to produce a descriptive, holistic analysis that brings the decision maker into the program experience so that informed judgments can be made by those decision makers; the connoisseur’s criticism and metaphors are both the data and the judgment. The danger of not clearly separating scientific inquiry using qualitative methods from evaluation connoisseurship based on expert opinion is that people who experience the expertise of the evaluation critic may think that they have also experienced the rigor of the evaluation scientist; thus, they judge the potential of the latter by the performance of the former.

The adversary model, on the other hand, is somewhat content-free insofar as type of data is concerned. Instead of specifying what data



to collect, the adversary model begins by designating the answers that are to result from the evaluation. One answer is that the program is excellent; the other answer is that the program is ineffective. Two evaluation teams are created. One is designated the advocacy team, the other the adversary team. The task of the advocacy team is to demonstrate positive program findings. The adversaries look for negative findings; they are charged with marshalling evidence to demonstrate that the program is inadequate or ought to be terminated. The advocacy-adversary model is a combination debate/courtroom approach to evaluation. As concern about the politics and pressures of evaluation grew in the early 1970s, and as critics of evaluation argued with increased vociferation that single evaluators could not maintain neutrality and objectivity throughout the evaluation process, support for the notion of the advocacy-adversary grew (Wolf, 1975; Levine, 1974; Kourilsky, 1974; Owens, 1973). The contrasting findings of the two teams are presented to some jury of decision makers for judgment.

The constraints placed on the two teams in advance of data collection to produce certain findings precludes in many ways a comprehensive qualitative evaluation strategy. It is difficult to be inductive and holistic when you are trying to prove a predetermined point. Moreover, ethical issues emerge in manipulating qualitative data to support a predetermined position if the evaluator uses the qualitative data collection strategy of establishing personal relationships built on closeness, openness, and the desire to achieve a holistic understanding of the complexities of special people in unique settings. To establish such relationships and then manipulate the data to present only part of the picture betrays the purpose of naturalistic inquiry. Thus, while descriptive observation and interview data may be gathered as part of an advocacy-adversary evaluation, a comprehensive qualitative methods strategy is largely incompatible with the adversary model. The qualitative evaluator seeks phenomenological understanding; the adversary evaluator seeks evidence to support a predetermined point of view.

## **EVALUATION MODELS COMPATIBLE WITH QUALITATIVE METHODS**

Five of the eight models identified by House I have judged to be largely incompatible with a comprehensive qualitative methods strategy; however, at least three of the models (accreditation, connoisseurship, and adversary) are amenable to the use of descrip-



tive data, though *not* to pure forms of naturalistic inquiry and holistic-inductive analysis. The remaining three evaluation models—the transaction model, goal-free evaluation, and the decision-making model—are highly compatible with a comprehensive qualitative methods strategy.

## THE TRANSACTION MODEL

The transaction model “concentrates on the educational [or program] processes themselves. . . . It uses various informal methods of investigation and has been drawn increasingly to the case study as the major methodology” (House, 1978:5). Robert Stake’s “responsive approach to evaluation” is the leading example of the transaction model.

Responsive evaluation is an alternative, an old alternative, based on what people do naturally to evaluate things, they observe and react. The approach is not new. But this alternative has been avoided in district, state, and federal planning documents and regulations because it is subjective and poorly suited to formal contracts. It is also capable of raising embarrassing questions. This chapter advocates technical steps (e.g., replication and nonverbal operationalization) to bolster the reliability of observation and opinion-gathering without sacrificing relevance [Stake, 1975:14].

The transactional model is derived for a “subjectivist epistemology” which “tends to be naturalistic” (House, 1978:5). It treats each case as unique and places prime emphasis on perception and knowing as a transactional process, thus the derivation of the label for this model.

One can study perceptions only by studying particular transactions in which the perceptions can be observed. All parts of the situation enter into the transaction as “active participants,” and do not appear as separate already-existing entities. . . . [The evaluator] affects and is affected by the situation, thus he is part of the transaction [House, 1978:9].

Stake describes in more concrete and practical terms how subjectivist epistemology is translated into the process of actually doing a “responsive evaluation.”

To do a responsive evaluation, the evaluator conceives of a plan of observations and negotiations. He arranges for various persons to observe the program, and with their help prepares brief narratives,



portrayals, product displays, graphs, etc. He finds out what is of value to his audiences, and gathers expressions of worth from various individuals whose points of view differ. Of course, he checks the quality of his records: he gets program personnel to react to the accuracy of his portrayals; and audience members to react to the relevance of his findings. He does most of this informally—iterating and keeping a record of action and reaction. He chooses media accessible to his audiences to increase the likelihood and fidelity of communication. He might prepare a final written report, he might not—depending on what he and his clients have agreed on [Stake, 1975:14].

Another variation on the transaction model is the “illuminative evaluation” approach of Parlett and Hamilton.

Illuminative evaluation takes account of the wider contexts in which educational programs function. Its primary concern is with description and interpretation rather than measurement and prediction. It stands unambiguously within the alternative anthropological paradigm. The aims of illuminative evaluation are to study the innovatory program: how it operates; how it is influenced by the various school situations in which it is applied; what those directly concerned regard as its advantages and disadvantages; and how students’ intellectual tasks and academic experiences are most affected. It aims to discover and document what it is like to be participating in the scheme, whether as teacher or pupil, and, in addition, to discern and discuss the innovation’s most significant features, recurring concomitants, and critical processes. In short, it seeks to address and to illuminate a complex array of questions. [Parlett and Hamilton, 1976:144].

The transition model is based on the same assumptions that undergird qualitative research: the importance of understanding people and programs in context; a commitment to study naturally occurring phenomena without introducing external controls or manipulation; and the assumption that understanding emerges most meaningfully from an inductive analysis of open-ended, detailed, descriptive, and quotive data gathered through direct contact with the program and its participants.

## GOAL-FREE EVALUATION

Philosopher-evaluator Michael Scriven first proposed the idea of goal-free evaluation. Essentially, goal-free evaluation means gathering data on a broad array of *actual effects* and evaluating the importance of these effects in meeting demonstrated needs. The



evaluator makes a deliberate attempt to avoid all rhetoric related to program goals; no discussion about goals is held with staff; no program brochures or proposals are read; only the program's outcomes and measurable effects are studied.

There are four reasons for doing goal-free evaluation: (1) to avoid the risk of narrowly studying stated program objectives and thereby missing important unanticipated outcomes; (2) to remove the negative connotations attached to the discovery of unanticipated effects—"The whole language of 'side-effect' or 'secondary effect' or even 'unanticipated effect' tended to be a put-down of what might well be the crucial achievement, especially in terms of new priorities" (Scriven, 1972:1-2); (3) to eliminate the perceptual biases introduced into an evaluation by knowledge of goals; and (4) to maintain evaluator objectivity and independence through goal-free conditions. In Scriven's own words:

It seemed to me, in short, that consideration and evaluation of goals was an unnecessary but also a possibly contaminating step. I began work on an alternative approach—simply the evaluation of *actual* effects against a profile of *demonstrated* needs. I call this Goal-Free Evaluation. . . .

The less the external evaluator hears about the goals of the project, the less tunnel-vision will develop, the more attention will be paid to *looking* for *actual* effects (rather than checking on *alleged* effects) [Scriven, 1972a:2].

Goal-free evaluation, in its search for "actual effects," is an inductive and holistic strategy aimed at countering the logical-deductive limitations inherent in the usual goals-based approach to evaluation. Evaluation questions are typically framed in relation to the formal, stated goals and objectives of a program. Peter Rossi states that "a social welfare program (or for that matter any program) which does not have clearly specified goals cannot be evaluated without specifying some measurable goals. This statement is obvious enough to be a truism" (Rossi and Williams, 1972:18). In a major review of the evaluation literature in education, Worthen and Sanders (1973) concluded that:

if evaluators agree in anything, it is that program objectives written in unambiguous terms are useful information for any evaluation study. Thus, program objectives and specifications become an extremely important consideration when an evaluation study is constructed [Worthen and Sanders, 1973:231].



Carol Weiss notes that “the traditional formulation of the evaluation question is: To what extent is the program succeeding in reaching its goals?” Weiss explains that “the goal must be clear so that the evaluator knows what to look for. . . . Thus begins the long, often painful process of getting people to state goals in terms that are *clear, specific, and measurable*” (Weiss, 1972b:24-26; italics in the original). *In effect these goal based, quantitative outcomes-oriented evaluations represent the extension of the behavioral objectives model beyond education to the full range of human service and social action programs.*

In contrast to the logically derived, measurable goals approach to evaluation, goal-free evaluation means gathering data directly on program effects and effectiveness without being constrained by a narrow focus on stated goals. Goal-free evaluation particularly lends itself to qualitative methods because it relies heavily on description and direct experience with the program. Moreover, and in particular, goal-free evaluation requires the evaluator to suspend judgment about what it is the program is trying to do and to focus instead on finding out what it is that actually happens in the program and as a result of the program. The evaluator thus can be open to whatever data emerge from the phenomena of the program itself.

## THE DECISION-MAKING MODEL

The primary characteristic of the decision-making model is that “the evaluation is structured by the decisions to be made. The evaluator is to supply information on these particular decisions” (House, 1978:4). Identification of future decisions to be made with evaluation findings constitutes the first step in Thompson’s (1975) “evaluation for decision” approach. The decision-making model, however, need not be narrowly construed as aimed at providing technical information about a narrow range of future actions—that is, to say that evaluation questions ought to be framed in a future action context does not mean that they need be aimed at some *single* future decision, though on occasion that may be possible and appropriate. Rather, more generally, decision makers and information users ought to be able to indicate where their knowledge uncertainties lie; what activities, actions, and options are clouded by those uncertainties; and how evaluative information would increase their potential for doing a better job and making the program more effective. In short, the evaluator attempts to “frame the decision context for the evaluation” (Alkin, 1975b).



The decision-making model does not imply any particular methodological stance. Indeed, it is the most open of all the models to a full variety of methodological strategies, both pure and mixed forms. The methods to be used would depend on what evaluative information is needed to help make specified decisions.

### BEYOND MODELS: THE PRACTICE OF EVALUATION

In 1977 a symposium session at the American Educational Research Association brought several major evaluation researchers (Worthen, Stake, Stufflebeam, Popham) together to consider the question "Are Synthesis and Resolution of Evaluation models possible?" The basic theme running through the comments of these evaluators was that their work is seldom guided by and directly built on specific evaluation models. Rather, each evaluation problem is approached as a problem to be solved—and the resulting design reflects their thinking about the problem as opposed to an attempt to carefully follow a prescriptive model. In effect, these experienced evaluators were describing how the *practice of evaluation research* requires more flexibility than is likely to be provided by any single model.

The previous chapter on qualitative research strategies began by describing the pure or ideal approach to naturalistic inquiry and then moved to discussion of how that ideal is altered and adapted in actual practice. This chapter has considered eight ideal-typical evaluation models and their relative compatibility with ideal-typical qualitative evaluation methods. This discussion of the match between models and methods has been aimed at getting a sense of how the field of evaluation research has been evolving conceptually, and the methodological implications of these evolving conceptual models. It is important to understand ideal conceptualizations so that we can be aware of when we have deviated from ideal-typical cases and the implications of such deviations. It is the discrepancy between actual practice and ideal conceptualizations that often leads to new, more meaningful and useful models.

In essence, the options open to evaluators have expanded tremendously in recent years. There are more models to choose from for those who like to follow models; there are legitimate variations in, deviations from, and combinations of models; and there is the somewhat model-free approach of problem-solving evaluators who are active, reactive, and adaptive in the context of specific evaluation



situations and information needs. Cutting across the evaluation model options are a full range of methods possibilities, the choice in any particular evaluation to be determined by the purpose of the evaluation, and the nature of the evaluation process.

## UTILIZATION-FOCUSED EVALUATION

The utilization-focused approach to evaluation (Patton, 1978) represents an attempt to move beyond models to the practice of evaluation; it is an explicit recognition of the expanded options available to active-reactive-adaptive evaluators. Utilization-focused evaluation describes an evaluative process for making decisions about the content of an evaluation—but the content itself is not specified or implied in advance. Thus, any of the eight models reviewed here, or adaptations and combinations of those models, might emerge as the guiding direction in utilization-focused evaluation. While the processes represented by the utilization-focused approach to evaluation are closest to those embodied in the decision-making and transaction models, utilization-focused evaluation does not preclude a priori any of the outcomes, methods, typical questions, assumptions, or audiences of any of the eight models identified by House (1978). Utilization-focused evaluation is not a recipe, or even a “model”—it is a strategy for making evaluation decisions.

Utilization-focused evaluation begins with *identification and organization of specific, relevant decision makers and information users* (not vague, passive audiences) who will use the information that the evaluation produces. The evaluator works with these persons (often an evaluation task force representing several constituencies; for example, program staff, clients, funders, administrators, board members, and community representatives) to *focus relevant evaluation questions*. From these questions flow the appropriate research methods and data analysis techniques. Utilization-focused evaluation plans for utilization before data is ever collected. The question that underlies the ongoing interactions between evaluators and decision makers is, “What difference would *that* information make?” The evaluator asks: “What would you do if you had an answer to *that* question?” In answering the evaluation questions of decision makers and information users, utilization-focused evaluation does not preclude the use of any of the full variety of methodological options open to evaluations. The ways in which qualitative methods strategies emerge as appropriate in particular utilization-focused evaluations is the subject of the remainder of this chapter.



Having reviewed the eight major evaluation models and the utilization-focused approach to evaluation, the problem remains of how to increase the competence of evaluators in the following cases:

Conceptualizing the research problem or defining the object of the evaluation

Selecting an appropriate inquiry strategy for addressing the research or evaluation problem

Selecting appropriate research and evaluation designs to collect data to . . . answer the question [Worthen, 1975:14].

The remainder of this chapter discusses evaluation situations, questions, and purposes for which qualitative research strategies can be particularly appropriate.

## PROCESS EVALUATION

Process evaluations are aimed at elucidating and understanding the internal dynamics of program operations. Process evaluations focus on the following kinds of questions: What are the factors that come together to make this program what it is? What are the strengths and weaknesses of the program? How are clients brought into the program and how do they move through the program once they are participants? What is the nature of staff-client interactions?

Process evaluations most typically require a detailed description of program operations. Such descriptions may be based on observations and/or interviews with staff, clients, and program administrators. Many process evaluations focus on how the program is perceived by participants and by staff. The mandate to generate an accurate and detailed description of program operations particularly lends itself to the use of qualitative methods.

The “process” focus in an evaluation implies an emphasis on looking at *how* a product or outcome is produced rather than looking at the product itself; that is, it is an analysis of the processes whereby a program produces the results it does. Process evaluation is developmental, descriptive, continuous, flexible, and inductive.

The process evaluator searches for explanations of the successes, failures, and changes in a program. Under field conditions in the real world, people and unforeseen circumstances shape programs and modify initial plans in ways that are rarely trivial. The process evaluator sets out to understand and document the day-to-day reality of the setting or settings under study. The evaluator tries to unravel what is actually happening in a program in a search for major patterns



and important nuances that give the program its character. A process evaluation requires sensitivity to both qualitative and quantitative changes in programs throughout their development; it means becoming intimately acquainted with the details of the program. Process evaluations look not only at formal activities and anticipated outcomes, but they also investigate informal patterns and unanticipated consequences in the full context of program implementation and development. Finally, process evaluations usually include perceptions of people close to the program about how things are going. A variety of perspectives may be sought from people with dissimilar relationships to the program—inside and outside sources.

Under what conditions are process evaluations undertaken? Process evaluations permit decision makers and information users to understand the dynamics of program operations. Such understanding permits decisions to be made about the extent to which the program is operating the way it is supposed to be operating. Process evaluations are particularly useful for revealing areas in which programs can be improved as well as highlighting those strengths of the program which should be preserved. Process evaluations also are useful in permitting people not intimately involved in the program—for example, external funders, public officials, and external agencies—to understand how a program operates. This permits such external persons to make more intelligent decisions about their own relationship and responsibilities with regard to such programs. Finally, process evaluations are particularly useful for dissemination and replication of programs under conditions where a program has served as a demonstration project or is considered to be a model worthy of replication at other sites. By understanding the dynamics of program processes and by studying descriptions of these program processes it is possible to isolate critical elements that have contributed to program successes and failures.

Qualitative methods are particularly appropriate for the conduct of process evaluations. To understand the unique, internal dynamics of a program it is best to approach that program without predetermined hypotheses about what those strengths and weaknesses are. Such an open-ended approach permits the strengths and weaknesses to emerge from the program observations and interviews rather than from the theories and expectations of the evaluator. An open-ended approach allows the evaluator to find out what is there rather than validating, confirming, or rejecting preordinate hypotheses about program strengths and weaknesses. Moreover, the nature of program processes is sufficiently complex and interdependent that they are seldom easily



represented along some set of unidimensional quantitative scales. Nor can quantitative dimensions and scales provide the kind of detail that is necessary for blueprints of program processes where the description of those processes are to be used in constructing models for purposes of replication and demonstration. Thus, qualitative methods are particularly appropriate for process evaluations.

## EVALUATING INDIVIDUALIZED OUTCOMES

A dominant pattern in the delivery of educational and social services in the 1970s has been a concern with individualization. Individualization means matching program services to the needs of individual clients. Highly individualized programs operate under the assumption that outcomes will be different for different clients. Not only will outcomes vary along specific common dimensions, but outcomes will be qualitatively different and will involve qualitatively different dimensions for different clients. Under such conditions program staff are justifiably reluctant to generate standardized criteria and scales against which all clients are compared. They argue that their evaluation needs are for documentation of the unique outcomes of individual clients rather than for measures of outcomes standardized across all clients.

There are numerous examples of such programs. Open education, for example, is partly a model of educational processes that assumes that the outcomes of education for each child are unique. Open education in particular, and experiential education in general, provide a set of activities in which students engage. Program activities and processes can be specified and even operationalized to monitor variations, but the outcomes of having engaged in those processes will be quite different for different students. Thus, a group of students in an open classroom may engage in some language experience that involves a field trip, dictating stories to the teachers and volunteers about that field trip, and then learning to read their stories. For some students such a process may involve learning about the mechanics of language: sentence construction, parts of speech, and verb conjugation, for example. For other students the major outcome of such a process may be learning about how to spell certain words. For other students the important outcome may be having generated an idea and conceptualized that idea based on a particular experience. For yet other students the important outcome may have been something that was learned in the exercise or experience itself, such as knowledge about the firehouse, or the farm that was visited.



Other students may become more articulate as a result of the dictation exercise. Still other students may have learned to read better as a result of the reading part of the exercise. The critical point is that a common process engaged in by all students can result in drastically different outcomes for different students depending on how they approach the process, what their unique needs were, and which part of the process they found most stimulating. For open educators, then, they need an evaluation process that permits the documentation of this variety of outcomes and they resist measuring the success of such a complex process by any limited set of outcome measures (for example, improved reading scores, better spelling, or more knowledge about some particular part of the world).

A similar case can be made with regard to the individualization of criminal justice, community mental health, job training, welfare, and health programs. Take, for example, the goal of increased independence among a group of clients receiving treatment in a community mental health center. It is possible to construct a test which can be administered to a large group of people measuring their relative degrees of independence. Indeed, such tests exist; these typically ask what kind of activities a person engages in and takes responsibility for, such as personal hygiene, transportation, initiatives in social interaction, food preparation, and so on. In many programs measuring such criteria in a standardized fashion provides the information that program staff would like to have. However, in programs that emphasize individualization of treatment and outcomes, program staff may argue, quite justifiably, that independence has a different meaning for different people under different life conditions. Thus, for example, for one person independence may have to do with a changing family dynamic and changed relationships with parents. For another person, independence may have to do with nonfamilial relationships—that is, interactions with persons of the opposite sex, social activities, and friendships. For still other clients the dominant motif in independence may have to do with employment and economic factors. For still others it has to do with learning to live alone. While clients in each case may experience a similar psychotherapeutic process, the meaning of the outcomes for their personal lives will be quite different. What program staff want to document under such conditions is the unique meaning of the outcomes for each client. What they want and need is descriptive information about how clients' lives change over the period of treatment and following treatment. They need descriptive information about what a client's life was like on entering treatment. They need descriptive information about



the client's response to treatment. They need descriptive information about what the client's life was like following treatment. Such descriptive information results in a set of individual case studies. By combining these case histories it is possible to construct an overview of the pattern of outcomes for a particular treatment facility or modality. Such clinical information is not obtained with the sole use of standardized tests and quantitative scales. Thus, qualitative methods and design strategies can be particularly useful for evaluation programs which emphasize individualized client outcomes.

## CASE STUDIES

The desire to document individualized client outcomes is one major reason why case studies may be more useful than measuring standardized, quantitative outcomes for all program participants. There are other strategic reasons why case studies may be preferred to standardized information. Sometimes staff or funders are puzzled by particular cases—unusual successes, unusual failures, or drop-outs. Detailed case studies of these extreme cases may generate particularly useful information.

In other situations a case study approach may be indicated by the critical nature of one or a few cases. For example, if a new type of clientele enters the program—for example, clients with different backgrounds or program histories—it may be useful to gather in-depth information about these new clients and their experiences in the program. The same need for case study data may be present at the program level. A statewide or national project may spin off an innovative local program that is of special interest to decision makers, thereby indicating the appropriateness of conducting a case study of that particular program.

The next chapter on qualitative research designs will discuss in detail sampling strategies for selecting critical cases for study. The point here is that many evaluation questions lend themselves to collection of qualitative case data.

A good example of some of the reasons for using qualitative methods based on case studies is the Northwest Regional Educational Laboratory evaluation of an Experience-Based Career Education (EBCE) program.

EBCE is a full-time alternative educational program for high school students that attempts to integrate a student's knowledge of a variety of careers with the acquisition of cognitive, interpersonal and affective skills. Emphasis is placed on the student's assumption of responsibility for his or her own learning through individually tailored learning



activities in the community. Working adults in the community become the student's role models and instructors. EBCE program staff members become facilitators of the learning process.

To evaluate this program, multiple evaluation strategies (including an experimental design, survey techniques and an ethnography) were combined into a comprehensive evaluation design (Owens, Haenn and Fehrenbacher, 1976). A case study approach was included as part of that design for a variety of reasons:

- a) The EBCE program is highly individualized. The case study approach, in keeping with the philosophy and practice of EBCE, also concentrates on the individual student. Thus outcomes directly related to individual needs are easily assessed. For example, it might be an EBCE learning goal to help a shy, withdrawn student become more outgoing and to help an overly aggressive person become more restrained. Changes in these two students would cancel each other out in a purely nomothetic approach. A case study, because it uses the student as the unit of analysis, can capture this individualization and can estimate the effectiveness of the program based on the experiences of sampled students.
- b) A systems approach to learning is employed in the design and operation of EBCE. The learning strategies and management techniques are highly interrelated and interdependent. The holistic nature of the case study approach (Glaser and Backer, 1972) also fits well with a systems approach. It does not reduce the learning processes to independent, isolated parts, but describes the Gestalt as it traces a student's progress, frustrations and challenges throughout the program year.
- c) The EBCE program is process oriented. It is the philosophy of the program that the medium really is the message. Therefore, experiential learning is employed to help students learn how to learn. Unlike traditional evaluation methodologies which focus on outcomes, the case-study approach zeroes in on the process of learning. Its primary focus is the student experiencing the learning situation. It describes the student, the situation, and the resulting interaction.
- d) Because both career education and EBCE are relatively new entities in the field of education, definitive descriptions of neither one are available. Both are evolving and changing. To make ultimate conclusions about their effects is not possible today. To fill the void in the interim, case studies do provide empirical definitions of EBCE. And, to generate working hypotheses, a case study can be a rich source of data [Fehrenbacher et al., 1976:3-5].



## COMPARING PROGRAMS: FOCUS ON DIVERSITY

Individualizing services to clients has been one major theme of social action and educational programs in the '70s. Another closely related theme has been the importance of adapting programs to local community needs and circumstances. While some basic framework of how programs should function may originate in Washington, D.C., or some state capitol, it is clear that program implementation at the local level seldom follows exactly the proposed design. When an evaluation project requires gathering data from several local sites, quantitative measures may be appropriate for comparing local programs along standardized dimensions, but qualitative methods are necessary to capture the unique diversities and contrasts that emerge as local programs adapt to local needs and circumstances.

Edwards et al. (1975) note in their introduction to the decision-theoretic approach to evaluation that it is "a common administrative fiction, especially in Washington" that national programs are comparable from place to place and from time to time; "we have frequently encountered the idea that a program is a fixed, unchanging object, observable at various times and places" (Edwards et al., 1975:142). The evidence is that local sites which are part of national programs show considerable variation in implementation and outcomes. These variations are not such that they can be captured and measured along standardized scales; they are differences in *kind*—differences in content, in process, in goals, in implementation, in politics, in context, in outcomes, and in program quality. To understand these differences a holistic evaluation picture of each local site is needed, a picture that captures each site in its own terms, not just in the limited terms of the national program which spawned it. As Malcolm Provus has argued, the evaluation of programs, even national programs, must begin at the local level because

it follows that if there are types of programs with different developmental characteristics, the development standards for these program types will vary also. . . . This local work is usually of the process assessment type in which evaluators systematically collect and weigh data descriptive of ongoing program activity [Provus, 1971:13].

Quantitative evaluations assume that local sites in a national program are attempting to implement processes or attain outcomes which can be measured along a standardized set of scales or dimensions. Such a measurement approach means that qualitative differences among programs will be disguised and diminished. Such



evaluation approaches, by their very design and measurement, force all programs into a common mold, whether that mold is appropriate or not. The use of such standardized measures can seriously distort what is actually occurring in a program.

A simple example of such distortion is data from a national educational program which measured staff-student ratios in the various programs. A few programs had student-staff ratios as high as 75:1 according to the uniform measures used; other programs had student-staff ratios as small as 15:1. However, what these data did not reveal is that in some of the programs with large staff-student ratios there was extensive use of volunteers. These regularly participating auxiliary staff made the effective and real adult-student ratios much smaller. The global and uniform reporting of the data, however, did not allow for that nuance to become evident.

A good example of the diversity that can emerge from attention to the qualitative differences among programs is Sharon Fieman's (1977) study of the national teacher center program. There are many standardized measures that can be used to monitor and evaluate teacher center training sites so that local programs can be compared with each other and with federal guidelines. But Fieman found that there were actually three quite different types of center programs which had emerged under the general label "teacher centers." The three types of teacher centers she identified were "behavioral" centers, "humanistic" centers, and "developmental" centers. The differences among the three types of centers are summarized in Table 4.1.

*Table 4.1* Types of Teacher Centers\*

<i>Type of Center</i>	<i>Primary Process of Affecting Teachers</i>	<i>Primary Outcomes of the Process</i>
1. Behavioral Centers	Curriculum specialists directly and formally instruct administrators and teachers.	Adoption of comprehensive curriculum systems, methods, and packages by teachers.
2. Humanistic Centers	Informal, nondirected teacher exploration; "teachers select their own treatment."	Teachers feel supported and important; pick up concrete and practical ideas and materials for immediate use in their classrooms.
3. Developmental Centers	Advisors establish warm, interpersonal, and directive relationship with teachers working with them over time.	Teachers' thinking about what they do and why they do it is changed over time; teacher personal development.

\*Based on Feiman (1977).



It is clear from Fieman's analysis that different teacher centers are trying to accomplish different outcomes. Moreover, to accomplish these divergent outcomes they have established different approaches to teacher center programming. Uniform, quantitative measures applied across all programs would be unlikely to capture and represent these critical differences.

While quantitative approaches tend to produce uniformity of measures which have the advantage of facilitating direct comparisons among programs, qualitative methods permit documentation of program differences, idiosyncracies, and uniquenesses. If decision makers and information users want to understand variations in program implementation and variations in the *nature* of program outcomes, qualitative case studies of local programs can provide such detailed information. Qualitative data are necessary to give a complete evaluation picture of national programs, a picture which is necessarily incomplete so long as the only data available are aggregated and standardized statistics from these diverse programs.

Data about site-to-site variability can also be useful in planning later comparison studies, in developing models of local program operations, and in understanding different dimensions of need, demand, and potential service. Thus, where the focus of the evaluation question is on understanding and documenting local adaptations of multiple-site programs, qualitative methods are highly appropriate—and potentially quite useful.

## IMPLEMENTATION EVALUATION

A prominent theme running through the preceding sections is that qualitative methods are particularly useful for capturing differences among people and programs. Evaluating individualized outcomes, developing unique case studies of people and programs, and documenting the local diversity within national programs—these are evaluation research issues for which qualitative strategies are particularly appropriate. This section looks more closely at the appropriateness of qualitative methods for evaluating program implementation.

It is important to know the extent to which a program is effective after it is fully implemented; but to answer that question it is first necessary to know how and the extent to which the program was actually implemented. In his important book on *Social Program Implementation*, editor Walter Williams concludes:

The underlying theme of this book is that the lack of concern for implementation is currently *the* crucial impediment to improving



complex operating programs, policy analysis, and experimentation in social policy areas [Williams, 1976:267; italics in the original].

In *Utilization-Focused Evaluation* (Patton, 1978) I suggested that *if* one had to choose between implementation information and outcomes information because of limited evaluation resources, there are many instances in which implementation information would be of greater value. A decision maker can use implementation information to make sure that a policy is being put into operation according to design—or to test the very feasibility of the policy. Unless one knows that a program is operating according to design, there may be little reason to expect it to produce the desired outcomes. Furthermore, until the program is implemented and a “treatment” is believed to be in operation, there may be little reason even to bother evaluating outcomes. Where outcomes are evaluated without knowledge of implementation, the results seldom provide a direction for action because the decision maker lacks information about what produced the observed outcomes (or lack of outcomes). This is the “black box” approach to evaluation.

One important way of studying program implementation is to gather detailed, descriptive information about what the program is doing. Implementation evaluations answer the following kinds of questions: What do clients in the program experience? What services are provided to clients? What do staff do? What is it like to be in the program? How is the program organized? Implementation evaluations tell decision makers what is going on in the program and how the program has developed.

It is important to study and evaluate program implementation in order to understand how and why programs deviate from initial plans and expectations. *Such deviations are quite common and natural*, as demonstrated in the findings of Rand’s “Change Agent Study” of 293 federal programs supporting educational change. That study found that national programs are implemented incrementally by adapting to local conditions, organizational dynamics, and programmatic uncertainties.

[W]here implementation was successful, and where significant change in participant attitudes, skills, and behavior occurred, implementation was characterized by a process of mutual adaptation in which project goals and methods were modified to suit the needs and interests of the local staff and in which the staff changed to meet the requirements of the project. This finding was true even for highly technological and initially well-specified projects; unless adaptations were made in the



original plans or technologies, implementation tended to be superficial or symbolic, and significant change in participants did not occur [McLaughlin, 1976:169].

If program implementation is characterized by a process of adaptation to local conditions, needs, and interests, then the methods used to study implementation must be open-ended, discovery oriented, and capable of describing developmental processes and program change. Qualitative methods are ideally suited to the task of describing program implementation.

Failure to monitor and describe the nature of implementation, case by case, program by program, can render useless standardized, quantitative measures of program outcomes. The national evaluation of Follow Through is a prime example of this point. Follow Through was a planned variation "experiment" in compensatory education featuring twenty-two different models of education to be tested in one hundred fifty-eight school districts on 70,000 children throughout the nation. The evaluation alone employed 3,000 people to collect data on program effectiveness. The multi-million-dollar evaluation focused almost entirely on standardized outcomes aimed at making possible comparisons of the effectiveness of the twenty-two models. It was assumed in the evaluation plan that models could be and would be implemented in some systematic, uniform fashion. Eugene Tucker of the U.S. Office of Education, however, has poignantly described the error of this assumption.

It is safe to say that evaluators did not know what was implemented in the various sites. Without knowing what was implemented it is virtually impossible to select valid effectiveness measures. . . . Hind-sight is a marvelous teacher and in large scale experimentations an expensive one [Tucker, 1977:11-12].

The Follow Through data analysis showed greater within-group variation than between-group variation; that is, the twenty-two models did not show treatment effects as such. Most effects were null; some were negative; but "of all our findings, the most pervasive, consistent, and suggestive is probably this: *The effectiveness of each Follow Through model depended more on local circumstances than on the nature of the model*" (Anderson, 1977:13; italics in the original). The evaluators, however, failed to study the local circumstances that affected variations in program implementation and outcomes.



Little remains in the existing Follow Through evaluation that specifically addresses the problem of how well, and by what process, program models are implemented [Elmore, 1976:119].

The study of these important program implementation questions requires case data rich with the details of program content and context. Because it is impossible to anticipate in advance how programs will adapt to local conditions, needs, and interests it is impossible to anticipate what standardized quantities could be used to capture the essence of each program's implementation. Under these evaluation conditions a strategy of naturalistic inquiry is particularly appropriate.

## FORMATIVE EVALUATION

Formative evaluations are conducted for the purpose of improving programs in contrast to those evaluations which are done for the purpose of making basic decisions about whether or not the program is effective, and whether or not the program should be continued or terminated. While formative evaluations can make use of both quantitative and qualitative data, it can be particularly useful to collect detailed, descriptive information about programs where the purpose of the evaluation is to improve the program operations and procedures.

It is important at the outset of the evaluation process to clarify the extent to which the primary purpose of an evaluation is to make an overall judgment about the effectiveness of a program (summative evaluation) *or* to collect information that can be used primarily for ongoing program development and improvement (formative evaluation). The labels for this distinction were introduced by Michael Scriven (1967:70-43) in discussing evaluation of educational curricula. Sanders and Cunningham (1973) extended the writing of Scriven on formative evaluation with particular emphasis on formative evaluation applied to the product development process.

Formative evaluation was defined as the process of judging an entity, or its components, that could be revised in form, for the expressed purpose of providing feedback to persons directly involved in the formation of the entity [Sanders and Cunningham, 1974:1].

Qualitative methods are included among the strategies suggested by Sanders and Cunningham for gathering formative program data, particularly "formative work dealing with the appraisal of early



product development efforts”—what they call “formative interim evaluation activities”:

Formative interim evaluation information can involve collecting internal information such as descriptive information and processing critical appraisals. Descriptive information refers to the objective information which can be generated by inspecting the pieces or preliminary versions of the product. Critical appraisals are judgments made concerning the pieces by representatives of concerned populations (e.g., experts, parents, students, etc.). . . .

The intent of collecting descriptive information is to describe fully and completely what *is*, not what should be. A comprehensive characterization of what is will aid greatly in making judgments and in determining where to revise once some deficit is identified [Sanders and Cunningham, 1974:23].

While the term “formative evaluation” initially referred to information gathered to improve curriculum products in education, the formative-summative distinction has since become a fundamental evaluation typology and the terms are applied more broadly than Scriven used them in the original. The purpose of any program evaluation can be examined in terms of the formative-summative distinction. Morris and Fitz-Gibbon (1978a) discuss in detail the role of formative evaluators in the development of educational programs generally. They suggest that

the key to an effective formative evaluation is good communication. Information about where the program is or is not working needs to be timely and clearly presented [Morris and Fitz-Gibbon, 1978a:66].

Formative evaluations often include a process evaluation strategy as described earlier in this chapter. Thus, such a formative evaluation would focus on identifying and elucidating the strengths and weaknesses of the program. Malcolm Provus, a prominent educational evaluator and originator of the “Discrepancy Model” of evaluation, emphasized the importance of formative evaluation using qualitative methods in the early stages of a program’s development.

An evaluation that begins with an experimental design denies to program staff what it needs most: information that can be used to make judgments about the program while it is in its dynamic stages of growth. . . . Evaluation must provide administrators and program staff with the information they need and the freedom to act on that information. . . .



We will not use the antiseptic assumptions of the research laboratory to compare children receiving new program assistance with those not receiving such aid. We recognize that these comparisons have never been productive, nor have they facilitated corrective action. The overwhelming number of evaluations conducted in this way show no significant differences between “experimental” and “control” groups [Provus, 1971:11-12].

A formative evaluation based on naturalistic inquiry methods can provide timely information about program dynamics without restricting the ability of administrators and program staff to act on that information. Feedback given by the formative evaluator as well as actions taken as a result of formative evaluation information become part of the developmental evaluation record; that is, formative evaluators must document and describe not only program developments but also *their role* as formative evaluators in those program developments.

Finally, formative evaluations may focus on gathering descriptive information about the *quality* of program activities and outcomes, not just levels or amounts of attainment. Formative evaluations are aimed at improving program quality. Judgments about quality often require data of considerable depth and detail—qualitative data. The next section discusses further the appropriateness of qualitative methods for evaluating program quality.

## A FOCUS ON QUALITY

There are many aspects of program operations, including implementation activities and client outcomes, that can be measured in terms of relative quantity. It makes sense to count the number of people who enter a program, the number who leave the program, and the number who receive or report some concrete benefit from the program. There are many attributes of programs, however, that do not lend themselves to counting. Even the scaling of quality attributes is an inadequate way of capturing either program quality or the effect of a program upon the quality of life experienced by participants following the program.

School outcomes can be looked at both in terms of quantity of change and quality of change. Quantity of change may involve the number of books read; a score on a standardized achievement test; the number of words spelled correctly; and the number of interactions with other students, the teacher, or people of a different race. Each of these outcomes has a corresponding quality dimension which requires description rather than scaling. Thus, to find out what



it means to a student to have read a certain number of books is an issue of quality. How those books affected the student personally and intellectually is a question of quality. In contrast to counting the correct number of words spelled, the quality issue focuses on what spelling *means* to the student. How is spelling integrated into the student's life? How does the student think about spelling, approach spelling, feel about spelling? The answer to such questions requires description of the student's perspective and situation such that the meaning of the experience for the student is recorded.

The same distinction holds with regard to programs that emphasize deinstitutionalization—for example, community mental health programs, community corrections, and community-based programs for the elderly. It is possible to count the number of people placed in the community. It is possible even to measure on standardized scales certain attributes of their livelihoods. It is possible to have them subjectively rate various aspects and dimensions of quality of life. However, to fully grasp the meaning of a change in life for particular clients and persons it is necessary to develop a description of life quality which allows the interdependent parts of quality to be integrated into a whole.

Quality has to do with nuance, with detail, with the subtle and unique things that make a difference beyond the points on a standardized scale. Quality is what separates and falls between those points on a standardized scale. Quality descriptions provide the detail to explain what the lives of two different people are like, one of whom responded on a scale of five points that he or she had a "highly" satisfactory experience, the other of whom responded that he or she had an "extremely" satisfying experience. This is not a question of interval versus ordinal scaling, but one of meanings. What do programs mean to participants? What is the quality of their experience? Answers to such questions require detailed, in-depth, and holistic descriptions that represent people in their own terms and that get close enough to the situation being studied to understand firsthand the nuances of quality.

The failure to find statistically significant differences in comparing people on some outcome measure does *not* mean that there are no important differences among those people on those outcomes. The differences may simply be qualitative rather than quantitative. A carpenter is reported to have explained this point to William James. The carpenter, having worked for many different people, observed: "There is very little difference between one man and another; but what little there is, is very important." Those differences are differences of quality.



## STATE OF THE ART CONSIDERATIONS: LACK OF PROVEN QUANTITATIVE INSTRUMENTATION

Another reason for using qualitative measurement is that for particular outcomes no acceptable, valid, and reliable quantitative measures exist. The extent to which one believes that quantitative measures in a particular instance and for a particular variable are useful, valid, and reliable is a matter of judgment. However, the state of the art in social science measurement is such that a number of desirable outcome measures still elude precise measurement. Where outcome measurements have not been carefully developed, it is more appropriate to gather descriptive information about what happens as a result of program activities than to use some scale which has the merit of being quantitative but whose validity and reliability are suspect.

Creativity is a prime example. While there are some instruments that purport to measure creativity, the applicability of those instruments in diverse situations is at least open to question. Thus, a program that was attempting to make students or clients more creative might do better to document in detail the activities, behaviors, thoughts, and feelings of participants rather than to administer some instrument. Documentation of this kind can then be inspected and judged by interested decision makers or information users to make their own interpretations of the extent to which creativity was exhibited by the products produced, activities undertaken, or statements made by participants.

Even such hallowed concepts as self-esteem are open to considerable controversy when it comes to specifying measurement criteria. In addition, for people whose self-esteem is already quite high, instruments that measure self-esteem are not very sensitive to incremental changes that may be important to the people involved. For staff development programs that include enhanced self-esteem as an outcome goal, it may be more important to use qualitative measures to document changes in clients or participants than to rely upon a particular measurement scale that was created for other conditions and situations.

The same point can be made with regard to controversy surrounding even long-standing measurement instruments. The use of standardized achievement tests to measure student learning is a prime example. Strong arguments have been made attacking the relevance of universal, standardized achievement tests for the evaluation of particular local programs (Perrone, 1977). The way in



which norm-referenced, standardized achievement tests are constructed reduce their relevance and validity for particular local programs, particularly those programs which serve populations where scores are likely to cluster at the lower or higher extremes of the normal curve. For such programs more accurate evaluation results can be produced through documentation of actual student work; that is, developing case histories of what students can do and have done rather than relying on their responses to a standardized instrument at a particular point in time.

## THE NEED FOR UNOBTRUSIVE MEASURES

Another condition under which qualitative strategies can be particularly appropriate is where an experimental design, the administration of standardized instruments and/or the collection of quantitative data would affect program operations by being overly intrusive. Observations of program activities and informal interviews with participants sometimes can be carried out in a less obtrusive fashion than having everyone complete some test or questionnaire. Indeed, administration of such an instrument may produce artificial results or affect program operations. The instrument itself can create a reaction which, because of its intrusiveness and interference with normal program operations and client functioning, fails to accurately reflect what has been achieved in the program.

Educational researcher Edna Shapiro (1973), to her surprise, found this to be precisely the case in her study of innovative Follow Through classrooms. She found that standardized tests can bias evaluation results by imposing an obtrusive and controlled stimulus in an environment where spontaneity, creativity, and freedom of expression are valued and encouraged. Shapiro found that the results of the test measured response to a stimulus (the test) which was essentially alien to the experience of the children. Because the classrooms she studied relied substantially less on paper-and-pencil skills than traditional schools, and because student progress was monitored daily on a personal basis without the use of written examinations, student outcomes in these classrooms could not be "objectively" measured by the sudden introduction of standardized tests.

I assumed that the internalized effects of different kinds of school experience could be observed and inferred only from responses in test situations, and that the observation of teaching and learning in the



classroom should be considered auxiliary information, useful chiefly to document the differences in the children's group learning experiences. . . .

The findings of this study, with the marked disparity between classroom responses and test responses, have led me to reevaluate this rationale. This requires reconsideration of the role of classroom data, individual test situation data, and the relation between them. . . .

The individual's responses in the test situation have conventionally been considered the primary means to truth about psychological functioning. Test behavior, whether considered as a sign or sample of underlying function, is treated as a pure measure. Yet the test situation is an unique interpersonal context in which what is permitted and encouraged, acceptable and unacceptable, is carefully defined, explicitly and implicitly. *Responses to tests are therefore made under very special circumstances. The variables that influence the outcome are different from those which operate in the classroom* [Shapiro, 1973:532-534; italics added].

In their imaginative book on *Unobtrusive Measures*, Webb et al. (1966) discuss at length the problems of "reactive measurement effect." A basic theme of their work is that the research subjects' knowledge and awareness that they are part of a study as they complete questionnaires or tests may distort and confound the study findings. Their documentation of the sources and nature of reactivity problems in scholarly social science research makes it highly likely that such problems are magnified in evaluation research. While qualitative methods are also subject to certain reactivity problems (to be discussed in later chapters), the less formal and obtrusive nature of qualitative strategies for conducting evaluations can sometimes serve to reduce distorting reactions to the evaluation on the part of the people being studied.

## LEGISLATIVE MONITORING

There are many occasions when some legislative body that has mandated and appropriated funds to a new program wants to have information about whether or not the program is operating in accordance with legislative intent. Legislative intent may involve achieving certain outcomes, but, more often, legislative intent focuses specifically on some certain kind of delivery system being provided. The precise nature of that delivery system is often not well articulated. Thus, such considerations as deinstitutionalization, decentralization, services integration, and community-based programs involve varied conceptualizations of legislative intent that do not



easily lend themselves to quantitative specification. Indeed, for the evaluator to unilaterally establish some quantitative measure of deinstitutionalization that provides a global, numerical summary of the nature of program operations may hide more than it reveals.

To monitor the complexities of program implementation in the delivery of human services it can be particularly helpful to decision makers to have detailed case descriptions of how programs are operating. Such legislative monitoring would include descriptions of program facilities, outreach efforts, staff selection procedures, the nature of services offered to clients, descriptions of actual service delivery activities, and descriptions from clients about the nature of their experiences and the results of their experiences. While busy legislators cannot be expected to read in detail a large number of such histories, legislators or funders are likely to be particularly interested in the case histories of those programs that are within their own jurisdiction or legislative district, and, more generally, certain legislative staff who are particularly interested in the program can be expected to read such case histories with some care. From a political point of view, programs are more likely to be in trouble or cause trouble for legislators because they fail to follow legislative intent in implementation rather than because they failed to achieve desired outcomes.

In this case, *the purpose of legislative monitoring is to become the eyes and ears of the legislature*. This means trying to provide descriptions of programs that are sufficiently detailed and elucidating that the legislator or legislative staff can read such descriptions and have a good idea of what that program is like. Having such descriptions enables legislators to decide whether or not their own interpretations of legislative intent are being met. Such case histories may also be of considerable service to programs being monitored, because it permits them to tell their own story in some detail. Thus, where they have deviated from legislative intent, such case histories would be expected to include information from program administrators and program staff about constraints under which the program operates and the decisions staff have made that give the program its character.

At the same time, the collection of such case histories through site visits and program monitoring need not neglect the need for more global statements about statewide patterns in programs, or even nationwide patterns. It is quite possible through content analysis to identify major patterns of program operations and outcomes for a number of separate cases. Thus, qualitative methods used for legislative monitoring allows one to document common patterns



across programs as well as unique developments within specific programs.

## BREAKING THE ROUTINE: GENERATING NEW INSIGHTS

There is another sense in which qualitative methods provide a real and useful option in program evaluation. Programs that have established ongoing evaluation systems or management information approaches may have become lulled into a routine of producing statistical tables that are no longer studied with any care. Inertia and boredom can seriously reduce the usefulness of program evaluation results. After program staff or other decision makers have seen the same statistical results used in the same kinds of statistical tables year after year, those results can begin to have a numbing effect. Even though the implications of those results may vary somewhat from year to year, the very format used to report the data can reduce the impact of the results. Mao Tse Tung commented on the tendency of human beings to settle into numbing routines when he said that a revolution would be needed every twenty years. Revolutions in the collection of evaluation data may be needed much more often.

One such revolution may be to introduce a totally new approach to evaluation simply for the purpose of attracting renewed attention to the evaluation process. At the same time, changing the method may produce new insights or at least force people to deal with the old insights in a new way. Of course, collection of qualitative data can also become routine. Programs of humanistic ideology and/or programs with an emphasis on individualization may find that the collection of qualitative data has become a routine and that new insights can be gained through even the temporary use of some quantitative measures.

This suggestion for a change in methods is not based on scientific prescription, but again derives from a concern with enhancing utilization of evaluation research. Given the ease with which human beings and social systems settle into inertia and routine, evaluators who want their results to make a difference need to find creative ways to get people to deal with issues of program effectiveness. Exploring methodological variations may be one such approach.

It is also worth noting that evaluators can settle into routine and inertia. Evaluators who have been using the same methods over and over may have lost the cutting edge of their own creativity. A paradigm of choice (Chapter One) requires evaluators to have at their disposal a large repertoire of possible data collection techniques



and approaches. Evaluators can be more useful to programs if they themselves are staying alive to the many possibilities available for looking at the world. Indeed, a change in methods may do as much or more to reenergize the evaluator as it does to renew the program evaluation process.

## GROUNDED EVALUATION THEORY

Most writing on social science methods is concerned with the generation and verification of theory. This concern for theory development is quite marked in the literature on qualitative methods. The writings of Glaser and Strauss (1967), Denzin (1978), Lofland (1971), Blumer (1969), and Webb et al. (1966), to name but a few well-known qualitative methodologists, take as their central focus the task of theory construction and verification. What distinguishes the discussion of theory in much of the literature on qualitative methods is the emphasis on inductive strategies of theory development in contrast to theory generated by logical deduction from a priori assumptions.

In contrasting grounded theory with logico-deductive theory and discussing and assessing their relative merits in ability to fit and work (predict, explain, and be relevant), we have taken the position that the adequacy of a theory for sociology today cannot be divorced from the process by which it is generated. Thus one canon for judging the usefulness of a theory is how it was generated—and we suggest that it is likely to be a better theory to the degree that it has been inductively developed from social research. . . . Generating a theory from data means that most hypotheses and concepts not only come from the data, but are systematically worked out in relation to the data during the course of the research. *Generating a theory involves a process of research* [Glaser and Strauss, 1967:5-6; italics in the original].

Evaluation research, particularly at the local program level, has been largely atheoretical—both in conception and in reporting findings. Evaluation research often ignores theoretical issues altogether. Evaluators are accused of being technicians who simply collect data without regard to the theoretical relevance of possible empirical generalizations. Certainly, pure outcomes evaluations are nontheoretical. Moreover, in many cases what decision makers need and want is specific data relevant to narrow, technical issues that are helpful in monitoring or fine-tuning program operations.

However, evaluation research is by no means inherently non-theoretical. It can be theoretical in the usual scientific sense that



deductive, logical systems are constructed to model causal linkages among general variables (Hage, 1972). Specific program operations are then modeled after the theory and monitored to test the theory. The deductive approach usually draws on the dominant theoretical traditions in specific scholarly disciplines to construct models of the relationship between program treatments and outcomes which are then verified through experimental research. In practice, the formality, complexity, and abstraction of most academic theories bear little relevance for practitioners caught up in the day-to-day realities of program functioning.

By way of contrast to logical, deductive theory construction, a grounded theory approach to evaluation research is inductive, pragmatic, and highly concrete (Patton, 1978:179-198). The evaluator's task is to generate program theory from holistic data gathered through naturalistic inquiry for the purpose of helping program staff and decision makers understand how the program functions, why it functions as it does, and the ways in which the impacts/consequences/outcomes of the program flow from program activities. Program staff and other program decision makers can use such grounded evaluation theory to *reality-test* their own theories of programmatic action, program effects, and the relationship between action and effects. Such grounded evaluation theory can serve to take decision makers *into the empirical world* so that they can discover whether what they think to be the nature of the empirical world is actually the case. Sociologist Herbert Blumer describes what it means to ground theory in the empirical world.

The prevailing disposition and practice is to allow the theory, the model, the concept, the technique, and the scientific protocol to coerce the research and thus to bend the resulting analytical depiction of the empirical world to suit their form. In this sense, much current scientific inquiry in the social and psychological sciences is actually social philosophizing.

I repeat once more that what is needed is to gain empirical validation of premises, the problems, the data, their lines of connection, concepts, and the interpretation involved in the act of scientific inquiry. The road to such empirical validation . . . lies in the examination of the empirical social world. It is not to be achieved by forming and elaborating catchy theories, by devising ingenious models, by seeking to emulate the advanced procedures of the physical sciences, by adopting the newest mathematical and statistical schemes, by coining new concepts, by developing more precise quantitative techniques, or by insisting on adherence to the canons of research design. Such



preoccupations, without prejudice to their merit in other respects, are just not headed in the direction that is called for here. What is needed is a return to the empirical social world [Blumer, 1978:35].

For Blumer, a return to the empirical social world involves the inductive process of naturalistic inquiry.

The empirical social world consists of on-going group life and one has to get close to this life to know what is going on in it.

The metaphor that I like is that of lifting the veils that obscure or hide what is going on. The task of scientific study is to lift the veils that cover the area of group life that one proposes to study. The veils are not lifted by substituting, in whatever degree, preformed images for first-hand knowledge. The veils are lifted by getting close to the area and by digging deep in it through careful study. Schemes of methodology that do not encourage or allow this betray the cardinal principle of respecting the nature of one's empirical world. . . . [T]he merit of naturalistic study is that it respects and stays close to the empirical domain [Blumer, 1978:38, 43].

Grounded evaluation theory provides a primary linkage between applied social science and basic social science research. More directly, grounded evaluation theory can provide relevant information which is useful to program staff and other decision makers in their efforts to understand and improve their programs. Grounded evaluation theory would be particularly useful in considerations of whether a program should be replicated in other settings and how such replication might occur. Thus, grounded evaluation theory would be a particularly important product of the evaluation of demonstration programs.

Finally, grounded theory represents a major strategy for studying the evaluation research process itself. As evaluators seek to understand the design, implementation, and utilization of evaluation research, grounded theory, based on the empirical evaluation world, can contribute to our understanding of how evaluation works and what the consequences are for an organization or program of engaging in evaluation. My own work on *Utilization-Focused Evaluation* (1978) is based on an inductive inquiry into the nature of the utilization of evaluation research. The work of Marvin Alkin and his associates (1979) in studying the utilization of evaluation research is an exemplary example of how naturalistic inquiry into cases of evaluation research can contribute to an understanding of the development, design, and utilization of evaluation research. Alkin et al. developed case histories of five educational evaluation •



projects. Their case materials provide a rich data source from which they were able to explore the nature of the development, implementation, and utilization of evaluation research. Their conclusions in the final chapter, which compares and contrasts cases, constitute a major contribution to grounded theory about evaluation research.

Blaine R. Worthen commented on the need for a program of research on evaluation as part of his analysis of the current state of evaluation models at the annual meeting of the American Educational Research Association in 1977.

Many calls have been issued over the years for a program of research on evaluation to be launched. . . . [T]he lack of an adequate empirical base is probably the single most important impediment to the development of a more adequate evaluation theory and model. In the absence of relevant evidence, adherence to any one model rather than another is much more a statement of philosophy or a profession of faith than a choice based on information about which model works best under which circumstances. . . . [T]he most important criterion in deciding whether or not we can live happily with extant models of evaluation is whether or not they are a good enough fit to reality to serve as trustworthy guides in the conduct of evaluation. . . . [T]he point of this whole discussion is that evaluation, as a phenomenon, must be researched [Worthen, 1977].

## PERSONALIZING EVALUATION

Another reason for using qualitative methods, particularly qualitative measurement, is that such strategies may be perceived by program staff and program clients as more personal in nature. Programs that are based on humanistic concerns and humanistic ideologies often resist any kind of quantification because of their perceptions about the impersonal nature of numbers and scientific categorization. The issue here is not whether or not such objections are reasonable. The point is that such objections are real and that in programs where staff, funders, and/or clients hold such views, evaluations that rely on quantitative measurement may be rejected out-of-hand. Whether the evaluator is right or wrong to believe that quantitative methods may be most appropriate to study the effects of such programs, if the primary decision makers and information users are going to dismiss the data simply because it is quantitative, then the evaluator may find that he or she has produced an excellent evaluation that is never used.

In evaluating such programs, then, qualitative methods can considerably enhance the utilization of evaluation findings because



the data are perceived as personal. The personal nature of the data derives from the fact that it is open-ended (and therefore does not willy-nilly categorize people), the evaluator has established close contact with the program and therefore has made it more personal, and the procedures of observation and in-depth interviewing, particularly the latter, communicate respect to respondents by making *their ideas and opinions stated in their own terms* the important data source for the evaluation. Qualitative methods may also be perceived as more personal because of their inductive strategy. This means that, again, rather than imposing on the program some predetermined model or hypotheses, the program picture unfolds in a way which takes into account idiosyncracies, uniquenesses, and complex dynamics. Finally, qualitative methods may be perceived as more humanistic and personal simply by avoiding numbers.

#### UTILIZATION-FOCUSED EVALUATION: DECISION MAKERS' PHILOSOPHICAL AND POLITICAL ORIENTATION

The suggestion that one reason for using qualitative methods is that such strategies may be perceived by program staff and program clients as more personal in nature opens up a whole range of potential philosophical and political orientations that can influence methods decisions. The argument here is that, in many instances, it may be appropriate and desirable to include in the making of methods decisions the philosophic orientations of decision makers and information users about the nature of data. This can work either for or against the selection of qualitative methods. Thus, there may be certain evaluation situations where qualitative methods seem particularly appropriate, but where decision makers and information users are philosophically predisposed in favor of quantitative measurement and automatically reject a qualitative approach, not because such an approach would be inappropriate, but because their perceptions and prejudices lead them to reject qualitative data.

In a utilization-focused approach to evaluation research the researcher has no intrinsic rights to unilaterally make critical design and data collection decisions. Quite the contrary, it is crucial that identified decision makers and information users participate in the making of measurement and methods decisions so that they understand the strengths and weaknesses of the data—and so that they believe in the data. Utilization potential can be severely diminished if decision makers are excluded at the critical operationalization stage when the evaluation comes down to the making of nitty-gritty choices.



Measurement and methods decisions are not simply a matter of expertly selecting the best techniques. Researchers and decision makers operate within quite narrow methodological paradigms about what constitutes valid and reliable data, rigorous and scientific design, and personal or impersonal research methods. Design and data collection decisions are far from being neutral, objective, or rational; such decisions are political, subjective, and satisfying.

The various scholarly disciplines have variable methodological standards. Most social scientists routinely apply those methods in which they have been trained with little sensitivity to the biases introduced by a particular data collection scheme. Social and behavioral scientists—experts in the ultimate subjectivity and arbitrariness of all human perception—are often least aware of their own sociomethodological biases and how these biases affect their view of the social program world. Yet, to be sure, social scientists are not the only participants in the evaluation process operating on the basis of selective perception; decision makers also hold conditioned views about the nature of social reality. One of the tasks during the active-reactive-adaptive interactions between evaluators and decision makers is to mutually explore design and data biases so that the evaluation generates information that is useful and believable to decision makers and information users.

One example of the result of such utilization-focused methodological discussions is a document that provided the framework within which the evaluation of Marcy Open School in Minneapolis was conducted. In September 1971 a system of alternative schools was established in Southeast Minneapolis. A basic premise behind that system was choice—that parents, staff, and children had a right to choose programs, goals, systems of governance, and evaluation methods which were compatible with their values and educational needs. One of those choices was Marcy Open School. The document establishing the framework for evaluation at Marcy was entitled “A Value-Perspective on Evaluation: Evaluation of Marcy Open School” (Olson, 1974). This framework document illustrates how *the decision to use qualitative methods in evaluation can flow from the values of the people who will use the evaluation information.*

### **EXCERPTS FROM “A VALUE-PERSPECTIVE ON EVALUATION”**

What are the values of the school? And what are the values of the evaluation? Throughout the life of the school, three strong themes



have emerged. The structures of the school organization, the formal goals of the school, emphases in classrooms, and investments of time and money in staff development activities have all placed strong value on (a) personalized curriculum, (b) the experiential nature of learning, and (c) a holistic nature of learning. These are not mutually exclusive, but are intricately intertwined. Their meaning for the school and for the evaluation of the school are described in the following excerpts from the Marcy Open School evaluation document.

#### MARCY OPEN SCHOOL:

*Personalized curriculum.* Curriculum will vary for each child as teaching extends from the interests, needs and abilities of each child or group of children. The school personnel seeks to be aware of each child as an individual, and of potential learning activities and materials. Decisions on the curriculum are made by the individual child, the teacher, and the parents.

*Experiential Nature of Learning.* The school seeks to have the children experience language rather than only learning to read, to experience computation rather than only to learn math, to be in and to learn from the community rather than only to learn about social studies. Participants in the school believe that experience is the best transmitter of knowledge. Further the child is expected to interact with his environment—to have an effect upon it in the process of experiencing it—to change it or to recognize ways in which he seeks to move toward change.

#### EVALUATION OF MARCY OPEN SCHOOL:

*Personalized Evaluation.* The determination of the success of the school will vary depending upon the values and perspectives of interested people. This evaluation will present a statement of what was made available by the school and of what was accomplished by children in the school. Decisions as to the success of the school and as to the validity of those activities must be left to the individual reading the evaluation report, according to his own perspective.

*Experiential Nature of Evaluation.* This evaluation will attempt to provide an opportunity for the reader to experience the school and its children. It will provide not only charts and statistics, but photographs, drawings and works of children and adults. Even so, it is acknowledged that, at best, any such report can only be an imperfect representation of the school and its processes. Further, this report is not presented as a final document representing the accomplishments of the school. It is, instead, presented as a report-in-process, to be reacted to and sent back for new descriptions.



new data, about which its readers may be concerned.

*Holistic Nature of Learning.* Much emphasis has been placed on the inter-relatedness of learning. Organizational structures, activities and materials are considered in terms of their multidimensionality of goals. Goal statements and staff development activities give conscious attention to a child's feelings about himself and his world, how those relate to his relationships with others and how those relate to his interest and ability in learning. The staff seeks activities which allow the child to experience the relationship between language, computation and other knowledges, rather than departmentalizing them into separate content areas.

*Holistic Evidence for Evaluation.* Three of the school's goals for children have been chosen for special attention in this evaluation. They include a range of process, content and context typical of the goal statement as a whole. The evidence presented attempts to observe a natural order of events as they happen in the school with a minimum of distortion through departmentalization. Both objective figures and subjective judgments are included and are considered to be valid. The provisioning of the school and the activities and products of the children are viewed, as much as possible, in terms of their multidimensionality of effect.

The involvement of identified decision makers and information users in the making of measurement and design decisions is based on the assumption that utilization is enhanced if users understand, *believe in*, and have a stake in the data. Understanding, belief, and interest are all increased when evaluators and decision makers share values about methods strategies. In any case, the making of decisions about what data to collect and how to collect it is a painstaking process. In part, the final design of an evaluation depends on calculated trade-offs, weighing options, and political/philosophic considerations; in part it depends on opportunity, resources, time constraints, and commitment. What is to be avoided is the routine selection of a design without consideration of its strengths and weaknesses in relation to this complex constellation of personal, political, and technical factors.

## EVALUATION RESEARCH STRATEGIES USING QUALITATIVE METHODS: SUMMARY AND OVERVIEW

If evaluation research is truly to become a paradigm of choices, then evaluators must understand the options available in matching



research methods to evaluation questions. The challenge is to select those methods most appropriate to particular decision maker concerns and specific evaluation situations. This chapter has identified those program circumstances and evaluation problems which are particularly suited to the gathering of qualitative data. The following is a checklist of questions which can be used to help decide if qualitative methods are an appropriate evaluation strategy. *If the answer to any of these questions is "yes," then the collection of some qualitative data is likely to be appropriate.*

### CHECKLIST OF EVALUATION SITUATIONS FOR WHICH QUALITATIVE METHODS ARE APPROPRIATE

1. Does the program emphasize individualized outcomes, i.e., different participants are expected to be affected in qualitatively different ways? And is there a need or desire to describe and evaluate these individualized client outcomes? Yes No
2. Are decision makers interested in elucidating and understanding the internal dynamics of programs—program strengths, program weaknesses, and overall program *processes*? Yes No
3. Is detailed, in-depth information needed about certain client cases or program sites, e.g., particularly successful cases; unusual failures; critically important cases for programmatic, financial, or political reasons? Yes No
4. Is there interest in focusing on the diversity among, idiosyncracies of, and unique qualities exhibited by individual clients or programs (as opposed to comparing all clients or programs on standardized, uniform measures)? Yes No
5. Is information needed about the details of program implementation—what clients in the program experience, what services are provided to clients, how the program is organized, what staff do, and basically inform decision makers as to what is going on in the program and how it has developed? Yes No
6. Are program staff and other decision makers interested in the collection of detailed, descriptive information about the program for the purpose of improving the program, i.e., is there interest in formative evaluation? Yes No
7. Is there a need for information about the nuances of program *quality*, i.e., descriptive information about the quality of program activities and outcomes, not just levels, amounts, or quantities of program activity and outcomes? Yes No
8. Will the administration of standardized measuring instruments (questionnaires and tests) be overly obtrusive in contrast to the gathering of data through natural observations and open-ended interviews, i.e., will the collection of qualitative data generate less reactivity among participants than the collection of quantitative data? Yes No



9. Is the state of measurement science such that no valid, reliable, and believable standardized instrument is available or readily capable of being developed to measure the particular program outcomes for which data are needed? Yes No
10. Are legislators or other decision makers/funders interested in having evaluators conduct program site visits such that the evaluators become the surrogate eyes and ears for decision makers who are too busy to make such site visits themselves and who lack the observing and listening skills of trained evaluators? Yes No
11. Are the goals of the program vague, general, and nonspecific, indicating the possible advantage of a goal-free evaluation approach to gather information about what effects the program is actually having? Yes No
12. Is there the possibility that the program may be affecting clients or participants in unanticipated ways and/or having unexpected side effects, indicating the need for a method of inquiry that can discover effects beyond those formally stated as desirable by program staff (again, an indication of the need for some form of goal-free evaluation)? Yes No
13. Has the collection of quantitative evaluation data become so routine that no one pays much attention to the results anymore, suggesting a possible need to break the old routine and use new methods to generate new insights about the program? Yes No
14. Is there a need and desire to *personalize* the evaluation process by using research methods that require personal, face-to-face contact with the program—methods that may be perceived as “humanistic” and personal because participants are not preordinately labeled and numbered, and methods that feel natural, informal, and understandable to participants? Yes No
15. Do decision makers and information users have philosophical or methodological biases that lead them to prefer qualitative methods, thus increasing the likelihood that they will find the results of a qualitative evaluation particularly believable, credible, understandable, and *useful*? Yes No
16. Are decision makers and evaluators interested in increasing their understanding of the program by developing a *grounded theory* of program actions and effects that is inductively derived from a holistic picture of the program? Yes No

## ON USING THE SUMMARY CHECKLIST

Matching research methods to evaluation questions is a complex, creative process. The questions above are meant to stimulate and guide in that process; they are not a mechanical tool for making routine decisions. The next chapter discusses some of the concrete and practical research design issues that must be addressed in using qualitative methods. In making the transition between the evaluation strategies in this chapter and the research design considerations in the next, it may be helpful to review some of *Halcolm's Laws of Evaluation Research Methods*.



1. *An evaluation not worth doing is not worth doing well.*
2. *Evaluation results always make clear to people what they had really wanted to know but forgot to ask.*
3. *Every evaluation serves a purpose, even if it is only to be a horrible example to others.*
4. *The perfect evaluation design isn't.*
5. *The Law of Divine Intervention: Alle Kunst ist unsunst Wenn ein Engel auf das Zündloch brunzt. (German version) All skill is in vain when an angel pees in the touchhole of your musket (English translation, Peers and Bennet, 1979).*



## EVALUATION DESIGNS

## THE FIRST EVALUATION

The young people gathered around Halcolm. “Tell us again, Teacher of Many Things, about the first evaluation.”

“The first evaluation was conducted a long, long time ago,” he began. “It happened in Ancient Babylon when Nebuchadnezzar was King. Nebuchadnezzar had just conquered Jerusalem in the third year of the reign of Jehoiakim, King of Judah. Now Nebuchadnezzar was a shrewd ruler. He decided to bring carefully selected children of Israel into the palace for special training so that they might be more easily integrated in Chaldean culture. This special program was the forerunner of the compensatory education programs that would become so popular in the twentieth century. The three-year program was royally funded with special allocations and scholarships provided by Nebuchadnezzar. The ancient text from the Great Book records that

the king spake unto Ashpenaz the master of his eunuchs that he should bring certain of the children of Israel, and of the King’s seed, and of the princes;

Children in whom was no blemish, but well-favored and skillful in all wisdom, and cunning in knowledge, and understanding science, and such as had ability in them to stand in the king’s palace, and whom they might teach the learning and the tongue of the Chaldeans.

And the king appointed them a daily provision of the king’s meat, and of the wine which he drank; so nourishing them for three years, that at the end thereof they might stand before the king [Daniel 1:3-5].

“Now this program had scarcely been established when the king found himself faced with a student rebellion led by a radical named Daniel who decided for religious reasons that he would not consume the king’s meat and wine. This created a real problem for the program



administrator who was responsible to the king. If Daniel and his co-conspirators did not eat their dormitory food they might fare poorly in the program and endanger not only future program funding but also the program director's head! The Great Book says,

But Daniel purposed in his heart that he would not defile himself with the portion of the king's meat, nor with the wine which he drank; therefore he requested of the prince of the eunuchs that he might not defile himself.

And the prince of the eunuchs said unto Daniel, I fear for my lord the king, who hath appointed your meat and your drink; for why should he see your faces worse liking than the children which are of your sort? Then shall ye make me endanger my head to the king [Daniel 1:8, 10].

“At this point, Daniel proposed history's first educational experiment and program evaluation. He and three friends would be placed on a strict vegetarian diet for ten days (nothing but pulse and water), while other students continued on the king's rich diet of meat and wine. At the end of ten days the program director would inspect the treatment group for any signs of physical deterioration and judge the productivity of Daniel's alternative diet plan. As Daniel described the experiment:

Prove thy servants, I beseech thee, ten days; and let them give us pulse to eat, and water to drink.

Then let our countenances be looked upon before thee, and the countenance of the children that eat of the portion of the king's meat: and as thou seest, deal with thy servants.

So he consented to them in this matter, and proved them ten days [Daniel 1:12-14].

“During the ten days of waiting Ashpenaz had a terrible time. He couldn't sleep; he had no appetite; and he had trouble working because he was preoccupied with worrying about how the evaluation would turn out. He had a lot at stake. Besides, in those days they hadn't quite worked out the proper division of labor so he had to play the roles of both program director and evaluator. You see . . .”

The young listeners interrupted Halcolm. They sensed that he was about to launch into a sermon on the origins of the division of labor when they still wanted to hear the end of the story about the origins of evaluation. “How did it turn out?” they asked. “Did Daniel end up looking better or worse from the new diet? Did Aspenaz lose his head?”



“Patience, patience,” Halcolm pleaded. “Aspenaz had no reason to worry. The results were quite amazing. The Great Book says that

at the end of ten days their countenances appeared fairer and fatter in flesh than all the children which did eat the portion of the king’s meat.

Thus Melzar took away the portion of their meat, and the wine that they should drink; and gave them pulse.

As for these four children, God gave them knowledge and skill in all learning and wisdom; and Daniel had understanding in all visions and dreams.

Now at the end of the days that the king had said he should bring them in, then the prince of the eunuchs brought them in before Nebuchadnezzar.

And in all matters of wisdom and understanding, that the king inquired of them, he found them ten times better than all the magicians and astrologers that were in all his realm [Daniel 1:15-18, 20].

“And that, my children, is the story of the first evaluation. Those were the good ole days when evaluations really got used. Made quite a difference to Aspenaz and Daniel. Now off with you—and see if you can do as well.”

*From Halcolm’s Evaluation Histories*

## A META-EVALUATION

A meta-evaluation is an evaluation of an evaluation. A great deal can be learned about evaluation designs by conducting a meta-evaluation of history’s first program evaluation. Let us imagine a panel of experts conducting a rigorous critique of this evaluation of Babylon’s compensatory education program for Israeli students.

- (1) Small sample size ( $N = 4$ ).
- (2) Selectivity bias since recruitment into the program was done by “creaming,” i.e., only the best prospects among the children of Israel were brought into the program to begin with.
- (3) Selectivity bias because students were self-selected into the treatment group.
- (4) Failure to clearly specify and control the nature of the treatment, thus allowing for the possibility of treatment contamination. We don’t know what other things besides a change in diet either group was involved in that might explain the outcomes observed.
- (5) Possibility of interaction effects between the diet and the students’ belief system and/or relationship to God.
- (6) Outcome criteria vague.
- (7) Measures of outcomes poorly operationalized and nonstandardized.



- (8) Single observer with deep personal involvement in the program introduces possibility of selective perception and bias in the observations.
- (9) Validity and reliability data are not reported for the instruments used to measure the final, summative outcome ("he found them ten times better than all the magicians and astrologers . . .").
- (10) Possible reactive effects from the students' knowledge that they were being evaluated.

Despite all of these threats to internal validity, not to mention external validity, the information generated by the evaluation appears to have been used to make a major decision about the program. Indeed, it is difficult to find a more exemplary model for the utilization of research in making educational policy decisions than that first evaluation conducted under the auspices of Nebuchadnezzar so many years ago. Immediately following determination of the evaluation results a policy decision was made to allow Daniel and friends to maintain their diet of pulse and water. The longitudinal indicators collected over the three-year period suggest that the decision was appropriate; Daniel did place first in his class.

*To my knowledge there is no better example of evaluation research having an immediate, decisive, and lasting impact on educational policy. Modern evaluation researchers, flailing away in seemingly futile efforts to affect contemporary governmental decision makers, can be forgiven a certain nostalgia for the "good old days" in Babylon when evaluation research really made a difference.* But should the results have been used? Given the apparent weakness of the evaluation design, was it appropriate to make a major policy decision on the basis of data generated by such a weak research design?

I would argue that not only was utilization exemplary in this case, but that *the research design was also exemplary.* The evaluation design was exemplary because the study was set up in such a way as to provide precisely the information needed by the program director to make the decision he needed to make. Certainly, it is a poor research design to study the relationship between nutrition and educational achievement. It is even a poor design to decide if all students should be placed on a vegetarian diet. But those were not the issues. The question that the program director had to deal with was whether or not to place four specific students on a special diet at their request. The information he needed concerned the consequences of that specific change and *only* that specific change. He showed no interest in generalizing the results beyond those four students, and he showed no interest in convincing others that the measures he made



were valid and reliable. He was the only person who had to trust the measures used, and so data collection was designed in such a way as to maximize his belief in the meaningfulness of the observations. If any bias existed in his observations, given what he had at stake, the bias would have operated *against* a demonstration of positive outcomes rather than in favor of such outcomes.

While there are hints of the whimsical in the suggestion that this first evaluation is exemplary, I do not mean to be facetious. I am absolutely serious in suggesting that the Babylonian example is an exemplar of utilization-focused evaluation. The decision maker who was to use information generated by the evaluation was clearly identified and deeply involved in every stage of the evaluation process. The evaluation question was carefully focused on needed information that could be used in the making of specific decisions. The evaluation methods and design were appropriately matched to the evaluation question. The results were understandable, credible, and relevant. Feedback was immediate and utilization was decisive. Few modern evaluations can meet the high standards for evaluation set by Aspenaz and Daniel several thousand years ago.

This chapter discusses some ways in which research designs can be appropriately matched to evaluation questions in an attempt to emulate the exemplary match between evaluation problem and research design achieved in the Babylonian evaluation. As with previous chapters, emphasis will be placed on the importance of being *strategic* in creating evaluation designs. Being strategic begins with recognizing the critical trade-offs involved in selecting from among a variety of possible research designs.

## CRITICAL TRADE-OFFS IN EVALUATION DESIGN

Strategies and trade-offs—these two themes go together. A discussion of design strategies and trade-offs is necessitated by the fact that there are no perfect research designs. There are always trade-offs. These trade-offs are necessitated by limited resources, limited time, and limits on the human ability to grasp the complex nature of social reality.

The very first trade-offs come in framing the evaluation question or questions to be studied. The problem here is to determine the extent to which it is desirable to study one or a few questions in great depth or to study many questions, but in less depth. This is what Guba (1978) calls the “boundary problem” in naturalistic inquiry evaluation.



Elsewhere I have discussed strategies for generating meaningful and relevant evaluation questions (Patton, 1978). Once a potential set of evaluation questions has been generated, it is necessary to begin the process of prioritizing those questions in order to decide which of them ought to be pursued at a particular point in time. Should all parts of the program be studied or only certain parts? Should all clients be studied or only some subset of clients? Should the evaluator aim at describing all program processes or is there reason to examine only certain selected processes in depth? Should all outcomes be examined or should the evaluation focus upon the attainment of only certain outcomes of particular interest at this point in time?

In my own experience the problem of establishing focus and priorities is much more difficult than the problem of generating potential questions at the beginning of the evaluation. Once a group of decision makers and information users begin to take seriously the notion that they can learn from the collection and analysis of evaluative information, they soon find that there are lots of things they would like to find out. The evaluator's role is to help decision makers and information users move from a rather extensive list of potential questions to a much shorter list of realistically possible questions, and finally to a focused list of essential and necessary questions.

An example of variations in evaluation focus may help illustrate the kinds of trade-offs involved. Suppose that a group of educators is interested in studying how a school program affects the social development of children of school age. They want to know how the interaction of children with others in the school setting contributes to the development of social skills. They believe that those social skills will be different for different children, and they are not sure of the range of social interactions that may occur, so they are interested in a naturalistic inquiry evaluation that will capture variations in program experience and individualized outcomes. Still, there are trade-offs in determining the final focus. It is clear that any given child has social interactions with a great many people. The problem in focusing our evaluation research endeavor is to determine how much of the social reality experienced by children we should attempt to describe. In a narrowly focused evaluation we might select one particular set of interactions and limit our study to those interactions—for example, the social interactions between teacher and children. Broadening the scope somewhat, we might decide to look at only those interactions that occur in the classroom, thereby increasing the scope of the study to include interactions not only between teacher



and child, but also among peers in the classroom and between any volunteers and visitors to the classroom and the children. Broadening the scope of the study still more, we might decide to look at all of the social relationships that children experience in schools; in this case we would move beyond the classroom to look at interactions with other teaching personnel in the school—for example, the librarian, school counselors, special subject teachers, the custodian and/or school administrative staff. Broadening the scope of the study still further, the educators might decide that it is important to look at the social relationships children experience in home and in school in order to understand how children experience those settings differently, and therefore to better understand the unique effects of the school. In this case we would include in our design interactions with parents, siblings, and other people in the home. Finally, one might look at the social relationships experienced throughout the full range of societal contacts that children have, including church, clubs, and even mass media contacts.

All of these are potentially important evaluation research questions. Suppose that we have a set amount of resources—for example, \$25,000—to conduct a study. At some level, any of these research endeavors could be undertaken for \$25,000. It is immediately clear, however, that there is a *trade-off between breadth and depth*. A highly focused question like the interactions between teacher and child could consume the entire amount of our resources and allow us to investigate the problem in great depth. On the other hand, we might attempt to look at all social relationships that children experience, but to look at each of them in a relatively cursory way in order, perhaps, to explore which of those relationships is primary. (If school relationships have very little impact on social development in comparison to relationships outside the school, decision makers could use that information to decide whether or not the school program ought to be redesigned to have greater impact on social development or if the school should forget about trying to directly affect social development at all.) The trade-offs involved are the classic trade-offs between breadth and depth.

In many ways the real trade-off between quantitative methods and qualitative methods is a trade-off between breadth and depth. Qualitative methods permit the evaluator to study selected issues in depth and detail; the fact that data collection is not constrained by predetermined categories of analysis contributes to the depth and detail of qualitative data. Quantitative methods, on the other hand, require the use of a standardized stimulus so that all experiences of people are limited to certain response categories. The advantage of



the quantitative approach is that it is possible to measure the reactions of many subjects to a limited set of questions, thus facilitating comparison and statistical aggregation of the data. By contrast, qualitative methods typically produce a wealth of detailed data about a much smaller number of people and cases.

## SAMPLE SIZE AND DEPTH OF INFORMATION

The breadth versus depth trade-off is applicable not only in comparing quantitative and qualitative methods; the same trade-off applies within qualitative methods. The human relations specialists tell us that we can never fully understand the experience of another person. The research question asks how much time and effort we are willing to invest in trying to increase our understanding about any single person's experience. Again, under conditions of limited resources, we can look at a narrow range of experiences for a larger number of people, or a broader range of experiences for a smaller number of people. Take the case of interviews. Interviewing with an instrument that provides a respondent with largely open-ended stimuli typically takes a great deal of time. In North Dakota when I was studying various aspects of open education we developed an open-ended interview consisting of 20 questions that were asked of children in grades one to eight in various open classrooms. Those questions consisted of items such as, "What do you like most about school?" and "What don't you like about school?" These interviews took between half an hour and two hours depending on how articulate students were and how old they were. It would certainly have been possible to have longer interviews. Indeed, I have conducted in-depth interviews with people that ran six to eight hours over a period of a couple of days. On the other hand, it would have been possible to ask fewer questions, to make the interviews shorter, and to obtain less depth.

To illustrate this trade-off between breadth and depth in sampling human behavior, let us consider the full range of possibilities. It is possible (and indeed it has been done) to study a single individual over an extended period of time—for example the study, in-depth, of one day in the life of one child. This necessitates gathering detailed information about every occurrence in that child's life and every interaction involving that child during some time period. With a more limited research question we might study several children during a more limited period of time. With still a more limited research question, or an interview of a half-hour, we could interview yet a



larger number of children on a smaller number of issues. The extreme case would be to spend all of our resources and time asking a single question of as many children as we could interview given the resource constraints.

*There is no rule of thumb that tells a researcher precisely how to focus an evaluation question. The extent to which a research question is broad or narrow depends on the resources available, the time available, and the needs of decision makers. In brief, these are not choices between good and bad, but choices among alternatives, all of which have merit.*

## UNITS OF ANALYSIS

The evaluation design specifies the unit or units of analysis to be studied. Decisions about samples, both sample size and sampling strategies, depend on prior decisions about the appropriate unit of analysis to study. Sometimes individual people, clients, or students are the unit of analysis. This means that the primary focus of data collection will be on what is happening to individuals in the program and how individuals are affected by the program. Individual variation would be the primary evaluation issue.

Focusing on and comparing groups of people in a program or across programs involve a different unit of analysis. One may be interested in comparing demographic groups (males compared with females, whites compared with blacks) or programmatic groups (dropouts versus people who complete the program, people who do well versus people who do poorly, people who experience group therapy versus people who experience individual therapy). One or more groups are selected as the unit of analysis when there is some important characteristic that separates people into groups and that characteristic has important implications for the program.

A different unit of analysis involves focusing on different parts of a program. Different classrooms within a school might be studied so that the classroom was a unit of analysis. Outpatient and inpatient programs in a medical facility might be studied. The intake part of a program might be studied separately from the service delivery part of a program as separate units of analysis.

Entire programs can become the unit of analysis. In state and national programs where there are a number of local sites the appropriate unit of analysis may be local projects. The focus in this case would be on variations among projects more than on variations among individuals within programs.



These different units of analysis are not mutually exclusive. However, each unit of analysis implies a different kind of data collection, a different focus for the analysis of data, and a different level at which statements about findings and conclusions would be made. Neighborhoods can be units of analysis, communities, cities, states, and even nations in the case of international programs. The key issue in selecting and making decisions about the appropriate unit of analysis is to decide what it is you want to be able to say something about at the end of the study. At what level do decision makers really need information? Do they want information about the different experiences of individuals in programs or do they want to know about variations in program processes at different sites? These are differences in nuance. The decision maker typically will be unable to say to the evaluator, "The unit of analysis we want to study is . . ." The evaluator must be able to hear the real issues involved in decision maker questions and translate those issues into the appropriate unit of analysis, then check out that translation with the decision makers.

## SAMPLING STRATEGIES

Once the unit or units of analysis have been identified and defined, decisions about sampling design can be made. A basic distinction involves the difference between *random* sampling and *purposeful* sampling. Random sampling is the appropriate strategy when one wants to generalize from the sample studied to some larger population. Having collected information from some sample of people who have experienced a program, it may be desirable to be able to generalize to all people who have experienced the program; or, having studied a few programs or program sites in a national operation, it may be desirable to generalize from that smaller number of programs studied to all programs. The reason for using random sampling is to increase the likelihood that the data collected are *representative* of the entire population of interest. Sample size is determined by the size of the population to which one wants to generalize, the expected amount of variation in that population, and the amount of error one is willing to accept. For a straightforward discussion of random sampling and sample size in evaluation, see Morris and Fitz-Gibbon (1978c).

Purposeful sampling is used as a strategy when one wants to learn something and come to understand something about certain select cases without needing to generalize to all such cases. In order to do purposeful sampling certain information must be known about variations among cases. Let us suppose that we are interested in



evaluating a national program with hundreds of local sites. We can assume that some programs are probably operating reasonably well, even quite well, and that other programs verge on being disasters. Perhaps we also know that most programs are doing “okay.” This information comes from decision makers who have traveled around to see enough programs that they have a basic idea about what the variation is. The question is how to sample programs for the evaluation. If one wanted to document more precisely the natural variation among programs, a random sample would be appropriate, preferably a random sample of sufficient size to be truly representative of and that would permit generalizations to the total population of programs. However, the decision makers already have a basic sense of what program variation is like. The more critical question may be for them to understand *the extreme cases*. With limited resources and limited time they might learn more by intensively studying one or more examples of poor programs and one or more examples of excellent programs. The evaluation focus, then, becomes a question of understanding under what conditions programs get into trouble and under what conditions programs exemplify excellence. It is not even necessary to randomly sample poor programs or excellent programs. Decision makers and evaluators think through *what cases they could learn the most from*, and those are the cases that are selected for study.

At the local project level the same strategy may apply. Instead of studying some representative sample of clients in a program, decision makers may decide that they can learn the most by studying and understanding the unusual cases in the program: a few of the people who are really struggling and a few of the people who are really doing well. In many cases more can be learned from intensively studying extreme cases than can be learned from trying to determine what the average case is like.

Purposeful sampling is also a strategy to be used to help manage the trade-off between the desire for in-depth, detailed information about cases and the desire to be able to generalize about the program. Limited resources may mean that it is not possible to get detailed information from a sufficiently large sample size to make generalizations. Indeed, this problem of small sample size is probably the most typical situation in the use of qualitative methods. Decision makers, under these conditions, may be able to identify on the basis of their experience and knowledge a few “typical” programs. The study of such typical programs does not, of course, permit generalizations in any rigorous sense. It does, however, mean that the



processes and effects described for the *typical* program need not be dismissed as peculiar to “poor” sites or “excellent” sites. When the typical site sampling strategy is used, the site is specifically selected because it is not in any major way atypical. Decision makers may have made their peace with the fact that there will always be some poor programs and some excellent programs, but the programs they really want more information about are those run-of-the-mill programs that are “hard to get a handle on.” It is important, when using this strategy, to attempt to get broad consensus about which programs are “typical.” If a number of such programs are identified, and only a few can be studied, and there is no other basis for selecting among them purposefully, then it is possible to randomly select from among all “typical” programs identified to select those few typical cases that actually will be included in the study.

A different strategy for dealing with the problem of representativeness under conditions of small sample size is to maximize the variation in site selection or case selection. Thus, if programs are spread around the state—some in rural areas, some in urban areas, some in suburbs—but sufficient resources are not available to randomly select a sufficient sample size to generalize to the state, the evaluator can at least be sure that the geographical variation among sites is represented in the study. By attempting to increase the diversity or variation in the sample, the evaluator will have more confidence in those patterns that emerge as common among sites, while at the same time being able to describe some of the variation that has emerged to make programs unique as they adapt to different settings. The same strategy could be used for a single program in selecting individuals for study. By including in the study sample individuals who staff-identify as having had quite different experiences it is possible to more accurately describe the variation in the program and to *understand* variations in experiences. Again, the evaluation using a *maximum variation sampling strategy* would not be attempting to generalize findings to all clients or all programs.

Another strategy for selecting purposeful samples is to look for *critical* cases. Critical cases are those that can make a point quite dramatically or are, for some reason, particularly important in the scheme of things. A clue to the existence of a critical case is a statement by a decision maker to the effect that “if the program doesn’t make it there, it won’t make it anywhere.” Perhaps, then, the focus of the evaluation should be on understanding what is happening in that critical program. Another kind of clue would be a statement to the effect that “if that program is having problems then we can be



sure all the programs are having problems.” Looking for the critical case is particularly important where resources may limit evaluation to the study of only a single site. Under such conditions it makes strategic sense to pick the site that would yield the most information and have the greatest impact on decision maker actions and understanding.

While studying one or a few critical cases does not technically permit broad generalizations to all possible cases, logical generalizations can often be made from the weight of evidence produced in studying a single, critical case. Physics provides a good example of such a critical case. In Galileo’s study of gravity he wanted to find out if the weight of an object affected the rate of speed at which it would fall. Rather than randomly sampling objects of different weights in order to generalize to all objects in the world, he selected a critical case—the feather. If in a vacuum, as he demonstrated, a feather fell at the same rate as some heavier object (a coin) then he could logically generalize from this one critical case to all objects. His findings were enormously useful *and* credible.

There are many comparable critical cases in social action programming—if one is creative in looking for them. For example, suppose national policy makers want to get local communities involved in making decisions about how their local program will be run, but they are not sure that the communities will understand the complex regulations governing their involvement. The first critical case is to evaluate the regulations in a community of well-educated citizens; if they cannot understand them, less educated folks are certain to find the regulations incomprehensible. Or conversely, one might consider the critical case to be a community consisting of people with quite low levels of education; “if *they* can understand the regulations, anyone can.”

A variation of the critical case strategy involves selecting (or sometimes avoiding) a politically sensitive site or unit of analysis. For example, a statewide program may have a local site in the district of a state legislator who is particularly influential. By studying carefully the program in that district evaluation data may be more likely to attract attention and get used. This does not mean that the evaluator then undertakes to make that site look either good or bad, depending on the politics of the moment. This is simply an additional sampling strategy for trying to increase the usefulness and utilization of information where resources permit the study of only a limited number of cases.



Identification of critical cases depends on recognition of the key dimensions that make for a critical case. As noted in the last paragraph, the critical dimension may be political sensitivity or visibility. A critical case might be indicated by the financial state of a program; a program with particularly high or particularly low cost-per-client ratios might suggest a critical case. A critical case might come from a particularly difficult program location. If the funders of a new program are worried about recruiting clients or participants into a program, it may make sense to study the site where resistance to the program is expected to be greater to provide the most rigorous test of the possibility of program recruitment. If the program works in that site, “it could work anywhere.”

Finally, there is the strategy of sampling by convenience: Which cases can be studied most easily? This is probably the most common sampling strategy—and the least desirable. Too often evaluators using qualitative methods think that since the sample size they can study is too small to permit generalizations it doesn’t matter how cases are picked, so they might as well pick ones that are easy to access and inexpensive to study. While convenience and cost are real considerations, they should be the last factors to be taken into account after the evaluator and decision makers have carefully considered how they can strategically get the most information of greatest utility from the limited number of cases to be sampled. Purposeful, strategic sampling can yield crucial information about critical cases.

Table 5.1 summarizes the sampling strategies discussed in this section. As this table shows, different sampling strategies serve different purposes.

## THE CREDIBILITY OF SAMPLING STRATEGIES

The evaluator, in this process of developing the evaluation design, is trying to consider and anticipate the kind of arguments that will lend credibility to the data and the kind of arguments that will be used to attack the data. Reasons that are used to make site selections or individual case selections need to be carefully articulated and made explicit. Moreover, it is also important to make explicit the reasons why any particular sampling strategy may lead to distortions in the data—that is, to anticipate criticisms that will be made of a particular sampling strategy. Having weighed the evidence and considered the alternatives, evaluators make the sampling decision, sometimes painfully, but always with the recognition that there are no perfect



Table 5.1 Sampling Strategies

<i>Type</i>	<i>Purpose</i>
A. Random sampling	Avoids systematic bias in the sample; large sample size is important for making generalizations.
1. simple random sample	Achieve a representative sample that permits generalizations to the whole population.
2. stratified random and cluster samples	Increase confidence in making generalizations to particular subgroups or areas.
B. Purposeful sampling	Increase the utility of information obtained from small samples; sampling criteria based on the reputation of programs among key decision makers and/or on previous data collected from programs.
1. sampling extreme or deviant cases	Provide decision makers with information about unusual cases that may be particularly troublesome or enlightening, e.g., outstanding successes/notable failures; programs with long waiting lists vs. programs with recruitment problems; unusually high morale and low morale programs, etc.
2. sampling typical case(s)	Avoid studying a program where the results would be dismissed outright because <i>that</i> program is known to be special, deviant, unusual, extreme, etc.
3. maximum variation sampling—picking three or four cases that represent a range on some dimension (e.g., size, location, budget)	Increase confidence in common patterns that cut across different programs; document unique program variations that have emerged in adapting to different conditions.
4. sampling critical cases	Permits <i>logical</i> generalization and maximum application of information to other cases because if it's true of this one case, it's likely to be true of all other cases.
5. sampling politically important or sensitive cases	Attracts attention to the study (or avoids attracting undesired attention by purposefully eliminating from the sample politically sensitive cases).
6. convenience sampling—take the easy cases	Saves time, money, and effort.

designs. The sampling strategy in the Babylonian evaluation worked because Daniel and his three friends were the critical cases.

The fact that a small sample size must be chosen does not automatically mean that the sampling strategy should be purposeful instead of random. For many audiences random sampling, even of small samples, will substantially increase the credibility of the data. I recently worked with a program that annually appears before the state legislature and tells “war stories” about client successes, also sometimes including a few stories about failures. They decided they



wanted to begin collecting evaluation information. Because they are striving for individualized outcomes they rejected the notion of basing the evaluation entirely on some standardized pre-post instrument. They wanted to collect case histories and do in-depth case studies of clients. Of course, they had very limited resources and time to devote to data collection. In effect, each program, many of which serve two or three hundred families a year, felt that they could only do ten or fifteen detailed, in-depth clinical case histories each year. We systematized the kind of information that would be going into the case histories at each program site and then set up a random procedure for selecting those clients whose case histories would be recorded in depth. These programs had systematized and randomized their collection of "war stories." While they cannot generalize to the entire client population on the basis of ten cases from their program, they will be able to tell legislators that the war stories they are reporting were randomly selected *in advance* of knowledge of how the outcomes would appear, and the information collected was comprehensive. The credibility of systematic and randomly selected war stories is considerably greater than the personal selection of cases to report after the fact—that is, after outcomes are known.

The analysis chapter will deal with the problem of generalization as one analyzes and tries to make sense out of the data collected. It is too late, however, to make crucial design decisions about generalizations once one has begun analyzing the data. *Decisions about what one wants to be able to say with the data, for what purpose, and with what degree of credibility are decisions that must be made in designing the evaluation.*

## DATA SAMPLES

The preceding discussion has concerned the issue of how one samples units of analysis, for example, programs, groups, or individuals. The conduct of naturalistic inquiry also involves a set of decisions about sampling during actual data collection. When doing observations it is not possible to capture everything. It is therefore necessary to make decisions about which activities to observe, which people to observe and interview, and what time periods will be selected to collect data. The strategies for making these decisions are actually the same as those used for sampling units of analysis. One can randomly sample time periods, activities, or people; or one can sample purposefully, deciding that certain activities are critical or that certain key informants are more knowledgeable than others. Likewise, the observer may look for extreme cases, typical cases, or



a variety of cases, activities, behaviors, or people. Indeed, once in the field, observers will frequently use all of these approaches at varying times for different parts of the data collection.

Time sampling can be an especially important issue because programs typically function in different ways at different times during the year. Of course, with some programs there is never a good time to collect data. I have learned that this is the case with schools. Educators will tell you that you don't want to collect data in the schools before Halloween because the school year is just getting started and the kids aren't quite fixed in the patterns that will be maintained later in the year. The period between Halloween and Thanksgiving is really too short to do very much, and then, of course, after Thanksgiving everybody's getting ready for Christmas, so that's not a typical or convenient period. It then takes students a few weeks after Christmas to get their attention focused back on school and then the winter malaise sets in and both teachers and students become deeply depressed with the endlessness of winter (at least in Minnesota). Then, of course, once spring hits, attention is focused on the close of school and the kids want to be outside, so that's not an effective time to gather data.

There are limits to how much one can apply logic and deduction in making sampling decisions, whether the decision is about which time periods to sample or which activities to observe. The arguments can rapidly become circular and at times solutions may appear hopeless. The trick is to keep coming back to the criterion of usefulness. What data collected during what time period describing what activities would make a difference? There are no perfect evaluation designs, only more and less useful ones.

## CONFIDENTIALITY

Confidentiality can be an especially difficult problem in evaluations of single programs, particularly if there are relatively few participants. Where confidentiality cannot be protected it is important that the evaluator make it clear to people being interviewed and participants being observed that while people will not be identified specifically in a report, it may be possible to identify them from descriptions of what they have done or quotations of what they have said. This is particularly true when reporting data from staff of a small program.

Unlike the social science tradition which changed the names of towns and places studied (although all the social scientists reading reports on these studies knew from informal sources where those places were), the purpose of program evaluation is to make decisions



about specific programs. It is therefore necessary to know what those programs are and where they are located, unless some sample of programs is studied in order to make generalizations about a larger group of programs. Under some circumstances it may be possible and desirable not to report the names of actual programs. Because each situation is different, how confidentiality will be handled is an issue to be negotiated with decision makers, information users, and funders *before* data are collected. This allows the evaluator to build in protections all along the way from data collection to analysis—protection for research subjects (program staff and participants) and protection for the evaluator.

## FORESIGHT IN EVALUATION DESIGNS

Designing an evaluation requires a great deal of foresight. The evaluator tries to anticipate how data will be used and what will be useful when the study is completed. At best this is a chancy operation, but the attempt requires making decisions about the appropriate evaluation focus, the appropriate methods, sampling strategies, design decisions, and analysis approaches. It is particularly important to include in the research design consideration of how the data will be analyzed. The chapters on interviewing, observation, and analysis will include information that will help in making design decisions. Before turning to those chapters, however, I want to briefly consider the design strategy of methodological mixes. Chapter Three, “The Strategy of Qualitative Methods,” emphasized a single, well-integrated, and comprehensive approach to naturalistic evaluations. In actuality the evaluation design for a particular situation may call for multiple methods or methodological mixes.

## METHODOLOGICAL MIXES

### TRIANGULATION

There are two ways in which methodological mixes are achieved. The first is through triangulation, or the combination of methodologies in the study of the same phenomena or programs. In the case of evaluation research this can mean using both quantitative and qualitative strategies to study the same program(s). Denzin (1978) has identified four basic types of triangulation: (1) *data triangulation*—the use of a variety of data sources in a study; (2) *investigator triangulation*—the use of several different researchers or evaluators; (3) *theory triangulation*—the use of multiple perspectives to inter-



pret a single set of data; and (4) *methodological triangulation*—the use of multiple methods to study a single problem or program.

Denzin explains that the logic of triangulation is based on the premise that

no single method ever adequately solves the problem of rival causal factors. . . . Because each method reveals different aspects of empirical reality, multiple methods of observations must be employed. This is termed triangulation. I now offer as a final methodological rule the principle that multiple methods should be used in every investigation [Denzin, 1978:28].

Triangulation is ideal. It is also very expensive. Most evaluations involve limited budgets, short time frames, and political constraints. Certainly, one important strategy for conducting evaluation research is to employ multiple strategies. However, in the real world of local program evaluation attempts at triangulation may mean a series of poorly implemented methods rather than one approach well executed. Since this book is aimed at enlarging the evaluator's repertoire to include qualitative methods, no further elaboration of triangulation strategies involving the use of several methods simultaneously will be undertaken. Where possible, triangulation is to be highly recommended. *Indeed, the capability to implement a strategy of triangulation means that evaluators must include in their repertoire of skills the ability to use qualitative methods.*

## MIXING DATA, DESIGN, AND ANALYSIS APPROACHES

While triangulation is one way of achieving methodological heterogeneity, a second approach is to borrow and combine parts from pure methodological strategies, thus creating mixed methodological strategies. To accomplish this it is necessary to separate the measurement, design, and analysis components of the hypothetico-deductive and holistic-inductive paradigms. The ideal-typical qualitative methods strategy is made up of three parts: (1) qualitative data, (2) a holistic-inductive design of naturalistic inquiry, and (3) content or case analysis. For the traditional hypothetico-deductive approach to scientific inquiry the ideal study would include (a) quantitative data from (b) experimental (or quasi-experimental) research designs and (c) statistical analysis. (For a systematic comparison of the holistic-inductive paradigm of research with the hypothetico-inductive paradigm see Patton, 1978:199-238.)



There are strong epistemological arguments for maintaining the integrity of a comprehensive and pure qualitative methods approach in research that integrates qualitative measurement, holistic-inductive designs based on naturalistic inquiry, and content analysis. Yet, in the real world of program evaluators it may be necessary and desirable to mix different types of measurement, design, and analysis. Thus, it is altogether possible, and even reasonable, to have an experimental design with randomized assignment of subjects to treatment and control groups and yet to collect qualitative data from those subjects. This means that experimental and quasi-experimental designs are entirely consonant with qualitative measurement if the evaluation researcher decides that the outcomes under study are best studied by observing and recording the open-ended behavior and expressions of program participants.

Nor does the collection of descriptive and narrative data through in-depth interviewing and detailed observation presuppose a content analysis strategy of data analysis. It is possible to superimpose quantitative scales and dimensions on qualitative data. Thus, in the data analysis phase of a project the researcher may decide to convert qualitative measures into quantitative scales that can be statistically manipulated.

Research designs allow for considerable mixing. This is true even within a particular design strategy. As noted in Chapter Three, the extent to which a study employs a naturalistic design is always a matter of degree. This applies particularly with regard to the extent to which the researcher places conceptual constraints on or makes presuppositions about the research setting. In practice the naturalistic approach to evaluation may often involve moving back and forth between inductive, open-ended, and phenomenological encounters with research settings to more hypothetical-deductive attempts to verify "hypotheses" or solidify ideas which emerged from those more open-ended experiences. Thus, naturalistic inquiry becomes a mixed strategy as the investigator moves back and forth between simply exploring and experiencing the research setting and studying that setting for purposes of verification and replication.

A variety of mixes, then, are possible—mixes of measurement, design, and analysis. At this point these mixes have been described in relatively abstract terms. In order to make the choices available to the active-reactive-adaptive evaluator more clear, and in order to illustrate the creative possibilities that can emerge out of flexible approaches to research design, a number of brief examples of research strategies which involve mixes of measurement, design, and analysis have been constructed in the pages which follow. By no



means are the examples provided meant to be exhaustive of the possibilities for constructing mixed methodological strategies. Moreover, these examples have been constructed under the artificial constraint that only one kind of measurement, design, and analysis could be used in each case. In practice, of course, the possible mixes are much more varied, because any given study could include several measurement approaches, design approaches, and analytical approaches.

### **THE CASE OF OPERATION REACH-OUT: VARIATIONS IN PROGRAM EVALUATION DESIGN**

With funds provided by United Way, local foundations, and the state government, a comprehensive program is established in a major city to serve high school age students who are high-risk educationally (poor grades, poor attendance, poor attitudes toward school), highly vulnerable in terms of their health (poor nutrition, sedentary lifestyle, high drug use), and likely candidates for contact with the criminal justice system (histories of juvenile delinquency, poor employment prospects, and alienation from dominant societal values). The program consists of experiential education internships through which these high risk students get individual attention in basic skills instruction, part-time job placements that permit them to earn income while gaining work exposure, and participation in peer group discussions aimed at changing health values, establishing a positive peer culture, and increasing social integration. The mandate for the program includes a requirement that the program be evaluated. Several evaluation scenarios are possible.

#### **PURE HYPOTHETICAL-DEDUCTIVE APPROACH TO EVALUATION: EXPERIMENTAL DESIGN, QUANTITATIVE DATA, AND CONTENT ANALYSIS**

It is determined by the evaluator that the program does not have sufficient resources to include all of the youth in the target population. Therefore, recruitment of youth is based on random selection with part of the group admitted into the program on a random basis and the other part receiving no treatment intervention. Before the beginning of the program and again one year later all youth, both those in the program and those in the control group, are administered standardized instruments measuring school achievement, self-esteem, anomie, alienation, and locus of control. Rates of school attendance, sick-



ness, drug use, and delinquency are obtained for each group. When all data have been collected at the end of the year, comparisons between the control group and experimental group are made using inferential statistics with an F-test significance level of .05 used as the criteria to determine significant differences between the two groups.

#### **PURE QUALITATIVE STRATEGY: NATURALISTIC INQUIRY, QUALITATIVE MEASUREMENT, AND CONTENT ANALYSIS**

Procedures for recruiting and selecting participants for the program are determined entirely by the staff. The evaluator finds a convenient time to conduct an in-depth interview with participants as soon as they are admitted into the program. These in-depth interviews ask students to describe what school is like for them, what they do in school, how they typically spend their time, what their family life is like, how they approach academic tasks, their views about health, and their behaviors/attitudes with regard to delinquent and criminal activity. In brief, participants are asked to describe themselves and their social world. The evaluator finds out from program staff when the program activities will be taking place and observes those activities, collecting detailed data about what happens during those activities: participant behaviors, participant conversations, staff behaviors, staff-participant interactions, and related phenomena. During the course of the program the evaluator finds convenient opportunities for conducting additional in-depth interviews with participants to find out how they view the program, what kind of experience they are having, and what they are doing. Near the end of the program, in-depth interviews are conducted with the participants to find out what behaviors they have changed, how they view the world at this point in time, and what their expectations are for the future. In-depth interviews are also conducted with program staff. These data are content analyzed to find out what patterns of experience participants bring to the program, what patterns characterize their participation in the program, and what patterns of change are reported by and observed in the participants.

#### **MIXED FORM: EXPERIMENTAL DESIGN, QUALITATIVE MEASUREMENT, AND CONTENT ANALYSIS**

As in the pure experimental form, potential participants are randomly assigned to treatment and control groups. In-depth inter-



views are conducted with all youth, both those in the treatment group and those in the control group, before the program begins. The focus of those interviews is similar to that in the pure qualitative approach. Interviews are conducted again at the end of the program. Content analysis is performed separately on the data from the control group and the experimental group. The patterns found in the control group and the experimental group are then compared and contrasted.

### **MIXED FORM: EXPERIMENTAL DESIGN, QUALITATIVE MEASUREMENT, AND STATISTICAL ANALYSIS**

Participants are randomly assigned to treatment and control groups, and in-depth interviews are conducted both before the program and at the end of the program. These interview data in raw form are then given to a panel of judges who rate each interview along several outcome dimensions operationalized as a ten-point scale. For both the preinterview and the postinterview the judges assign ratings on such dimensions as likelihood of success in school (low = 1, high = 10), likelihood of committing criminal offenses (low = 1, high = 10), commitment to education, commitment to engaging in productive work, self-esteem, and manifestation of desired nutritional and health habits. Inferential statistics are then used to compare these two groups and the ratings received by participants in the two groups. Judges make the ratings without knowledge of which participants were in the treatment groups and the control group. Outcomes on the rated scales are also statistically related to background characteristics of participants.

### **MIXED FORM: NATURALISTIC INQUIRY, QUALITATIVE MEASUREMENT, STATISTICAL ANALYSIS**

As in the pure qualitative form, students are selected for the program on the basis of whatever criteria staff choose to apply. In-depth interviews are conducted with all students before the program and at the end of the program. These data are then submitted to a panel of judges who rate them on a series of dimensions similar to those listed in the previous example. Change scores are computed for each individual, and changes are statistically related to background characteristics of the students to determine in a regression format which characteristics of students are likely to predict success in the program. In addition, observations of program activities are rated on a set of scales developed to quantify the organizational attributes of



activities: for example, the extent to which the activity involved active or passive participation, the extent to which student-teacher interaction was high or low, the extent to which interactions were formal or informal, and the extent to which participants had input into program activities. Ratings of activities based on qualitative descriptions are then aggregated to provide an overview of the treatment environment of the program.

### **MIXED FORM: NATURALISTIC INQUIRY, QUANTITATIVE MEASUREMENT, STATISTICAL ANALYSIS**

Students are selected for the program according to staff criteria. The evaluator enters the program setting without any predetermined categories of analysis or presuppositions about important variables or variable relationships. The evaluator observes important activities and events in the program, looking for the types of behaviors and interactions that will emerge. For each new type of behavior or interaction, the evaluator creates a category and then uses a time and space sampling design to count the frequency with which those categories of behavior and interaction are exhibited. The frequency of the manifestation of observed behaviors and interactions are then statistically related to such characteristics as group size, duration of the activity, staff-student ratios, and social/physical density.

### **MAKING CHOICES ABOUT RESEARCH METHODS**

The examples just listed provide only a few illustrations of possible research strategies. Figure 5.1 summarizes these six possibilities. It is also possible to combine strategies, although the same evaluator cannot employ an experimental design and naturalistic inquiry at the same time. Thus, certain designs pose constraints that exclude other possibilities. It is not possible to create an experimental situation with treatment and control groups while at the same time studying the natural evolution of a program, including the natural development by program staff of criteria for selection into the program. The examples also illustrate another limitation on mixes. It is possible to convert detailed, qualitative descriptions into quantitative scales for purposes of statistical analysis. It is not possible, however, to work the other way around and convert purely quantitative measures into detailed, qualitative descriptions.

Which research design is best? Which strategy will provide the most useful answers to decision makers? There is no simple,



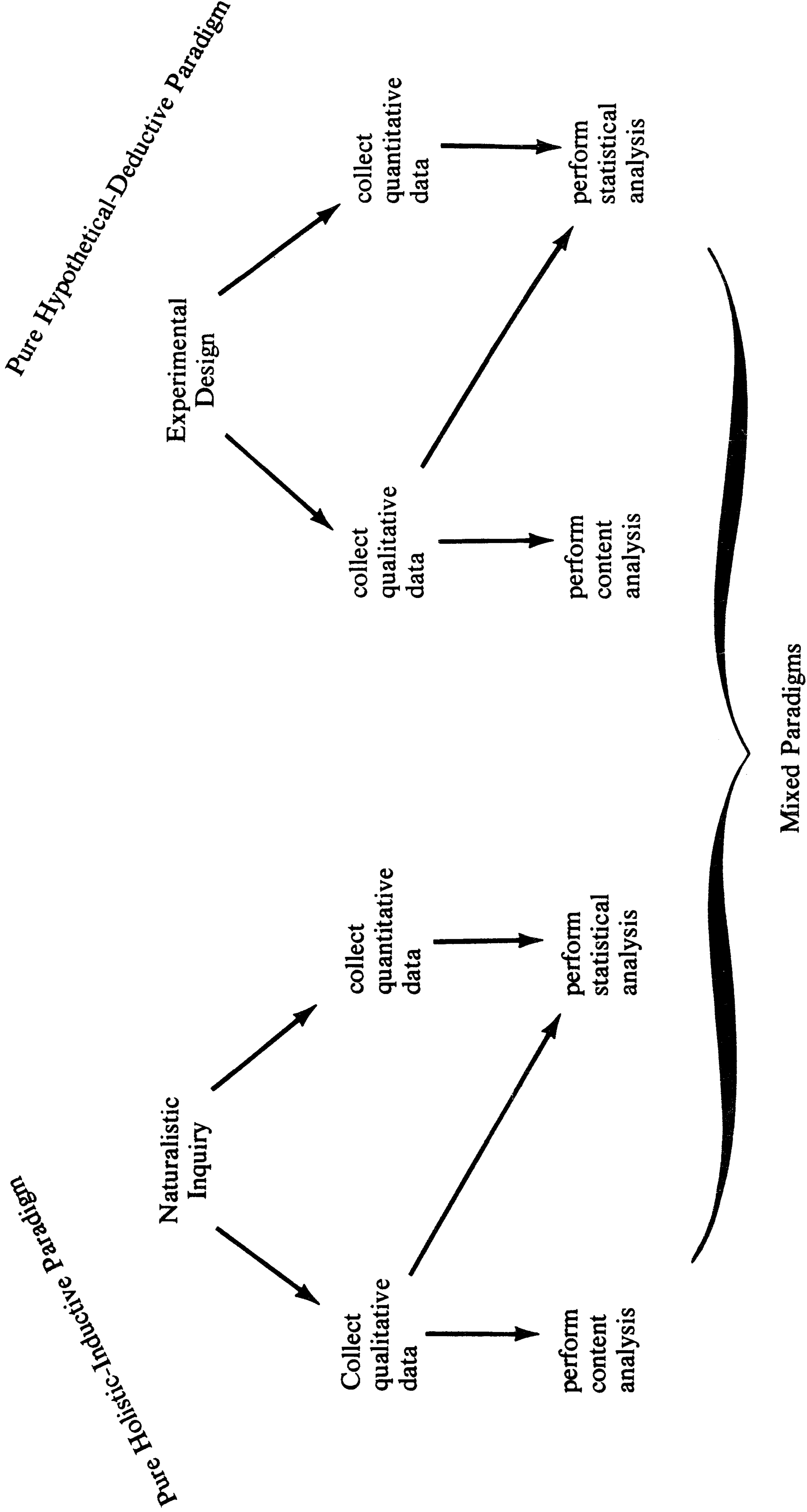


Figure 5.1 Measurement, Design, and Analysis: Pure and Mixed Combinations



immediate, and universal answer to that question. The answer in each case will depend on what decision makers want to know, the purpose of the evaluation, the evaluation funds available, the political context, and the interests/abilities/biases of the evaluators and decision makers.

What is certain is that different methods produce quite different information. The challenge is to find out which information is most needed and most useful in a given situation, and then to employ those methods best suited to producing the needed information. Martin Trow (1970) points out (quite nicely, I think) the difference between arguments about which methods are most appropriate for studying a particular problem as opposed to arguments about the intrinsic and universal superiority of one method over another.

Every cobbler thinks leather is the only thing. Most social scientists, including the present writer, have their favorite research methods with which they are familiar and have some skill in using. And I suspect we mostly choose to investigate problems that seem vulnerable to attack through these methods. But we should at least try to be less parochial than cobblers. Let us be done with the arguments of “participant observation” *versus* interviewing—as we have largely dispensed with the arguments for psychology *versus* sociology—and get on with the business of attacking our problems with the widest array of conceptual and methodological tools that we possess and they demand. This does not preclude discussion and debate regarding the relative usefulness of different methods for the study of specific problems or types of problems. But that is very different from the assertion of the general and inherent superiority of one method over another on the basis of some intrinsic qualities it presumably possesses [Trow, 1970:149; italics in the original].

The first chapter suggested that evaluation research should be built on the foundation of a “paradigm of choices” rather than become the handmaiden of any single, narrow disciplinary or methodological paradigm. But be careful, the Sufis would warn us, for the exercise of *real* choice can be elusive. Trow admonishes us to “at least try to be less parochial than cobblers.” The Sufi teacher, Mulla Nasrudin, might suggest that all too often the methods choices made by evaluators are like the bear’s decision to like honey.

One day in a sudden impulse of generosity a bear decided to enlighten the other animals in the forest about the marvelous properties of honey. The bear assembled all the other animals together for his momentous announcement.



“I have studied the matter at great length,” began the bear, “and I have decided that honey is the best of all foods. Therefore, I have chosen to like honey. I am going to describe to you the perfect qualities of honey which, due to your past prejudices and lack of experience, you have ignored. Then you will be able to make the same rational decision that I have made.

“Honey comes conveniently packaged in beautifully shaped prisms of the most delicate texture. It’s ready to eat, slides down the throat ever so easily, is a highly nutritious source of energy, digests smoothly, and leaves a lingering taste of sweetness on the palate that provides pleasure for hours. Honey is readily available and requires no special labor to produce since bees do all the work. Its pleasing aroma, light weight, resistance to spoilage, and uniformly high quality make it a food beyond compare. It comes ready to consume—no peeling, no killing, no tearing open—and there’s no waste. What’s more, it has so many uses; it can be eaten alone or added as an improvement to any other food.

“I could go on and on but suffice it to say that I have studied the situation quite objectively and at great length. A fair and rational analysis leads to only one conclusion. Honey is the supreme food and any reasonable animal will undoubtedly make the same conscious decision I have made. I have chosen to like honey.”



## *Collecting Qualitative Data*

- Always be suspicious of data collection that goes according to plan.
- Research subjects have also been known to be people.
- The evaluator's scientific observation is some person's real-life experience. Respect for the latter must precede respect for the former.
- Total trust and complete skepticism are twin losers in the field. All things in moderation, especially trust and skepticism.
- Evaluators are presumed guilty until proven innocent.
- Make sure when you yield to temptation in the field that it appears to have something to do with what you are studying.
- A fieldworker should be able to sweep the floor, carry out the garbage, carry in the laundry, cook for large groups, go without food and sleep, read and write by candlelight, see in the dark, see in the light, cooperate without offending, suppress sarcastic remarks, smile to express both pain and hurt, experience both pain and hurt, spend time alone, respond to orders, take sides, stay neutral, take risks, avoid harm, be confused, seem confused, care terribly, become attached to nothing. . . . The nine-to-five set need not apply.
- Always carry extra batteries and getaway money.

*From Halcolm's Evaluation Laws*



## EVALUATION THROUGH OBSERVATION

And the children said unto Halcolm, "We want to understand the world. Tell us, O Sage, what we must do to know the world."

"Have you read the works of our great thinkers?"

"Yes, Master, every one of them as we were instructed."

"And have you practiced diligently your meditations so as to become One with the infinity of the universe?"

"We have, Master, with devotion and discipline."

"Have you studied the experiments, the surveys, and the mathematical models of the Sciences?"

"Beyond even the examinations, Master, we have studied in the innermost chambers where the experiments and surveys are analyzed, and where the mathematical models are developed and tested."

"Still you are not satisfied? You would know more?"

"Yes, Master. We want to understand the world."

"Then, my children, you must go out into the world. Live among the peoples of the world as they live. Learn their language. Participate in their rituals and routines. Taste of the world; smell it. Watch and listen. Touch and be touched. Write down what you see and hear, how they think and how you feel.

"Enter into the world. Observe and wonder; experience and reflect. To understand a world you must become part of that world while at the same time remaining separate, a part of and apart from.

"Go then, and return to tell me what you see and hear, what you learn, and what you come to understand."

*From Halcolm's Methodological Chronicle*

FOLK WISDOM ABOUT  
HUMAN OBSERVATION

Every student who takes an introductory psychology or introductory sociology course learns that human perception is highly



selective. When looking at the same physical design or object different people will see different things. What people “see” is highly dependent on their backgrounds. Our culture tells us what to see; our early childhood socialization instructs us in how to look at the world; and our value systems tell us how to interpret what passes before our eyes. How, then, can one trust observational data?

In their popular “guide for users of social science research” Kitzer et al. (1978) entitle their chapter on observation “Seeing is Not Believing.” In that chapter they tell an oft-repeated story that demonstrates the problem with observational data.

Once at a scientific meeting, a man suddenly rushed into the midst of one of the sessions. He was being chased by another man with a revolver. They scuffled in plain view of the assembled researchers, a shot was fired, and they rushed out. About twenty seconds had elapsed. The chairperson of the session immediately asked all present to write down an account of what they had seen. The observers did not know that the ruckus had been planned, rehearsed, and photographed. Of the forty reports turned in, only one was less than 20-percent mistaken about the principal facts, and most were more than 40-percent mistaken. The event surely drew the undivided attention of the observers, was in full view at close range, and lasted only twenty seconds. But the observers could not observe all that happened. Some readers chuckled because the observers were researchers but similar experiments have been reported numerous times. They are alike for all kinds of people (Kitzer et al., 1978:21-22).

Research and experimentation on selective perception and the incomplete nature of human observation certainly cast doubt on the validity and reliability of observation as a major method of scientific inquiry. Yet, what is often overlooked in stories like the one about the inaccurate observations of researchers at the scientific meeting is that (1) these researchers were not *trained* as social science observers and (2) they had not prepared themselves to make observations at that particular moment in time. Scientific inquiry using observational methods requires disciplined training and rigorous preparation. The simple fact that a person is equipped with five functioning senses—sight, taste, hearing, smell, and touch—does not make that person a skilled observer. The fact that ordinary persons experiencing any particular situation will experience and perceive that situation differently does not mean that *trained and prepared observers* cannot report with accuracy, validity, and reliability the nature of that situation.



Training includes learning how to write descriptively; practicing the disciplined recording of field notes; knowing how to separate detail from trivia in order to achieve the former without being overwhelmed by the latter, and using rigorous methods to validate observations. Training researchers to become astute and skilled observers is particularly difficult because so many people think that they are “natural” observers and therefore have very little to learn. Training to become a skilled observer is a no less rigorous process than the training necessary to become a skilled quantitative social scientist.

Careful preparation for making observations is as important as disciplined training. While I have invested considerable time and effort in becoming a trained observer, I am confident that had I been present at the scientific meeting where the shooting scene occurred my recorded observations would not have been significantly more accurate than those of my less trained colleagues. The reason is that I would not have been prepared to observe what occurred, and lacking that preparation would have been seeing things through my ordinary participant’s eyes rather than my scientific observer’s eyes.

Preparation has material, physical, intellectual, and psychological dimensions. Pasteur said: “In the fields of observation, chance favors the prepared mind.” Part of preparing the mind is learning how to concentrate during the observation. Observation, for me, involves enormous energy and concentration. I have to “turn on” that concentration; “turn on” my scientific eyes, ears, and taste, touch, and smell mechanisms. A scientific observer cannot be expected to engage in scientific observation on the spur of the moment any more than a world class boxer can be expected to defend his title spontaneously on a street corner or an olympic runner can be asked to dash off at record speed because someone suddenly thinks it would be nice to test the runner’s time. Athletes, artists, musicians, dancers, engineers, *and* scientists require training and mental preparation to do their best. Experiments and simulations that document the inaccuracy of spontaneous observations made by untrained and unprepared observers are no more indicative of the potential quality of observation than an amateur community talent show is indicative of what professional performers can do.

Two points are critical, then, in this introductory section. First, the folk wisdom about observation being nothing more than selective perception is true in the ordinary course of participating in day-to-day events. Second, the skilled observer is able to improve the accuracy, validity, and reliability of observations through intensive



training and rigorous preparation. The rest of this chapter is devoted to helping evaluators move their observations from the level of ordinary looking to that of the rigor of scientific inquiry. Decision makers and information users who believe that their evaluation problems would be appropriately studied using observational methods bear responsibility to make sure that the evaluators they select to make observations are able to move beyond ordinary looking to scientific seeing.

## THE VALUE OF OBSERVATIONAL DATA

The purpose of observational data is to *describe* the setting that was observed; the activities that took place in that setting; the people who participated in those activities; and the meanings of the setting, the activities, and their participation to those people. Observational reports must include sufficient descriptive detail to allow one to know what has occurred and how it has occurred. The descriptions must be factual, accurate, and thorough without being cluttered by irrelevant minutiae and trivia. The basic criterion to apply to a recorded observation is the extent to which that observation permits the reader to enter the situation. The value of observational data in evaluation research is that decision makers and information users can come to *understand* program activities and impacts through detailed descriptive information about what has occurred in a program and how the people in the program have reacted to what has occurred.

Naturalistic observations take place in the *field*. For evaluation researchers, the *field* is the program being studied. Field methods in evaluation research take the evaluator close enough to the situation to understand in a direct and personal fashion what the program is all about.

There are many ways of talking about the methods for gathering observational data, including

participant observation, field observation, qualitative observation, direct observation, or field research. All these terms refer to the circumstance of being in or around an on-going social setting for the purpose of making a qualitative analysis of that setting [Lofland, 1971:93].

Direct, personal contact with and observations of a program have several advantages for evaluators. First, by directly observing program operations and activities the evaluator is better able to



understand the *context* within which the program operates. Understanding the program context is essential to a holistic perspective.

Second, firsthand experience with a program allows an evaluator to be inductive in approach. This is the case because the observer, by being on-site, has less need to rely on prior conceptualizations of the program, whether those prior conceptualizations originate with the evaluator or others. In short, the evaluator can directly experience the program as a phenomenon unto itself, thereby making the most of an inductive, discovery-oriented approach.

A third strength of observational methods is that the evaluator has the opportunity to see things that may routinely escape conscious awareness among participants and staff. In order for someone to report information in an interview they must be aware that they have the desired information. Because all social systems involve routines, participants in those routines may take them so much for granted that they cease to be aware of important nuances that are apparent only to an observer who has not become fully immersed in those routines.

A fourth value of direct observational approaches is the extent to which the evaluator can learn about things program participants and/or staff may be unwilling to talk about in an interview. The sensitivity of some subjects in an interview combined with the interviewer's lack of appreciation for that sensitivity because of lack of experience in the program may make interviewees unwilling to provide information on sensitive topics. Through direct experience with and observation of programs evaluators can gain information that otherwise would not become available.

A fifth and closely related point is that observations permit the evaluator to move beyond the selective perceptions of others. Interviews present the understandings of the people being interviewed. Those understandings constitute important, indeed critical, information. However, it is necessary for the researcher to keep in mind that interviewees are always reporting perceptions—selective perceptions. Evaluators as field observers will also have selective perceptions. By making their own perceptions part of the data available in a program, evaluators are able to present a more comprehensive view of the program being studied.

Finally, getting close to a program through firsthand experience permits the evaluator to access personal knowledge and direct experience as resources to aid in understanding and interpreting the program being evaluated. Reflection and introspection are important parts of field research. The impressions and feelings of the observer become part of the data to be used in attempting to understand a



program and its effects. The observer takes in information and forms impressions that go beyond what can be fully recorded in even the most detailed field notes.

Because he sees and hears the people he studies in many situations of the kind that normally occur for them, rather than just in an isolated and formal interview, he builds an ever-growing fund of impressions, many of them at the subliminal level, which give him an extensive base for the interpretation and analytic use of any particular datum. This wealth of information and impression sensitizes him to subtleties which might pass unnoticed in an interview and forces him to raise continually new and different questions, which he brings to and tries to answer in succeeding observations [Becker and Geer, 1970:32].

## VARIATIONS IN OBSERVATIONAL METHODS

Observational research is not a single thing. The decision to employ field methods in order to gather information about some evaluation question is only the first step in a decision process that involves a large number of options and possibilities. Making the choice to employ field methods involves a commitment to get close to the program, to be factual and descriptive in reporting what is observed, and to find out what is happening in the program from the point of view of program participants and staff. Once these fundamental commitments have been made, it is necessary to make additional decisions about which particular observational approaches are appropriate for the evaluation situation at hand.

Decisions concerning which observational approaches are appropriate for particular evaluation situations involves criteria that are different than those same decisions made for purposes of conducting basic social scientific research. These differences emerge from the nature of evaluation research, the politics of evaluation, the nature of contract funding in most evaluations, and the accountability of evaluators to decision makers and information users. Thus, while evaluation research field methods are derived from basic anthropological field methods and qualitative approaches in sociology and phenomenology, the adaptation of these methods to evaluation settings will involve different criteria than are typically discussed in the classic and traditional field methods literature. The sections which follow will discuss both the similarities between evaluation field methods and basic research field methods, as well as the differences that affect the conduct of research for these varying purposes.



## VARIATIONS IN OBSERVER INVOLVEMENT: PARTICIPANT OR ONLOOKER?

The first and most fundamental distinction that differentiates observational strategies concerns the extent to which the observer is a participant in the activities or program being studied. This is not really a simple choice between participation and nonparticipation. The extent of participation is a continuum which varies from complete immersion in the program as full participant to complete separation from the program as spectator; there is a great deal of variation along the continuum between these two extremes. Nor is it simply a matter of deciding once and for all in a study how much the observer will participate. The extent of participation can change over time. In some cases the evaluator may begin as an onlooker and gradually become a participant as the study progresses. In other cases the evaluator may begin as a complete participant in order to experience what it is like to be initially immersed in the program and then gradually withdraw participation over the period of the study until finally taking the role of occasional observer from an onlooker stance.

Participant observation is an omnibus field strategy in that it “simultaneously combines document analysis, interviewing of respondents and informants, direct participation and observation, and introspection” (Denzin, 1978:183). If, on the other hand, the evaluator enters the program as an onlooker to make direct observations of program activities, the processes of observation are separate from the processes of data collection through interviewing. In participant observation there is no such separation. Typically, anthropological fieldworkers combine in their field notes data from personal, eye-witness observation with information gained from informal, natural interviews and informants’ descriptions (Pelto and Pelto, 1978:5). Thus, the participant observer is fully engaged in experiencing the setting under study while at the same time trying to understand that setting through personal experience, observations, and talking with other participants about what is happening.

In participant observation the evaluator shares as intimately as possible in the life and activities of the program under study. The purpose of such participation is to develop an *insider’s* view of what is happening. This means that the evaluator not only sees what is happening but *feels* what it is like to be a part of the program. Anthropologist Hortense Powdermaker has described the basic assumption undergirding participant observation:



[T]o understand a society, the anthropologist has traditionally immersed himself in it, learning, as far as possible, to think, see, feel and sometimes act as a member of its culture and at the same time as a trained anthropologist from another culture [Powdermaker, 1966:9].

Experiencing the program as an insider is what necessitates the *participant* part of participant observation. At the same time, however, there is clearly an *observer* side to this process. The challenge is to combine participation and observation so as to become capable of understanding the program as an insider while describing the program for outsiders.

Obtaining something of the understanding of an insider is, for most researchers, only a first step. They expect, in time, to become capable of thinking and acting within the perspective of two quite different groups, the one in which they were reared and—to some degree—the one they are studying. They will also, at times, be able to assume a mental position peripheral to both, a position from which they will be able to perceive and, hopefully, describe those relationships, systems and patterns of which an inextricably involved insider is not likely to be consciously aware. For what the social scientist realizes is that while the outsider simply does not know the meanings or the patterns, the insider is so immersed that he may be oblivious to the fact that patterns exist. . . . What field workers eventually produce out of the tension developed by this ability to shift their point of view depends upon their sophistication, ability, and training. Their task, in any case, is to realize what they have experienced and learned and to communicate this in terms that will illumine [Wax, 1971:3].

The extent to which it is possible for an evaluator to become a participant in a program will depend partly on the nature of the program. Human service and education programs that serve children do not lend themselves to the evaluator becoming a student and therefore experiencing the program as a child; it may be possible, however, for the evaluator to participate as a volunteer, parent, or staff person in such a program and thereby develop the perspective of an insider in one of those adult roles. Programs that serve special populations may also involve natural limitations on the extent to which the evaluator can become a *full* participant. For example, an evaluator who is not chemically dependent will not be able to become a full participant in a chemical dependency program, though it is possible for the evaluator to actually experience the treatment and join the program as a client. Such participation in a treatment program can lead to important insights and understanding about



what it is like to be in the program; however, the evaluator must avoid the delusion that participation has been complete. This point is illustrated by an exchange between a young evaluator who was doing participant observation in a prison.

INMATE: What you here for, man?

EVALUATOR: I'm here for a while to find out what it's like to be in prison.

INMATE: What you mean—"find out what it's like"?

EVALUATOR: I'm here so that I can experience prison from the inside instead of just studying what it's like from out there.

INMATE: You got to be jerkin' me off, man. Experience from the inside . . . ? Shit, man, you can go home when you decide you've had enough can't you?

EVALUATOR: Yeah.

INMATE: Then you ain't never gonna know what it's like from the inside.

There are also social and political factors which can limit participation. If the participants in a program all know each other intimately they may object to an outsider trying to become part of their close circle. Where there are marked social class differences between the evaluator and program clients, the participants in the program may object to the ruse of full participation. Program staff may object to the additional burden for them of having to include an evaluator in a program where resources are limited and staff-client ratios would be unbalanced by an additional participant. The extent to which full participation is possible and desirable will depend on the precise nature of the program, the political context, and the nature of the evaluation questions being asked. Adult training programs, for example, may easily lend themselves to full participation by evaluators. Evaluators must therefore be flexible, active, reactive, and adaptive in negotiating with decision makers and information users the precise degree of participation that is appropriate in any particular observational study.

The researcher who insists on equating understanding with intense and intimate participation—who believes he can do field research only if he lives with his hosts, shares in all their activities, and refrains absolutely from asking questions—may find himself thrown out on his ear, or, a less harsh lot, simply unable to carry on his work. Had Evans-Pritchard insisted on following to the letter that dictum of Malinowski's that an anthropologist must camp right in the native village, he would



not have produced his great work on the Azande, as these people would not permit him to live with them as an equal. . . . Or again, a dedicated would-be participant may find that the powers-that-be will not permit him to live with or even near the people he wishes to study [Wax, 1971:7].

The point here is that evaluators who adopt qualitative research strategies must avoid the fallacy of thinking that the *ideal* is full and complete participation in the program—“going native.” *The ideal is to negotiate and adopt that degree of participation which will yield the most meaningful data about the program given the characteristics of the participants, the nature of staff-participant interactions, and the sociopolitical context of the program.* These factors and their effects on the type of participation and observation undertaken will be discussed in more depth when data collection procedures are described.

One final caution: The evaluator’s plans and intentions regarding the degree of program involvement to be experienced may not be the way things actually turn out. Lang and Lang (1960) report that two scientific participant observers who were studying audience behavior at a Billy Graham evangelical crusade made *their* “decision for Christ” and left their observer posts to walk down the aisle and join the Reverend Graham’s campaign. Such are the occupational hazards (or benefits, depending on your perspective) of observational research.

## OVERT AND COVERT OBSERVATIONS

A major concern about the validity and reliability of observational data concerns the effects of the observer on what is observed. The basic notion here is that people may behave quite differently when they know they are being observed compared with how they behave if they are not aware of being observed. Thus, the argument goes, covert observations are more likely to capture what is really happening than are overt observations where the people in the setting are aware they are being studied.

There are a full range of opinions concerning the ethics and morality of conducting covert research. On one end of the continuum is the absolute opposition by Edward Shils (1959) to all forms of covert research. He opposes any “observations of private behavior, however technically feasible, without the explicit and fully informed permission of the person to be observed”; he argues that there should be full disclosure of the purpose of any research project, and argues



that even the technique of participant observation is “morally obnoxious . . . manipulation” unless the observer makes explicit his or her research questions at the very beginning of the observation (Shils, 1959; quoted in Webb et al., 1966:vi). At the other end of the continuum is the “investigative social research” approach of Jack Douglas (1976). Douglas argues that conventional anthropological field methods have been based on a consensus view of society which assumes that people are basically cooperative and helpful and willing to have their points of view understood and shared with the rest of the world. In contrast to the consensus model, Douglas adopts a conflict paradigm of society which leads him to believe that any and all covert methods of research should be considered acceptable options in a search for truth.

The investigative paradigm is based on the assumption that profound conflicts of interest, values, feelings and actions pervade social life. It is taken for granted that many of the people one deals with, perhaps all people to some extent, have good reason to hide from others what they are doing and even to lie to them. Instead of trusting people and expecting trust in return, one suspects others and expects others to suspect him. Conflict is the reality of life; suspicion is the guiding principle. . . . It's a war of all against all and no one gives anyone anything for nothing, especially truth. . . .

All competent adults are assumed to know that there are at least four major problems lying in the way of getting at social reality by asking people what is going on and that these problems must be dealt with if one is to avoid being taken in, duped, deceived, used, put on, fooled, suckered, made the patsy, left holding the bag, fronted out and so on. These four problems are (1) misinformation; (2) evasions; (3) lies; (4) fronts [Douglas 1976:55, 57].

Just as participation is not an either-or proposition in observational research, so the question of how explicit to be about observations and the purpose of research is not an either-or proposition. The extent to which participants in a program under study are informed that they are being observed and are told the purpose of the research varies from full disclosure to no disclosure, with a great deal of variation along the middle of this continuum. Buford Junker developed a typology of participant observation which describes four points along this continuum.

1. *Complete Participant*. In this role, the observer's activities as such are wholly concealed. The fieldworker is or becomes a complete member of an in-group, thus sharing secret information guarded from



outsiders. The fieldworker's freedom to observe outside the in-group system of relationships may be severely limited, and in such a role tends to block perception of the workings of the reciprocal relations between the in-group and the larger social system, nor is it easy to switch from this to another role permitting observation of the details of the larger system.

2. *Participant as Observer.* In this role, the field-worker's observer activities are *not* wholly concealed, but are "kept under wraps" as it were, or subordinated to activities as participant, activities which give the people in the situation their main bases for evaluating the fieldworker in his role. This role may limit access to some kinds of information, perhaps especially at the secret level; precisely how he "rates" as a pseudo-"Member of the Wedding" will affect the fieldworker's ability to communicate below the level of public information.

3. *Observer as Participant.* This is the role in which the observer's activities as such are made publicly known at the outset, are more or less publicly sponsored by people in the situation studied, and are intentionally *not* "kept under wraps." The role may provide access to a wide range of information and even secrets may be given to the fieldworker when he becomes known for keeping them, as well as for guarding confidential information. In this role the social scientist might conceivably achieve maximum freedom to gather information but only at the price of accepting maximum constraints upon his reporting.

4. *Complete Observer.* This describes a range of roles in which, at one extreme, the observer hides behind a one-way mirror, perhaps equipped with sound film facilities, and at the other extreme, his activities are completely public in a special kind of theoretical group where there are, by consensus, "no secrets" and "nothing sacred" [Junker, 1960: 35-38].

In traditional field work for the purpose of basic research, the decision about the extent to which observations will be covert is made in the context of the researcher's search for truth. The researcher alone bears responsibility for deciding how scientific truth can best be discovered. In evaluation research the investigator is not involved in a search for basic generalizations about the nature of human behavior and society. Rather (at least from my point of view), evaluation researchers are involved in the more modest task of generating information that can be used for decision-making. From the perspective of utilization-focused evaluation, the decision makers and information users for whom the evaluation is done have a stake in what kind of methods are used. This means that *the evaluator alone*



*cannot make the decision about the extent to which observations and research purposes will be kept secret.*

The complexities of program evaluation mean that there are several levels at which decisions about the covert-overt nature of the evaluation observations must be made. Sometimes only the funders of the program or of the evaluation know the full extent and purpose of observations. On occasion, program staff may be informed that evaluators will be participating in the program, but clients will not be so informed. In other cases a researcher may reveal the purpose and nature of program participation to fellow program participants and ask for their cooperation in keeping the evaluation secret from program staff. On still other occasions a variety of people intimately associated with the program may be informed of the evaluation, but public officials who are less closely associated with the program may be kept "in the dark" about the fact that observations are underway. Sometimes the situation becomes so complex that the evaluator may lose track of who knows and who does not know, and, of course, there are the classic situations where everyone involved knows that a study is being done and who the evaluator is—but the evaluator doesn't know that everyone else knows.

The nature of the questions being studied in any particular evaluation will have a primary effect on the decision about who will be told that an evaluation is underway. In formative evaluations where staff members and/or program participants are anxious to have information that will help them improve their program, the quality of the data gathered may be enhanced by overtly soliciting the cooperation of everyone associated with the program. Indeed, the ultimate acceptance and usefulness of formative information may depend upon such prior disclosure and agreement that a formative evaluation is appropriate. On the other hand, where program funders have reason to believe that a program is corrupt, abusive, incompetently administered, and/or highly negative in impact on clients, it may be decided that an external, covert evaluation is necessary to find out what is really happening in the program.

Finally, there is the related issue of confidentiality. Those who advocate covert research usually do so with the condition that reports conceal names, locations, and other identifying information so that the people who have been observed will be protected from harm or punitive action. Because the basic researcher is interested in truth rather than action, it is easier to protect the identity of participants in the program. In evaluation research, however, while the identity of who said what may be possible to keep secret, it is



more difficult, and indeed may defeat the purpose of the evaluation, to conceal the identity of a program.

Evaluators and decision makers will have to resolve these issues in each case in accordance with their own consciences, evaluation purposes, political realities, and ethical sensitivities.

## VARIATIONS IN DURATION OF OBSERVATIONS

Another important dimension along which observational studies vary is the length of time devoted to data-gathering. In the anthropological tradition of field research a participant observer expects to spend six months at a minimum and often years living in the culture being observed. In sociological studies of subcultures studies vary in length from months to years. To develop a holistic view of an entire culture or subculture takes a great deal of time. The purpose of basic research in the social sciences using field methods is to unveil the basic complexities and patterns of social reality. The social scientist engaged in the conduct of basic research hopes to generate and verify theoretical truths and empirical generalizations.

The purpose of evaluation research is more modest: generating useful information for action. To be useful evaluation information must be timely. Decision makers cannot wait for years while field workers sift through mountains of field notes. Many evaluations are conducted under enormous pressures of time and limited resources. Where observational data have been selected as the appropriate kind of information for a particular evaluation problem the duration of the observations will depend to a considerable extent on the time and resources available in relation to the information needs of decision makers.

Evaluation research projects may involve years of study accumulating detailed data throughout the life of a program. The work of Patricia Carini at the Prospect School in North Bennington, Vermont, is exemplary in this regard. She has worked with the staff of the school to collect detailed case records on students of the school. She has established an archive with as much as twelve years of detailed documentation about the learning histories of individual students and the nature of the school programs they experienced.

On the other end of the continuum are studies that involve observations of a single one- or two-hour segment of a program. Evaluations that include brief site visits to a number of program locations may serve the purpose of simply establishing the existence of various levels of program operations at different sites.



Sometimes an entire segment of a program may be of sufficiently short duration that the evaluator can participate in the complete program; for example, a weekend or week-long training program, a treatment program that meets on a weekly basis for two hours over some period of time, a three-month pilot program in community organizing, and the like.

The critical point about the duration of observational studies is that the length of time during which observations take place should follow from the nature of the evaluation question being studied and not from some ideal about what a typical participant observation must necessarily involve. Field studies do not have to be massive efforts with a team of people participating in a program for a year. At times and for certain programs long-term field work is essential. At other times and for other programs it may be helpful for a program staff to have an evaluator serve the purpose of providing feedback based on one hour of onlooker observation.

My answer to students who ask me how long they have to observe a program to do a good evaluation follows the line of thought developed by Abraham Lincoln in responding to a question posed by a member of the audience at one of the Douglas-Lincoln debates. In an obvious reference to the difference in stature between Douglas and Lincoln a heckler asked, "Tell us, Mr. Lincoln, how long do you think a man's legs ought to be?"

Lincoln replied, "Long enough to reach the ground."

Observations in evaluation research should last long enough to get the job done—to answer the basic evaluation questions being asked.

## VARIATIONS IN OBSERVATIONAL FOCUS

The preceding three sections have discussed how observational evaluations vary in the extent to which the observer participates in the program being evaluated, the extent to which the evaluation is overt or covert, the extent to which the purpose of the evaluation is made explicit, and the duration of the observations. A major factor affecting each of these other dimensions is the scope or focus of the evaluation. The scope of an evaluation can be broad, encompassing virtually all aspects of the program, or it can be narrow, involving a look at only some small part of program activities.

The tradition of anthropological field studies emphasizes the importance of trying to capture the holistic essence of cultural systems. The various subsystems of a society were seen as interdependent parts so that the economic system, the cultural system, the political system, the kinship system, and other specialized sub-



systems could only be understood in relation to each other. In reality, fieldwork and observations often tended to focus on a particular part of the society or culture because of investigator interests and the practicalities of allocating time differentially to those things which the researcher considered most important. Thus, a particular study might present an overview of a particular culture but then go on to report in greatest detail the religious system of that culture.

In the evaluation of programs there is a broad range of possible options which can serve as the focus for a study. One way of thinking about different parts of a program that can be observed is to think separately about the processes by which program participants come into a program (outreach and recruitment component of a program), the processes by which participants are oriented to and socialized into the program (the initiation period into the program), the basic activities that comprise program implementation (the service delivery system), and the activities that go on around program termination including follow-up activities and client impacts over time. It would be possible to engage in observational analysis of any of these program components, any combination of components, or all of the components together. Which parts of the program and how many parts are to be the subjects of study will clearly affect such issues as the extent to which the observer is a participant, who will know about the evaluation and its purpose, and the duration of observations.

Decisions about the focus and scope of a study necessarily involve trade-offs. These trade-offs are necessitated by limited resources, limited time, and limits in the human ability to grasp the complex nature of social reality. The very first trade-off comes in framing the research question or questions to be studied: whether the options involve selecting different components of the program to study or different research foci within components. The problem is to determine the extent to which it is desirable and useful to study one or few questions in great depth or to study many questions but with less depth.

As noted in Chapter Five (where the problem of focus is discussed at greater length), there is no rule of thumb that tells the evaluator precisely how to focus research questions. The extent to which the focus of observations is broad or narrow depends on the resources available, the time available, and the needs of decision makers and information users. In brief, these are not choices between good and bad, but choices among alternatives, all of which have merit.



## DIMENSIONS ALONG WHICH FIELDWORK VARIES: AN OVERVIEW

Five primary dimensions can be used to describe the variation in approaches to observations. Those dimensions, discussed in the previous sections, are graphically summarized in Figure 6.1.

These graphic dimensions can be used with decision makers and information users to help make decisions about the parameters of the evaluation. They can also be used to review how the evaluation is proceeding along each dimension during the course of the evaluation and following the completion of observations.

### WHAT TO OBSERVE

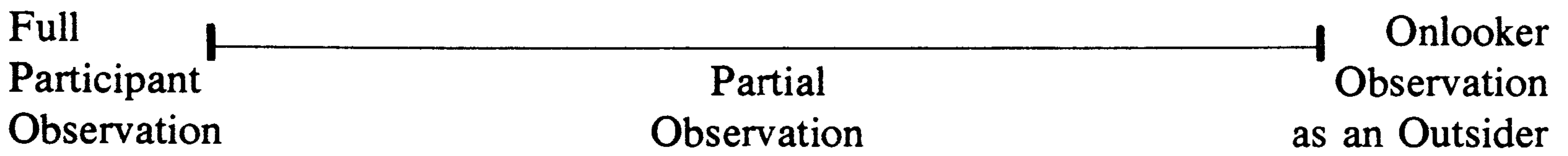
It is not possible to observe everything. The human observer is not a movie camera, and even a movie camera has to be pointed in the right direction to capture what is happening. Moreover, a movie camera has a limited field of vision, taking in only those activities that can be seen within that field accessible to it. For both the human observer and the camera there must be focus. In evaluation observations this focus is provided in part by the nature of the evaluation questions being asked. Once in the field, however, the observer must somehow organize the complex reality represented by the program so that observing that reality becomes manageable.

The purpose of this section is to help evaluators identify some areas within which observations can be focused. These areas are neither theoretical nor conceptual; rather, they represent descriptive cues that can help sensitize the observer to information that is needed for a complete view of the situation and setting being observed.

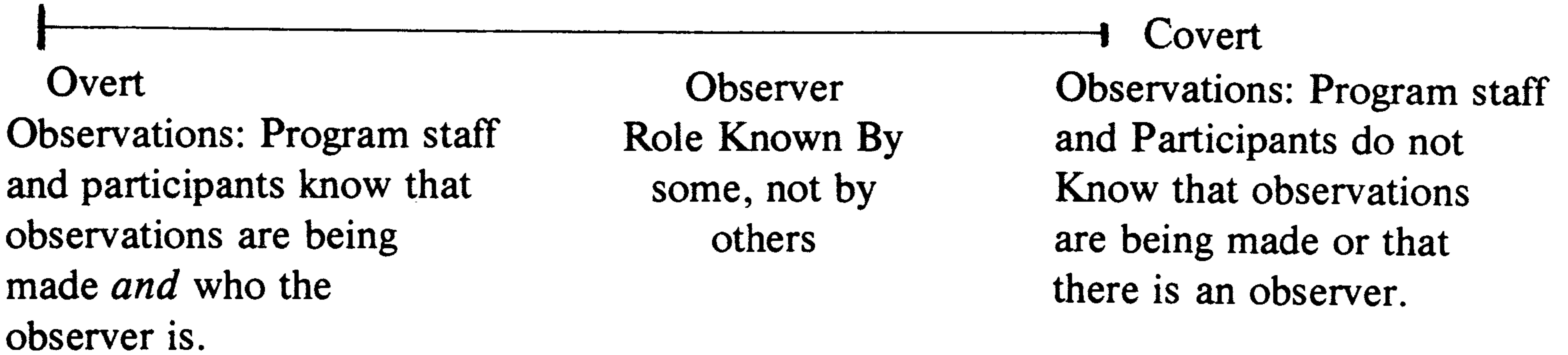
Experienced observers learn that certain kinds of activities and events are likely to yield particularly useful information and insights. When a strategy for placing particular emphasis on certain kinds of observations is made explicit, it often involves what qualitative methodologists call "sensitizing concepts." Rather than being pre-ordinate categories or operationalized variables, sensitizing concepts provide a basic framework highlighting the importance of certain kinds of events, activities, and behaviors. Group process is a sensitizing concept; kinship, leadership, socialization experiences, power, and similar notions are sensitizing in that they alert us to ways of organizing the experience and making decisions about what to record. The observer



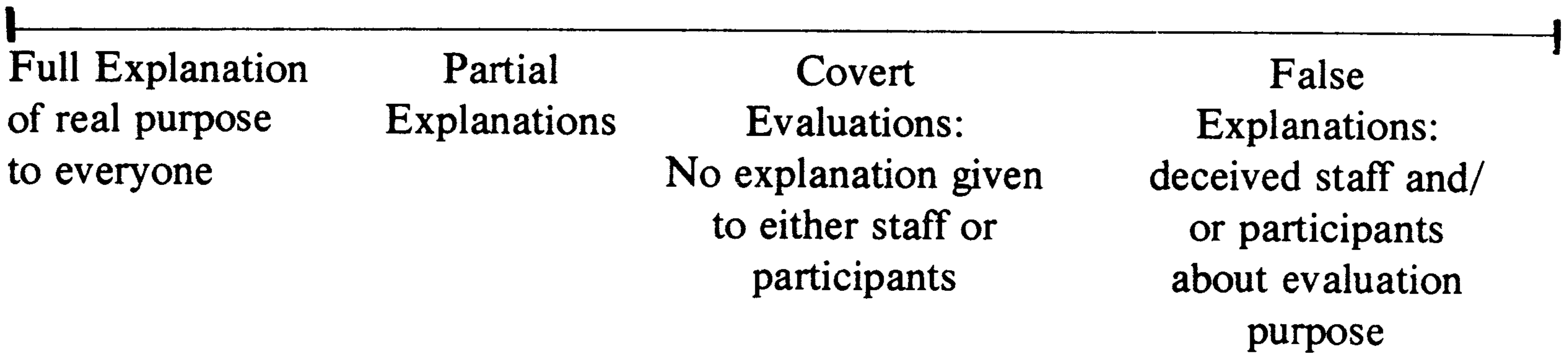
I. Role of the Evaluator-Observer



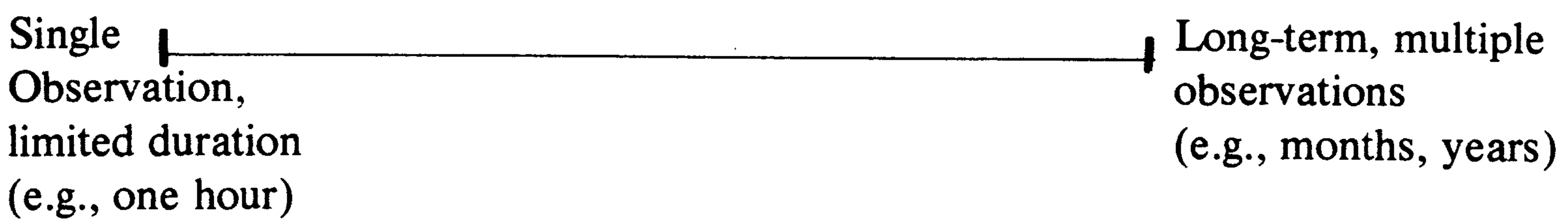
II. Portrayal of the Evaluator Role to Others



III. Portrayal of the Purpose of the Evaluation to Others



IV. Duration of the Evaluation Observations



V. Focus of the Observations

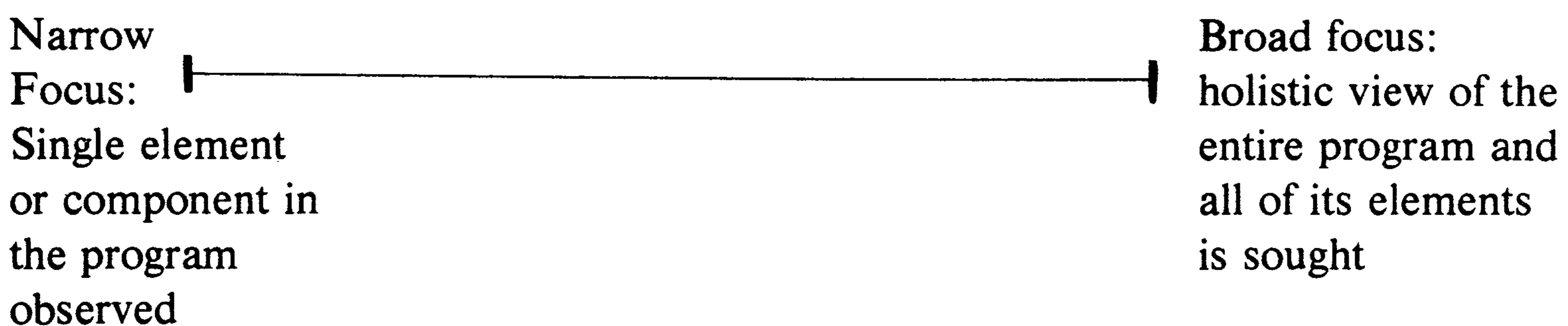


Figure 6.1 Five Dimensions of Variations in Approaches to Observations



moves from sensitizing concepts to the immediate world of social experience and permits that world to shape and modify his conceptual framework. In this way he moves continually between the realm of more general social theory and the worlds of native people. Such an approach recognizes that social phenomena, while displaying regularities, vary by time, space, and circumstance. The observer, then, looks for repeatable regularities. He uses ritual patterns of dress and body-spacing as indicators of self-image. He takes special languages, codes and dialects as indicators of group boundaries. He studies his subject's prized social objects as indicators of prestige, dignity and esteem hierarchies. He studies moments of interrogation and derogation as indicators of socialization strategies. He attempts to enter his subject's closed world of interaction so as to examine the character of private versus public acts and attitudes [Denzin, 1978:9].

The extent to which sensitizing concepts are available to the field worker will vary from situation to situation depending, in particular, on the purpose of the evaluation. In the pages which follow I will present a very broad overview of some categories within which observations can take place in doing evaluation field work. These categories of observational data are not based upon any theoretical framework, but are a simple organizing framework for thinking about the variety of things that one experiences in observing a program.

## THE PROGRAM SETTING

The program setting is the physical environment within which the program takes place. The description of the program setting should be sufficiently detailed to permit the reader to visualize that setting. The observer need not attempt to rival the novelist in writing program descriptions. Interpretive adjectives should be avoided in the observer's own description of the environment, but should be included insofar as they represent quotes from participants about their reactions to and perceptions of that environment. Such adjectives as "comfortable," "beautiful," and "stimulating" are interpretive adjectives rather than descriptive adjectives, such as colors, ("a room painted blue with a blackboard at one end"), measurements ("a forty-foot by twenty-foot classroom"), and purpose ("a library with the walls lined with books").

A good exercise for beginning observers is to write a description of some setting and share it with another person, asking them if they can visualize the setting described. Likewise, it is helpful to have two people observe the same environment and share descriptions, watching in particular for the use of interpretive adjectives instead of



descriptive ones. Learning to be descriptive means providing sufficient information that the reader does not have to guess what the observer means or supply a great deal of additional information to interpret the description. For example, "a crowded room" is an interpretive evaluation statement. What the observer should have reported is: "The room was large enough for a three-person couch across one side, six chairs on the walls next to the couch, and three chairs by the wall facing the couch, which included the door. With twenty people in the room, each person had space to fit, but when everyone was standing there was very little space between people. Several participants were overheard to say, 'this room is really crowded.' "

The physical environment of a program can be important to what happens in that environment. The way the walls look in rooms, the amount of space available, how the space is used, the nature of the lighting, how people are organized in the space, and the interpretive reactions of program participants to the physical setting can be important information about both program implementation and the effects of the program on participants.

A common mistake among observers is to take the physical environment for granted. Thus, an evaluator may report that the program took place in "a school." The evaluator may have a mental image of "school" that matches what was observed, but schools vary considerably in size, appearance, and neighborhood setting. Even more so, the interiors of schools vary considerably. The same can be said for criminal justice settings, health settings, community mental health programs, and any other human service activity.

The variations in physical environments in a wilderness training program for which I served as participant observer provides an interesting example of how physical environments affect a program. The explicit purpose of holding the "field conferences" in the wilderness was to remove people from their everyday settings in largely urban environments surrounded by human-made buildings and the paraphernalia of modern industrial society. Yet, wilderness environments are no more uniform than the environments of human service programs. During the year-long program, participants were exposed to three different wilderness environments: the autumn forest in the Gila Wilderness of New Mexico, the desert terrain of the Kofa Mountains in Arizona during winter, and the San Juan River in the canyon lands of Utah. One focus of the evaluation, then, was to watch how participants responded to the opportunities and constraints presented by these different environments: forest, desert mountains, and canyon-lined river. In addition, the weather imposed



itself to create considerable differences in the environment. Program activities were clearly affected by the extent to which there was rain, cold, wind, and shelter. The weather was expected to provide one of the stress factors which would be a natural part of the program. Program activities, meanwhile, included group discussions and exercises. During one field conference when a large part of the week was cold and wet, it became very difficult to carry out these group activities, thus reducing considerably the amount of group process time that was available and hurrying much of the group process that did occur because of participants' discomfort. It became clear from these experiences that program staff needed to anticipate more clearly the possible variations in physical environments, plan for those variations, and include the participants in that planning so as to increase their commitment to continuing the process under different physical conditions.

## THE HUMAN, SOCIAL ENVIRONMENT

Just as physical environments vary, so too social environments vary. The ways in which human beings interact create social ecological constellations that affect how participants behave toward each other in those environments. Rudolf Moos (1975) describes the social ecological view of programs as follows:

The social climate perspective assumes that environments have unique "personalities," just like people. Personality tests assess personality traits or needs and provide information about the characteristic ways in which people behave. Social environments can be similarly portrayed with a great deal of accuracy and detail. Some people are more supportive than others. Likewise, some social environments are more supportive than others. Some people feel a strong need to control others. Similarly, some social environments are extremely rigid, autocratic, and controlling. Order, clarity, and structure are important to many people. Correspondingly, many social environments strongly emphasize order, clarity and control [Moos, 1975:4].

In describing the social environment the observer looks for the ways in which people organize themselves into groups and subgroups. Patterns of interaction, frequency of interactions, the direction of communication patterns (from staff to participants and participants to staff), and changes in these patterns tell us things about the social environment. The characteristics of people in these different groupings are also illuminative. Male groupings, female groupings,



male-female interactions, and interactions among people with different background characteristics, different racial characteristics, and of different ages alert the observer to patterns in the social ecology of the program.

Decision-making patterns can be a particularly important part of the social environment of a program. Who makes decisions about the activities that take place? To what extent are decisions made openly, so that participants are aware of the decision-making process? How are decisions by staff presented to the full group? How are decisions communicated? Answers to these questions are an important part of the description of a program's social ecology.

As with the physical environment, it is important that the observer maintain a distinction between a description of what has happened in the program and reporting on the perceptions of participants about what has happened. The observer's descriptions of a program's social environment will not necessarily be the same as a perception of that environment expressed by participants. Indeed, it is unlikely that all participants will perceive the social environment in the same way. At all times it is critical that the observer record participant perceptions in quotation marks, indicating the source of those perceptions, so as to keep such observations separate from the evaluator's own interpretations and descriptions of the situation.

The shape of a program emerges from the interaction of physical setting, the social environment of people in the program, the nature of staff leadership and administration, and the activities provided for in the program. The "personality" of the human, social environment will affect how program activities are perceived by participants and experienced by participants. Thus, to understand the activities of the program and the behaviors of participants in those activities it is necessary to frame observations in the context of both the physical and social environments of the program.

## **PROGRAM ACTIVITIES AND PARTICIPANT BEHAVIORS**

The most central focus in most program observations is on program activities and participant behaviors in the program. What do people do in the program? How do they experience the program? What is it like to be a participant in the program? What would one see if one were watching the program in progress? These are the kinds of questions evaluators bring to the program setting as they begin to observe program implementation.



In describing program activities it is usually necessary to find units of activity that have a kind of unity about them. Activities involve some chronological sequence of events; they typically have a beginning, some middle point, and a closure point. In observing a program the evaluator looks for these units of activity. They will usually have specially designated labels given to them either by staff or participants, but sometimes the evaluator will discover units of activity embedded within larger events. Units of activity are things like a class session, a counseling session, meal time in the residential facility, a meeting of some kind, a home visit in an outreach program, a consultation, or a registration procedure. In brief, these are the formal activities of the program.

In observing formal activities the evaluator attempts to capture a comprehensive overview of what takes place in that activity. In order for the descriptions of activities to be comprehensive, information about the full sequence of events must be included: How is the activity introduced or begun? Who is present at the beginning? What exactly was said at the beginning? How did participants respond or react to what was said? These are the questions that focus the beginning of an activity. These same kinds of questions may sensitize the evaluator throughout the full sequence of observation. Who is involved? What is being said by staff? What is being said by participants? What are participants doing? What are the variations in how participants are engaging in the activity being observed? How does it feel to be engaged in this activity? (The observer records his or her own feelings as part of the data of observation.) How did behaviors and feelings change over the course of the activity? Finally, the observer looks for closure points. What are the signals that the activity unit is being ended? Who is present at that time? What is said? How do participants react to the ending of the activity? How is the completion of this unit of activity related to other program activities and future plans?

It is helpful to treat units of activity as self-contained events for the purpose of observation. The process of looking for patterns across units of activity is the process of analysis. During the initial stages of an observation the evaluator will be kept busy just trying to capture self-contained units of activity without looking for the patterns and relationships among those activities that will emerge later from the analysis. The observation of a parent education session presented in Chapter Two is an example of what a single activity observation can contain (pp. 31-35).

This section has been concerned with observing formal program activities that have marked beginnings and recognizable points of



closure. Such formal activities constitute the main programmatic content of human service programs. The implementation of a program typically consists of a series of formal, planned activities. To understand a program, however, and its effects on participants the evaluation observer cannot be restricted to formal, planned activities. The next section discusses the observation of the things that go on between and around formal, planned program activities.

## INFORMAL INTERACTIONS AND UNPLANNED ACTIVITIES

If evaluation observers put away their seeing and observing selves as soon as a program activity ends, they will miss a great deal of data. Some programs build in "free time," or unstructured time, between activities with the clear recognition that such periods of time provide opportunities for participants to assimilate what has occurred during formal programmatic activities and to provide participants with necessary breathing space. Other programs do not recognize the programmatic potential for unstructured time, but it is the rare program or institution that can plan every moment of participants' time.

During periods of informal interaction and unplanned activity it can be particularly difficult to organize observations because people are likely to be milling around, coming and going, moving in and out of small groups, with some sitting alone, some writing, some seeking refreshments, and all otherwise engaging in a full range of what appear to be random behaviors. How, then, can the evaluator-observer collect data during such a time?

This situation illustrates beautifully the importance of staying open to the data. It is impossible to anticipate what kinds of things might emerge during observation of unplanned activity time. Without attempting to interpret in advance or attach significance to participant behaviors at a moment in time, the observer simply continues to gather descriptive information about what people do and, in particular, what people are saying to each other. This last point is particularly important. It is during periods of unplanned activity time that participants have the greatest opportunity to exchange views and to talk with each other about what they are experiencing in the program. In some cases the evaluator will simply listen in on the conversation of others. In some cases it may be appropriate to conduct informal interviews, either with a single participant in natural conversation or with some small group of people. At such times the evaluator asks normal, conversational



questions: "So what did you think of what went on this morning? Was it clear to you what they were trying to get at? What did you think of the session today? How do you think what went on today fits into this whole thing that we're involved in?"

At the same time it is important to remember that everything that goes on in or around the program is data. The meaning, pattern, and significance of the data will vary, but there is still activity there, human experience, and program dynamics. The fact that none of the participants talk about a session when it is over is data. The fact that people immediately split in different directions when a session is over is data. The fact that people talk about personal interests and share gossip that has nothing to do with the program is data. It is also not at all unusual in many kinds of programs for the most significant participant learnings to go on during unstructured program time. A personal interaction with another participant may be the most important event that occurs during a program. To capture a holistic view of the program the evaluator-observer must stay alert to these informal, unplanned activity periods.

As happens in many programs, the participants in the wilderness project I was observing/evaluating began expressing a desire among themselves early in the program, and later formally to staff, to have more free, unstructured time. Many people spent such time writing. Some simply wanted the time to reflect. Most of all, people expressed a desire to have time when they could have one-to-one personal interactions with other participants with whom they shared interests or with whom they wanted to share ideas. The content of such one-to-one sessions was not something I typically attempted to listen to or eavesdrop on directly. I would, however, watch for such interactions and, judging by the body language and the nonverbal cues of the participants, try to tell when serious interpersonal exchanges were taking place. I would then look for natural opportunities to engage each of the participants in informal interviews to find out from them what had happened and what significance they attached to the conversation.

Participant observation is necessarily a combination of observing and informal interviewing. It is important that evaluator-observers not make assumptions about the meaning of what they observe without including the perspectives of participants about their own behaviors. During one period of unstructured time in the wilderness program following a fairly intensive group activity in which a great deal of interpersonal sharing had taken place, I decided to pay particular attention to one of the older males in the group who had



resisted involvement in the group and who had been expressing throughout the week the extent to which he was unimpressed with the program and its potential for impact on him. When the session ended he immediately walked over to his backpack, pulled out his writing materials, and went off to a quiet spot where he could write. He continued writing until dinnertime an hour later. During that time he was very much absorbed in the writing. No one interrupted him. With legs folded, notebook in lap, and his head and shoulders bent over the notebook he gave off clear signals that he was involved, concentrating, and working on something to which he was giving a great deal of effort.

I suspected as I watched him that he was venting his rage and dissatisfaction with the program. I tried to figure out how I might read what he had written. I considered several covert means of getting my hands on his notebook and finally dismissed those alternatives as unethical and an invasion of his privacy. I decided to look for a natural opportunity to initiate a conversation about his writing. During the evening meal around the campfire I moved over next to him, made some small talk about the weather, and then began the following conversation:

“You know one of the things I’m trying to do in documenting the kinds of experiences people are having in this program is keep track of some of the different kinds of things that people do during the week. The staff have encouraged people to keep journals and do writing, and I noticed that you were writing fairly intensely before dinner. It would be helpful to me to know how you see the writing fitting in to your whole experience with the program.”

He hesitated, moved his food about in his bowl a little bit, and then said: “I’m not sure about the program or how it fits in or any of that but I will tell you what I was writing. I was writing,” and he hesitated because his voice cracked, “a letter to my teen-age son trying to tell him how I feel about him and make contact with him about some things. I don’t know if I’ll give the letter to him. The letter may have been more for me than for him. But the most important thing that’s been happening for me during this week is the time to think about my family and how important it is to me and I haven’t been having a very good relationship with my son. In fact, it’s been pretty shitty and so I wrote him a letter. That’s all.”

Having observed this participant’s informal, planned activities and having recorded his perceptions about those activities, most of which were aimed at showing that the program was not having any impact on him, this opportunity to observe his behavior during an



unstructured period and then to talk with him about how that period of time fit into the total program for him revealed a very different side of his participation and an important impact of the program on his personal life.

*In observing programs evaluators must avoid equating the program treatment with the planned, formal activities of the program. A major part of a program's impact may take place on the periphery of structured activity.*

## THE NATIVE LANGUAGE OF PROGRAM PARTICIPANTS

It is an axiom of anthropology that one cannot understand another culture without understanding the language of the people in that culture. Language is a way of organizing the world. The things for which people have words tell others what is important to that culture and that group. Thus, the Eskimos have many words for snow and Arabs have many words for camel. Likewise, the artist has many words for red and many words to describe different kinds of brushes.

The same phenomenon is observable in human service and education programs. Different program areas have their own language to describe the problems they deal with in their work. People who work with the mentally retarded have a complex system of language to distinguish different types of retardation. People in the criminal justice area have their own language for differentiating different types of offenders and different kinds of treatment. Part of the task of observation in a program is learning the native language of the program. This means learning not only the literal meanings of the words used but the connotations and symbolism involved in those words for people in the program. It is not at all unusual in a program for participants to create their own words to describe particular aspects of their experience or of the program. The field notes of the observer should include the exact language used by participants to describe their experiences so that patterns of word usage can emerge in the analysis and so that the reader of the evaluator's observations can be given the flavor of native program language.

In the wilderness education project I observed it became clear that language distinctions were quite important. The program staff wanted to communicate to participants that the time spent together in the wilderness constituted a program. For that reason, the staff decided to call each week in the wilderness a "field conference" to communicate to participants the notion that learning was to be taking place and interactions of both a personal and professional nature



were part of the program experience. The program was a “conference” held in the “field.” The participants, however, never really adopted this language for themselves. Almost universally they referred to the weeks in the wilderness as “trips.” During one week of the program the staff fell into the habit of referring to the week as a “trip” rather than as a “field conference.” Interestingly enough, during that week the planned program activities had not gone very well and the “conference” emphasis seemed to be secondary to the wilderness “trip.” Staff language reflected that change.

Other language developed that was important in identifying what was happening in the program. One of the participants said that he came to the wilderness to “detoxify” and that returning to his everyday world was a “process of retoxification.” The group immediately adopted the language of “detoxification” and “retoxification” to refer to wilderness time versus ordinary urban civilization time. This language was gradually used on a regular basis during discussion sessions and was maintained over the full series of field conferences.

The language of participants in the wilderness education program often reflected the physical environment in which program activities took place. There was a marked pattern of allegory construction in group discussion sessions during which participants would focus on the contrasts between their professional work lives and their experiences in the wilderness. During a backpacking field conference, participants could be heard talking about learning how to “pace myself in my work,” “shifting the burdens of responsibilities that I carry so that the load is more evenly balanced” (a reference to the experience of adjusting the weight of the backpack), and “the long march of career development.” In the mountains when engaging in rock climbing participants referred to “the danger of trying to take risks in my work without having any support down below” (a reference to the *balay* system of climbing where someone supports the climber with a safety rope below). There were references to “finding toeholds and handholds” to bring about change in institutions. There were references to “how steep the wall of resistance is in my institution” and even the assigning of numbers to degrees of back-home institutional resistance corresponding to the numbers used to describe the degree of difficulty of various rock climbs.

On the river field conference participant language was filled with phrases like “going with the flow,” “learning to read the course of program change like you learn to read the current in the water,” and “trying to find my way out of the eddies of life.” (An eddy is an area of



backwash where the current reverses itself and water moves upstream as the onrushing current circles in behind some obstruction; an eddy is both a place of safety from the current of the river and also a place where the water will hold a boat against the shore so that it is often difficult to move out of an eddy and on down the river.)

Using the precise language of participants is an important way to record participants' own understanding of their experiences. Observers must learn the language of participants in the program they are observing and record that language and its patterns in order to represent participants in their own terms and be true to the world view of participants.

## NONVERBAL COMMUNICATION

Social and behavioral scientists have reported at length the importance of both verbal and nonverbal communication in human groups. While recording the language of participants it is important that the evaluator-observer not overlook nonverbal forms of communication. In educational settings nonverbal communications would include patterns that are established for the participants or students to get the attention of or otherwise approach instructors. In group settings a great deal of fidgeting, moving about, and trying to get comfortable can communicate things about attention to and concentration on the group process. The way in which participants dress, express affection, physically space themselves in discussions, and arrange themselves in their physical setting are nonverbal cues about what is happening in the program.

Again, the wilderness education program can serve as an example. In that program hugging emerged as a nonverbal way of providing support for people who were having or had had a particularly difficult time with some experience, and to provide recognition to people who had overcome some particularly difficult challenge. Physical contact and hugging among individuals both in pairs and in larger groups was a direct indicator of the level of affective sharing and caring that had emerged in the group. Different subgroups experienced different degrees of hugging and different field conferences manifested varying degrees of such physical contact. When the group felt disparate, separated, with people on their own "trips," isolated from each other, there was relatively little hugging either in pairs or around the group campfire. When the level of group affect was high, there was shoulder-to-shoulder contact around the campfire, sometimes singing, and a much greater incidence of interpersonal expressions of warmth through physical contact. Over time it became possible to read the



tenor of the group by observing the amount and nature of physical contact participants were having with each other.

The evaluator-observer's own feelings about and reactions to nonverbal cues are an important part of the data in observations. The evaluator-observer should record his or her own reactions to nonverbal communications as well as describing the reactions of others. Also, combining observing with interviewing, it is often appropriate to ask people about their nonverbal behaviors and reactions to the nonverbal behaviors of others. I confirmed with other participants in the wilderness training project the importance of physical contact as a mechanism that they themselves used to sense the tenor of the group.

There is a wealth of nonverbal behavior presented to the evaluator-observer. By watching for patterns of behavior and *describing* what people are doing in different situations, the evaluator-observer will be able to isolate those nonverbal behaviors that have particular significance in a particular program setting.

## UNOBTRUSIVE MEASURES

Program evaluation creates considerable anxiety among many program staff and participants. Regardless of how humanely the evaluation is conducted, regardless of the amount of preparation and process involved in preparing for the evaluation, there is always the possibility that people will behave differently under conditions where an evaluation is taking place than they would if no evaluation were taking place. In their well-known book, *Unobtrusive Measures: Non-reactive Research in the Social Sciences*, Webb et al. (1966:13) suggest:

Even when he is well-intentioned and cooperative, the research subject's knowledge that he is participating in a scholarly search may confound the investigator's data. . . . It is important to note early that the awareness of testing need not, by itself, contaminate responses. It is a question of probabilities, but the probability of bias is high in any study in which a respondent is aware of his subject status.

It is concern about participants' reactions to evaluation observations that leads some evaluators to recommend covert evaluations, as discussed in an earlier section of this chapter. If participants and staff are not aware that an evaluation is taking place and are not aware that



they are being observed, then reactivity is less of a threat to the validity of the evaluation. Even when participants and staff are aware, however, that evaluation observations are taking place, there are often opportunities to make unobtrusive measurements. Unobtrusive measures are those made without the knowledge of the people being observed.

Robert L. Wolf and Barbara L. Tymitz, for example, included unobtrusive measures in their natural inquiry evaluation of the National Museum of Natural History at the Smithsonian Institute. They looked for “wear spots” as indicators of use of particular exhibit areas. They felt that worn rugs would indicate the popularity of particular areas in the museum.

The creative evaluator, once the possibility of looking for unobtrusive measures becomes part of his or her repertoire, can learn a number of things about a program by looking for physical clues about activities in the program. Dusty equipment or records may indicate things that are not used. Areas that are used a great deal by children in a school will look different—that is, more worn—than areas that are little used.

In the wilderness program evaluation one kind of unobtrusive measure was the thickness of notebooks, called “learning logs,” furnished to program participants by the staff. These three-ring binders contained almost no paper when first given to participants on the first field conference. Participants brought back with them the “learning logs” each time they returned to the wilderness. The extent to which paper had been added to the notebooks was one indicator of the extent to which the logs were being used, not only for journal-keeping but also to keep other relevant materials. It was possible to tell which participants were making use of the learning logs by observing how the thickness and composition of the logs changed over the course of the year.

The personnel of the National Forest Service and the Bureau of Land Management have a kind of unobtrusive measure they can use in “evaluating” the wilderness habits of groups that go through an area such as the San Juan River in Utah. The canyons along the San Juan River are a very fragile environment. The regulations for use of that land are essentially “take only photographs, leave only footprints.” This means that all garbage, including human waste and feces, are to be carried out. It takes several days to go down the river. By observing groups at the take-out point and the amount of garbage they carry out as well as the *types* of waste they have with them, one can learn a great deal about the wilderness habits of those groups.



The point here is that fieldwork is not a routine activity. The creative observer, aware of the variety of things to be learned from studying physical and social settings, may learn a great deal through attention to unobtrusive indicators of human activity.

## PROGRAM DOCUMENTS

It has already been noted that observation in the field includes looking and listening, observing and interviewing. The purpose of fieldwork is to find out as much as possible about what is happening in the program. Depending on the focus of any particular evaluation, there is a variety of possible things to look at. One particularly rich source of information about many programs is program records and documents. The nature of program records and documents will vary from program to program, but in contemporary society all programs leave a trail of paper that the evaluator can follow and use to increase knowledge and understanding about the program.

It is important at the very beginning of the evaluation to negotiate access to program documents and records. The evaluator should attempt to anticipate as many different sources of information as possible. The ideal situation would include access to all routine records on clients, all correspondence from and to program staff, financial and budget records, organizational rules, regulations, memoranda, charts, and any other official or unofficial documents generated by or for the program. These kinds of program documents provide the evaluator with information about many things that cannot be observed because they may have taken place before the evaluation began, because they include private interchanges to which the evaluator is not directly privy, and because they reflect aspects of the organization that may be idealized in formal documents but which, because those ideals are not realized in actual program performance, might be unknown to the evaluator. Program documents provide valuable information because of what the evaluator can learn directly by reading them; but they also provide stimulus for generating questions that can only be pursued through direct observation and interviewing. Thus, program records and documents serve a dual purpose: (1) they are a basic source of information about program activities and processes, and (2) they can give the evaluator ideas about important questions to pursue through more direct observations and interviewing. As with all information to which an evaluator has access during observations, the confidentiality of program records—particularly client records—must be respected.



The extent to which an evaluator will include actual references to and quotations from program records and documents in a final report needs to be negotiated on the basis of which documents ought to be considered part of the public record of the program being studied and therefore able to be publicized without breach of confidentiality.

One of the best uses of program records and documents is to get a behind-the-scenes look at program processes and how those came into being. In the wilderness education program evaluation the files of program staff were made available to me as part of the record of the program. In those files was a great deal of information about which program participants had not the slightest idea: letters detailing both conceptual and financial debates between the technical staff who managed the wilderness part of the program and the project directors who had responsibility for the overall conceptualization of the program. Without knowledge of those arguments it would have been impossible to fully understand the nature of the interactions between field staff and program staff in the project. The conflict about program finances and organization that went on between field conferences was reflected in some of the difficulties in communication that emerged between field staff and program staff during field conferences. Interviews with both field staff and program staff revealed quite different perceptions of the nature of those conflicts, their intensity, and their potential for resolution. While participants knew that there were arguments going on among staff, for the most part they were unaware of the origins of those conflicts and the extent to which the smoothness of program development was hampered by them.

Another part of the wilderness program that was revealed in the program documents and budget information was the enormous complexity of the logistics of the program. Participants, including this participant observer-evaluator, arrived at an airport for each field conference, where they were picked up in vans and driven to the wilderness location where the field conference would take place. Participants were supplied with the necessary gear for surviving in the wilderness. Staff were assembled and, for the most part, a basic route plan had already been established which participants would follow. Only through study of program documents and interviews with program staff did a picture of the logistical challenge of the program emerge. Decisions that had to be made about location, arrangements that had to be made for gear, negotiations that took place with the field staff, and exchanges with individual participants



about their particular needs and difficulties took place prior to the actual field conference. Without having looked over the program documents and correspondence I would have missed the extent to which preparation for the one-week experiences in the wilderness consumed the time and energy of program staff. The intensity of work involved before the field conference helped explain the behavior of program staff once the field conference got underway. So much had gone into the preparations, virtually none of which was appreciated by or known to program participants, that program staff would sometimes experience a let-down effect and have difficulty energizing themselves for the actual experience itself.

The paper traces that are the spoor of contemporary organizations are part of the information resources available to the field evaluator. Learning how to use, study, and understand these traces is part of the repertoire of skills needed by the evaluator-observer who does fieldwork.

### OBSERVING WHAT DOES NOT HAPPEN

The preceding sections have been concerned with the things one can observe in a program. Observing what happens in a program, variations in program activities, what people say, what they do, how they interact, and the nature of the physical setting are all important in a comprehensive approach to fieldwork. It is also often important to observe *what does not happen* in the program. Observing what does not happen is a precarious venture because it can take the evaluator-observer into areas of speculation about “what might have been” when such speculation may be off-target.

The potential absurdity of speculating about what does not occur is illustrated by a Sufi story. During a plague of locusts, the wise-fool Mulla Nasrudin, always looking on the bright side, went from village to village encouraging the people by observing how fortunate they were that elephants had no wings. “You people don’t realize how lucky you are. Imagine what life would be like with elephants flying overhead. These locusts are nothing.”

To observe that elephants have no wings is indeed data. Moreover, elephants have no fins, claws, feathers, or leaves. Clearly, once one ventures into the area of observing what does not happen, there are a near-infinite number of things one could point out in the area of “absence of occurrence.” It is therefore with some caution that I include among the tasks of the evaluator-observer that of noting what does not occur.

There are two conditions under which it is appropriate and helpful to point out what has not occurred in a program. First, if program



goals, implementation designs, and/or proposals suggest that certain things ought to happen or are expected to happen, then it is appropriate for the evaluator to note that those things did not happen. If a school program is supposed to, according to its funding mandate, provide children with opportunities to explore the community and no such explorations of the community occur, then it is appropriate for the evaluator to note that no such community activities occurred. If the evaluator reported only what occurred, a question might be left in the mind of the reader about whether or not the other activities had occurred but had not been observed. Likewise, if a criminal justice program is supposed to provide one-to-one counseling to juveniles and no such counseling takes place, it is entirely appropriate for the evaluator to note the absence of counseling.

The second condition under which it is appropriate to note that something did *not* occur is when the evaluator's basic experience with programs suggests to him or her that the absence of some particular activity or factor is noteworthy. This clearly calls for judgment, common sense, and experience on the part of the evaluator-observer. Yet, such judgments are often among the most important contributions an evaluator can make, because these kinds of observations can provide program staff or other decision makers and information users with information that they may not have thought to request, or they may simply have lacked a basis from their own experience for identifying the importance of the omitted activity or condition.

In many such cases the observation about what did not occur is simply a restatement, in the opposite, of what did occur. That restatement, however, will attract attention in a way that the initial observation might not. For example, if one were observing a program being conducted in a multiracial community, it is possible that program goals would include statements about the necessity of staff being sensitive to the needs, interests, and cultural differences of minorities, but there may not be specific mention about the desired racial composition of program staff. If, then, the evaluator observes that the staff of the program consists entirely of Caucasians, it is appropriate to make two kinds of observations: (1) it is appropriate to describe the staff as "Caucasian"; and (2) it is also appropriate to point out that "there were no representatives of minorities on the program staff." The second observation states something about what did *not* occur.

Observations of staff interactions or program planning and decision processes also provide opportunities for evaluators to note both



things that happen and things that do not happen. If, over time, the observer notes ongoing harmony and consensus among a program staff, it is worth making clear in the observations that “no conflicts or personality hassles were observed in the interactions among program staff.” The reason for reporting this observation about what did not occur is to make it clear to the reader that the data have not been selected only to reflect interactions of consensus and harmony. Likewise, if program planning processes never include participants’ input in any systematic or direct way, it may well be appropriate for the evaluator to point out the absence of such input based on experience indicating the significance of participant input in the planning processes of other programs.

In my evaluation of the wilderness education program there were a number of important observations about things that did not occur. There were no serious injuries on any of the six field conferences in the wilderness. This is important information for someone who is considering the possible risks to themselves and others of participating in such a program. There was never a participant who refused to shoulder his or her share of the work that had to be done in order for the group to live in the wilderness. This observation emerged from discussions with technical field staff members who often worked with juveniles in wilderness settings where uneven sharing of cooking, cleaning, and related responsibilities was often a major issue. The fact that the group never had to deal with some one or two people who were not helping out was worth noting. Perhaps the most important observation about what did not happen came out of observations of staff meetings. Over time I observed a pattern in which meetings among the technical and program staff were held to make decisions about what was supposed to happen in the program, and no such decisions were made. Staff sometimes thought that a decision had been made, but in observing those sessions it became clear that closure was not being brought to the decision-making process and responsibility was not being sufficiently clearly assigned to constitute the making of an actual decision. Many subsequent failures of communication and ambiguities about responsibilities could be traced to the absence of real decision-making at staff meetings. By hearing described what was *not* occurring, it was possible for staff to become more explicit about the making of decisions. Reporting what did happen in staff meetings was important; but it was also extremely important to observe what did not happen.



## OBSERVATIONS, INTERVIEWS, AND DOCUMENTATION: BRINGING TOGETHER MULTIPLE PERSPECTIVES ON A PROGRAM

Fieldwork is not a single method or technique. Evaluation fieldwork means that the evaluator is on-site (where the program is happening), observing, talking with people, and going through program records or documentation to gather information for decision makers and information users. Multiple sources of information are sought and multiple resources are used because no single source of information can be trusted to provide a comprehensive perspective on the program. By using a combination of observations, interviewing, and document analysis, the evaluation fieldworker is able to use different data sources to validate and cross-check evaluation findings.

Observations are a limited method of data collection because the evaluator may affect the situation being observed in unknown ways; program staff and participants may behave in some atypical fashion when they know they are being observed; and a selective or partial perception of the evaluator-observer may distort observational data. Observations are also a limited form of data collection because they focus only on external behaviors; the observer cannot see what is happening inside people. Moreover, observational data are often constrained because of the limited sample of program activities actually observed. If only part of the program is observed, the evaluator-observer must use other data sources to find out the extent to which those activities observed are typical or atypical of other program operations.

Interviews are a limited source of data because participants and staff can only report their perceptions of and perspectives on what has happened. Those perspectives and perceptions are subject to distortion due to personal bias, anger, anxiety, politics, and simple lack of awareness. Interview data can be greatly affected by the emotional state of the interviewee at the time the interview takes place. Interview data are also subject to recall error, reactivity of the interviewee to the interviewer, and self-serving responses. Observations provide a check on what is reported in interviews; interviews, on the other hand, permit the observer to go beyond external behaviors to explore the internal states of persons who have been observed.

Program documents are subject to a variety of measurement errors: They may be incomplete; inaccurate; selective, in that only



certain aspects of a program (that is, positive aspects) are documented; and they are often highly uneven in quality, with great detail on some occasions and virtually nothing for other programmatic components. Document analysis, however, provides a behind-the-scenes look at the program that may not be directly observable and about which the interviewer might not ask appropriate questions without the leads provided through documents.

Each data source has strengths and weaknesses. By using a variety of sources and resources, the evaluator-observer can build on the strengths of each type of data collection while minimizing the weaknesses of any single approach. A multimethods approach to fieldwork increases both the validity and the reliability of evaluation data. An example of how observation, interviewing, and documentation supplement each other in fieldwork is provided by my attempt to understand some of the problems involved in staff communication during the wilderness education program. There were two kinds of staff in the program: (1) the program staff who conceptualized and had responsibility for the overall program, and (2) the technical wilderness staff who had responsibility for wilderness skills training, field logistics, and safety. This latter group, the technical staff, was composed of people who typically directed their own wilderness programs and who had the skills to be involved in a number of program processes, including the group processes and personal development activities of the program. The lines of responsibility between technical staff and overall program staff were often blurred and, on occasion, gave rise to conflicts and communication difficulties among staff members. I observed the conflicts and communication difficulties but did not understand the background of staff relationships. All of the staff people, both program and technical staff, had known each other prior to this wilderness education program. In order to learn how those prior relationships affected current program operations I found it necessary to talk with each member in informal, casual, and natural conversations, learning about their former relationships with each other and their perceptions about those relationships. As it turned out, the overall program staff had been the college professors of the technical staff while the latter were still undergraduate students. However, the technical staff had taught the program staff about the wilderness and had introduced them to the wilderness as an environment for experiential education. Each of the staff members described in interviews their perceptions of how these former relationships affected the current setting in the program. In addition, they



described the communications they had had about the program prior to the actual field conferences. Much of those communications had occurred in the form of letters and written correspondence. Reading that correspondence gave me a much better understanding of some of the different assumptions that were operating for various staff members. But *the documentation would not have made sense without the interviews, and the focus of the interviews came from the field observations. Taken together, these diverse sources of information and data gave me a complete picture of staff relationships.* Working back and forth among individual staff members and group staff meetings, I was able to use this information to assist the staff in their efforts to improve their communications during the final field conference. All three sources of information were critical in giving me a full understanding of the nature of the situation and providing me with information that could be used in my role as formative evaluator.

## CREATIVITY IN FIELDWORK

It is my hope that this last section on observing what does *not* happen in a program has made it clear, if the earlier sections did not, that fieldwork is a creative process. It is not possible to begin fieldwork with a comprehensive checklist of what is going to happen and expect that checklist to guide all aspects of the ensuing observations. The evaluator-observer is constantly making judgments about what is worth noting. It is impossible to note everything—the evaluator who attempts to capture everything that occurs in the program is rapidly overwhelmed. Some process of selection must go on. Making such decisions requires judgment, experience, and creativity in attempting to provide useful and relevant information to decision makers and information users. I don't know how to teach someone to be creative about fieldwork. In later sections of this chapter I will suggest some ways of conducting fieldwork that undergrid the rigor, validity, and reliability of the data collection, but I am less sure what to say about creativity. Perhaps all I can do is note that creative fieldwork means using every part of oneself to experience and understand what is happening.

Creativity has to do with the insights that come from being directly involved and personally in contact with the program being evaluated. The usefulness of qualitative evaluation methods depends very directly on the insights generated by the fieldworker. The skilled observer learns to separate description from interpretation, for such separation is the basis of rigor. But to separate these two processes



does not mean that their interdependence is to be ignored. The evaluator-observer works back and forth between descriptions and interpretations to generate insight that, merged with creativity, provides new direction for the next set of observations.

I shall return to the issue of creativity in considering the interpretation of field notes later in this chapter, and again in the analysis chapter. For the moment it is sufficient to acknowledge the centrality of creativity in naturalistic inquiry, and to exclaim with Virginia Woolf:

Odd how the creative power at once brings the whole universe to order. . . . I mark Henry James' sentence: observe perpetually. Observe the oncome of age. Observe greed. Observe my own despondency. By that means it becomes serviceable [quoted by Partnow, 1978:185].

## **DOING FIELDWORK: THE DATA-GATHERING PROCESS**

The decision makers and information users for the evaluation have been identified; the evaluation research questions have been focused; a naturalistic inquiry through field observations has been selected as one of the appropriate methods of data-gathering. It is time to enter the field. Now begins the arduous task of taking field notes.

### **FIELD NOTES**

There are many options in the mechanics of taking field notes: the kind of writing materials used, the time and place at which field notes are taken, the symbols developed by observers as their own method of shorthand, and how field notes are stored. It is impossible to provide universal prescriptions about the mechanics of and procedures for taking field notes because different settings lend themselves to different ways of proceeding, and the precise organization of fieldwork is very much a matter of personal style and individual work habits. *What is not optional is the taking of field notes!*

Aside from getting along in the setting, the fundamental concrete task of the observer is the taking of field notes. Whether or not he performs this task is perhaps the most important determinant of later bringing off a qualitative analysis. Field notes provide the observer's *raison d'etre*. If he is not doing them, he might as well not be in the setting [Lofland, 1971:102].



Field notes contain the description of what has been observed: they should contain everything that the observer believes to be worth noting. Don't trust anything to future recall. At the moment one is writing it is very tempting, because the situation is still fresh, to believe that the details or particular elements of the situation can be recalled later. If it's important to be part of your consciousness as an observer, if it's information that has helped you understand the context, the setting, what went on, and so on, then as soon as possible that information should be put into the field notes.

First and foremost, field notes are descriptive. They should be dated and should record such basic information as where the observation took place, who was present, what the physical setting was like, what social interactions occurred, what activities took place, and other descriptive information that will permit the evaluator-observer to return to that setting later through the field notes, and eventually permit the reader of the evaluation findings to experience the activity observed through edited and organized field notes.

The following passages illustrate different kinds of descriptive field notes. On the left side are field notes which are vague and overgeneralized. On the right side are field notes, from the same observation, that are detailed and concrete.

*Vague and Overgeneralized Notes*

1. The new client was uneasy waiting for her intake interview.

*Detailed and Concrete Notes*

1. At first the new client sat very stiffly on the chair next to the receptionist's desk. She picked up a magazine and let the pages flutter through her fingers very quickly without really looking at any of the pages. She set the magazine down, looked at her watch, pulled her skirt down, and picked up the magazine again. This time she didn't look at the magazine. She set it back down, took out a cigarette, and began smoking. She would watch the receptionist out of the corner of her eye, and then look down at the magazine, and back up at the two or three other people waiting in the room. Her eyes moved from people to the magazine to the



cigarette to the people to the magazine in rapid succession. She avoided eye contact. When her name was finally called she jumped like she was startled.

2. The client was quite hostile toward the staff person.

2. When the staff member told her that she could not do what she wanted to do, the client began to yell at the staff member, telling her that she couldn't control her life, that she was nothing but on a "power trip," that she'd "like to beat the shit out of her," and that she could just "go to hell." She shook her fist in her face and stomped out of the room, leaving the staff person standing there with her mouth open, looking amazed.

3. The next student who came in to take the test was very poorly dressed.

3. The next student who came into the room was wearing clothes quite different from the three students who'd been in previously. The three previous students looked like they had been groomed before they came to the test. Their hair was combed, their clothes were clean and pressed, the colors of their clothes matched, their clothes were in good repair. This new student had on pants that were soiled, with a hole or tear in one knee and threadbare seat. The flannel shirt was wrinkled with one tail tucked into the pants and the other tail hanging out. Hair was disheveled and the boy's hands looked like he'd been playing in the engine of a car.

These examples illustrate the problem of using general terms to describe specific actions and conditions. Words like "poor," "anger," and "uneasy" are not descriptive. Such interpretive words conceal what actually went on rather than reveal the details of the situation.



Such terms have very little meaning for the person who is not present. Moreover, the use of such terms in field notes without the accompanying detailed description means that the fieldworker is recording primarily interpretations rather than description. Particularly revealing are terms that can only make sense in comparison to something else. To describe someone as “poorly” dressed it is necessary to have a frame of reference about what constitutes “good” dress. No skill is more critical in fieldwork than learning to be descriptive, concrete, and detailed.

Field notes also contain what people said. Direct quotations, or as near as possible recall of direct quotations, should be included in the field notes. These quotations will come from what people said during activities as well as what they said during interviews, both informal and formal.

Third, field notes contain the evaluator-observer’s own feelings, reactions to the experience, and reflections about the personal meaning and significance of what has occurred for the observer. Again, it is critical that one not deceive oneself that those feelings can be conjured up again simply by reading the descriptions of what took place. Feelings and reactions should be recorded at the time they are experienced, while the evaluator-observer is in the field. The nature and intensity of those feelings should be recorded as well. In naturalistic inquiry the observer’s own experience is a crucial part of the data. The purpose of getting close to the program and close to the participants in the program through fieldwork is to permit the observer to experience what it is like to be in that setting. If the information about what it is like for the observer is not recorded in the field notes, then much of the purpose of being there is lost.

Finally, the field notes include the observer’s insights, interpretations, beginning analyses, and working hypotheses about what is happening in the setting. The observer’s primary responsibilities are to experience and describe what is going on in the program. The observer approaches fieldwork with a disciplined intention not to impose preconceptions and early judgments on the phenomenon being experienced and observed. Nevertheless, the evaluator-observer does not simply become a recording machine on entering the field. Insights, ideas, inspirations—and yes, judgments, too!—will occur while making observations and while recording field notes. It is not that the observer sits down early on and begins to try to perform the analysis and make judgments. Rather it is in the nature of our intellects that ideas about the meaning, causes and significance of what we experience will find their way into our minds. These insights



and inspirations are also part of the data of fieldwork. They should be recorded just as the observer's feelings are recorded. Interpretations should be clearly set off by some routine symbolization, such as brackets, parentheses, or asterisks. Interpretations should be understood to be just that, interpretations, and labeled as such. But insights are sufficiently precious that the evaluator-observer need not ignore them in the hopes that they will return later.

Field notes, then, contain the ongoing data that are being collected. They consist of descriptions of what is being experienced and observed, quotations from the people observed, the observer's feelings and reactions to what is observed, and field-generated insights and interpretations.

### PROCEDURALLY SPEAKING . . .

When field notes are written will depend on where the evaluation falls along the dimensions discussed earlier in this chapter. Taking field notes is a different process in programs where the observer participates secretly as an evaluator compared with programs in which observations are overt and the observer is an external onlooker. In the evaluation of early childhood education programs aimed at increasing the skills of parents (from which the observation in the second chapter was taken), I openly took extensive notes without participating in the discussions at which I was present. Immediately following those sessions I would go back over the notes to fill in the details and check on the comprehensiveness of the observation. By way of contrast, in the wilderness education program I was a full participant engaged often in full days of hiking, rock climbing, and rafting/kayaking. I was sufficiently exhausted by the end of the day that I did not use my flashlight batteries and stay awake into the wee hours making detailed field notes. I jotted down basic notes that I could expand during the time that others were writing in their journals, but much of the expansion had to be completed after the week-long field conference.

The extent to which notes are openly recorded during the activities being observed is again a function of the evaluator's role and purpose, as well as the stage of development of the participation observation. If the evaluator is openly identified as a short-term, external, nonparticipant observer, participants may expect him or her to write down what is going on. If, on the other hand, the evaluator is engaged in longer-term participant observation, the early part of the evaluation process may be devoted to establishing the participant observer role with emphasis on participation so that open taking of



notes is deferred until the evaluator's role has been firmly established with the group. At that point it is often possible, since the evaluator is then a known entity to the group, to openly take field notes. The wilderness program evaluation involved three full weeks with participants at different times during the year. During the first week I never took notes openly. The only time I wrote was when others were also writing. During the second field conference I began to openly make observations at certain times when discussions were going on and taking notes did not interfere with my participation. By the third week I felt free to take notes whenever I wanted to, and I had no indication from anyone that they even paid attention to the fact that I was taking notes. By that time I had already established myself as a participant, and my participant role was more primary than my evaluator role. The point here is that evaluator-observers must be strategic about taking field notes, timing their writing and recording in such a way that they are able to get their work done without unduly affecting either their participation or their observations. Given those constraints, *the basic rule of thumb is to write promptly*, to complete field notes as soon and as often as physically and programmatically possible.

The writing of field notes is rigorous and demanding work. Lofland describes this rigor quite forcefully.

Let me not deceive the reader. The writing of field notes takes *personal discipline* and *time*. It is all too easy to put off actually writing notes for a given day and to skip one or more days. For the actual writing of the notes may take as long or longer than did the observation! Indeed, a reasonable rule of thumb here is to expect and plan to spend as much time writing notes as one spent in observing. This is, of course, not invariant . . . but one point is inescapable. All the fun of actually being out and about monkeying around in some setting must also be met by cloistered rigor in committing to paper—and therefore to future usefulness—what has taken place [Lofland, 1971:104].

## THE TECHNOLOGY OF FIELDWORK AND OBSERVATION

The basic image of the anthropological fieldworker is of someone huddled in an African hut writing voluminously by lantern. Contemporary evaluation researchers, however, have available to them a number of technological innovations which, when used judiciously, can make fieldwork more efficient and comprehensive. First and foremost are the battery-operated tape recorder and dictaphone. For some people, myself included, dictating field notes saves a great deal



of time while increasing the comprehensiveness of the report. Learning to dictate takes practice, effort, and critical review of early attempts. Tape recorders must be used judiciously so as not to become obtrusive and inhibit program processes or participant responses. A tape recorder is much more useful for recording field notes in private than it is as an instrument to be carried about at all times, available to put a quick end to any conversation into which the evaluator enters.

Typewriters are another tool that can facilitate the writing of field notes. Traditional anthropological training often involves learning how to type. For people who learn how to think and type at the same time, writing field notes with a typewriter, and taking steps to make sure that a typewriter is available during field work, can make a measurable difference in the quality of what is recorded.

Photographic materials can also become part of the repertoire of tools available to the fieldworker. Photographs can help the evaluator-observer recall things that have happened as well as vividly capture the setting for others. Advances in printing and photocopying now make it possible to economically reproduce photographs in evaluation reports. I expect photographic materials to begin to play a much larger part in qualitative methods over time. In the wilderness education program I officially took over the role of being group photographer and making photographs available for reproduction to all of the participants. This helped legitimize the taking of photographs and reduced the extent to which other people felt it necessary to carry their own cameras at all times, particularly at times when it was possible that the equipment might be damaged. Looking at selected photographs during analysis helped me recall the details of certain activities which I had not fully recorded in my written notes.

Videotape equipment is another technological innovation that has become readily accessible and that can sometimes be used unobtrusively. For example, in a formative evaluation of a staff training program I used videotape equipment as part of the visual feedback process with staff. Videotapes of activities, classrooms, training sessions, therapeutic interactions, and a host of other observational interests can sometimes substitute for the physical presence of the evaluator when that would be more intrusive than running a videotape machine. Of course, use of such equipment must be negotiated with program staff and participants, but the creative and judicious use of technological innovations can greatly increase the quality of field observations and can increase the utility of the observational record to others.



The cost of using technological conveniences is rapidly changing as mass production and economies of scale are realized. Moreover, greater familiarity with tape recorders and videotape machines has made it increasingly possible to use such technology without causing great discomfort among professional people who may be the subjects of observation. Costs, of course, are always relative to use; thus, the cost of videotaping might be prohibitive simply to gather evaluation data. However, when such videotapes can also be used for future training, program development, and public relations, the relative costs change. Likewise, transcription of tape-recorded field notes are quite expensive. However, it is less expensive to have a secretary type up what I have dictated than to pay me to handwrite the same amount of material. Evaluators must learn to balance such costs and look for multiple uses of more expensive techniques where there is a need to make judicious decisions about evaluation costs.

Perhaps the ultimate in observer technology for fieldwork is the Stenomask, a sound-shielded microphone attached to a portable tape recorder that is worn on a shoulder strap. The handle of the Stenomask contains the microphone switch. The Stenomask allows the observer to talk into the recorder while an activity is occurring without people in the area being able to hear the dictation. Its use is limited to external, onlooker observations, as the following passage makes clear.

Two procedures precede any data taking. The first is orientation of the subject and as many other persons in the environment as are likely to be present during observations. During this phase the study is explained to the subject and others, the equipment is demonstrated, a sample record is presented and any questions answered. Methods typically found helpful in reducing any potential effects of the observer to as near zero as possible are also discussed. Once this orientation has been accomplished, adaptation begins. During this phase, the observer goes into the habitat and behaves exactly as he or she will during the actual recording. They wear the Stenomask, follow the subject about and run the machine, taking mock records. The purpose of these activities is exactly what is implied in the title, to adapt the subject and others in the environment to the presence of the observer and to reduce the effects of that presence to as near zero as possible. The cardinal rule for the observer during this time is to be *completely* nonresponding. It has been demonstrated over and over again that if the observer continues to resist all social stimuli from the subject and others (and some will occur despite the most careful orientation) by simply keeping the mask in place, looking busily at work and remaining nonresponding, both subjects and others soon cease emitting



stimuli to the observer and come to truly accept him or her as a present and sometimes mobile but completely nonresponding part of the environment, perhaps somewhat like a rolling chair [Scott and Eklund, 1979:9-11].

The Stenomask, of course, is most appropriate for overt, onlooker observations. The imagery of a field worker following a subject around through a day wearing a Stenomask provides a stark contrast to that of the traditional anthropologist doing participant observation and trying covertly to write notes during informal field interviews. The potential for the taking of field notes to be intrusive is present whether one is wearing a Stenomask or simply trying to write notes secretly, as illustrated in the excerpt below from the fieldwork of Carlos Castenada.

In the following passage Castenada is negotiating with Don Juan, his Indian key informant on sorcery and indigenous drugs, in order to learn about the use of plants for medicinal and magical purposes. While they were discussing Castenada's interest in working with Don Juan, the young anthropologist records that "he looked at me piercingly."

"What are you doing in your pocket?" he asked, frowning. "Are you playing with your whanger?"

He was referring to my taking notes on a minute pad inside the enormous pockets of my windbreaker.

When I told him what I was doing he laughed heartily.

I said that I did not want to disturb him by writing in front of him.

"If you want to write, write," he said. "You don't disturb me" [Castenada, 1973:21-22].

Whether one uses some instrument of modern technology to take field notes or simply writes down what is occurring, some method of keeping track of what is observed must be established. In addition, the nature of the recording system must be worked out in accordance with the evaluator's role, the purpose of the evaluation, and consideration of how the data-gathering process will affect the activities and persons being observed. Many of these procedures and questions must be worked out during the initial phases (entry period) of fieldwork. The next section considers some of the issues that must be addressed during the initial phase of fieldwork.



## THE STAGES OF FIELDWORK

Thus far fieldwork has been described as if it was a single, integrated experience. When fieldwork goes well there is a certain continuity to the experience, but it is useful to look at the evolution of fieldwork through identifiable stages. Three stages are most often discussed in the anthropological literature: the entry stage, the basic data-gathering period of fieldwork, and the closing stage. In the sections which follow these stages of fieldwork will be discussed, with particular emphasis on the conduct of evaluative research. The conditions for entry into the field and the responsibilities that must be carried out before leaving the field are often different when fieldwork is for the purpose of evaluation research in contrast to basic anthropological fieldwork. Those differences have important implications for evaluators doing observational work.

### ENTRY INTO THE FIELD

The writings of anthropologists sometimes present a picture of the early period of fieldwork that reminds me of the character in Franz Kafka's haunting novel, *The Castle*. The central character is a wandering stranger, K., with no more identity than that initial. He does not belong anywhere; thus, when he arrives at the Castle he wants to become part of that world. His efforts to make contact with the faceless authorities who run the Castle lead to frustration and anxiety. He can't quite figure out what is going on, can't break through their vagueness and impersonal nature. He doubts himself; then he gets angry at the way he is treated; then he feels guilty, blaming himself for his inability to break through the ambiguous procedures. Yet, he is determined to make sense out of the incomprehensible regulations of the Castle. He is convinced that, after all, where there are rules—and he does find that there are rules—they must fit together somehow, they must have some meaning. There must be some way to make contact, to satisfy the needs of the authorities, to find some pattern of behavior that will permit him to be accepted. If only he could figure out what to do, if only he could understand the rules, then he would happily do what he was supposed to do. Such are the trials of entry into the field.

Entry into the field in evaluation research involves two separate parts: (1) negotiation with decision makers and information users about the nature of the fieldwork to be done and how fieldwork is to be conducted, and (2) actual physical entry into the field setting to begin collecting data. These two parts are closely related, for the



negotiations with decision makers will establish the rules and conditions for how one goes about playing the role of observer and how that role is defined for the people being observed.

In traditional fieldwork for the purpose of basic research the investigator unilaterally decides how best to conduct the fieldwork. From that point on, interactions with those who control entry into the field are primarily manipulative, involving figuring out how to gain entry while preserving the integrity of the study and the investigator's interests. The degree of manipulation involved varies depending on the purpose of the fieldwork and the expected or real degree of resistance to the study. Where the field researcher expects cooperation, gaining entry may be largely a matter of establishing trust and rapport. At the other end of the continuum are those research settings where considerable resistance, even hostility, is expected, in which case gaining entry becomes a matter of "infiltrating the setting" (Douglas, 1976:167).

A major difference between the entry process in traditional anthropological research and the entry process for evaluation research is the extent to which fieldworkers are free to make up whatever story they want to about the purpose of the study. In traditional anthropological research the investigators represent only themselves and so they are free to say whatever they want to say about why they are doing the research. The usual explanation is some variation of, "I'm here because I would like to understand you better and learn about your way of life because the people from my culture would like to know more about you." While anthropologists admit that such an explanation almost never makes sense to indigenous peoples in other cultures, it remains a mainstay initial explanation until mutual reciprocities can be established with enough local people for the observation process to become established and accepted in its own right.

Evaluators, however, are not just doing fieldwork out of personal interest; they are doing the fieldwork for some decision makers and information users who may be either known or unknown to the people being studied. It is critical, then, that evaluators and decision makers give careful thought to how the evaluation is going to be presented.

Because the word "evaluation" has such negative connotations for many people, it is often best to find some other term to describe the fieldwork. In our onlooker, nonparticipatory observations of early childhood programs in Minnesota, we described our role to local program participants and staff as follows:



We're here to be the eyes and ears for state legislators. They can't get around and visit all the programs, and so they've asked us to come out and describe for them what you're doing so that they can better understand the programs they have funded. We're not here to make any judgments about whether your programs are good or bad, we are here to be the eyes and ears for the legislature. This is your chance to inform them, and give them your point of view.

Other settings lend themselves to other terms that are less threatening than "evaluator." Sometimes it is appropriate to describe a fieldwork project as one of "documentation." In the analysis chapter, Beth Alberty of the Workshop Center for Open Education at City College School of Education describes "the documentor's perspective." Another term that is less threatening than evaluator is that of "process historian." In the wilderness education program I was a full participant observer and described my role to the groups as "keeper of the community record." The staff of the project explained that they had asked me to join the project because they wanted someone who did not have direct ego involvement in the success or outcomes of the program to observe and describe what went on, both because they were too busy running the program to keep detailed notes about what occurred and because they were too involved with what happened to be able to look at things relatively dispassionately. There was agreement from the beginning that the community record I produced would be accessible to everyone.

In none of these cases did changing the language automatically make the entry process smooth and easy. Regardless of the story told or the terms used, the entry period of fieldwork is likely to remain "the first and most uncomfortable stage of field work" (Wax, 1971:15). It is a time when the observer is getting used to the new setting and the people in that setting are getting used to the observer. Johnson (1975) suggests that there are two reasons why the entry stage is both so important and so difficult.

First, the achievement of successful entree is a precondition for doing the research. Put simply, no entree, no research. . . . [P]ublished reports of researcher's entree experiences describe seemingly unlimited contingencies which may be encountered, ranging from being gleefully accepted to being thrown out on one's ear. But there is a more subtle reason why the matter of one's entrance to a research setting is seen as so important. This concerns the relationship between the initial entree to the setting and the validity of the data that is subsequently collected. The conditions under which an initial entree is



negotiated may have important consequences for how the research is socially defined by the members of the setting. These social definitions will have a bearing on the extent to which the members trust a social researcher, and the existence of relations of trust between an observer and the members of a setting is essential to the production of an objective report, one which retains the integrity of the actor's perspective and its social context [Johnson, 1975:50-51].

Anthropologist Rosalie H. Wax (1971) has written at length on the problems and processes involved in the first stage of fieldwork. She emphasizes the importance of establishing *reciprocal* relationships during entry. This means that while the observer must learn how to behave in the new setting, the participants in that setting are deciding how to behave toward the observer. Mutual trust, respect, and cooperation are dependent on the emergence of an exchange relationship in which the observer obtains data and the people being observed find something that makes their cooperation worthwhile, whether that something is feelings of importance from being observed, feedback that helps them understand their world better, pleasure from interactions with the observer, or assistance in the activities going on in the observational setting. The reciprocity model of gaining entry assumes that some reason can be found for participants to cooperate in the research and that some kind of mutual exchange can indeed occur.

The *infiltration approaches* to entry described by Douglas emerged in reaction to the fact that many field settings are not open to observation, mutual cooperation, and an explicit model of exchange relationships. If the evaluator and decision makers decide that the information needed is likely to be sufficiently threatening that participants may resist cooperation, some form of manipulation may be necessary. Douglas (1976:167-71) has described a number of infiltration strategies, including "worming one's way in," "using the crowbar to pry them open for our observations," showing enough "saintly submissiveness" to make members guilty enough to provide help, or playing the role of a "spineless boob" who could never possibly hurt the people being observed. He has also suggested using various "ploys of indirection" where the researcher diverts people away from the real purpose of the study and convinces them that he is studying something else. There is also the "phased-entree tactic" by which the researcher who is refused entree to one group begins by studying another group until it becomes possible to get into the group that is the real focus of the researcher's attention.



For many evaluators doing fieldwork the best approach in gaining entree may be the *known sponsor approach*. By this tactic observers use the legitimacy and credibility of another person to establish their own legitimacy and credibility in the setting to be observed. When using this approach it is important to make sure that the known sponsor is indeed a source of legitimacy and credibility. Using program administrators or funders as known sources may increase suspicion and distrust among program participants and staff. Before using the known sponsor approach some prior assessment must be made of the extent to which that person can provide carry-over feelings that will be positive and helpful.

There is a great deal of wisdom in the saying, “it’s who you know that counts.” Upon becoming interested in a particular setting or a particular type of setting, professional sociologists typically seem to begin *not* by going directly to the people of the setting—if they know no one there. Rather, they cast about among their friends, acquaintances, colleagues, and the like, for one or more persons who are either already members or are already favorably regarded by members in the setting of interest. That is, there is an attempt to use *pre-existing* relations of trust as a route into the setting, rather than “going in cold.” Gatekeepers of the setting itself then feel more assured as to the trustworthiness of this newly appeared, would-be observer [Lofland, 1971:95].

It is impossible to provide a universal prescription about how to enter fieldwork settings. The nature of the evaluation, the nature of the program, and the observer’s skills will all affect entree. In selecting a strategy the evaluator will need to use a variety of social skills, psychological sensitivities, and political awarenesses. The demands on the observer to be sensitive and aware can become so great that this initial period of the observation process can give rise to a great deal of frustration and self-doubt. The fieldworker may lie awake at night worrying about some mistake, some *faux pas*, made during the day. There will be times of embarrassment, of feeling foolish, of feeling uncertain, of questioning the whole purpose of the project, and even of paranoia. The fact that one is an evaluation researcher does not mean that one is immune to all the normal pains of learning in and about new situations. But those pains are also part of the data of the study. Moreover, this initial period of fieldwork is also an exhilarating time, a time of rapid new learning, a time when the senses are heightened by their exposure to new stimuli, and a time



of testing one's social, intellectual, emotional, and physical capabilities. The entry stage of fieldwork magnifies both the joys and the pains of doing fieldwork.

Evaluators can often reduce the extent of their "stick-out-like-a-sore-thumb" syndrome by beginning their observations and participation in a program at the same time that participants are beginning the program. In traditional fieldwork anthropologists cannot become children again and experience the same socialization into the culture that children experience. Evaluators, however, can often experience the same socialization process that regular participants experience by becoming part of the initiation process and timing their observations to coincide with the beginning of a program. Such timing makes the evaluator one among a number of novices, and substantially reduces the disparity between the evaluator's knowledge and the knowledge of other participants.

Just beginning the program with other participants, however, does not assure the evaluator of equal status. Some participants may be suspicious that the real difficulties experienced by the evaluator as a novice participant are phony—that the evaluator is play-acting and only pretending to have difficulty. On the first day of my participation in the wilderness education program we had our first backpacking experience. The staff leader began by explaining that "your backpack is your friend." I managed to both pack and adjust my "friend" incorrectly. As a result, as soon as we hit the trail, I found that the belt around my waist holding the backpack on my hips was so tight that my "friend" was making my legs fall asleep. I had to stop several times to try to adjust the pack. Because of these delays and because of the difficulties I was having with the weight and carriage of the pack, I ended up as the last participant along the trail. The next morning when the group was deciding who should carry the map and walk at the front of the group to learn map reading, one of the participants immediately volunteered my name. "Let Patton do it. That way he can't hang back at the end of the group to just observe the rest of us." No amount of protestation on my part seemed to convince the participants that I ended up at the back of the line because I was having trouble hiking (working out my "friendship" with my backpack), rather than because I had taken that position as a strategic place from which to evaluate what was happening.

It is well to remember that regardless of the nature of the fieldwork, during the entry stage more than at any other time, *the observer is also the observed*.



## WHAT YOU SAY AND WHAT YOU DO

Halcolm says: "Evaluation actions speak louder than evaluation words." Evaluators necessarily plan strategies to present themselves and their function, but participant reactions to statements about the evaluator's role are quickly superseded by judgments based on how the evaluator actually behaves in the field setting.

The relative importance of words versus deeds in establishing credibility is partly a function of the length of time the observer expects to be in a setting. For some direct onlooker observations the fieldworker may be present in a particular program for only a few hours or a day. The entry problem in such cases is quite different than the situation where the observer expects to be participating in the program over some longer period of time. Rosalie Wax describes this difference with considerable insight:

All field workers are concerned about explaining their presence and their work to a host of people. "How shall I introduce myself?" they wonder, or, "what shall I say I am doing?"

If the field worker plans to do a very rapid and efficient survey, questions like these are extremely important. The manner in which an interviewer introduces himself, the precise words he uses, may mean the difference between a first-rate job and a failure. . . .

But if the field worker expects to engage in some variety of participant observation, to develop and maintain long-term relationships, to do a study that involves the enlargement of his own understanding, the best thing he can do is relax and remember that most sensible people do not believe what a stranger tells them. In the long run, his host will judge and trust him, not because of what he says about himself or about his research, but by the style in which he lives and acts, by the way in which he treats them. In a somewhat shorter run, they will accept or tolerate him because some relative, friend, or person they respect has recommended him to them [Wax, 1971:365].

Wax argues that over the long run the people being observed will respond to the observer more on the basis of what the observer does than what the observer says about what he or she does. While it is necessary to make some kind of statement about the nature of the research or evaluation being done, such statements are more a matter of formality and courtesy than they are complete determinants of how the evaluator-observer will be received. The same thing is true for the reaction of program participants and staff to the experience with evaluators and evaluation. While evaluators may become



practiced in the *rhetoric* of doing research that is relevant, useful, and aimed at representing participants in their own terms, what the evaluator actually does and the kind of relationships the evaluator establishes with others will determine how people respond to the evaluation and how evaluation findings are used.

## AN ENTRY CASE EXAMPLE: THE PART-TIME OBSERVER

by Joyce Keller

*Introductory Note: The contrast presented in the previous section was between the one-shot onlooker observer and the long-term participant observer. As noted at the beginning of this chapter, there is a great deal of middle ground between these two extremes. In this section Joyce Keller, a senior staff member of the Minnesota Center for Social Research, describes her entry into fieldwork as a part-time observer. Because limitations of time and resources are common in evaluation research, many evaluation situations call for a part-time observer. Joyce's reflections capture some of the special entry problems associated with this "now you're here, now you're gone" role.*

One word can describe my role, at least initially, in a recent evaluation assignment: ambiguous. I was to be neither a participant-observer nor an outsider coming in for a brief but intensive stint. I was to allocate approximately six hours a week for seven months to observing the team development of a group of 23 professionals in an educational setting. At first, the ambiguity was solely on my side: what, really, was I to do? The team, too busy in the beginning with defining their own roles, had little time to consider mine. Later on, as I became accustomed to my task, the team's curiosity about my function began to grow.

In their eyes, I had very little function; I was in the way a great deal of the time inhibiting their private conversations: I served no useful purpose that they could see. On the other hand, they appeared to be concerned about what I was thinking. Some of them—most of them—began to be friendly, to greet me as I came in, to comment when I had missed a team meeting. They came to see me as I saw myself: neither really part of the group nor a separate, removed force.

Observing their interaction perhaps six hours a week out of their 40-hour work week obviously meant that I missed a great deal. I needed to develop a sense of when to be present, to choose among group meetings, subgroup meetings, and activities when all the members were to come together. At the same time, I was working on other contracts which limited the amount of adjustable time available. "Flexible" was the way I came to define my weekly schedule; others, not as charitable, would probably have defined it as "shifty."



A hazard that I encountered as I filled my ambiguous, flexible role was that I soon discovered I was not high on the priority list to be notified in the event of schedule changes. I would have firmly in mind that a subgroup was to meet on Tuesday at 10:00 a.m. in a certain place. I would arrive to find no one there. Later, I would discover that on Monday the meeting had been changed to Wednesday afternoon and no one had been delegated to tell me. At no time did I seriously feel that the changes were planned to exclude me; on the contrary, the members' contrition about their oversight seemed quite genuine. They had simply forgotten me.

Another area of sudden change which caused me difficulty was in policy and procedure. What had seemed to be firm commitments on ways to proceed or tasks to be tackled were being ignored. I came to realize that while a certain amount of this instability was inherent in the program itself, other shifts in direction were outgrowths of planning sessions I had not attended or had not heard the results from after they had occurred. Therefore, keeping current became for me a high-priority activity. Not to do so would have added to my feeling of ambiguity. Also, if I had not operated with a certain degree of self-confidence, I would have felt somehow at fault for coming to a meeting at the wrong time or place or assuming that a certain decision, which the team had previously made, was still valid.

I began my observation of this team in its formative stage. Had I begun after the team was well established, my difficulties would have been greater. Nevertheless, many of the team members were already well acquainted with each other; all had been employees of the same school district over a period of time. They were much better versed in what they had come together to accomplish than I, whose only orientation was reading the proposal which, upon acceptance, had brought them together. I found also that the proposal and the way they planned to proceed were, in actuality, far from being identical.

With my observer role to continue over many months, I realized that I must maintain the difficult position of being impartial. I could not be thought of by the team members as being closely aligned with their leaders, nor could I expect the leaders to talk candidly and openly with me if they believed that I would repeat their confidences to the group members. Reluctantly, for I discovered several team members with whom friendships could easily have developed, I declined invitations to social activities outside of working hours.

When I met with the group for the first time, I directed most of my energies to matching names and faces. I would be taking notes at most of the sessions and it was essential that I could record not only what was said but who was saying it. At the first session everyone, including me, wore a name tag. But within a few days, they were all well acquainted and had discarded their name tags; I was the only one still fumbling for names. While being able to greet each member by name was important, so was knowing something about each one's background. Coffee breaks allowed me to circulate among the group and carry on short conversations with as many as possible to try to fix



in my mind who they were and where they came from, which provided insights into why they behaved in the group as they did.

Team members at first expressed a certain amount of enthusiasm for minutes to be taken of their meetings. This enthusiasm was short-lived, for willing volunteers to serve as secretary did not emerge. I was disappointed, for, had minutes been kept of the meetings and had I been able to rely on receiving copies, I would have concentrated solely on observing the interactions and would not have had to keep track of what they were interacting about. I noted (and ignored) a few passing suggestions that since *I* was obviously taking notes maybe I could. . . A newsletter appeared at the end of the first month and I was again hopeful that some of my informational needs would be met in print by the program. After the second issue, it was seen no more and it was quite apparent that I needed to listen carefully and fill in the holes by asking innumerable questions.

I took copious notes before I began to develop a sense of what was or was not important to record. When I relaxed more and aimed for the tone of the meeting my understanding of the group increased. I had to realize that, as a part-time observer, it was impossible for me to understand all of what was said. My decision frequently was to let this portion of the meeting pass or to jot down a reminder to myself to ask clarifying questions later.

Side-stepping sensitive questions from both leaders and team members had to be developed into a fine art. As I became more finely tuned to the interactions, and most became aware that I was, I was frequently queried as to my perceptions of a particular individual or situation. On one occasion, I found a team member jumping into an elevator to ride two floors with me in a direction he didn't want to go so that he could ask me privately what I thought of another team member. My response was "I think she's a very interesting person," or something equally innocuous, and received from him a highly raised eyebrow, since the woman in question had just behaved in a very peculiar manner at the meeting we had both just attended.

In-depth interviews with each team member began in the fourth month of my observation and was the mechanism which filled in many of the gaps in my understanding. The timing was perfect: I had gained enough familiarity with both personnel and project by that time so that I was knowledgeable, they had come to trust me, and they still cared deeply about the project. (This caring diminished for some as the project year drew to a close without any real hopes of refunding for a second year.) My interview design was intentionally simple and open-ended. What I wanted most was for them to talk about their experiences in terms of strengths and weaknesses. I presented a sign-up sheet at a team meeting asking for specific appointments of about half an hour, although, ultimately, the sessions ran much longer. The dynamic, verbal members were the first to commit themselves. The last few interviews were with those who had to be prodded to set a time and who also had a tendency to "forget" the appointment. The amount of information, therefore, that I gained at the interviews diminished throughout the six weeks or so that was required to meet with all team members. My own performance



unquestionably diminished too as the weeks went on. It was difficult to be as animated and as interesting a person myself as I asked the same questions over and over, devised strategies with which to probe, and recorded perceptions and incidents which I had heard many times before.

Nevertheless, the interview appears in retrospect to be a necessary tool of the part-time observer. Bit by bit team members filled in holes in my information and their repeated references to particular situations and conditions reinforced for me what were sometimes at best only vague perceptions. Team members who appeared to be passive and quiet when I saw them at group meetings were often referred to by their team members as hard-working and creative when they were out in the field. The interviews also helped me become aware of misconceptions on my part caused by seeing only part of the picture, due to time constraints.

The experience had been a new one for me, that of part-time observer. Quite frankly, this mode of evaluation probably will never be a favorite one. On the other hand, it provided a picture that no "snap-shot" evaluation method could have accomplished as interactions changed over time and in a situation where the participant observer role was clearly not appropriate.

## ROUTINIZATION OF FIELDWORK: THE SOCIAL DYNAMICS OF THE SECOND STAGE

During the second stage of fieldwork the evaluator has established a role and a purpose and is able to concentrate on and carry out the tasks of gathering data. The observer is no longer caught up in adjustments to the newness of the field setting. The observer begins to really *see* what is going on instead of just looking around. As Florence Nightingale said, "Merely looking at the sick is not observing" (quoted by Partnow, 1978:44).

One of the things that can happen in the course of fieldwork is the emergence of a strong identity with the people being observed. As you come to understand the behaviors, ideals, anxieties, and feelings of program participants and staff, you may find yourself identifying with their lives, their hopes, and their pain. Such an identification can be a natural part and a logical consequence of having established relationships of rapport, trust, and mutuality with participants. For me, that awakening identification involves some realization of how much I have in common with these people whose world I have been permitted to enter. At times during fieldwork I feel a great separation from the people being observed, and then at other times I feel a strong sense of identification, a sense of our common humanity. To identify, however briefly, with the clients in a program can be a startling experience because evaluator-observers are often quite separated



from clients by education, experience, confidence, and income. It is those differences which sometimes make the world of programs as exotic to evaluators as nonliterate cultures are exotic to anthropologists.

There is a time, then, during fieldwork when the evaluator-observer must deal with his or her own feelings about and perspectives on the people being observed. Part of the sorting-out process of fieldwork is establishing an understanding of the relationship between the observed and the observer. When that happens, and as it happens, the person involved in fieldwork may be no less startled than Joseph Conrad's character Marlowe in *Heart of Darkness*. Marlowe had followed Kurtz, the European ivory trader, up river deep into the Congo where Kurtz established himself as a man-god to the tribal people there. He used his position to acquire ivory, but to maintain his position he had to perform the indigenous rituals of human sacrifice and cannibalism. Marlowe was initially horrified by the darkness of the jungle and its peoples, but as he watched the rituals of those seeming savages, he found an emergent identification with them and even entertained the suspicion that they were *not* inhuman. He became aware of a linkage between himself and them:

They howled and leaped and spun, and made horrid faces; but what thrilled you was just the thought of their humanity—like yours—the thought of your remote kinship with this wild and passionate uproar. Ugly. Yes, it was ugly enough; but if you were man enough you would admit to yourself that there was in you just the faintest trace of a response to the terrible frankness of that noise, a dim suspicion of there being a meaning in it which you—you so remote from the night of the first ages—comprehend. And why not? [Conrad, 1960:70].

In many ways it is our common humanity, whether we are fully aware of it at any given moment or not, that makes fieldwork possible. As human beings we have the amazing capability of becoming part of other people's experiences, and through watching and reflecting we can come to understand something about those experiences.

Beginning to identify—and be identified with—people being studied creates its own new problems. New social definition concerns may emerge during the second stage of fieldwork. Social situations are seldom very simple. The evaluator-observer is not immune to the political dynamics of the settings being observed. Virtually any setting is likely to include subgroups of people who may be in conflict with other subgroups. These factions or cliques may either woo or reject the evaluator-observer, but they are seldom



neutral. There are often deep divisions among the staff and/or the participants in a program, and the evaluator will discover frequent attempts to align one or the other group with the evaluation. Indeed, the evaluator may want to become part of a particular subgroup in order to gain further insight into and understanding of that subgroup. How such an alliance occurs, and how it is interpreted by others, can greatly affect the course of the evaluation. When the evaluation is overt, the evaluator can sometimes use the responsibilities attached to the position of evaluator to gain access to a faction or clique without becoming part of the subgroup, using as a basis for some distancing the necessity of maintaining basic neutrality for the greater good of the total evaluation process. Any point at which the evaluator-observer becomes overidentified with a single subgroup may threaten access to other participants or staff.

At the same time, my experience suggests that it is impossible—or at least impractical—to expect to have the same degree of closeness or distance with every group or faction. Evaluators, human beings with their own personalities and interests, will be naturally attracted to some people more than others. Indeed, to resist those attractions may hinder the observer from acting naturally and being more thoroughly integrated into the program. Recognizing this, the evaluator-observer will be faced with ongoing decisions about personal relationships, group involvement, and how to manage differential associations without losing perspective on what the experience is like for those with whom the evaluator is less directly involved.

Perhaps the most basic division that will always be experienced in program evaluation is the separation of staff and participants. While the rhetoric of many programs attempts to reduce the distinction between staff and participants, there is almost always a distinction between those who are paid for their responsibilities in the program (staff), and those who are primarily recipients of what the program has to offer (participants). In addition, there are some natural reasons why in most programs there is some distance between staff and participants, often a distance that evolves into conflict or distrust. It is very easy for the evaluator to appear to be part of the staff, or the administration, or the funding sources—virtually any group except part of the participants. If the evaluator-observer is attempting to experience the program as a participant, a special effort is required to focus attention on that participant role and to make a concerted effort *not* to be identified as a staff person. Lofland (1971) suggests that it is typical for participant observers to reduce suspicion and fear about a study by becoming aligned with a single



broad grouping within a setting while remaining aloof from that grouping's own internal disputes.

Thus, known observers of medical schools have aligned themselves only with the medical students, rather than attempting to participate extensively with both faculty and students. In mental hospitals, known observers have confined themselves largely to mental patients and restricted their participation with staff. To attempt to participate with both, extensively and simultaneously, would probably have generated suspicion about the observers among people on both sides of those fences [Lofland, 1971:96-97].

In the evaluation of the wilderness education program I found myself moving back and forth between a full participant role, where I was identified primarily as a participant, and a full staff role, where I was identified primarily with those who carried responsibility for directing the program. During the earlier parts of the program, (that is, during the earlier field conferences), I took on the complete role of participant and made as visible as possible my allegiance to fellow participants while maintaining distance from the staff. Over time, however, as my personal relationships with the staff increased, I became more and more aligned with the staff. This coincided with a change of emphasis in the evaluation itself, with the earlier part of the fieldwork being directed at describing the participant experience and the later part of the fieldwork being aimed at describing the workings of the staff. However, there was always a tension both within myself and within the group at large about the extent to which I was a participant or a staff member. I found that as my observational skills became increasingly valued by the program staff I had to more consciously and actively resist their desires to have me take on a more active and explicit staff role. The ambiguities of my role were never fully resolved. I suspect that such ambiguities were inherent in the situation, and are to be expected in many evaluation fieldwork experiences.

Another aspect of the routinization of fieldwork is the cultivation and use of key informants. One of the mainstays of fieldwork is the use of key informants as sources of information about what the observer has not or cannot experience, as well as a source of explanation for events the evaluator-observer has actually witnessed. Key informants are people who are particularly knowledgeable and articulate, people whose insights can prove particularly useful in helping an observer understand what is happening. The selection of key informants must be done carefully so as to avoid arousing



political hostility or personal antagonisms. It is not necessary to formally announce that the "position" of key informant has been filled; the key informant is simply that person or those persons with whom the observer is likely to spend considerable time talking about what is happening in the program. Key informants often must be trained in their role. Anthropologists Pelto and Pelto made this point in reflecting on their own fieldwork.

We noticed that humans differ in their willingness as well as their capabilities for verbally expressing cultural information. Consequently, the anthropologist usually finds that only a small number of individuals in any community are good key informants. Some of the capabilities of key informants are systematically developed by the field workers, as they train the informants to conceptualize cultural data in the frame of reference employed by anthropologists. . . . The key informant gradually learns the rules of behavior in a role vis-a-vis the interviewer-anthropologist [Pelto and Pelto, 1978:72].

The danger in using key informants is that their perspectives will be distorted and biased, thus giving an inaccurate picture of what is happening. It is important that notes obtained from key informants *be clearly specified as such* in the evaluator-observer's fieldwork notes. Data obtained from informants represent perceptions, not truths.

Key informants can provide particularly useful information about what is happening in subgroups to which the evaluator-observer does not or cannot have direct access. During the second year of the wilderness education program one group of participants formed a subgroup and called themselves the "turtles" to set themselves apart from participants with more experience in the wilderness and those who wanted to hike at a fast pace, climb the highest peaks, or otherwise demonstrate their prowess. Because of my wilderness experiences the first year, it would have been difficult for me to have become an intimate part of the "turtles." I therefore established an informant relationship with one of the "turtles" who willingly kept me informed about the details of what went on in that group. Without that key informant relationship, I would have missed some very important information about the kinds of experiences the "turtle" participants were having and the significance of the project to them.

While being part of any setting necessarily involves personal choices about social relationships and political choices about group alliances, the emphasis on planning strategies in fieldwork should not be interpreted as suggesting that the conduct of qualitative research in naturalistic settings is an ever-exciting game of chess or



war in which players and pieces are manipulated to accomplish some ultimate goal. Fieldwork certainly involves times of both exhilaration and frustration, but the dominant motifs in fieldwork are hard work, enormous discipline, and concentration on the mundane, often to the point of boredom. The routinization of fieldwork is a time of concentrated effort and hard work in gathering data. Alas, let the truth be told: The gathering of field data involves very little glory and an abundance of nose-to-the-grindstone drudgery.

## BRINGING FIELDWORK TO A CLOSE

In traditional fieldwork for purposes of basic research and contributing to “truths” about society, it is often difficult to predict how long the fieldwork will go on. The major determinant of the length of the fieldwork is the investigator’s own resources, interests, and needs. Evaluation research is quite different. The major determinant of the length of fieldwork is the evaluation contract, the budget for the evaluation, and the negotiated needs of decision makers for information, including the timetable for feedback and reporting the results of the fieldwork.

Over the course of fieldwork, as one nears completion of data-gathering, more and more attention is devoted to matters of interpretation. As the observer becomes more knowledgeable about the setting being observed, as information increases, more and more ideas about things to check out occur to the evaluator. Possible explanations for what is happening show up in the field notes. Some of these explanations have been offered by others; others occur directly to the evaluator-observer. In short, data analysis has begun, even before the observer has left the field. Chapter Nine deals in depth with how the analysis emerges from the data. At this point, I simply want to recognize the fact that there is no definite point at which data collection stops and analysis begins. Over the course of the fieldwork one process flows into another. As the evaluator-observer gains confidence in the field and sophistication about the nature of the program being studied, it is possible to become increasingly strategic about what goes into the field notes. “As a field researcher develops a better understanding of activities in a given setting, the observational records will change to reflect the observer’s changing understanding” (Johnson, 1975:187).

As fieldwork draws to a close, the researcher is increasingly concerned with *verification* of data and less concerned with the generation of new data. While the evaluator engaged in naturalistic



inquiry avoids imposing preordained conceptual categories on situations being studied, experience with the setting being observed gives rise to categories and dimensions that help organize what has been experienced and observed. These emergent concepts and dimensions generated by the fieldwork are also verified by the fieldwork. Thus, as noted in Chapter Three, “*naturalistic inquiry is always a matter of degree*” (Guba, 1978:6; italics in the original).

Naturalistic inquiry cannot be “pure” in the sense of being absolutely free of constraints placed on either antecedents or responses; initial efforts by the investigator to discover the meaning of what he has observed will cause him to propose certain categories in which to assimilate and account for the noted responses. These categories lead him further into a verification mode, so that on subsequent observations he is more likely to select situations that elicit the response categories of interest. Most likely, the investigator will *cycle* through a series of observations that are, alternately, directed at discovery and then at verification; some initial verification leads to a reorientation to further discovery, and so on [Guba, 1978:6].

Guba goes on to describe this moving back and forth between the discovery mode and the verification mode as a kind of “wave”: The ebb and flow of research involves moving in and out of periods when the investigator is open to new inputs in data and periods when the investigator is testing out hunches, ideas, and explanations.

When fieldwork has gone well the evaluator-observer grows increasingly confident that things make sense and begins to believe in the data. Glaser and Strauss (1967) have described the feelings that the traditional field observer has as fieldwork moves toward a close and data begin to fall into place.

The continual intermeshing of data collection and analysis has direct bearing on how the research is brought to a close. When the researcher is convinced that his conceptual framework forms a systematic theory, that it is a reasonably accurate statement of the matter studied, that it is couched in a form possible for others to use in studying a similar area, and that he can publish his results with confidence, then he has neared the end of his research. He believes in his own knowledgeability and sees no reason to change that belief. He believes not because of an arbitrary judgment but because he has taken very special pains to discover what he thinks he may know, every step of the way from the beginning of his investigation until its publishable conclusion. . . .



The theory that emerges from the researcher's collection and analysis of qualitative data is in one sense equivalent to what he *knows systematically* about his data. Why does the researcher trust what he knows? . . . They are his perceptions, his personal experiences, and his own hard-won analyses. A field worker knows that he knows, not only because he has been in the field and because he has carefully discovered and generated hypotheses, but also because "in his bones" he feels the worth of his final analysis. He has been living with partial analyses for many months, testing them each step of the way, until he has built his theory. What is more, if he has participated in the social life of his subject, then he has been living by his analyses, testing them not only by observation and interview but also by daily living [Glaser and Strauss, 1967:224-225].

This passage from Glaser and Strauss represents the ideal. In the real world of evaluation research, with limited time and resources and reporting schedules that may not permit as long a time of fieldwork as is desirable, the evaluator-observer may have to bring the fieldwork to a close before that state of real confidence has fully emerged. Nevertheless, I find that there is a kind of Parkinson's Law in fieldwork that, as time runs out, the investigator feels more and more the pressure of making sense out of things, and some form of order does indeed begin to emerge from the observations. Alexander Pope noted the tendency of observers to become committed to their observations in his "Epistle I. To Lord Cobham," *Moral Essays*: "To observations which ourselves we make, We grow more partial for the observer's sake."

What is distinct about evaluation research, in contrast to traditional field research, is that the evaluator-observer must also be concerned about feedback. The purpose of evaluation research is not simply to publish an academic treatise on the life of the observed. The purpose of evaluation research is to make a difference in decision-making and programmatic action. Thus, as the research draws to a close, the evaluator-observer must begin to give attention to the question of how feedback is to be given, to whom, and of what nature.

*The giving of feedback can be a major part of the verification process in fieldwork.* My own preference is to provide the participants and staff who have been studied with descriptions and analysis, verbally and informally, and to include their reactions as part of the data. Part of the reciprocity of fieldwork can be an agreement to provide participants with descriptive information about what has been observed. I find that participants and staff are hungry for such



information and fascinated by it. I also find that I learn a great deal from their reactions to my descriptions and analyses. Of course, it is not possible to report everything one has observed. Moreover, the informal feedback that occurs at the end of fieldwork is very different from the rigorous analysis that must go on once the investigator leaves the field. But that analysis may take a great deal of time, and while one is still in the field it is possible to share at least some evidence of what the data look like and to learn from the reactions of those who are described in the data.

The timing of feedback is particularly important—and difficult—in formative evaluations. When the purpose of the evaluation is to provide data for program improvement, the program staff are often anxious to get that information as soon as possible. In some cases their desire to learn will pressure the evaluator-observer to report findings prematurely before there is reason for confidence in the patterns that have emerged or are emerging. I experienced this problem throughout the evaluation of the wilderness education program. During the first year we met with the staff at the end of each week of the three-week program (the three weeks were spread out over the course of the year) to report to them what we had observed and to share interpretations about those observations. At the very first feedback session, at the end of the first week of the program, the initial reaction of the staff to the fieldwork observations and interpretations, particularly about the nature of the staff role in the program, was: “I wish you’d told me that in the middle of the week, when I could have done something about it. I wish you’d told me that that was what was going on. We could have used that information to change the program right then and there.”

I tried to explain that the implications of the descriptions had only become clear to me an hour before we met together when I sat down with my field notes, looked them over, and discussed their significance with my co-evaluator. Nevertheless, there was a lingering distrust that we had purposely withheld information that would have been useful to program staff. Throughout the two years of the project there was an ongoing discussion between myself as the evaluator and the program staff about the timing of feedback. As they increasingly came to value the observations and interpretations they received from the fieldwork, they wanted that information to come earlier and earlier during each field conference week. During the second field conference in the second year, when a number of factors had combined to make the program quite different from what the staff had hoped for, the end-of-the-week evaluation feedback session generated



an unusual amount of frustration from the staff because my analyses of what had happened had not been shared earlier. Again, I found some distrust of my insistence that those interpretations had emerged later rather than sooner as the patterns became clear to me.

It is important that evaluators who are providing formative feedback on an ongoing basis not yield to pressures to make interpretations and report analyses before they have confidence that they have observed and sorted out important patterns. Yet, there typically is one clear moment in time at which the evaluator-observer knows that he or she now has something to report. The evaluator is caught in a dilemma: Reporting patterns before they are clearly established may lead program staff to intervene to change those patterns inappropriately; on the other hand, giving feedback too late may mean that patterns are so established that they are difficult, if not nearly impossible, to change.

No ideal balance between observing from a distance versus making interpretations and providing feedback has ever emerged for me. Feedback is a matter of judgment and dependent on the nature of the relationship between program staff and the evaluator. When in doubt, and where the relationship between the evaluator and program staff has not stabilized into one of long-term trust, I counsel evaluator-observers to err on the side of less feedback rather than more. As often happens in social relationships, negative feedback that was wrong is long remembered and often recounted. On the other hand, it may be a measure of the success of the feedback that program staff so fully adopt it that they make it their own and cease to credit the insights of the evaluator.

Learning how to give feedback in the field is a skill that was not required of the traditional fieldworker. Once feedback is given, the role of the evaluator changes. Those to whom the feedback was presented are likely to become much more conscious of how their behavior and language is being observed. Thus, added to the usual effect of the fieldworker on the setting being observed, this feedback dimension of fieldwork increases the impact of the evaluator-observer on the situation in which he or she is involved.

As the evaluation comes to a close, as the researcher prepares to leave the field, and as he or she organizes notes and thoughts to provide feedback, the impact of that person's presence on the setting may become particularly clear. Providing feedback merely heightens and directs the inevitable effects of having been present in the setting. Because those effects have been of such major concern to people who engage in qualitative methods and their opponents, the final section



in this chapter considers this question of how the observer affects what is observed.

### **THE OBSERVER AND WHAT IS OBSERVED: UNITY AND SEPARATION**

The problem of how the observer affects what is observed is not unique to qualitative research methods. The Heisenberg Uncertainty Principle in physics expresses the same problem from the perspective of natural science. The Heisenberg Uncertainty Principle states that the instruments used to measure velocity and position of an electron alter the accuracy of measurement. When the scientist measures the position of an electron, its velocity is changed, and when the focus of measurement is on the velocity, it becomes more difficult to measure accurately the electron's position. The process of observing affects what is observed. These are real effects, not just errors of perception or measurement. The situation is changed by the intrusion of the observer.

Situations are changed by the intrusion of fieldworkers. How much a situation is changed will depend on the nature of the evaluation, the type of program being studied, the personality and procedures of the evaluator, and a host of unanticipated conditions. Nor is it simply in fieldwork involving naturalistic inquiry that scientific observers affect what is observed. Experimentalists, survey researchers, cost-benefit analysts, and psychologists who administer standardized tests all affect the situations into which they introduce data-collection procedures. The issue is not whether or not such effects occur; rather, the issue is how to monitor those effects and take them into consideration when interpreting data.

The strength of naturalistic inquiry is that the observer is sufficiently a part of the situation to be able to understand personally what is happening. The fact that the presence of the observer will change a situation is something that must be discussed and made clear to decision makers and program staff when planning the evaluation. It is not possible to anticipate exactly how the observer will make a difference. It is possible, when making decisions about what role the evaluator-observer will play, to anticipate certain of the situations that may arise and to establish agreement about how those situations will be handled. By reviewing activities it is possible to decide to what extent the evaluator-observer will participate in those activities and how participation may affect what occurs.



I have evaluated a number of different kinds of programs where group discussions among participants include the expectation that participants will take increasing control in determining the activities of the program. Because of the fundamental nature of those discussions, and because my full participation in such discussions had the potential for greatly affecting the direction of the program, I decided with the program staff in each case not to participate actively in those discussions I might have had I not been involved in the role of evaluator-observer. The ideology of these programs was that each participant had a responsibility to place his or her agenda before the group and to make happen those things that he or she wanted to have happen. In my role as evaluator-observer I had to reduce consciously to extent to which I acted out that ideology, so as to limit my impact on the direction of the group. I calculated my involvement in such a way that I would not appear to be completely withdrawn from the process, yet at the same time I attempted to minimize my impact on the direction of the discussions.

Often the role and impact of the evaluator-observer changes over the course of conducting fieldwork. The evaluation of the wilderness education program discussed throughout this chapter involved six weeks spread over two years. Early in the project I was entirely withdrawn from participant planning discussions, and my withdrawal was observed by others and believed by the staff to have placed a damper on those discussions. Later in the program, particularly the final field conference of the second year, I became quite active in discussions about the direction of the project. That decision was based partly on the fact that certain directions had not occurred in the program over the course of the two years, and the only way I would have a chance to observe what happened if those directions were attempted was to intervene in the program sufficiently to move the program in those directions. This was done with staff knowledge and cooperation, but it involved me as a very active participant and a major force in determining what happened during that project week.

The responsibilities of an evaluator-observer to provide useful information for decision-making means that the impact of a participant evaluator is likely to be greater than the impact of the field anthropologist in traditional cultures. Even in traditional contexts, however, it is clear that the personality of the observer was a major factor in determining what occurred and what was observed. As Patricia Carini (1975) has explained, the person of the observer is

---



essential to an understanding of the point of view that emerges from naturalistic inquiry.

The observer has a point of view, is central to the datum and it is in the articulation—in the revelation of his point of view—that the datum of inquiry is assumed to emerge. In effect the observer is here construed as one moment of the datum and as such the fabric of his thought is inextricably woven into the datum as he is assumed to be constituent of its meaning. From this assumption it is possible to consider the relationship of the observer to the phenomenon under inquiry.

Relatedness can be stated in many ways: opposition, identity, proximity, interpenetration, isolation, to name only a few. All imply that the way in which a person construes his relationship to the phenomenal world is a function of his *point of view* about it. That is, relationship is not a given nor an absolute, but depends upon a personal perspective. It is also true that perspective can shift, the only necessity of a person's humanity being that he takes some stance in relationship to the events about him [Carini, 1975:8-9].

There is, then, an interdependence between the observer and what is observed. This is not, however, a random interdependence—at least, it should not be. The fieldworker calculates the nature of that interdependence and develops a strategy aimed at permitting the observer to experience the phenomenon being observed, while at the same time maintaining sufficient separation from the phenomenon to permit the observer to be an observer—to abstract the experience and the phenomenon.

It is in this sense that Bruyn (1966:14) suggests that a basic corollary of participant observation is that the “role of the participant observer requires both detachment and personal involvement.” To be sure, there is both tension and ambiguity in this corollary. Its actual manifestation in any given situation will depend on both the observer and the phenomenon being observed.

Thus, we may observe at the outset that while the traditional role of the scientist is that of a neutral observer who remains unmoved, unchanged, and untouched in his examination of phenomena, the role of the participant observer requires sharing the sentiments of people in social situations; as a consequence he himself is changed as well as changing to some degree the situation in which he is a participant. . . . The effects are reciprocal for observer and observed. The participant observer seeks, on the one hand, to take advantage of the changes due to his presence in the group by recording these changes as part of his study, and on the other hand, to reduce the changes to a minimum by the manner in which he enters in the life of the group [Bruyn, 1966:14].



Whether one is engaged in participant observation or onlooker observation, what happens in the setting being observed will, to some extent, be dependent on the role assumed by the observer. Likewise, the nature of the data collected will, to some extent, be dependent on the role and perspective of the observer. The personal nature of observations is both their strength and weakness; their strength in that personal involvement permits firsthand experience and understanding, and their weakness in that personal involvement permits the possible introduction of bias and distortion. The interdependence of the observer and what is observed gives naturalistic inquiry its perspective. I shall have a great deal more to say about the *perspective* attained through observations in the chapter on analysis.

## THE RULES AND PROCEDURES OF OBSERVATIONAL EVALUATION RESEARCH

The reader who came to this chapter looking for precise rules and procedures of observation by now will have been disappointed. Looking back over this chapter (even, perhaps, doing a content analysis of it), the major theme seems to be: “What you do depends on the situation; the nature of the evaluation; the nature of the program; and the skills, interests, needs, and point of view of the evaluator-observer.” Yet, the conduct of observational research is not without direction. The mandates of field research include being careful to be descriptive in taking field notes; gathering a variety of information from different perspectives; cross-validating emergent patterns by gathering data from multiple sources and by gathering different kinds of data—observations, interviews, and documentation; being quotive and representing participants in their own terms; reporting on the observer’s own experience, location, and feelings; and clearly separating description from interpretation as one puts together a comprehensive, holistic, and sufficiently detailed picture of what has been observed to allow the reader of that observation to enter into the situation.

Despite these prescriptions, the point remains that what one does depends on the situation; the nature of the evaluation; the nature of the program; and the skills, interests, needs, and point of view of the evaluator-observer. Once all the rules and procedures of observational research have been specified and considered as best one can, and after the situational constraints on and variations in the conduct of fieldwork have been properly recognized, it is helpful to return to the core elements of qualitative methods. Those core elements consist of the research methodology that Nicholas Tinbergen described as



“watching and wondering” in his acceptance speech for the 1975 Nobel Prize in physiology and medicine. It was by “watching and wondering” that Tinbergen, who is neither a physiologist nor a medical doctor, made a major breakthrough in our understanding of autism. He found in his observations that the major clinical research on autism did not hold up outside clinical settings. His “watching and wondering” allowed him to see that normal individuals, those not clinically labeled as autistic, exhibited under a variety of circumstances all of the behaviors described as autistic in the clinical research. He also noted that children diagnosed as autistic responded in nonautistic ways outside the clinical setting. By observing people in a variety of settings and in a full range of behaviors, he was able to make a major scientific breakthrough. His research methodology: “watching and wondering.”

## THE PERSONAL EXPERIENCE OF FIELDWORK

The intersection of social science procedures with individual capabilities and situational variation is what makes fieldwork a highly personal experience. At the end of her book on *Doing Field Work*, Wax reflects on “how field work changed me.”

A colleague has suggested that I reflect on the extent to which I was changed as a person by doing field work. I reflected and the result astonished me. For what I realized was that I had not been greatly changed by the things I suffered, enjoyed or endured; nor was I greatly changed by the things I did (though they strengthened my confidence in myself). What changed me irrevocably and beyond repair were the things *I learned*. More specifically, these irrevocable changes involved replacing mythical or ideological assumptions with the correct (though often painful) facts of the situation [Wax, 1971:363].

Fieldwork is not for everyone. Some, like Henry James, will find that “innocent and infinite are the pleasures of observation.” Others find observational research anything but pleasurable. Some students have described their experiences to me as tedious, frightening, boring, and “a waste of time,” while others have experienced challenge, exhilaration, personal learning, and intellectual insight. More than once the same student has experienced both the tedium and the exhilaration, the fright and the growth, the boredom and the insight. Whatever the adjectives used to describe any particular individual’s fieldwork, of this much we are assured: The experience



of observing provides the observer with both experience and observations. No less an evaluation authority than William Shakespeare gives us this assurance.

*Armado.* How hast thou purchased this experience?

*Moth.* By my penny of observation.

—*Love's Labor Lost*

## AND FOR THE METAPHOR-MINDED: A FINAL LOOK AT EVALUATION THROUGH OBSERVATION

Some of the most delightful, entertaining, and suspenseful fairy tales and fables concern tales of kings who discard their royal robes to take on the apparel of peasants so that they can move freely among their people to really understand what is happening in their kingdom. Our modern-day kings and political figures are more likely to take television crews with them when they make excursions among the people than they are to secretly disguise themselves so that they can move through the streets anonymously. It is left, then, to the evaluators to play out the fable, to take on the appropriate appearance and mannerisms that will permit them to move easily among the people, sometimes secretly, sometimes openly, but always with the purpose of better understanding what the world of programs is really like. They are then able to report those understandings to our modern-day version of kings so that political wisdom can be enhanced and programmatic decisions enlightened. That, at least, is one notion of how to think about evaluation through observation.



## QUALITATIVE INTERVIEWING

After much study of the evaluation masters three youths came before Halcolm to ask how they might further increase their knowledge and wisdom. Halcolm sensed that they lacked experience in the real world, but he wanted to have them make the transition from the seclusion of their studies to the outside world in stages. During the first stage he sent them forth under a six-month vow of silence. During those six months they wore the identifying garments of the muted truth-seekers so that people would know that they were forbidden to speak. Each day, according to their instructions, they sat at the market in whatever village they entered, watching but never speaking. After six months in this fashion they returned to Halcolm.

“So,” Halcolm began, “you have returned to us from your journey. Your period of silence is over. Your transition to the world beyond our walls of study has begun. What have you learned on this your first journey?”

The first youth answered, “In every village the patterns are the same. People come to the market. They buy the goods they need, talk with friends, and leave. I have learned that all markets are alike and the people in markets always the same. I have learned that all things are ultimately the same from place to place.”

Then the second youth reported, “I too watched the people come and go in the markets. I have learned that all life is coming and going, people forever moving to and fro in search of food and basic material things. I understand now the simplicity of human life.”

Halcolm looked at the third youth: “And what do you have to tell us?”

“I saw the same markets and the same people as my fellow-travelers, yet I know not what they know. My mind is filled with questions. I kept wondering where the people came from and where they went. I pondered what they might be thinking and feeling as they



came and went. I reflected on how they happened to be at this market on this day, who they left behind, and who came with them. I wondered how today was the same or different for them. I have failed, Master, for I am filled with questions rather than answers, questions for the people I saw. I do not know what I have learned.”

Halcolm smiled. “You have learned most of all. You have learned the value of being able to ask questions. You have learned the importance of finding out what people have to say. You are ready now to return to the world, this time without the vow of silence.”

“Go forth now. Go forth and question. Ask and listen. The world is just beginning to open up to you. Each person you question can take you into a new part of the world. For the person who is willing to ask and listen the world will always be new. The skilled questioner and attentive listener knows how to enter into another’s experience.”

From: *Halcolm: Biography of a Master Evaluator*

## INNER PERSPECTIVES

The purpose of interviewing is to find out what is in and on someone else’s mind. The purpose of open-ended interviewing is *not* to put things in someone’s mind (for example, the interviewer’s preconceived categories for organizing the world) but rather to access the perspective of the person being interviewed. We interview people to find out from them those things we cannot directly observe. The issue is not whether observational data is more desirable, valid, or meaningful than self-report data. The fact of the matter is that we cannot observe everything. We cannot observe feelings, thoughts, and intentions. We cannot observe behaviors that took place at some previous point in time. We cannot observe situations that preclude the presence of an observer. We cannot observe how people have organized the world and the meanings they attach to what goes on in the world—we have to ask people questions about those things. The purpose of interviewing, then, is to allow us to enter into the other person’s perspective. The assumption is that that perspective is meaningful, knowable, and able to be made explicit.

Interview data for program evaluation purposes allow the evaluator to capture the perspectives of program participants, staff, and others associated with the program. What does the program look like and feel like to the people involved? What are the experiences of program participants? What thoughts do people knowledgeable about the program have concerning program operations, processes, and out-



comes? What do people know about the program? What are their expectations? What features of the program are most salient to the people involved? What changes do participants perceive in themselves as a result of their involvement in the program? It is the responsibility of the evaluator to provide a framework within which people can respond comfortably, accurately, and honestly to these kinds of questions. The task undertaken by the interviewer is to make it possible for the person being interviewed to bring the interviewer into his or her world. The quality of the information obtained during an interview is largely dependent on the interviewer. The purpose of this chapter is to discuss ways of obtaining high quality evaluative information by talking with people who have that information.

Evaluators can enhance the utilization potential of the information they collect by making sure they take the necessary steps to increase the quality of their findings. As Hermann Sudermann said in *Es Lebe das Leben I*, "I know how to listen when clever men are talking. That is the secret of what you call my influence." Evaluators must learn how to listen when knowledgeable people are talking. That may be the secret of their influence.

This chapter begins by discussing three different types of interviews, three basic approaches to qualitative interviewing. Later sections consider the content of interviews: what to ask questions about and ways of phrasing interview questions. The chapter ends with a discussion of how to record the responses obtained during interviews.

## VARIATIONS IN QUALITATIVE INTERVIEWING

There are three basic approaches to collecting qualitative data through open-ended interviews. The three approaches involve different types of preparation, conceptualization, and instrumentation. Each approach has strengths and weaknesses, and each serves a somewhat different purpose. The three choices are:

- (1) the informal conversational interview;
- (2) the general interview guide approach; and
- (3) the standardized open-ended interview.

The differences among these three approaches to the design of the interview is the extent to which interview questions are determined and standardized *before* the interview occurs. The *informal conversational interview* relies entirely on the spontaneous generation



of questions in the natural flow of an interaction, typically an interview that occurs as part of ongoing participant observation fieldwork. During an informal conversational interview, the persons being talked with may not even realize they are being interviewed.

The *general interview guide approach* involves outlining a set of issues that are to be explored with each respondent before interviewing begins. The issues in the outline need not be taken in any particular order and the actual wording of questions to elicit responses about those issues is not determined in advance. The interview guide simply serves as a basic checklist during the interview to make sure that all relevant topics are covered. The interview guide presumes that there is common information that should be obtained from each person interviewed, but no set of standardized questions are written in advance. The interviewer is thus required to adapt both the wording and sequence of questions to specific respondents in the context of the actual interview.

The *standardized open-ended interview* consists of a set of questions carefully worded and arranged with the intention of taking each respondent through the same sequence and asking each respondent the same questions with essentially the same words. Flexibility in probing is more or less limited, depending on the nature of the interview and the skills of interviewers. The standardized open-ended interview is used when it is important to minimize variation in the questions posed to interviewees. This reduces the possibility of bias that comes from having different interviews for different people, including the problem of obtaining more comprehensive data from certain persons while getting less systematic information from others. A standardized open-ended interview may be particularly appropriate when a large number of people are to conduct interviews on the same topic and the evaluator wishes to reduce the variation in responses due to the fact that, left to themselves, different interviewers will ask questions on a single topic in different ways. By controlling and standardizing the open-ended interview the evaluator obtains data that are systematic and thorough for each respondent but that reduce flexibility and spontaneity.

## THE INFORMAL CONVERSATIONAL INTERVIEW

The informal conversational interview is the phenomenological approach to interviewing. A phenomenological approach is used when the researcher has no presuppositions about what of importance may be learned by talking to people in the program. The phenomono-

---



logical interviewer wants to maintain maximum flexibility to be able to pursue information in whatever direction appears to be appropriate, depending on the information that emerges from observing a particular setting or from talking to one or more individuals in that setting. Most of the questions will flow from the immediate context. Thus, the conversational interview is a major tool used in combination with participant observation to permit the evaluator who is participating in some programmatic activity to understand other participants' reactions to what is happening. No predetermined set of questions is possible under such circumstances, because the evaluator does not know beforehand what is going to happen and what it will be important to ask questions about.

The data gathered from informal conversational interviews will be different for each person interviewed. In many cases, the same person may be interviewed on a number of different occasions using an informal, conversational approach. The phenomenological approach is particularly useful where the evaluator can stay in the situation for some period of time, so that he or she is not dependent on a single interview to collect information about the program. Interview questions will change over time, and each interview builds upon the other, expanding information that was picked up previously, moving in new directions and seeking elucidations and elaborations from various participants in their own terms. The phenomenological interviewer must "go with the flow." Depending on how the interviewer's or evaluator's role has been defined, the people being interviewed may not know during any particular informal conversation that the purpose of the conversation is the collection of data. This means that in many cases phenomenological interviewers do not take notes during the interview; rather, they write down what they learned after they have left the interview/observation situation. In other cases, it can be both appropriate and comfortable to take notes or even use a tape recorder.

The strength of the phenomenological approach to interviewing is that it allows the interviewer/evaluator to be highly responsive to individual differences and situational changes. Questions can be individualized to establish in-depth communication with the person being interviewed and to make use of the immediate surroundings and situation to increase the concreteness and immediacy of the interview questions and responses. The informal, conversational interview is a mainstay of participant observation. It is particularly useful when the interviewer/evaluator is able to explore a field setting or program over a fairly long period of time so that a



comprehensive data base is accumulated through in-depth interviewing (by which later interviews build on information obtained in earlier interviews), thus establishing a holistic picture of program change and development.

The weakness of the informal conversational interview is that it requires a greater amount of time to collect systematic information because it may take several conversations with different people before a similar set of questions has been posed to each participant in the program. The informal conversational interview is also more open to interviewer effects in that it depends on the conversational skills of the interviewer/evaluator to a greater extent than do more formal, standardized formats. The phenomenological interviewer must be able to interact easily with people in a variety of settings, generate rapid insights, formulate questions quickly and smoothly, and guard against asking questions that impose interpretations on the situation by the structure of the questions. Data obtained from informal conversational interviews are also difficult to pull together and analyze. Because different questions will generate different responses, the phenomenologist has to spend a great deal of time sifting through responses to find patterns that have emerged at different points in different interviews with different people. By contrast, interviews that are more systematized and standardized facilitate analysis but provide less flexibility in terms of being able to be responsive to individual and situational differences.

## THE INTERVIEW GUIDE

An interview guide is a list of questions or issues that are to be explored in the course of an interview. An interview guide is prepared in order to make sure that basically the same information is obtained from a number of people by covering the same material. The interview guide provides topics or subject areas within which the interviewer is free to explore, probe, and ask questions that will elucidate and illuminate that particular subject. Thus, the interviewer remains free to build a conversation within a particular subject area, to word questions spontaneously, and to establish a conversational style—but with the focus on a particular subject that has been predetermined.

The advantage of an interview guide is that it makes sure that the interviewer/evaluator has carefully decided how best to use the limited time available in an interview situation. The interview guide helps make interviewing across a number of different people more



systematic and comprehensive by delimiting the issues to be discussed in the interview. The interview guide approach is especially useful in conducting group interviews: A guide keeps the interaction focused, but allows individual perspectives and experiences to emerge. Interview guides can be developed in more or less detail, depending on the extent to which the researcher is able to specify important issues in advance and the extent to which it is felt that a particular sequence of questions is important to ask in the same way or the same order to all respondents. Lofland (1971), in his book, *Analyzing Social Settings*, provides a number of examples of interview guides that have been used in the conduct of sociological research. What follows is an example of an interview guide used with participants in a manpower training program.

#### Interview Guideline For Manpower Program Evaluation

What has the trainee done in the program: activities? interactions? products? work performed?

What are the trainee's current work skills? What things can the trainee do that are marketable?

How has the trainee been affected by the program in areas other than job skills—feelings about self? attitudes toward work? aspirations? interpersonal skills? spinoffs?

What are the trainee's plans for the future—work plans? income expectations? lifestyle expectations/plans?

What does the trainee think of the program—strengths? weaknesses? things liked? things disliked? best components? poor components? things that should be changed?

This interview guideline provides a framework within which the interviewer would develop questions, sequence those questions, and make decisions about which information to pursue in greater depth. The interviewer normally would not be expected, however, to go into totally new subjects that are not covered within the framework of the interview guide. The interviewer does not ask questions, for example, about previous employment or education, how the person got into the program, how this program compares with other programs the trainee has experienced, and the trainee's health. Other topics might still emerge during the interview, topics of importance to the respondent that are not listed explicitly on the guide and, therefore, would not normally be explored with each person interviewed. For



example, respondents might comment on their reactions to staff, reactions to written materials, and reactions to specific program components. Comments on these concerns might emerge when, in accordance with the interview guide, the trainee is asked for reactions to program strengths, weaknesses, and so on, but if staff are not mentioned by the respondent, the interviewer would not raise that issue.

An additional, more detailed example of the interview guide approach is included as Appendix 7.1. The example in the appendix illustrates how it is possible to use a detailed outline guide to conduct a series of interviews with the same respondents over the course of a year. The guide in the appendix is the outline for a “descriptive interview” developed by the Educational Testing Service Collaborative Research Project on Reading.

The flexibility permitted by the interview guide approach will become clearer after reviewing the third strategy of qualitative interviewing in the next section.

## THE STANDARDIZED OPEN-ENDED INTERVIEW

In many cases, when conducting a program evaluation, it is only possible to interview participants for a limited period of time. Sometimes it is only possible to interview each participant once. At other times it is possible and desirable to interview participants before they enter the program, when they leave the program, and again after some period of time (for example, six months) after they have left the program. Because of limited time, and because it is desirable to have the same information from each person interviewed, a standardized open-ended format may be used in which each person is asked essentially the same questions. The interview questions are written out in advance *exactly* the way they are to be asked during the interview. Careful consideration is given before the interview about how to word each question. Any clarifications or elaborations that are to be used are written into the interview itself. Probing questions are placed in the interview at appropriate places. The basic purpose of the standardized open-ended interview is to minimize interviewer effects by asking the same question of each respondent. Moreover, the interview is systematic and the necessity for interviewer judgment during the interview is reduced. The standardized open-ended interview also makes data analysis easier because it is possible to locate each respondent’s answer to the same question rather quickly and to organize questions and answers that are similar.



There are three major reasons for using standardized open-ended interviews as part of an evaluation:

- (1) the exact instrument used in the evaluation is available for inspection by decision makers and information users;
- (2) variation among interviewers can be minimized where a number of different interviewers must be used; and
- (3) the interview is highly focused so that interviewee time is carefully used.

In many cases it is sufficient to make available a topical interview guide for decision makers and information users to inspect. However, the problems of legitimacy and credibility for qualitative data can make it politically wise to produce an exact interview form that one can show to decision makers and information users, telling them with certainty that these are the exact questions that will be asked of clients or others who are interviewed. By generating a standardized form decision makers and information users can participate more completely in writing the interview instrument before the interview is used. They will then know exactly what is going to be asked and what is not going to be asked. This reduces the likelihood of the data being attacked later because certain questions were missed or asked in the wrong way. By making it clear, in advance of data collection, exactly what questions will be asked, the limitations of the data can be known and discussed beforehand.

A related political problem is asking different questions of different clients. While a phenomenological approach, and even the interview guideline approach, have the strengths of permitting greater flexibility and individualization, these approaches also open up the possibility that more information will be collected from some people than from others. When analyzing the data it becomes difficult to be certain how the findings are influenced by these qualitative differences in the depth and breadth of information received from different people. For the conduct of basic research, when one is attempting to understand the holistic world view of a group of people it is not necessary to collect the same information from each person. The political credibility of the data collected is less of an issue under basic research conditions. However, when using qualitative data-collection procedures for evaluation purposes, it is often helpful to minimize issues of legitimacy and credibility by carefully collecting the same information from everyone who is interviewed.



The standardized open-ended interview also reduces variation among interviewers. Some evaluations rely on volunteers to do interviewing; at other times program staff may be involved in doing some interviewing; and in still other instances interviewers may be novices, students, or others who are not professional social scientists/evaluators. When a number of different interviewers are used, variations in data created by differences among interviewers will become particularly apparent if an informal conversational approach to data-gathering is used or even if each interviewer uses a basic guide. The best way to guard against variations among interviewers is to carefully word questions in advance and train the interviewers not to deviate from the precise forms. The data collected are still open-ended, in the sense that the respondent supplies his or her own words, thoughts, and insights in answering the questions, but the precise wording of the questions is determined ahead of time.

The weakness of this approach is that it does not permit the interviewer to pursue topics or issues that were not anticipated when the interview was written. Constraints are also placed on the use of different lines of questioning with different people based on their unique experiences. Therefore, a standardized open-ended interview approach will reduce the extent to which individual differences and circumstances can be taken into account; on the other hand, this approach can reduce individual interviewer effects and facilitate data analysis.

Just as it was possible to some extent to combine a phenomenological approach with an interview guide approach, it is also possible to combine an interview guide approach with a standardized open-ended approach. Thus, a number of basic questions may be worded precisely in a predetermined fashion, while permitting the interviewer more flexibility in probing and more decision-making flexibility in determining when it is appropriate to explore certain subjects in greater depth, or even to undertake whole new areas of inquiry that were not originally included in the interview instrument. It is even possible to adopt a standardized open-ended interview format in the early part of an interview and then leave the interviewer free to pursue any subjects of interest during the latter parts of the interview. Another combination would include using a phenomenological approach (the informal conversational interview) early in the evaluation project, followed midway through by an interview guide, and then closing the program evaluation with a standardized open-ended interview to give systematic information from a sample of



participants at the end of the program or when conducting follow-up studies of participants.

To illustrate the standardized open-ended interview three interviews have been reproduced in Appendix 7.2. These interviews were used to gather information from participants in an Outward Bound wilderness program for disabled persons. The first interview was conducted at the beginning of the program; the second interview was used at the end of the ten-day experience; and the third interview took place six months after the program.

## SUMMARY OF INTERVIEWING STRATEGIES

The common characteristic of all three qualitative approaches to interviewing is that the persons being interviewed respond in their own words to express their own personal perspectives. While there are variations in strategy concerning the extent to which the wording and sequencing of questions ought to be predetermined, there is no variation in the principle that the response format should be open-ended. The interviewer never supplies and predetermines the phrases or categories that must be used by respondents to express themselves. The purpose of qualitative interviewing in evaluation is to understand how program staff and participants view the program, to learn their terminology and judgments, and to capture the complexities of their individual perceptions and experiences. This is what distinguishes qualitative interviewing from the closed interview, questionnaire, or test typically used in quantitative evaluations. Such closed instruments force program participants to fit their knowledge, experiences, and feelings into the evaluator's categories. *The fundamental principle of qualitative interviewing is to provide a framework within which respondents can express their own understandings in their own terms.*

Table 7.1 summarizes the basic variations in evaluation research interview instrumentation. In reviewing this summary table it is important to keep in mind that these are presented as pure types. In practice any particular evaluation may employ several of these strategies or combinations of approaches.



Table 7.1 Variations in Evaluation Research Interview Instrumentation

<i>Type of Interview</i>	<i>Characteristics</i>	<i>Strengths</i>	<i>Weaknesses</i>
1. Informal conversational interview	Questions emerge from the immediate context and are asked in the natural course of things; there is no predetermination of question topics or wording.	Increases the salience and relevance of questions; interviews are built on and emerge from observations; the interview can be matched to individuals and circumstances.	Different information collected from different people with different questions. Less systematic and comprehensive if certain questions don't arise "naturally." Data organization and analysis can be quite difficult.
2. Interview guide approach	Topics and issues to be covered are specified in advance, in outline form; interviewer decides sequence and wording of questions in the course of the interview.	The outline increases the comprehensiveness of the data and makes data collection somewhat systematic for each respondent. Logical gaps in data can be anticipated and closed. Interviews remain fairly conversational and situational.	Important and salient topics may be inadvertently omitted. Interviewer flexibility in sequencing and wording questions can result in substantially different respondents, thus reducing the comparability of responses.
3. Standardized open-ended interview	The exact wording and sequence of questions are determined in advance. All interviewees are asked the same basic questions in the same order.	Respondents answer the same questions, thus increasing comparability of responses; data are complete for each person on the topics addressed in the interview. Reduces interviewer effects and bias when several interviewers are used. Permits decision makers to see and review the instrumentation used in the evaluation. Facilitates organization and analysis of the data.	Little flexibility in relating the interview to particular individuals and circumstances; standardized wording of questions may constrain and limit naturalness and relevance of questions and answers.
4. Closed quantitative interviews	Questions and response categories are determined in advance. Responses are fixed; respondent chooses from among these fixed responses.	Data analysis is simple; responses can be directly compared and easily aggregated; many questions can be asked in a short time.	Respondents must fit their experiences and feelings into the researcher's categories; may be perceived as impersonal, irrelevant, and mechanistic. Can distort what respondents really mean or experienced by so completely limiting their response choices.



## THE CONTENT OF INTERVIEWS: WHAT QUESTIONS TO ASK

A number of decisions must be made in conceptualizing an interview, whether the interview takes place spontaneously in the field or is carefully prepared as a standardized open-ended instrument. The evaluator must decide what questions to ask, how to sequence questions, how much detail to solicit, how long to make the interview, and how to word the actual questions. These are all measurement questions that will affect the quality of interview responses. There are basically six kinds of questions that can be asked of people. On any given topic it is possible to ask any of these questions.

### EXPERIENCE/BEHAVIOR QUESTIONS

These are questions about what a person does or has done. These questions are aimed at eliciting descriptions of experiences, behaviors, actions, and activities that would have been observable had the observer been present. "If I had been in the program with you, what would I have seen you doing?" "If I followed you through a typical day, what would I see you doing, what experiences would I observe you having?"

### OPINION/VALUE QUESTIONS

These are questions aimed at understanding the cognitive and interpretive processes of people. Answers to these questions tell us what people *think* about the world or about a specific program. They tell us about people's goals, intentions, desires, and values. These questions typically carry an implication of respondent rationality and decision-making. "What do you believe?" "What do you think about?" "What would you like to see happen?" "What is your opinion of \_\_\_\_\_?"

### FEELING QUESTIONS

These are questions aimed at understanding the emotional responses of people to their experiences and thoughts. There is an implicit assumption of spontaneity about the origin of emotional responses. Feelings occur inside people; they are their natural, emotional responses to what happens around them or to them. Feelings tap the affective dimension of human life. In asking feeling



questions, the interviewer is looking for adjective responses, for example, “Do you feel anxious, happy, afraid, intimidated, confident, . . .?”

Opinions and feelings are often confused. It is critical that interviewers understand the distinction between the two in order to know when they have the kind of answer they want to the question they are asking. Suppose an interviewer asks: “How do you feel about that?” The response is: “I think it’s probably the best that we can do under the circumstances.” The question about *feelings* has not really been answered. Analytical, interpretive, and opinion statements are not answers to questions about feelings.

This confusion sometimes occurs because interviewers give the wrong cues when asking questions—for example, by asking opinion questions using the format “How do you feel about that?” instead of “What is your opinion about that?” or “What do you think about it?” When one wants to understand the respondents’ emotional reactions it is appropriate to ask about feelings. When one wants to understand what they think about something, the question should explicitly ask about opinions, beliefs, and considered judgments—not about feelings.

## KNOWLEDGE QUESTIONS

Knowledge questions are asked to find out what factual information the respondent has. The assumption here is that certain things are considered to be *known*—these things are not opinions, they are not feelings; rather, they are the things that one *knows*, the facts of the case. Knowledge about a program may consist of reporting on what services are available, who is eligible, the characteristics of clients, who the program serves, how long people spend in the program, what the rules and regulations of the program are, how one enrolls in the program, and so on. While from a philosophical point of view it is possible to argue that all knowledge is merely a set of beliefs rather than facts, the issue here is to find out what the person being interviewed considers to be factual. It is the respondent’s perspective on the empirical nature of the world that is being elicited.

## SENSORY QUESTIONS

These are questions about what is seen, heard, touched, tasted, and smelled. The purpose of these questions is to allow the interviewer to enter into the sensory apparatus of the respondent. “When you walk through the doors of the program, what do you see?”



Describe to me what I would see if I walked through the doors into the program.” Or again: “What does the counselor ask you when you meet with him? What does he actually say?” Sensory questions attempt to have interviewees describe the stimuli to which they are subject.

## BACKGROUND/DEMOGRAPHIC QUESTIONS

These questions concern the identifying characteristics of the person being interviewed. Answers to these questions help the interviewer locate the respondent in relation to other people. Age, education, occupation, residence/mobility questions, and the like are standard background questions. They are distinguishable from knowledge questions primarily because of their routine nature.

Behaviors, opinions, feelings, knowledge, sensations, and demographics: these are the kinds of questions that it is possible to ask in an interview. Any kind of question one might want to ask can be subsumed in one of these categories. Keeping these types of questions in mind can be particularly helpful when it comes to planning the comprehensiveness of the interview and ordering the questions in some sequence. Before considering the sequence of questions, however, it is important to consider how the time dimension intersects with the different kinds of questions.

## THE TIME FRAME OF QUESTIONS

Any of the questions described above can be asked in the present tense, past tense, or future tense. For example, it is possible to ask a person what they are doing now, what they have done in the past, and what they plan to do in the future. Likewise, one might be interested in present attitudes, past attitudes, or future attitudes. By combining the time frame of questions with the different type of questions it is possible to construct a matrix which generates eighteen different types of questions. Table 7.2 shows that matrix.

Asking all eighteen questions about any particular situation, event, or programmatic activity may become somewhat tedious, especially if the sequence is repeated over and over throughout the interview for different program elements. The matrix constitutes a set of options from which one can select which pieces of information are most important to obtain. In order to understand how these options are applied in an actual interview situation it may be helpful to review an actual interview. The Outward Bound standardized



Table 7.2 A Matrix of Question Options

	<i>Past</i>	<i>Present</i>	<i>Future</i>
Behavior/Experience Questions			
Opinion/Value Questions			
Feeling Questions			
Knowledge Questions			
Sensory Questions			
Demographic/Background Questions			

interview in Appendix 7.2 might be used for this purpose. Try identifying which cell in the matrix (Table 7.2) is represented by *each* question in the Outward Bound interviews.

## THE SEQUENCING OF QUESTIONS

There are no fixed rules of sequence in organizing an interview. Informal conversational interviewing is flexible and responsive so that a fixed sequence is seldom possible. However, standardized open-ended interviews must establish a fixed sequence of questions due to their structured format. I offer, then, some suggestions about sequencing.

I prefer to begin the interview with questions about noncontroversial present behaviors, activities, and experiences. Such questions ask for relatively straightforward descriptions; they require minimal recall and interpretation. Such questions are therefore fairly easy to answer. They encourage the respondent to talk descriptively. Probes should focus on eliciting greater detail—filling out the descriptive picture.

Once some experience or activity has been described it is appropriate to ask about interpretations, opinions, and feelings about the behaviors and actions described. Opinions and feelings are likely to be more accurate at this point in the interview because the respondent has just verbally relived the experience. Thus, a context is established for expressing feelings and opinions.

Knowledge and skill questions also typically need a context. These questions can be quite threatening. It is helpful to ask them in conjunction with specific questions about program activities and experiences that have a bearing on knowledge and skills. Finding out from people what they know and what skills they possess works best



once some rapport and trust have been established in the interview. Relating knowledge and skills to descriptions of program activity can help provide a concrete context for these kinds of questions.

Questions about the present tend to be easier for respondents than questions about the past. Future-oriented questions involve considerable speculation, and responses to questions about future actions or attitudes are typically less reliable than questions about the present or past. I generally prefer to begin by asking questions about the present, then, using the present as a baseline, ask questions about the same activity or attitude in the past. Only then will I broach questions about the future.

Background and demographic questions are basically boring; they epitomize what people don't like about interviews. They can also be somewhat uncomfortable for the respondent, depending on how personal they are. I keep such questions to an absolute minimum and prefer to space them strategically and unobtrusively throughout the interview. I advise never beginning an interview with a long list of routine demographic questions. In qualitative interviewing the interviewee needs to become actively involved in providing descriptive information as soon as possible instead of becoming conditioned to providing short-answer, routine responses to uninteresting categorical questions. Some background information may be necessary at the beginning to make sense out of the rest of the interview, but such questions should be tied to descriptive information about present program experience as much as possible. Otherwise, save the sociological-demographic inquiries (age, socioeconomic status, birth order, and the like) for the end.

## THE WORDING OF QUESTIONS

An interview question is a stimulus that is aimed at creating or generating a response from the person being interviewed. The way a question is worded is one of the most important elements determining how the interviewee will respond. As Payne (1951) put it, asking questions is an art. For purposes of qualitative measurement, good questions should, at a minimum, be open-ended, neutral, singular, and clear. Each of these criteria will be discussed in some detail.

## ASKING TRULY OPEN-ENDED QUESTIONS

The basic thrust of qualitative measurement is to minimize the imposition of predetermined responses when gathering data. When



using qualitative interviewing strategies for data collection it is critical that questions be asked in a truly open-ended fashion. This means that the question should permit respondents to respond in their own terms.

The standard questionnaire item in quantitative measurement provides the respondent with a categorical list of response possibilities: "How do you feel about the program? Would you say that you are (a) very satisfied, (b) somewhat satisfied, (c) not too satisfied, (d) not at all satisfied." It is clear in this instance that the question is closed and that the respondent has been provided with a limited and predetermined set of alternatives. The response possibilities are clearly stated and made *explicit* in the way in which the question is asked. Many interviewers think that the way to make a question open-ended is simply to leave out the structured response categories. Such an approach does *not*, however, make a question truly open-ended. It merely makes the predetermined response categories implicit and disguised. Consider the following "open-ended" question: "How satisfied are you with this program?" On the surface this appears to be an open-ended question. On close inspection, however, it is clear that the dimension along which the respondent can answer the question has already been identified. The respondent is being asked for some degree of satisfaction. It is true that the interviewee can use a variety of modifiers for the word satisfaction—for example, "pretty satisfied," "kind of satisfied," "mostly satisfied," and so on. But in effect the response set has been narrowly limited by the wording of the question. The desired dimension of response is identified in the wording of the question such that the typical answers are only slightly different from those that would have been obtained had the categories been made explicit from the start.

The truly open-ended question does not presuppose which dimensions of feeling, analysis, or thought will be salient for the interviewee. The truly open-ended question allows the person being interviewed to select from among that person's full repertoire of possible responses. Indeed, in qualitative measurement one of the things the evaluator is trying to determine is what dimensions, themes, and images/words people associate with the program use among themselves to describe their feelings, thoughts, and experiences. Examples, then, of truly open-ended questions would take the following format:



How do you feel about the program?

What is your opinion of the program?

What do you think of the program?

The truly open-ended question permits persons being interviewed to take whatever direction and use whatever words they want in order to represent what they have to say.

To be truly open-ended a question cannot be phrased as a dichotomy. In the next section we shall consider the problem of dichotomous questions in interviews.

## THE HORNS OF A DICHOTOMY

Dichotomous response questions provide the interviewee with a grammatical structure suggesting a “yes” or “no” answer.

Are you satisfied with the program?

Have you changed as a result of your participation in this program?

Was this an important experience for you?

Do you know the procedures for enrolling in the program?

Have you interacted much with the staff in the program?

The object of an in-depth interview is to get the person being interviewed to talk about their experiences, feelings, opinions, and knowledge. Far from encouraging the respondent to talk, dichotomous response questions create a dilemma for the respondent because they frequently are not sure whether they are being asked a simple yes-no question or if, indeed, the interviewer expects a more elaborate response. I have found in many cases that interviewers who report that they have difficulty getting respondents to talk are using a string of dichotomous response questions to guide the interview and thereby have programmed the respondent to be entirely reactive in a binary way, allowing the interviewer to supply the content to the interview. Perhaps the classic example is a conversation between a parent and a teenager.

(Teenager returns home from a date.)

Oh, you're home a bit late?

Yeah.

Did you have a good time?

Yeah.

Did you go to a movie?



Yeah.

Was it a good movie?

Yeah, it was ok.

So, it was worth seeing?

Yeah, it was worth seeing.

I've heard a lot about it. Do you think I would like it?

I don't know. Maybe.

Anything else you'd like to tell me about your evening?

No, I guess that's it.

(Teenager goes upstairs to bed. One parent turns to the other and says:  
It sure is hard to get him to talk to us. I guess he's at that age where kids  
just don't want to tell their parents anything.)

Dichotomous response questions give an interview the aura of an interrogation or a quiz rather than an in-depth conversation. In everyday conversation our interactions with each other are filled with dichotomous response questions which we unconsciously ignore and treat as if they were open-ended questions. In a more formal interview setting, however, the interviewee will become more conscious of the grammatical structure of questions and is less likely to ignore questions that pose dichotomous alternatives. Indeed, the more intense and concentrated the interview situation, the more likely the respondent is to pay close attention to the structure of questions and to take questions literally.

In training interviewers I like to play a game where I will only respond literally to the questions asked without volunteering any information that is not clearly demanded in the question. I do this before explaining the difficulties involved in asking dichotomous questions. I have played this game hundreds of times, and the reaction is typically the same. When getting dichotomous responses to general questions, the interviewer will begin to rely on more and more specific dichotomous response questions, thereby digging a deeper and deeper hole which makes it difficult to pull the interview out of the dichotomous response pattern. Transcribed below is an actual interview from a training workshop. In the left column I have recorded the interview that took place; the right column records a truly open-ended alternative to the dichotomous response question that was asked.

### INTERVIEW DEMONSTRATION

Instruction: Okay, now we're going to play an interviewing game. I want you to take turns asking me questions about an evaluation I just



completed. The program being evaluated was a staff development demonstration project that involved taking professionals into a wilderness setting for a week. That's all I'm going to tell you at this point. I'll answer your questions as precisely as I can, but I'll only answer what you ask. I won't volunteer any information that isn't directly asked for by your questions.

	<i>Actual interview</i>	<i>What the interviewer really wanted to know</i>
<i>Question:</i>	Were you the evaluator of this program?	What was your role in this program?
<i>Answer:</i>	Yes.	
<i>Q:</i>	Were you doing a formative evaluation?	What was the purpose of the evaluation?
<i>A:</i>	Mostly.	
<i>Q:</i>	Were you trying to find out if the people changed from being in the wilderness?	What were you trying to find out in doing the evaluation?
<i>A:</i>	That was part of it.	
<i>Q:</i>	Did they change?	How did participation in the program affect participants?
<i>A:</i>	Some of them did.	
<i>Q:</i>	Did you interview people both before and after the program?	What kinds of information did you collect for the evaluation?
<i>A:</i>	Yes.	
<i>Q:</i>	Did you also go along as a participant in the program?	How were you personally involved in the program?
<i>A:</i>	Yes.	
<i>Q:</i>	Did you find that being in the program affected what happened?	How do you think your participation in the program affected what happened?
<i>A:</i>	Yes.	
<i>Q:</i>	Did you have a good time?	What was the wilderness experience like for you?
<i>A:</i>	Yes.	



<i>Actual interview</i>	<i>What the interviewer really wanted to know</i>
Q: Are you reluctant to tell us about the program?	I'd like to find out more about the program. What would be the best way for me to learn more from you about it?
A: No.	

This is clearly an extreme example of using dichotomous response questions in an interview. It should be clear, however, that the truly open-ended questions would have generated quite different information than was being generated, and was likely to be generated, by the dichotomous response questions. In addition, dichotomous response questions can easily become leading questions. Once the interviewer begins to deal with what appears to be a reluctant or timid interviewee, by asking more and more detailed dichotomous response questions he or she can easily begin guessing at possible responses and actually impose those responses on the person being interviewed. One sure sign that this is happening is when the interviewer is doing more talking than the person being interviewed. Consider the following excerpt from an actual interview. This occurred with a teenager who was participating in a chemical dependency program. The interview took place during the time the teenager was involved in the program.

<i>Interview</i>	<i>Comments</i>
Q: Hello, John. It's nice to see you again. I'm anxious to find out what's been happening with you. Can I ask you some questions about your experience?	The opening is dominated by the interviewer. No informal give-and-take. The interviewee is set up to take a passive/reactive role.
A: Okay.	
Q: I'd like you to think about some of the really important experiences you've had here. Can you think of something that stands out in your mind?	Introductory cue sentence is immediately followed by a dichotomous response question.
A: Yeah, . . . the hot seat.	John goes beyond the dichotomous response.



- Q: The hot seat is when one person is the focus of attention for the whole group, right?
- A: Right.
- Q: So, what was it like . . . ? Was this the first time you've seen the "hot seat" used?
- A: One person does it every day.
- Q: Is it different with different people?
- A: Yeah, it depends.
- Q: Well, how about telling me about one that really stands out in your mind.
- A: Okay, let's see, hmm . . . there was this guy yesterday who really got nailed. I mean, he really caught a lot of crap from the group. It was really heavy.
- Q: Did you say anything?
- A: No, it was them others.
- Q: So what was it like for you? Did you get caught up in it? You said it was really heavy. Was it heavy for you or just him or the group?
- A: Yeah, right, and it really got to him.
- The interviewer has provided the definition, rather than getting John's own definition of the hot seat.
- Began open-ended, then changed the question and made a dichotomous response question. The question is no longer singular or open.
- Answer goes beyond the question.
- Question follows previous answer but still a dichotomous response format.
- Spoken as a statement but has the structure of a dichotomous response question.
- Before responding to the open request John reacts to the dichotomous response format.
- Dichotomous response question.
- Multiple questions. Unclear connections. Ambiguous, multiple-choice format at the end.
- John's positive answer ("Yeah, right") is actually uninterpretable, given the questions asked.



- Q: Did you think it was good for him? Did it help him?      Dichotomous response question.
- A: He started crying and got mad and one guy really came down on him and afterwards they were talking, and it seemed to be okay for him.      John wants to describe what happened. The narrowness of the interview questions are limiting his description.
- Q: So it was really intense?      Leading question, setting up an easy acquiescence response.
- A: Yeah, it really was.
- Q: And you got really involved.      Same as previous question.
- A: It was pretty heavy.      John doesn't actually respond to the question. Ambiguous response.
- Q: Okay, I want to ask you some about the lecture part of the program. Anything else you want to say about the hot seat?      Transition. John is cued that the hot seat questions are over.
- (John doesn't answer verbally. Sits and waits for the next questions.)

The person conducting this interview said that she wanted to find out two things in this portion of the interview: what experiences were most salient for John; and how personally involved John was becoming in the experience. She has learned that the "hot seat" was highly salient for John, but she really knows very little about the reasons for that salience. With regard to the second question of his personal involvement, the only data she has comes from his acquiescence to leading questions. In fact, if one lists the data from the interview, there is very little there:

Okay.

Yeah, . . . the hot seat.

Right.

One person does it every day.

Yeah, it depends.



Okay, let's see, hmmm . . . there was this guy yesterday who really got nailed. I mean he really caught a lot of crap from the group. It was really heavy.

No, it was them others.

Yeah, right, and it really got to him.

He started crying and got mad and one guy really came down on him and afterwards they were talking, and it seemed to be okay for him.

Yeah, it really was.

It was pretty heavy.

In looking over the transcript of this portion of the interview, it is clear that the interviewer is talking more than the interviewee. The questions put the interviewee in a passive stance, able to confirm or deny the substance provided by the interviewer but not really given the opportunity to provide in-depth, descriptive detail.

## PRESUPPOSITION QUESTIONS

Presuppositions are a major focus of study for many linguists (Karttunen, 1973; Bandler and Grinder, 1975). Natural language is filled with presuppositions. In the course of communicating as we go about our day-to-day activities, it would be impossible to interact with other people without relying heavily on presuppositions. The dominance of presupposition structures in language has important implications for interviewing. By becoming aware of the effects of presupposition structures in interviewing situations, it is possible for the skillful interviewer to use presuppositions to increase the richness and depth of responses and data obtained. What then, are presuppositions? Linguists Grinder and Bandler define presuppositions as follows:

[W]hen each of us uses a natural language system to communicate, we assume that the listener can decode complex sound structures into meanings, i.e., the listener has the ability to derive the Deep-Structure meaning from the Surface-Structure we present to him auditorily. . . .

[W]e also assume the complex skill of listeners to derive extra meaning from some Surface-Structures by the nature of their form. Even though neither the speaker nor the listener may be aware of this process, it goes on all the time. For example, if someone says:

*I want to watch Kung Fu tonight on TV*

we must understand that Kung Fu is on TV tonight in order to process the sentence *I want to watch* . . . to make any sense. These processes are called presuppositions of natural language [Bandler and Grinder, 1975: 241].



Presuppositions are particularly useful in interviewing because the interviewer presupposes that the respondent has something to say. Such a presupposition increases the likelihood that the person being interviewed will, indeed, have something to say. Consider the following question: "What is the most important experience you had in the program?" This question presupposes that the respondent has had an important experience. The person of whom the question is asked, of course, has the option of responding, "I haven't had any important experiences." However, it is more likely that the interviewee will go directly to the issue of which experience to report as important, rather than dealing first with the question of whether or not an important experience has occurred. Contrast the presupposition format of the open-ended question to the format of the following dichotomous response question: "Have you had any experiences in the program so far that you would call really important?" This dichotomous response question requires the person to make a decision about what an important experience is and whether or not an important experience has occurred. By raising the question at all, the interviewer focuses on the decision about whether or not something important has occurred, rather than finding out what has occurred. The presupposition format, then, bypasses this initial step by asking directly for description rather than asking for an affirmation of the existence of the phenomenon in question. Listed below on the left are typical dichotomous response questions that are used to introduce a longer series of questions. On the right are presuppositions asked in a truly open-ended format that bypass the dichotomous response questions.

### ALTERNATIVE QUESTION FORMATS

*Dichotomous response  
lead-in question*

Do you feel like you know enough about the program to assess its effectiveness?

*Presupposition  
lead-in question*

How effective do you think the program is? (Presupposes that a judgment can be made)

What do you know about the program that leads you to say that? (Presupposes some knowledge of the program)



Have you learned anything from this program?	What have you learned from this program? (Presupposes some learning)
Do you do anything now in your work that you didn't do before the program began?	What do you do now that you didn't do before the program began (Presupposes change)
Is there any misuse of funds in this program?	What kinds of misuse of funds have occurred in this program? (Presupposes at least some misuse of funds)
Are there any conflicts among the staff?	What kinds of staff conflicts have occurred here? (Presupposes conflicts)

There is often a naturalness about the use of presuppositions that makes more comfortable what might be otherwise embarrassing questions. The presupposition includes the implication that what is presupposed is the natural way things occur: It is natural for there to be conflict in programs; it is natural for there to be some misuse of funds in programs; and it is natural for people to have learned something from participation in a program. The presupposition provides a stimulus that asks the respondent to assess the answer to the question directly without making a decision about whether or not something has actually occurred.

I first learned about presuppositions in interviewing from a friend who worked with the agency in New York City that had responsibility for interviewing carriers of venereal disease. The purpose of the interviews was to find out about the carrier's previous sexual contacts so that those persons could be informed that they might have venereal disease. He had learned from experience that there was all the difference in the world between asking a man, "Have you had any sexual relationships with other men?" and asking him, "How many sexual contacts with other men have you had?" The dichotomous response question requires a decision about some admission of homosexuality. The presupposition form of the open-ended question suggests that some sexual contacts with other men might be quite natural, and focuses on the frequency of occurrence rather than whether or not the event has occurred at all. The venereal disease interviewers found that they were much more likely to generate



responses with the presupposition format than with the dichotomous response format.

The real point here is that the purpose of in-depth interviews is to find out what someone has to say. By presupposing that the person being interviewed does, indeed, have something to say, the quality of the descriptions received is likely to be enhanced.

## ASKING SINGULAR QUESTIONS

One of the basic rules of questionnaire writing is that each item must be singular—that is, no more than one idea should be contained in any given question. Consider this example: “How well do you know and like the staff in this program? (a) a lot; (b) pretty much; (c) not too much; (d) not at all.” This item is impossible to interpret in analysis because it asks two questions: (1) How well do you know the staff? (2) How much do you like the staff? Therefore, this is a poor questionnaire item.

When one turns to open-ended interviewing, however, many people think there is no longer a need for the same precision in asking questions. I have seen transcripts of interviews conducted by experienced and well-known field researchers in which several questions have been thrown together which they might think are related but which are likely to confuse the person being interviewed about what is really being asked.

In order to help the staff improve the program, we'd like to ask you to talk about your opinion of the program. What you think are the strengths and weaknesses of the program? What you like? What you don't like? What you think could be improved or should stay the same?

The evaluator who used this question regularly in interviewing argued that by asking a series of questions it was possible to find out what was most salient to the person being interviewed because the interviewee was forced to choose what he or she most cared about in order to respond to the question. The evaluator would then probe more specifically in those areas which were not answered in the initial question.

My own experience is that multiple questions create tension and confusion because the person being interviewed doesn't really know what is being asked. An analysis of the strengths and weaknesses of a program is not the same as reporting what one likes and dislikes about a program. Likewise, recommendations for change may be unrelated to strengths, weaknesses, likes, and dislikes. The following



is an excerpt from an interview with a parent participating in a family education program aimed at helping parents become more effective as parents.

Q: Based on your experience, what would you say are the strengths of this program?

A: The other parents. Different parents can get together and talk about what being a parent is like for them. The program is really parents with parents. Parents really need to talk to other parents about what to do and what they do do and what works and doesn't work. It's the parents, it really is.

Q: What about weaknesses?

A: I don't know . . . I guess I'm not always sure that the program is really getting to the parents who need it the most. I don't really know how you do that, but I just think there are probably a lot of parents out there who need the program and . . . especially maybe single-parent families. And fathers. It's really hard to get fathers into something like this. It should just get to everybody and that's real hard.

Q: Let me ask you now about some of your feelings about the program. What are some of the things that you really have liked about the program?

A: I'd like to put the staff right at the top of that. I really like the program director. She's a really well-educated person and knows a lot, but she never makes us feel dumb. We can say anything or ask anything. She treats us like people, like equals even. I like the other parents. And I like being able to bring my daughter along. They take her into the child's part of the program, but we come together. It's something for us to do together and she has her time and I have my time.

Q: What about dislikes? What are some things you don't like so much about the program?

A: I don't like the time that we meet. We meet in the afternoons after lunch and it kind of breaks into the day at a bad time for me, but there isn't any really good time for all the parents and I know they've tried different times. Time is always going to be a hassle for people. Maybe they could just offer different things at different times. The room we meet in isn't too great but that's no big deal.

Q: Okay, you've given us a lot of information about your experiences in the program, strengths and weaknesses you've observed, and some of the things you've liked and haven't liked so much. Now I'd like to ask you about your recommendations for the program. If you had the power to change things about the program, what would you make different?



A: Well, I guess the first thing is money. It's always money. I just think they should put, you know, the legislature should put more money into programs like this. I don't know how much the director gets paid, but I hear that she's not even getting paid as much as school teachers. She should get paid like a professional. I think there should be more of these programs and more money in them. Oh, I know what I'd recommend. We talked about it one time in our group. It would be neat to have some parents who have already been through the program come back and talk with new groups about what they've done with their kids since they've been in the program, you know, like problems that they didn't expect or things that didn't work out, or just getting experience of parents who've already been through the program to help new parents. We talked about that one day and thought that would be a neat thing to do. I don't know if it would work, but it would be a neat thing. I wouldn't mind doing it, I guess.

Each of these questions solicited a different response. Qualitative measurement through in-depth interviewing requires no less precision in asking questions than is demanded by questionnaires constructed for quantitative measurement. The most important theme running through this discussion of question formulation is that the wording used in asking questions makes a tremendous difference in the kind of response that is received. The interviewer who throws out a bunch of questions all at once to see which one takes hold puts an unnecessary burden on the interviewee to decipher what is being asked. In addition, multiple questions asked at the same time usually mean that the interviewer hasn't figured out what question should be asked at that juncture in the interview, so the interviewer takes the easy way out by asking several questions at once.

Asking several questions at once can also cause the interviewer to lose control of the interview. Given multiple stimuli, the interviewee, not being sure of the focus of the question, is free to go off in any direction at all, including providing information that is irrelevant to the issues under examination. In conducting evaluation interviews there is virtually always a limited amount of time available; both interviewers and respondents have only so much time to give to an interview. To make the best use of that time, it is helpful to prepare highly focused questions that elicit genuine and relevant responses. This means that the interviewer must know what issues are important enough to ask questions about, and to ask those questions in a way that the person being interviewed can clearly identify what it is they are being asked.



## CLARITY OF QUESTIONS

It is the responsibility of the interviewer to make it clear to the interviewee what is being asked. Asking questions that are understandable is an important part of establishing rapport. Unclear questions can make the person being interviewed feel uncomfortable, ignorant, confused, or hostile. Asking singular questions helps a great deal to make things clear. There are a number of other factors that contribute to clarity.

First, in preparing to do an interview, one should find out what terms are being used by respondents when they refer to the program being evaluated. State and national programs are often labeled differently at the local level than they are in the higher offices which fund them. In evaluating local CETA programs (Comprehensive Employment and Training Act Programs) local contractors are funded to establish and implement services in their area. Participants know those programs by the name of the local contractor, such as "Youth Employment Services," "Work for Youth," and "Working Opportunities for Women." Many participants in these programs did not know they were in CETA programs. Conducting an interview with these participants where the word CETA was used would have been confusing and disruptive to the interview.

This can also occur within an agency. Agencies that provide multiple services typically have subunits with their own programmatic names and identities. Participants may identify only with subunit names and not with the larger agency. In other instances the agency may be identified completely with the subunit program. In still other cases the same participants may have participated in a number of subunit programs and therefore find it difficult to respond to questions aimed at evaluating the overall agency. In short, the interviewer carries the burden of (1) deciding which is the appropriate unit about which to question the program participant and (2) learning the language that participants use in talking about that particular unit of analysis.

Second, the clarity of interview questions will depend on understanding what language participants use among themselves in talking about program activities or other aspects of program life. This is a different issue from the question of what the program is labeled. The kind of issue that arises here, for example, concerns how participants refer to program staff. When we interviewed juveniles who had been placed in foster group homes by juvenile courts we had to spend a good deal of preparatory time trying to find out how the juveniles typically referred to the group home parents, to their natural parents,



to probation officers, and to each other in order to ask questions *clearly* about each of those sets of people. For example, when asking about relationships with peers, should we use the word “juveniles,” “adolescents,” “youth,” “teenagers,” or what? In preparation for the interviews we checked with a number of juveniles, group home parents, and court authorities about the proper language to use. There was complete consensus that the best language was to talk about “the other kids in the group home.” There was no consensus at all about how “kids in the group home” referred to group home parents. Thus, one of the questions we had to ask in each interview was: “What do you usually call Mr. and Mrs. \_\_\_\_\_?” We then used the language given to us by that youth throughout the rest of the interview to refer to group home parents.

Third, providing clarity in interview questions may mean avoiding using labels altogether. This means that when asking about a particular program component it may be better to first find out what the interviewee believes that component to be and then to ask questions about the descriptions provided by the person being interviewed. An evaluation of open classrooms in North Dakota included interviews with parents. All of the parents interviewed had children who were participating in an open classroom. However, many of the teachers and local school officials did not use the term “open” to refer to these classrooms because they wanted to avoid political conflicts and stereotypes that were sometimes associated with the notion of “open education.” Thus, when interviewing parents we found that we could not ask general questions concerning their opinions about and feelings toward “open education.” Rather, we had to begin with a sequence of questions like the following:

What kinds of differences have you noticed between your child’s classroom last year and the classroom this year?

Ok, you’ve mentioned several differences. Let me ask you your opinion about each of the things you’ve mentioned. What do you think about \_\_\_\_\_?

This strategy in questioning avoids the problem of collecting data which later turns out to be uninterpretable because the evaluator is not sure what persons being interviewed meant by their responses.

A related problem emerged in interviewing children about their classrooms. The decision makers and information users for the evaluation were interested in, among other things, finding out the basic skill activities of children in the open classrooms. In preparing



for the interviews we learned that many teachers avoided the use of terms like “math time” or “reading time” because they wanted to integrate math and reading into other activities. This meant that in many cases children did not report to parents that they did any “math” in school. These same children would be working on projects, such as the construction of a model of their town using milk cartons that required geometry, fractions, and reductions to scale, but they did not perceive of these activities as “math” because they associated math with worksheets and workbooks. Thus, to find out what kind of math activities children were doing, it was necessary to talk with them in detail about specific projects and work they were engaged in without asking them the simple question, “What kind of math do you do in the classroom?”

The theme running through these suggestions for increasing the clarity of questions relates to the importance of using language that is understandable and part of the frame of reference of the person being interviewed. It means taking special care to find out what language the interviewee uses to describe the program, the staff, program activities, or whatever else the evaluator is interested in talking about, and then using that language provided by the interviewee in the rest of the interview. Those questions which use the respondent’s own language are questions which are most likely to be *clear* to the respondent.

Being clear about what you are asking contributes to the process of establishing and maintaining rapport during an interview. Using words that make sense to the interviewee, words that reflect the respondent’s world view, will improve the quality of data obtained during the interview. In many cases, without sensitivity to the impact of particular words on the person being interviewed, the answer may make no sense at all—or there may be no answer. A Sufi story makes this point quite nicely.

A man had fallen between the rails in a subway station when Nasrudin came along one afternoon. People were all crowding around trying to get him out before the train ran him over. They were all shouting, “Give me your hand!” but the man would not reach up. Mulla Nasrudin elbowed his way through the crowd and leant over the man. “Friend,” he asked, “what is your profession?”

“I am an income tax inspector,” gasped the man. “In that case,” said Nasrudin, “*take* my hand!” The man immediately grasped Mulla’s hand and was hauled to safety.

Nasrudin turned to the open-mouthed audience. “Never ask a tax man to *give* you anything, you fools,” he said [Shah, 1973:68].



Before leaving the issue of clarity there is one other major suggestion that I would make. It is a suggestion about which there is not consensus—indeed, most researchers who use interviews would likely disagree. Nevertheless, my own experience in interviewing leads me to at least offer this suggestion: Avoid “why” questions.

### WHY TO AVOID ASKING “WHY?”

“Why?” questions presume cause-effect relationships, an ordered world, perfect knowledge, and rationality. “Why?” questions presuppose that there are reasons why things occur and that those reasons are knowable. “Why?” questions move beyond what has happened, what one has experienced, how one feels, what one opines, and what one knows to the making of analytical and deductive inferences.

The difficulty of making causal inferences has been thoroughly explored at great length by philosophers of science (Bunge, 1957; Nagel, 1961). Reports from parents about “Why?” conversations with their children also document the difficulty of providing causal explanations about the world. The infinite regression quality of “Why?” questions is part of the difficulty engendered by using them as part of an interview.

Dad, why does it get dark at night?

Because our side of the earth turns away from the sun.

Dad, why does our side of the earth turn away from the sun?

Because that's the way the world was made.

Dad, why was the world made that way?

So that there would be light and dark.

Dad, why should there be dark? Why can't it just be light all the time?

Because then we would get too hot.

Why would we get too hot?

Because the sun would be shining on us all the time.

Why can't the sun be cooler sometimes?

It is, that's why we have night.

But why can't we just have a cooler sun?

Because that's the way the world is.

Why is the world like that?

It just is. Because.

Because why?

Just because.

Oh.

Daddy?

Yes.

Why don't you know why it gets dark?



In a program evaluation interview it might seem that the context for asking a “Why?” question would be clearer. However, if a precise reason for a particular activity is what is wanted, it is usually possible to ask that question in a way that does not involve using the word “why.” Let’s look first at the difficulty posed for the interviewee by the “Why?” question, and then look at some alternative phrases.

“Why did you join this program?” The actual reason for joining the program is probably made up of a constellation of factors, including the influences of other people, the nature of the program, the nature of the person being interviewed, the interviewee’s expectations, and practical considerations. It is unlikely that an interviewee can sort through all of these levels of possibility at once, so the person to whom the question is posed must pick out some level at which to respond.

“Because it was at a convenient time.” (programmatic reason)

“Because I’m a joiner.” (personality reason)

“Because a friend told me about the program.” (information reason)

“Because my priest told me about the program and said he thought it would be good for me.” (social influence reason)

“Because it was inexpensive.” (economic reason)

“Because I wanted to learn about the things they’re teaching in the program.” (outcomes reason)

“Because God directed me to join the program.” (personal motivation reason)

“Because it was there.” (philosophical reason)

Anyone being interviewed could respond at any or all of these levels. The question that the evaluator must decide before conducting the interview is which of these levels is of sufficient importance that it is worth asking a question about. If the primary evaluation question concerns characteristics of the program that attracted participants, then instead of asking, “Why did you join?” the interviewer should ask something like the following: “What was it about the program that attracted you to it?” If the evaluator is interested in learning about social influences that led to participation in a program, either voluntary or involuntary participation, a question like the following could be used:

In most of the decisions we make there are other people who have some influence on what we do. In terms of your participation in this program, what other people played a role in your becoming part of the program?



In some cases the evaluator may be particularly interested in the characteristics of participants, so the question might be phrased in the following fashion:

I'm interested in learning more about you as a person and your personal involvement in this program. What is it about you—your situation, your personality, your desires, whatever—what is it about you that you think led you to become part of this program?

Depending on the depth to which an evaluator wants to explore a particular situation, it might be appropriate to ask all of these questions as well as others. The point is that by thinking carefully about what one wants to know, there is a greater likelihood that respondents will supply answers that make sense and that are usable. My feelings about the difficulties raised with “Why?” questions come from trying to analyze such questions when responses cover such a multitude of dimensions that it was clear that different people were responding to different things. This makes analysis extremely difficult, and often leads to data that simply are unusable. By thinking carefully about exactly what information is needed and how it will be used, the interviewer can focus questions to make them clear to the interviewee as well as to make the responses across interviewees more systematic and comprehensive.

Even with more precise focus, questions that require the interviewer to make deductions and provide explanations are sufficiently taxing on the energy of the interviewee that such questions should be used sparingly. Social scientists in particular, given that they have so much trouble sorting out causes and effects in their own analyses, should be particularly sensitive to the difficulty posed by questions that ask for explanations.

Perhaps my reservations about the use of “Why?” questions come from having so often appeared the fool when asking such questions during interviews with children. In the open classroom interviews we were trying to find out the extent to which children chose to spend time doing interesting things in the room when children in other classrooms might be playing outside. Several teachers during their interviews had mentioned that children in open classrooms often become involved in what they were doing and chose not to go outside for recess.

What's your favorite time in school?

Recess.

Why do you like recess?



Because we go outside and play on the swings.

Why do you go outside?

Because that's where the swings are! (She replied with a look of incredulity that adults could ask such stupid questions.)

Children take interview questions quite literally, and so it rapidly becomes clear when a question is not well thought out. It was during those days of interviewing children in North Dakota that I learned about the problems with “Why?” questions.

## NEUTRAL QUESTIONS

As an interviewer I want to establish rapport with the person I am questioning, but that rapport must be established in such a way that it does not undermine my neutrality concerning what the person tells me. Neutrality means that the person being interviewed can tell me anything without engendering either my favor or disfavor with regard to the content of their response. I cannot be shocked; I cannot be angered; I cannot be embarrassed; I cannot be saddened—indeed, nothing the person tells me will make me think more or less of them.

At the same time that I am neutral with regard to the *content* of what is being said to me, I care very much that that person is willing to share with me what they are saying. *Rapport is a stance vis-à-vis the person being interviewed. Neutrality is a stance vis-à-vis the content of what that person says.* Rapport means that I respect the people being interviewed, so what they say is important because of who is saying it. I want to convey to them that their knowledge, experiences, attitudes, and feelings are important. Yet, the content of what they say to me is not important.

Rapport is built on the ability to convey empathy and understanding without judgment. Throughout this chapter we have been considering ways of phrasing questions that facilitate the establishment of rapport. In this section I want to focus on ways of wording questions that are particularly aimed at conveying that important sense of neutrality.

One kind of question wording that can help establish neutrality is the *illustrative examples format*. When phrasing questions in this way I want to let the person I'm interviewing know that I have pretty much heard it all—the bad things and the good things—and so I'm not interested in something that is particularly sensational, particularly negative, or especially positive. I'm really only interested in what that person's experience has been like. An example of the illustrative examples format is provided by a question taken from interviews we



conducted with juvenile delinquents who had been placed in foster group homes. One section of the interview was aimed at finding out how the juveniles were treated by group home parents.

Ok, now I'd like to ask you to tell me how you were treated in the group home by the parents. Some kids have told us that they felt that they were treated like one of the family in the group home; some kids have told us that they got knocked around and beat up by the group home parents; some kids have told us about sexual things that were done to them; some of the kids have told us about a lot of recreational and hobby kinds of things; some kids have felt they have been treated really good and some kids have been treated really bad. When you think about how *you* are treated in the group home, what kind of things come to mind?

A closely related format is the *illustrative extremes format*. With this format I attempt to let the interviewee know that I have heard it all by giving examples only of extreme responses.

How much dope did you use while you were in the group home? I know that some kids have told me they were doped up the whole time they were in the home, they smoked or dropped stuff every day and every night, while other kids have said that they decided to stay completely straight while they were in the home. How about you?

It is critical to avoid in both the illustrative examples format and the illustrative extremes format asking a *leading* question. Leading questions are the opposite of neutral questions; they give the interviewee hints about what would be a desirable or appropriate kind of answer. Leading questions "lead" the respondent in a certain direction.

An example of a typical leading question that might be asked of juveniles is the following:

We know that most kids use a lot of dope because that's part of what it means to be young, so we figure you use it too—right? So what do you think about everybody using dope?

This question has a built-in response bias that communicates the interviewer's belief that drug use among the young is legitimate and universal. The question is "leading" because the interviewee is led into acquiescence with the interviewer's point of view.

It is important in giving examples that the examples cover several dimensions and be balanced between what might be construed as



positive and negative kinds of responses. My own preference is to use these illustrative formats only as clarifying questions after having begun with a simple, straightforward, and truly open-ended question where the response was not constrained or influenced by any kinds of examples: “What has been your experience with the use of drugs in the group home?”

## ROLE-PLAYING AND SIMULATION QUESTIONS

It is sometimes helpful to provide the interviewee with a context for responding to a question. This context provides cues about the level at which a response is expected. One way of providing such a context is to role play with persons being interviewed, asking them to respond to the interviewer as if he or she were someone else.

Suppose I was a new person who just came into this program, and I asked you what I should do to really do well in the program. What would you tell me?

or

Suppose I was a new kid in this group home, and I didn't know anything about what goes on around here. What would you tell me about the rules that I have to follow?

The effect of these questions is to provide a context for what would otherwise be quite difficult questions. “How does one get the most out of this program?” “What are the rules of this group home?” The role-playing question, in this format, also puts interviewees in the role of expert: they know something of value to someone else. This places the interviewer in the position of a novice, an apprentice. The “expert” is being asked to share his or her expertise with the novice. I have often observed a marked change in animation and enthusiasm on the part of interviewees when role-playing kinds of questions have been used.

Another variation of the role-playing format is a question whereby the interviewer dissociates himself or herself somewhat from asking the question. This has the effect of making the question less personal and probing. Consider these two questions: “How do you sneak dope into the prison?” “Suppose someone you trusted asked you how to sneak dope into the prison. What would you tell him?”

The first question comes across like an interrogation or inquisition. The second question is softened and has more of an informal and informative tone. Despite the fact that the content is the same for



both questions, the second question has the psychological effect on the interviewee of permitting the interviewer to be dissociated from the question. While this technique can be overused and can sound like a phony or trick question if the intonation with which it is asked is hesitating or implies awkwardness, used sparingly and with subtlety the role-playing format can ease the asking of difficult questions and can permit the interviewer to obtain high quality information.

Simulation questions provide a context in a different way. The simulation question asks the person being interviewed to imagine himself or herself in some situation about which the interviewer is interested.

Suppose I was present with you during one of your group therapy sessions. What would I see happening? What would be going on? Describe to me what one of those sessions is like.

or

Suppose I was in your classroom at the beginning of the day when the students first come in. What would I see happening as the students came in? Take me there. Take me to your classroom and let me see what happens during the first ten to fifteen minutes as the students arrive, what you'd be doing, what they'd be doing, what those first fifteen minutes are like.

In effect, these questions ask the interviewer to become an observer. The observer is asked to simulate for the interviewer some situation that has been experienced. In most cases, a response to this question will require the interviewee to visualize the situation to be described. When the interviewee is able to fully move into and experience the simulated situation through a visualization, the interviewer may observe that persons being interviewed take on an abstracted expression. As the purpose of the question is to achieve that abstraction, the interviewer should not try to bring respondents back, but rather encourage them to describe what is happening in the simulation. I frequently find that the richest and most detailed descriptions come from a series of questions that ask a respondent to re-experience and/or simulate some aspect of a program.

## PREFATORY STATEMENTS AND ANNOUNCEMENTS

The purpose of prefatory statements is to let the person being interviewed know what is going to be asked before it is asked. This can serve two functions. First, it alerts the interviewee to the nature



of the question that is coming; it directs their awareness; and it focuses their attention. Second, an introductory announcement about subject matter about to be broached gives the person being interviewed a few seconds to organize his or her thoughts before the question is actually asked. Such questions can help the flow of the interview and reduce the amount of time taken up in what is sometimes an awkward silence while the interviewee is reflecting on or remembering the information necessary to answer a question. There are several different formats that can be used as prefaces to asking specific questions.

The *transition format* announces that one section of the interview has been completed and the new section is about to begin. The transition format tells the respondent that closure has been reached on one topic and a new topic is about to be introduced.

We've been talking about the goals and objectives of the program. Now I'd like to ask you some questions about actual program activities. What are the major activities offered to clients in this program?

or

We've been talking about your personal experiences with this program. Now I'd like to ask you some questions concerning your opinions about the program. First, I'd like to ask you to think about the program's strengths and weaknesses. Let's begin with strengths. What would you say are the basic strengths of this program, from your point of view?

The transition format essentially says to the interviewee: "This is where we've been . . . and this is where we're going . . ." Questions prefaced by transition format help maintain the smooth flow of an interview.

An alternative form of transition is the *summarizing transition format*. This format brings closure to a section of the interview by repeating to the person interviewed what it is they have said in that section of the interview and then asking them if they have anything to add or to clarify before moving on to a new subject. The summarizing transition format announces to the respondent that the interviewer is ready to bring closure to one section of the interview and to begin a new section. However, first the interviewer should make sure that he or she is not cutting off any final comments from the person being interviewed.

Before we move on to the next set of questions, let me make sure I've got everything you said about the program goals and objectives. You



said the program had five goals. First, . . . Second, . . . Before I ask you some questions about program activities related to these goals, are there any additional goals or objectives that I haven't got down here?

The summarizing transition format lets the person being interviewed know that the interviewer is listening and is recording what is being said. The summary allows the interviewee to make clarifications, corrections, and additions in order to bring closure to one section of the interview. This format also announces that it is time to move on to other questions and lets the respondent know what is coming up next.

The *direct announcement format* is a simple statement telling the interviewee what will be asked next. A preface to a question that announces its content softens the harshness or abruptness of the question itself. Direct prefatory statements can make an interview more conversational and easy-flowing, less like an interrogation. The transcriptions below show two interview sequences, one without prefatory statements and the other with prefatory statements.

### DEMONSTRATION OF THE DIRECT ANNOUNCEMENT FORMAT

#### *Interview without direct preface*

A: . . . so I guess I'd say that's what the program has done for me.

Q: How have you changed as a result of the program?

#### *Interview with direct preface*

A: . . . so I guess I'd say that's what the program has done for me.

Q: Let me ask you to think now about what changes you see in yourself because of this program. (pause) How have you changed since you began the program?

There are times when the flow of the interview makes it imperative that direct, follow-up questions be asked without preface or announcement. There are other times when the flow of the interview is made more conversational by the insertion of direct announcements about the content of a question before it is asked. All of these formats must be used selectively and strategically. Constant repetition of the same format or mechanical use of a particular format will make the interview more, rather than less, awkward.



The *attention-getting preface* makes a comment about the question that is going to be asked. The comment may concern the importance of the question, the difficulty of the question, the openness of the question, or any other characteristic of the question the interviewer thinks should be called to the attention of the respondent. Several such prefaces are illustrated in the following questions.

This next question is particularly important to the program staff. How do you feel the program could be improved?

or

This next question is purposefully vague so that you can respond in any way that makes sense to you. What difference has this program made to the larger community?

or

This next question may be particularly difficult to answer with certainty, but I'd like to get your thoughts on it. In thinking about how you've changed during the last year, how much has this program caused those changes compared to other things that were happening in your life at this time?

or

This next question is aimed directly at getting *your* perspective. What's it like to be a client in this program?

or

As you will recognize, this next question has been particularly controversial. What kind of staff are needed to run a program like this?

The common element in each of these examples is that some prefatory comment is made about the question to alert the interviewee to the nature of the question. The attention-getting format tells the person being interviewed that the question about to be asked has some unique quality that makes it particularly worthy of being answered.

Making statements about the questions being asked is a way for the interviewer to engage in some conversation during the interview without commenting on the answers being provided by the interviewee. Thus, the interviewer is given something to say that goes beyond a pure interrogation function, but what is said concerns the questions and not the respondent's answers. In this fashion the interview can be made more interesting, more conversational, and interactive.



## PROBES AND FOLLOW-UP QUESTIONS

*Probes* are used to deepen the response to a question, to increase the richness of the data being obtained, and to give cues to the interviewee about the level of response that is desired. The word “probe” itself is usually best avoided in interviews. The expression, “Let me probe that further” can sound as if the interviewer is about to perform surgery on the respondent or conducting an investigation of something illicit or illegal. Quite simply, a probe is an interview tool used to go deeper into the interview responses. As such, probes should be conversational, offered in a natural style and voice, and used to follow up initial responses.

One natural set of conversational probes consists of *detail-oriented* questions. These are the basic questions that fill in the blank spaces of a response.

When did that happen?

Who else was involved?

Where were you during that time?

What was your involvement in that situation?

How did that come about?

Where did that happen?

These *detail-oriented* probes are the basic “who,” “where,” “what,” “when,” and “how” questions that are used to obtain a complete and detailed picture of some activity or experience. There are times, as in the probes suggested above, when *particular* details are elicited through follow-up questions.

At other times an interviewer may want to keep a respondent talking more about a subject. In such cases *elaboration* probes are used. Elaboration probes encompass a variety of ways to cue the person being interviewed that they should keep talking. The best cue an interviewer can use to encourage continued talking is to gently nod his or her head. (Overenthusiastic head-nodding will often be perceived as endorsement of the content of a response or as wanting the person to stop talking because the interviewer has already understood what the respondent has to say. Gentle and strategic head-nodding is aimed at communicating that the interviewer is listening and wants to go on listening.) The verbal corollary of head-nodding is the quiet “uh-huh.” A combination may be necessary; when the respondent seems about to stop talking and the interviewer would like to encourage him or her to continue, a combined “uh-huh”



with a gentle rocking of the whole upper body can communicate interest in having the interviewee elaborate.

*Elaboration probes* also have direct verbal forms. These consist of any statement or request that the person keep talking.

Would you elaborate on that?

Could you say some more about that?

That's helpful. I'd appreciate it if you could give me more detail.

I'm beginning to get the picture. (The implication is that I don't have the full picture yet, so please keep talking.)

I think I'm beginning to understand.

Let me make sure I've got down exactly what you said, then I'd like to ask you to say some more on that.

There are times when the interviewee should be encouraged to say more because the interviewer has not fully understood an answer. If something has been said that is ambiguous or an apparent non-sequitur, a *clarification probe* may be useful. Clarification probes tell the interviewee that the interviewer needs more information, a restatement of the answer, or more context.

You said the program is a "success." What do you mean by "success?"

I'm not sure I understand what you meant by that. Could you elaborate, please.

I want to make sure I understand what you're saying. I think it would help me if you could say some more about that.

What you're saying now is very important and I want to make sure that I get it in exactly the way you mean it. Would you repeat what you said so that I can get your exact thoughts?

I'm not sure I understand exactly what you mean.

I didn't quite catch your full meaning. Would you run that by me again?

A clarification probe should be used naturally and gently. It is best for the interviewer to convey the notion that the failure to understand is the fault of the interviewer and not a failure by the person being interviewed. The interviewer does not want to make the respondent feel inarticulate, stupid, or muddled. After one or two attempts at achieving clarification, it is sometimes best to leave the particular



topic that is causing the confusion and move on to other questions, perhaps returning to that topic at a later point.

A major characteristic that separates probes from general interview questions is that probes are seldom written out in an interview. Probing is a skill that comes from knowing what to look for in the interview, listening carefully to what is said and what is not said, and being sensitive to the feedback needs of the person being interviewed. Probes are always a combination of verbal and nonverbal cues. Silence at the end of a response can indicate as effectively as anything else that the interviewer would like the person to continue. Probes are used to communicate with the interviewee about what the interviewer wants. More detail? Elaboration? Clarity? Probes, then, provide guidance to the person interviewed. They also provide the interviewer with a way to maintain control of the flow of the interview, a subject discussed in more detail in a later section.

## SUPPORT AND RECOGNITION RESPONSES

Effective interviewing should cause both the interviewer and the interviewee to feel that a two-way flow of communication is going on. Interviews should not be simply interrogations in which the interviewer intensively pursues a set of questions and the respondent provides the answers. The interviewer has a responsibility to communicate clearly what information is desired, why that information is important, and to let the interviewee know how the interview is progressing.

Previous sections have emphasized the wording of questions so that interview questions are clear and detailed responses can be obtained from persons being interviewed. The purpose of the overall interview and the relationship of particular questions to that overall purpose are important pieces of information that go beyond simply asking questions. While the reason for asking a particular question may be absolutely clear to the interviewer, such purposes are not always clear to the respondent. The interviewer communicates respect for persons being interviewed by giving them the courtesy of explaining why questions are being asked. Understanding the purpose of the interview will increase the motivation of the interviewee to respond openly and in detail.

The overall purpose of the interview is conveyed in an opening statement. The most important elements to communicate in this opening statement, at least when interviewing is being done as part of a program evaluation process, are all of the following:



What will be asked in the interview?

Who is the information for?

How will the information be handled, including confidentiality?

What is the purpose of collecting the information?

How will it be used?

The interviewer has an obligation to provide this information at the beginning of the interview. Providing such information does *not*, however, require making long and elaborate speeches. Statements of purpose should be simple, straightforward, and understandable. Long statements about what the interview is going to be like and how it will be used, when such statements are made at the beginning of the interview, are usually either boring or anxiety-producing. The interviewee will find out soon enough what kinds of questions are going to be asked, and, from the nature of the questions, will make judgments about the likely use of such information. The basic message to be communicated in the opening statement is (1) that the information is important, (2) the reasons for that importance, and (3) the willingness of the interviewer to explain the purpose of the interview out of respect for the interviewee.

The purpose of this interview is to get information that will help the program staff improve the program. As someone who has been in the program, you are in a unique position to describe what the program does and how it affects people. And that's what the interview is about: your experiences with the program and your thoughts about your experiences. The answers from all the people we interview, and we're interviewing about 25 people, will be combined into an overview before the program staff see what people said. Nothing you say will ever be identified with you personally. As we go through the interview, if you have any questions about why I'm asking some particular things, please feel free to ask. Or if there's anything you don't want to answer, just say so. The purpose of the interview is to get your insights about how the program operates and how it affects people. Any questions about that before we begin?

While this overview gives a basic notion about the purpose of the interview, it will still be appropriate and important to explain the purpose of particular questions at strategic points throughout the interview. Explaining the purpose of particular questions is a form of prefatory statement that tells the respondent why the interviewer is asking what he or she is about to ask.



This next set of questions is about your own personal background and experiences. The purpose of these background questions is to help us find out how different kinds of people have experienced the program.

or

This next set of questions is about the program staff. The staff are particularly interested in your answers to these questions because they want to know how they come across to participants in the program so that they can become more effective in working with people. The staff has told us that they don't really get a chance to find out how people in the program feel about what they do, so this part of the interview is aimed at giving them some direct feedback.

The other part of this process of maintaining communication with the interviewee is giving out clues about how the interview is going. One of the most common mistakes in interviewing is a failure to provide reinforcement and feedback to the person being interviewed about how the interviewer perceives the interview is progressing. This involves letting the interviewee know from time to time that the purpose of the interview is being fulfilled. Words of thanks, support, and praise will help make the interviewee feel that the interview process is worthwhile.

Your comments about program weaknesses are particularly helpful, I think, because identification of the kind of weaknesses you describe can really help in making changes in the program.

or

It's really helpful to get such a clear statement of what the program is like. That's just the kind of thing we're trying to get at.

or

We are about half-way through the interview now and I think a lot of really important things are coming out of what you're saying.

or

I really appreciate your willingness to express your feelings about that. That's really helpful.

The interviewer can often get clues about what kind of reinforcement is appropriate by watching the interviewee. When the verbal and nonverbal behaviors of the person indicate that he or she is really struggling with the question, going deep within himself or herself trying to form an answer, after his or her response it is entirely appropriate for the interviewer to say: "I know that was a difficult



question and I really appreciate your working with it because what you said came out very clearly.” At other times the interviewer may perceive that only a surface or shallow answer has been provided. It may then be appropriate to say something like the following: “I don’t want to let that question go by without asking you to think about it just a little bit more, because I feel you’ve really given some important detail and insights on the other questions and I’d like to get more of your reflections about this question.”

The point here is that the interview is an interaction. The interviewer provides stimuli to generate a reaction. *That reaction from the interviewee, however, is also a stimulus to which the interviewer responds.* The interviewer must maintain awareness of how the interview is flowing, how the interviewee is reacting to questions, and what kinds of feedback are appropriate and helpful to maintain the flow of communication.

## MAINTAINING CONTROL OF THE INTERVIEW

Time is precious in an interview. Long-winded responses, irrelevant remarks, and digressions in the interview will reduce the amount of time available to focus on critical questions. The interviewer must maintain control of the interview; that control is maintained by (1) knowing what one wants to find out, (2) asking the right questions to get the desired answers, and (3) giving appropriate verbal and nonverbal feedback to the person being interviewed.

Knowing what one wants to find out in the interview means that one is able to recognize and distinguish appropriate from inappropriate responses. It is not enough just to ask the right questions. The interviewer must listen carefully to make sure that the responses received provide answers to the questions that are asked. Consider the following exchange:

Q: What happens in a typical interviewer training session that you lead?

A: I try to be sensitive to where each person is at with interviewing. I try to make sure that I am able to touch base with each person so that I can find out how they’re responding to their training, to get some notion of how each person is doing.

Q: How do you begin a session, a training session?

A: I believe it’s important to begin with enthusiasm, to generate some excitement about interviewing.



In this interaction the interviewer is asking descriptive, behavioral questions. The responses, however, are about beliefs and hopes—they have not actually described what happens. Rather, the responses describe what the interviewee thinks *ought* to happen. Since the interviewer is waiting for behavioral data, it is necessary to first recognize that the responses are not providing the kind of data desired, and then to ask appropriate questions that will lead to behavioral responses.

INTERVIEWER: Okay, you try to establish contact with each person, and you try to generate enthusiasm at the beginning. What I'd like you to do now is to actually take me to a training session. Describe for me what the room looks like, where the trainees are, where you are, and tell me what I would see and hear if I were right there in that session. What would I see you doing? What would I hear you saying? What would I see the trainees doing? What would I hear the trainees saying? Take me into a session so that I can actually experience it.

It is the interviewer's responsibility to work with the person being interviewed to facilitate the desired responses. At times it may be necessary to give more direct feedback about the kind of information that has been received and the kind of information that is desired.

INTERVIEWER: I think I understand now what it is you try to do during an interview training session. You've explained to me what you hope to accomplish and stimulate, now I'd like you to describe to me what you actually do, not what you expect, but what I would actually see happening if I was present at the session.

It is not enough to simply ask the right initial question. Neither is it enough to have a well-planned interview with good, on-target basic questions. The interviewer must listen carefully to the kinds of responses supplied to make sure that the interview is working according to plan. I've seen many well-written interviews that have resulted in largely useless data because the interviewer did not listen carefully to the responses being received and did not recognize that the responses were not providing the kind of information needed. The first responsibility, then, in maintaining control of the interview is knowing what kind of data one is looking for and directing the interview in order to collect that data.

Giving appropriate feedback to the interviewee is essential in pacing an interview and maintaining control of the interview process. Head-nodding, taking notes, "uh-huhs," and silent probes (remaining



quiet when a person stops talking to let them know the interviewer is waiting for more) are all signals that the person being interviewed is on the right track. On the other hand, it is often necessary to stop a highly verbal respondent who gets off the track. The first step in stopping the long-winded respondent is to cease giving the usual cues mentioned above that encourage talking: stop nodding the head; interject a new question as soon as the respondent pauses for breath; stop taking notes, or call attention to the fact that one has stopped taking notes by flipping the page of the writing pad and sitting back, waiting. When these nonverbal cues do not work, it becomes necessary to interrupt the long-winded respondent.

Let me stop you here, for a moment. I want to make sure I fully understand something you said earlier. (Then ask the question aimed at getting the response more targeted.)

or

Let me ask you to stop for a moment because some of what you're talking about now I want to get later in the interview. First I need to find out from you . . .

Interviewers are sometimes concerned that it is impolite to interrupt an interviewee. It certainly can be awkward, but when done with respect and sensitivity, the interruption can actually help the interview. It is both patronizing and disrespectful to let the respondent run on when no attention is being paid to what he or she is saying. It is respectful of both the person being interviewed, and the interviewer, to make good use of the short time available to talk. It is the responsibility of the interviewer to help the interviewee understand what kind of information is being requested and to establish a framework and context that makes it possible to collect the right kind of information.

Asking focused questions in an appropriate style to get relevant answers that are useful in understanding the interviewee's world is what interviewing is all about. Yet, maintaining focus on information that is useful, relevant, and appropriate requires concentration, practice, and *the ability to separate that which is foolish from that which is important*. In his classic *Don Quixote*, Cervantes describes a scene in which Sancho is rebuked by Don Quixote for trying to impress his cousin by repeating deeply philosophical questions and answers that he has heard from other people, all the while trying to make the cousin think that these philosophical discourses were Sancho's own insights.



“That question and answer,” said Don Quixote, “are not yours, Sancho. You have heard them from someone else.”

“Whist, sir,” answered Sancho, “if I start questioning and answering, I shan’t be done til tomorrow morning. Yes, for if it’s just a matter of asking idiotic questions and giving silly replies, I needn’t go begging help from the neighbors.”

“You have said more than you know, Sancho,” said Don Quixote, “for there are some people who tire themselves out learning and proving things that, once learned and proved, don’t matter a straw as far as the mind or memory is concerned” [Cervantes, 1964:682].

Regardless of which interview strategy is used—the informal conversational interviews, the interview guide approach, or standardized open-ended interviews—the wording of questions will affect the nature and quality of responses received. Constant attention to the purpose of specific interviews and to the ways in which questions can be worded to achieve that evaluation purpose will reduce the extent to which, in Cervantes’ words, evaluators “tire themselves out learning and proving things that, once learned and proved, don’t matter a straw as far as the mind or memory is concerned.”

## RECORDING THE DATA

The primary data of in-depth, open-ended interviews are quotations. What people say, what they think, how they feel, what they’ve done, and what they know—these are the things one can learn from talking to people in interviews. The purpose of qualitative evaluation methods is to understand the perspective and the experience of people associated with a program. But no matter what style of interviewing is used, and no matter how careful one words interview questions, it all comes to naught if the interviewer fails to capture the actual words of the person being interviewed. The raw data of interviews are the actual quotations spoken by interviewees. There is not substitute for these data.

Data interpretation and analysis involve making sense out of what people have said, looking for patterns, putting together what is said in one place with what is said in another place, and integrating what different people have said.

These are processes that belong primarily to the analysis phase of qualitative evaluations *after* the data are collected. During the interviewing process itself—that is, during the data collection phase



of evaluation—the purpose of each interview is to record as fully and fairly as possible that particular interviewee’s perspective. Some method for recording the verbatim responses of people being interviewed is essential.

## TAPE RECORDING INTERVIEWS

A tape recorder is part of the indispensable equipment of evaluators using qualitative methods. Tape recorders do not “tune out” conversations, change what has been said because of interpretation (either conscious or unconscious), or record words more slowly than they are spoken. (Tape recorders, do, however, break down and malfunction—a point addressed in the next section.) In addition to increasing the accuracy of data collection, the use of a tape recorder permits the interviewer to be more attentive to the interviewee. The interviewer who tries to write down every word will have a difficult time responding appropriately to interviewee needs and cues. The pace of the interview can become decidedly nonconversational. In brief, the interactive nature of in-depth interviewing is seriously affected by the attempt to take verbatim notes during the interview.

The major justification for using a tape recorder should be made clear to the interviewee.

I’d like to tape record what you have to say so that I don’t miss any of it. I don’t want to take the chance of relying on my notes and thereby miss something that you say or inadvertently change your words somehow. So, if you don’t mind, I’d very much like to use the recorder. If at any time during the interview you would like to turn the tape recorder off, all you have to do is press this button on the microphone, and the recorder will stop.

The use of the tape recorder does not eliminate the need for taking notes. Notes can serve at least two purposes: (1) notes taken during the interview can help the interviewer formulate new questions as the interview moves along, particularly where it may be appropriate to check out something that was said earlier; and (2) taking notes about what is said will facilitate later analysis, including locating important quotations from the tape itself. In addition, note-taking is one of the nonverbal behaviors that helps pace the interview. Note-taking becomes a kind of nonverbal feedback to the interviewee about when something sufficiently important to have written down has been said; conversely, the failure to take notes will often indicate to the respondent that nothing of particular importance is being said.



It should be obvious from my earlier remarks about the nature of the interactions that take place in an interview that the use of a tape recorder does *not* mean that the interviewer can become less attentive to the respondent. This is important regardless of whether a standardized open-ended interview format is used or the more informal conversational approach is the basis for data collection.

One's full attention must be focused upon the interviewee. One must be thinking about probing for further explication or clarification of what he is now saying; formulating probes linking up current talk with what he has already said; thinking ahead to putting in a *new* question that has now arisen and was not taken account of in the standing guide (plus making a note at that moment so one will not forget the question); and attending to the interviewee in a manner that communicates to him that you are indeed listening. All of this is hard enough simply in itself. Add to that the problem of writing it down—even if one takes shorthand in an expert fashion—and one can see that the process of note-taking in the interview decreases one's interviewing capacity. Therefore, if conceivably possible, *tape record*; then one can interview [Lofland, 1971:89].

## TRANSCRIBING INTERVIEWS

Since the raw data of interviews are quotations, the most desirable data to obtain would be full transcription of interviews. Unfortunately, transcribing is enormously expensive. At the Minnesota Center for Social Research, we found that the ratio of transcribing time to tape time was typically 4:1—on the average, it took four hours to transcribe one hour of tape. Despite these costs, full transcriptions are the most desirable data to obtain. Transcripts can be enormously useful in data analysis and later in replications or independent analyses of the data.

Where resources are not sufficient to permit full transcriptions, the interviewer can work back and forth between interview notes and sections of the tape; only those quotations that are particularly important to take from the tape for data analysis and reporting need be transcribed. In either case, whether the full tape is transcribed or only parts of the tape are used to preserve exact quotations, it is critical that the tape recording be of high technical quality. Few things are more distressing in collecting qualitative data than finding that the tape is blank or that background noise is so severe that the tape is virtually worthless. In the first large-scale interviewing project with which I was involved, *nearly twenty percent of the data was lost because of poor-quality recordings*. Transcribers are



particularly sensitive to the quality of tapes, and costs vary directly with tape recording quality. Because of the continuing problem of poor-quality recording, transcribers at the Minnesota Center for Social Research, under the supervision of Neala Schleuning, put together the following suggestions for interviewers using tape recorders.

### How to Keep Transcribers Sane

#### I. *Equipment*

- a. Use electrical outlet and outside mike whenever possible.
- b. If you use batteries check them.
- c. Recorder should be clean and in good condition—check *before* going to an interview.
- d. Take along extra tape cassettes.

#### II. *Before Interview*

- a. Choose a place that's quiet and free from interruptions.
- b. Place microphone close to respondent, then speak loud enough so we hear what you're saying; most important, we want to hear the answer.
- c. Set recorder on stable surface.
- d. *Test the recording system.*

#### III. *During Interview*

- a. Speak clearly and not too fast—respondent is likely to do the same.
- b. Ask respondent to speak clearly.
- c. *Make test with respondent:* Then rewind and listen so respondent can hear whether she/he is speaking distinctly; *if not*, say, "The recorder does not seem to be picking up well. Could you speak up a little?" Whether the problem is mechanical or personal, *correct it before continuing.*
- d. Don't rustle papers, cups, bottles, etc., near the mike.
- e. Turn off recorder during irrelevant discussion.
- f. Watch for tape breakage and tangling.
- g. Follow all cassette recorder instructions.
- h. Repeat test if tape change is necessary.
- i. At end of interview, say "This is the end of interview with

#### IV. *After Interview*

- a. Listen to tape—make notes and erase irrelevant discussion (make note of this for transcribers); list proper names and unfamiliar terminology.
- b. Label tapes and return them to appropriate containers.
- c. Keep tapes and recorder in good condition—do not touch tape or expose it to extreme temperatures.

## NOTE-TAKING DURING INTERVIEWS

When a tape recorder is being used during the interview, notes will consist primarily of key phrases, lists of major points made by the



respondent, and key terms or words shown in quotation marks that capture the interviewee's own language. While most interviewers will not know how to take technical shorthand, it is enormously useful to develop some system of abbreviations and informal shorthand to facilitate note-taking. Some important conventions along this line include: (1) use quotation marks during note-taking only to indicate full and actual quotations; (2) develop some mechanism for indicating interpretations, thoughts, or ideas that may come to mind during the interview—for example, the use of brackets to set off one's own ideas from those of the interviewee; and (3) keep track of questions asked and answers received.

When it is not possible to use a tape recorder because of some sensitive situation, interviewee request, or tape recorder malfunction, note-taking must become much more thorough and comprehensive. Again, it is critical to gather actual quotations as often as possible; when the interviewee has said something that seems particularly important or insightful, it may be necessary to say: "I'm afraid I need to stop you at this point so that I can get down exactly what you said, because I don't want to lose that particular quote. Let me read back to you what I have and make sure it is exactly what you said."

With practice and training, an interviewer can learn to expand notes into more comprehensive detail of what was said in the interview. To do this with accuracy and reliability requires expanding the notes taken during the interview *immediately following the interview*. It is necessary to go through the entire interview afterwards and make extensive notes and comments, elaborating the phrases and outline that was obtained during the interview. This must be done while the responses are still fresh in the interviewer's mind and before other conversations intervene to cloud the memory. This elaboration will consist largely of summaries of responses to each question and integrating actual quotations obtained during the interview into those summaries. On occasion this process of immediately elaborating the interview will reveal areas of ambiguity or uncertainty where the interviewer is not really sure what the person said or meant. As soon as these areas of vagueness are found, the interviewer should check back with the respondent to clarify the meaning. This can often be done over the telephone. In my experience people who are interviewed appreciate such a follow-up because it indicates the seriousness with which the interviewer is taking their responses. Guessing the meaning of a response is absolutely unacceptable; if there is no way of following up the comments with the



respondent, then those areas of vagueness and uncertainty must simply become missing data.

## AFTER THE INTERVIEW

The period after the interview is critical to the rigor and validity of qualitative measurement whether the methods used involved interviewing or observation. The period following an interview is a time for guaranteeing the quality of the data. The first thing to be done after an interview that has been recorded on tape is to check the tape to make sure it was functioning properly. If for some reason a malfunction occurred, the interviewer should immediately make extensive notes of everything that he or she can remember. Even if the tape functioned properly, the interviewer should go over the interview notes to make certain that what is written makes sense, to uncover areas of ambiguity or uncertainty, and to review the quality of information received from the respondent. Did the interviewer find out what he or she really wanted to find out in the interview? If not, what was the problem? Poorly worded questions? Wrong topics? Poor rapport?

At this point immediately following the interview observations should be written down about the interview itself. The interviewer should note where the interview occurred, who was present, observations about how the interviewee reacted to the interview, observations about the interviewer's own role in the interview, and any additional information that would help establish a context for interpreting and making sense out of the interview. If a tape recorder is available, the interviewer may simply want to talk into the microphone. In any case, this period after the interview is a critical time of reflection and elaboration. *It is a time of quality control to guarantee that the data obtained will be useful, reliable, and valid.*

This period after an interview or observation requires great discipline. Interviewing can be exhausting, and it is easy to forego this time of reflection and elaboration, put it off, or neglect it altogether. To do so is to seriously undermine the rigor of qualitative methods. Interviews and observations should be scheduled so that sufficient time is available for data clarification, elaboration, and evaluation. As a rule of thumb I expect to spend at least as much time after the interview going over notes and making observations as I spent in the interview itself. This is also the beginning of analysis because while the situation and data are fresh, insights can occur that might otherwise have been lost. Thus, ideas and interpretations that emerge following an interview or observation should be written down and clearly marked as such.



## PERSONAL REFLECTIONS ON INTERVIEWING

This chapter has attempted to suggest some ideas about how to go about doing interviews. There is no single right way of interviewing, no single correct format that is appropriate for all situations, and no single way of wording questions that will always work. The particular evaluation situation, the needs of the interviewee, and the personal style of the interviewer all come together to create a unique situation for each interview. Therein lies the challenge of in-depth interviewing.

I find that interviewing people can be invigorating and stimulating. It is a chance for a short period of time to try to get inside another person's world. If participant observation means "walk a mile in my shoes," then in-depth interviewing means "walk a mile in my head." A good interview lays open thoughts, feelings, knowledge, and experiences not only to the interviewer, but also to the interviewee. The process of being taken through a directed, reflective process affects the persons being interviewed and leaves them knowing things about themselves that they didn't know—or at least were not aware of—before the interview.

I'm personally convinced that to be a good interviewer you must like doing it. This means taking an interest in what people have to say. You must yourself believe that the thoughts and experiences of the people being interviewed are worth knowing. In short, you must have the utmost respect for these persons who are willing to share with you some of their time to help you understand their world. There is a Sufi story that describes what happens when the interviewer loses this basic sensitivity to and respect for the person being interviewed.

### An Interview with the King of the Monkeys

A man once spent years of his life learning the language of monkeys so that he could personally interview the king of monkeys. Having completed his studies he made careful inquiries to find the king of the monkeys. In the course of searching for the king of the monkeys he had to talk to a number of monkey underlings. He found that the monkeys he spoke to were generally, to his mind, neither very interesting nor very clever. He began to doubt whether he could learn very much from the king of the monkeys either.

Finally he located the king of the monkeys and arranged for an interview. Because of his doubts, however, he decided to begin with a few basic questions before moving on to the deeper questions in which he was really interested. "What is a tree?" he asked.



“It is what it is,” said the king of the monkeys. “We use trees to swing on.”

“And what is the purpose of the banana?”

“They are to eat.”

“How do animals find pleasure?”

“By doing things they enjoy.”

At this point the man decided that the monkey’s responses were rather shallow and uninteresting, and went on his way, severely disappointed. Soon afterwards an owl flew into the tree next to the king of the monkeys. “What was that man doing here?” the owl asked.

“Oh, he was only another silly human,” said the king of the monkeys. “He asked a bunch of simple and meaningless questions so I gave him simple and meaningless answers.”

There have been a number of scholarly studies of the dynamics of interviewing, different types of respondents, and the problems that can emerge in attempting to obtain valid and reliable data from interviewees (Richardson et al., 1965; Hyman, 1954). Certainly there are uncooperative respondents, people who are paranoid, respondents who seem overly sensitive and easily embarrassed, aggressive and hostile interviewees, timid people, and the endlessly verbose who go on at great length about very little. When an interview is going badly it is easy to call forth one of these stereotypes to explain how the interviewee is ruining the interview. Such blaming of the victim (the interviewee), however, does little to improve the quality of the data. Nor does it improve interviewing skills.

A different approach is to believe that there is a way to unlock the internal perspectives of every interviewee. It is the task and responsibility of the interviewer to find which interviewing style and which question format will work with a particular respondent. It is the responsibility of the interviewer to establish an interview climate that facilitates open responses. When the interview goes badly, it is the responsibility of the interviewer, not the fault of the interviewee.

#### Evaluation Interviewing Beatitudes

Ask.

Listen and record.

Ask.

Listen and record.

Ask.

Listen and record.

It is a privilege to listen. To ask is a grave responsibility. Evaluators, listen. Do you not know that you shall be evaluated by your questions?



To ask is to seek entry into another's world. Therefore, ask respectfully and with sincerity. Do not waste questions on trivia and tricks, for the value of the answering gift you receive will be a reflection of the value of your question.

Blessed are the skilled questioners, for they shall be given mountains of words to ascend.

Blessed are the wise questioners, for they shall unlock hidden corridors of knowledge.

Blessed are the listening questioners, for they shall gain perspective.

*From Halcolm's Evaluation Beatitudes*



**Appendix 7.1**  
**Sample of a Detailed Interview Guide\***  
**GUIDELINES FOR THE DESCRIPTIVE INTERVIEW**

ETS COLLABORATIVE RESEARCH  
PROJECT ON READING—SEPTEMBER, 1978

*Spirit of the Guidelines.* This set of guidelines is not a checklist. If it were, it would defeat the basic strategy of the study—which is to make full use of the observations and thought of the teacher and other team members. The guidelines are intended as an index of topics that should be discussed over the course of the year.

*Organization and Use of the Guidelines.* The guidelines are divided into three broad categories of topics for discussion:

- I. Salient Observations
- II. General Behavior Topics
- III. Language and Reading Topics

This roughly corresponds to the organization of each interview, though not necessarily in the sequence given above. That is, each interview will begin with the teacher's salient impressions derived from observation—what the teacher thinks is important to report about the child. Depending on what those impressions are, the interviewer will pick up on related topics within the guidelines. For example, if the teacher begins the interview with a description of some interesting work the child has done, the interviewer might pick up his/her end of the conversation by asking questions about the child's method of work (Topic E, p. 4). After exploring other topics on page 4 that seem pertinent to the sense of the discussion, the interviewer would then move on to talk about some topics in category III. If, on the other hand, the teacher's salient impressions were mainly concerned with reading, the interviewer would move directly to related topics in category III and eventually wind backwards into category II topics.

Teachers should strive to be as descriptive as possible throughout the interview, and interviewers should strive to facilitate description by asking for concrete instances and examples.

*Coverage of Topics in the Guidelines.* No one interview could possibly aspire to cover all topics in the guidelines. Throughout the course of five interviews over the year, however, we will be able to obtain information relevant to each topic.

Some topics (e.g., Physical/Gestural Characteristics) may only be discussed once, assuming the child does not change. Other topics (e.g., Activities and Reading Competence) will undoubtedly be touched on at every interview to update the child's documentary record. Again, it is the



judgment of teacher and interviewer alike that will determine the most relevant topics of discussion for any given interview.

## **I. TEACHER'S SALIENT OBSERVATIONS ABOUT CHILD'S FUNCTIONING**

Basically cover impressions gained through teacher's own observations of the children during the normal course of instruction.

Where appropriate include:

- comments about continuities/changes/fluctuations
- comments about child's work samples

Organization of the Day (first interview only)

- any changes in organization (subsequent interviews)

## **II. GENERAL BEHAVIOR TOPICS**

### **A. PHYSICAL/GESTURAL CHARACTERISTICS**

- typical posture, bearing
- pace of movement
- forcefulness/impact of physical presence
- gestural characteristics
- eye contact
- voice qualities (e.g., loud, soft, fluent, halting)
- voice tone/inflection

### **B. AFFECTIVE EXPRESSION**

- characteristic disposition and how expressed
- how is anger expressed, controlled
- how is affection expressed
- general level of energy

### **C. RELATIONSHIPS**

- how does child relate to (fit in with) whole class
- what social situations does child seek in work/play
- do other children seek out child
- relationship to adults
- does approach/interaction vary in different settings? at different times?

### **D. ACTIVITIES**

- what does child do in classrooms when there is an opportunity to choose?
- breadth and depth of activities
- what are unusual activities for the child to engage in?
- what are things child has never engaged/attempted in classroom?



## E. METHOD OF WORKING

- how does child organize self for work?
- how does child carry through on work?
- does child seek feedback about work? when? from whom?
- does child ask for help with work? when? from whom?
- does child use help that is offered? how?
- evidence that child “knows what he knows”—can gauge own capabilities
- how does child demonstrate capabilities?

## F. SUMMARY OF PROGRESS IN SCHOOL-RELATED WORK (OTHER THAN READING)

- differential/even progress
- unusual accomplishments, activities
- unusual difficulties, blockings

*(The remaining parts of the guide are omitted because of length.)*

*This example of the guide approach to interviewing makes it clear that a great deal of preparation, effort, and concentration is required of the interviewer in using the guide. The interviewer must be thoroughly familiar with the details of the outline so that the interview flows smoothly. After any one interview session the interviewer would compare the data actually obtained in the interview to the data desired as specified in the guide in order to begin planning for the next interview.*

\*My thanks to Ann Bussis and Ted Chittenden of ETS for permission to include this guide.



## APPENDIX 7.2

**EXAMPLES OF STANDARDIZED OPEN-ENDED INTERVIEWS**

*The attached edited interviews were used in evaluation of an Outward Bound program for the disabled. Outward Bound is an organization that uses the wilderness as an experiential education medium. This particular program consisted of a ten-day experience in the Boundary Water Canoe Area of Minnesota. The group consisted of half able-bodied participants and half disabled participants including: paraplegics; persons with cerebral palsy, epilepsy, or other developmental disabilities; blind and deaf participants; and, on one occasion, a quadriplegic. The first interview was conducted at the beginning of the program; the second interview was used at the end of the ten-day experience; and the third interview took place six months later.*

*To save space, many of the probes and elaboration questions have been deleted and space for writing notes has been eliminated. The overall thrust and format of the interviews have, however, been retained.*

**PRE-COURSE INTERVIEW MINNESOTA OUTWARD BOUND  
SCHOOL COURSE FOR THE DISABLED**

This interview is being conducted before the course as part of an evaluation process to help us plan future courses. You have received a consent form to sign, which indicates your consent to this interview. The interview will be recorded.

1. First, we'd be interested in knowing how you became involved in the course. How did you find out about it?

a. What about the course appealed to you?

b. What previous experiences have you had in the outdoors?

2. Some people have difficulty deciding to participate in an Outward Bound course, and others decide fairly easily. What kind of decision process did you go through in thinking about whether or not to participate?

a. What particular things were you concerned about?

b. What is happening in your life right now that stimulated your decision to take the course?

3. Now that you've made the decision to go on the course, how do you feel about it?

a. How would you describe your feelings right now?

b. What lingering doubts or concerns do you have?

4. What are your expectations about how the course will affect you personally?

a. What changes in yourself do you hope will result from the experience?

b. What do you hope to get out of the experience?

5. During the course you'll be with the same group of people for an extended period of time. What feelings do you have about being part of a group like that for nine full days?

a. Based on your past experience with groups, how do you see yourself fitting into your group at Outward Bound?



**FOR DISABLED**

6. One of the things we're interested in understanding better as a result of these courses is the everyday experience of disabled people. Some of the things we are interested in are:

- a. How does your disability affect the types of activities you engage in?
- b. What are the things that you don't do that you wish you could do?
- c. How does your disability affect the kinds of people you associate with?

(Clarification): Some people find that their disability means that they associate mainly with other disabled persons. Others find that their disability does not effect their contacts with people. What has your experience been along these lines?

- d. Sometimes people with disabilities find that their participation in groups is limited. What has been your experience in this regard?

7. About half of the participants on the course are disabled people and about half are people without disabilities. How would you expect your relationship with the disabled people to be different from your relationship with course participants who are not disabled?

8. We'd like to know something about how you typically face new situations. Some people kind of like to jump into new situations, whether or not some risk may be involved. Other people are more cautious about entering situations until they know more about them. Between these two, how would you describe yourself?

9. Okay, you've been very helpful. Are there other thoughts or feelings you'd like to share with us to help us understand how you're seeing the course right now. Anything at all you'd like to add?

**FOR ABLE-BODIED**

6. One of the things we're interested in understanding better as a result of these courses is feelings able-bodied people have about being with disabled folks. What kinds of experiences with disabled people have you had in the past?

- a. What do you personally feel you get out of working with disabled people?
- b. In what ways do you find yourself being different from your usual self when you're with disabled people?
- c. What role do you expect to play with disabled people on the Outward Bound course?

(Clarification): Are there any particular things you expect to have to do?

- d. As you think about your participation in this course, what particular feelings do you have about being part of an outdoor course with disabled people?

**POST-COURSE INTERVIEW**

We're conducting this interview right at the end of your course at Minnesota Outward Bound. We hope this will help us better understand what you've



experienced so that we can improve future courses. You have signed a form giving your consent for material from this interview to be used in a written evaluation of the course. This interview is being tape-recorded.

1. To what extent was the course what you expected it to be?
  - a. How was it different from what you expected?
  - b. To what extent did the things you were concerned about before the course come true?
    - b-1. Which things came true?
    - b-2. Which didn't come true?
2. How did the course affect you personally?
  - a. What changes in yourself do you see or feel as a result of the course?
  - b. What would you say you got out of the experience?
3. During the last nine days you've been with the same group of people constantly. What kind of feelings do you have about having been a part of the same group for that time?
  - a. What feelings do you have about the group?
  - b. What role do you feel you played in the group?
  - c. How was your experience with this group different from your experiences with other groups?
  - d. How did the group affect you?
  - e. How did you affect the group?
  - f. In what ways did you relate differently to the able-bodied and disabled people in your group?
4. What is it about the course that makes it have the effects it has? What happens on the course that makes a difference?
  - a. What do you see as the important parts of the course, that make an Outward Bound course what it is?
  - b. What was the high point of the course for you?
  - c. What was the low point?
5. How do you think this course will affect you when you return to your home?
  - a. Which of the things you experienced this week will carry over to your normal life?
  - b. What plans do you have to change anything or do anything differently as a result of this course?

#### FOR DISABLED

6. We asked you before the course about your experience of being disabled. What are your feelings about what it's like to be disabled now?
  - a. How did your disability affect the type of activities you engaged in on the course?

#### FOR ABLE-BODIED

6. We asked you before the course your feelings about being with disabled people. As a result of the experiences of the last nine days, how have your feelings about disabled people changed?
  - a. How have your feelings *about*



(Clarification): What things didn't you do because of your disability?

b. How was your participation in the group affected by your disability?

*yourself* in relation to disabled persons changed?

b. What did you personally get out of being/working with disabled people on this course?

c. What role did you play with the disabled people?

d. How was this role different from the role you usually play with disabled people?

7. Before the course we asked you how you typically faced a variety of new situations. During the last nine days you have faced a variety of new situations. How would you describe yourself in terms of how you approached these new experiences?

a. How was this different from the way you usually approach things?

b. How do you think this experience will affect how you approach new situations in the future?

8. Suppose you were being asked by a government agency whether or not they should sponsor a course like this. What would you say?

a. What arguments would you give to support your opinion?

9. Okay, you've been very helpful. We'd be very interested in any other feelings and thoughts you'd like to share with us to help us understand your experience of the course and how it affected you.

## SIX MONTH FOLLOW-UP INTERVIEW

This interview is being conducted about six months after your Outward Bound course to help us better understand what participants experience so that we can improve future courses.

1. Looking back on your Outward Bound experience, I'd like to ask you to begin by describing for me what you see as the main components of the course? What makes an Outward Bound course what it is?

a. What do you remember as the highlight of the course for you?

b. What was the low point?

2. How did the course affect you personally?

a. What kinds of changes in yourself do you see or feel as a result of your participation in the course?

b. What would you say you got out of the experience?

3. For nine days you were with the same group of people, how has your experience with the Outward Bound group affected your involvement with groups since then?



## FOR DISABLED

(\*Check previous responses before interview. *If person's attitude appears to have changed, ask if they perceive a change in attitude.*)

4. We asked you before the course to tell us what it's like to be disabled. What are your feelings about what it's like to be disabled now?

- a. How does your disability affect the types of activities you engage in?

(Clarification): What are some of the things you don't do because you're disabled?

- b. How does your disability affect the kinds of people you associate with?

(Clarification): Some people find that their disability means they associate mainly with other disabled persons. Other people with disabilities find that their disability in no way limits their contacts with people. What has been your experience?

- c. As a result of your participation in Outward Bound, how do you believe you've changed the way you handle your disability?

5. About half of the people on the course were disabled people and about half were people without disabilities. To what extent did you find yourself acting differently with disabled people compared to the way you acted with able-bodied participants?

6. Before this course we asked you how you typically face new situations. For example, some people kind of like to jump into new situations even if some risks are involved. Other people are more cautious, etc. How would you describe yourself along these lines right now?

- a. To what extent, if at all, has the way you have approached new situations since the course been a result of your Outward Bound experience?

## FOR ABLE-BODIED

4. We asked you before the course to tell us what it's like to work with the disabled. What are your feelings about what it's like to work with the disabled now?

- a. What do you personally feel you get out of working with disabled persons?
- b. In what ways do you find yourself being different from your usual self when you are with disabled people?
- c. As you think about your participation in the course, what particular feelings do you have about having been part of a course with disabled people?



7. Have there been any ways in which the Outward Bound course affected you that we haven't discussed?

(If Yes): How? Would you elaborate on that?

- a. What things that you experienced during that week carried over to your life since the course?
- b. What plans have you made, if any, to change anything or do anything differently as a result of the course?

8. Suppose you were being asked by a government agency whether or not they should support a course like this. What would you say?

- a. Who shouldn't take a course like this?

9. Okay, you've been very helpful. Any other thoughts or feelings you might share with us to help us understand your reactions to the course and how it affected you?

- a. Anything at all you'd like to add?







## *Data Analysis*

Halcolm will tell you this:

“Because you can name something does not mean you understand it.  
Because you understand it does not mean it can be named.”

And this:

“What you do not see you cannot describe. What you cannot  
describe you cannot interpret.

“But because you can describe something does not mean you can  
interpret it.”

*From Halcolm's Evaluation Proverbs*

The riddle about the sound of one hand clapping arose from watching  
the first decision-maker reading the first evaluation report.

*From Halcolm's Evaluation Koans*







## THE PURPOSE OF QUALITATIVE ANALYSIS: A Perspective

A young man traveling through a new country heard that a great Sufi Mulla, also traveling in that region, had unequaled insight into the mysteries of the world. The young man determined to become his disciple. He found his way to the Mulla and said, "I wish to place my education in your hands that I might learn to interpret what I see as I travel through the world."

After six months of traveling from village to village with the Mulla the young man was confused and disheartened. He decided to reveal his frustration to the Mulla.

"For six months I have observed the services you provide to the people along our route. In one village you tell the hungry that they must work harder in their fields; in another village you tell the hungry to give up their preoccupation with food; in yet another village you tell the people to pray for a richer harvest. In each village the problem is the same, but always your message is different. I can find no pattern of Truth in your teachings."

The Mulla looked piercingly at the young man.

"Truth? When you came here you did not tell me you wanted to learn Truth. You ask for Truth. Truth is like the Buddha. When met on the road it should be killed. If there were only one Truth to be applied to all villages there would be no need of Mullas to travel from village to village.

"When you first came to me you said you wanted to 'learn how to interpret' what you see as you travel through the world. Your confusion is simple. To interpret and to state Truths are two quite different things."

Having finished his story Halcolm smiled at the youths. "Go, my children. Seek what you will, do what you must."

*From Halcolm's Evaluation Parables*



## THE PURPOSE OF EVALUATION RESEARCH

In order to analyze and interpret qualitative data the evaluator must have some sense of purpose and direction. What is the end result of doing an evaluation study? What is evaluation research supposed to produce?

Analysis, interpretation, and evaluation are not simple, technical processes. There are no formal, universal rules to follow in analyzing, interpreting, and evaluating qualitative data. *Analysis* is the process of bringing order to the data, organizing what is there into patterns, categories, and basic descriptive units. *Interpretation* involves attaching meaning and significance to the analysis, explaining descriptive patterns, and looking for relationships and linkages among descriptive dimensions. *Evaluation* involves making judgments about and assigning value to what has been analyzed and interpreted: Is it “good” or “bad?”; should something be done, and if so, what? Evaluation research is thus the systematic collection, analysis, and interpretation of information about the activities and outcomes of actual programs in order for interested persons to make judgments about specific aspects of what the program is doing and affecting (Patton, 1978).

What makes it difficult to get a firm fix on the purpose of evaluation is that evaluators are typically trained in behavioral and social science disciplines, whose socialization processes often include commitment to scientific purposes that go beyond the simple definition of evaluation research offered above. In addition, there is a long-standing debate in the social sciences about the particular purpose of qualitative methods and how qualitative approaches to the world fit into the larger purposes of social science. Because I find students are often experiencing considerable conflict and uncertainty about these issues, and because those conflicts and uncertainties interfere with their ability to analyze, interpret, and report evaluation findings, I think it is appropriate to begin our discussion of data analysis by attempting to provide a perspective on the purpose of qualitative methods in evaluation research.

## TRUTH

Let me first remove from the shoulders of evaluators the burden of having to generate Truth. I recently had a student who was virtually paralyzed in writing a final report because he wasn't *sure* if the



patterns he thought he had uncovered were *really true*. I suggested to him that he not try to convince himself or others that his findings were really true, but that he do the best job he could in describing the patterns that appeared to him to be present in the data, and that he present those patterns as his *perspective* on the program based on his analysis and interpretation of the data he had collected. Even if *he* believed that what he eventually produced was Truth, any sophisticated program person reading the report would know that what he presented was no more than a perspective on the program, and they would judge that perspective by their own common-sense understandings of the program and the world and use the information according to how it contributed to their own perspectives.

## TRUTH AND POSITIVISM

“What is truth?” Pontius Pilate is recorded as having put this question to Jesus at his trial, and then immediately left the hall of judgment without waiting for an answer. Lancelot Andrews, a seventeenth-century priest, suggested in a moment of reflection in his Sermon on the Resurrection that this was simply one of those rhetorical questions which, unfortunately, is pushed aside by the press of business. Pilate asked his question, Andrews observed, “and then some other matter took him in the head, and so up he rose and went his way before he had his answer.” Before the business of doing evaluation research and analyzing qualitative data presses us to consider more practical issues, let us consider for a moment. What is truth?

Our modern concern with and confusion about truth is nicely captured by that great contemporary observer of human behavior, comedienne Lily Tomlin:

Lady, I do not make up things. That is lies. Lies is not true. But the truth could be made up if you know how. And that's the truth.

Johnson, in his book *Doing Field Research* (1975), discusses extensively the issues of objectivity and truth in social science research. He notes that social and behavioral scientists really have had only two points of view from which to choose on these issues, both derived from positivism: the correspondence theory of positivism, and the coherence theory of positivism.

The correspondence theory of positivism asserts the existence of one absolute, physical-material reality from which there are no variations.



The coherence theory posits an infinite array of social realities made up of combinations of material objects and individual knowing minds. Despite these differences, however, both express a faith in the existence of an objectively factual social world [Johnson, 1975:185].

Johnson goes on to note, however, that there has been a great gap between the ideals of what social science is supposed to produce and the actual practice of social science. The confusion for students arises because they are taught the ideals without being given insight into actual practice.

Those advancing the traditional theories of truth have commonly understood those abstract notions as idealizations. If one is realistic about the conduct of science, one would not anticipate their complete realization in actual practice. In reviewing the traditional methodological writings, one learns that traditional practitioners have ways to manage the gap between the ideals and realities, ways which lie beyond the parameters of the theoretical formulation itself. These include appeals to readers to accept the factual claim of the researchers on the basis of the professional standing of the claimer, the powerful rhetoric of science, the magic of numbers, the moral omniscience of the observer, or his political sentiments [Johnson, 1975:185-186].

Johnson then builds a strong case for why research inevitably involves personal perspective. Reporting on his own research practices and the practices of others, he concludes that research findings are not *just* personal, but they are also not just objective or truthful. The mix will vary from one research problem to another, from one study to another, but always there will be a mix.

Douglas (1976) is another social scientist involved in individual and team field research who has been greatly concerned with the issue of truth. His concern for truth permeates his writings on field research.

The goal of all social research is to discover, understand and communicate truth about human beings in society. Sociologists, anthropologists, political scientists, economists, journalists and others involved in doing social research may differ over whether these truths should be about matters of immediate practical concern to society or whether they should be more abstract truths that can become useful only over the long run. But there has never been much argument over whether our goal is truth about humans in society. There has, however, been considerable and growing argument over the nature of truth and how we know what is truth [Douglas, 1976:1].



Douglas discusses, and rejects, theories of “absolute truth” or “the theory of absolute objectivity.” He contrasts this view to the emergence of absolute relativism among logical positivists who succeeded in divorcing questions of truth from questions of method. He suggests a new approach to truth:

Social research methods must always be constructed in accord with the basic ideas of truth and the basic goal of achieving truth in this kind of social world. Our analysis of the basic ideas of truth and of the problems of achieving truth in this society lead us to see truth about the social world as being highly problematic and partially dependent upon the goals and methods of research. We eliminate the idea of absolute truth and substitute a more problematic, multiperspectival conception of truth; we eliminate the idea of absolute methods, substituting a multiperspectival conception of methods which argues that our choice of methods must always be made in light of the degree of reliable truth we are seeking and the problems we face in the concrete settings we are studying. We shall see that this method makes the researcher, the live and socially situated individual, the ultimate “measure of all things” [Douglas, 1976:3-4].

Douglas continues to talk about truth throughout his book, but his truth eventually becomes something like the ability of the individual to use trusted evidence in trying to make sense of the world. “Truth” becomes sufficiently watered down that it ends up being the same thing as perspective.

### TRUE AND USEFUL INFORMATION: EMPHASIS ON THE USEFUL

Pelto and Pelto (1978), in their large volume on research methods in anthropology, begin by asserting that “the essence of research methodology lies in seeking answers to the following basic questions: How can we find ‘true and useful information’ about a particular domain of phenomena in our universe?” (p. 1). No sooner have they introduced the notion of finding “true information” than they qualify it, in a footnote, to the point where the search for truth ceases to be a real part of the process.

We use the expression “true and useful information” in quotation marks in order to indicate that, although we generally assume the presence of a concrete, real world, “the truth” or “the facts” about the real world are always seen and interpreted by means of our observational equipment, our perceptual categories, and our general



theoretical outlook. Hence, we can never establish any final "absolute truth." On the other hand, scientific information varies with regard to its degree of approximation to some postulated absolute truth. In general, though, the truth value of our information is best measured by criteria of usefulness—in predicting and explaining our experience in the natural world. Criteria of usefulness are derivable both from theoretical domains of science and from people's practical experience and problems [Pelto and Pelto, 1978:1].

Evaluators have certainly not escaped engaging in battles over truth. Gephart (1978), in an essay "On Truth," examined the two dominant perspectives on truth in evaluation which had led to what he called the "what is evaluation?" war." One side of the war consists of those who believe that the logical positivism tradition with its emphasis on truth as being singular and external to the individual is the path to truth; the other side represents the phenomenological tradition which views truth as multiple and internal to individuals. Gephart asked:

What is truth? Is it unitary, objective, outside of the being of individuals? Or, is it multiple and varied, colored by our layers and layers of experience? Or, is it both? The latter "feels" better to me [Gephart, 1978:2].

Smith, in responding to Gephart's question, suggested that evaluators replace their concern for truth with a concern for practical utility. He argued that in order to act in the world we often accept either approximations to truth or even untruths. For example, when one drives from city to city, one acts as if the earth is flat and does not try to calculate the earth's curvature in planning the trip, even though acting as if the earth is flat means acting on an untruth.

Therefore, in our study of evaluation methodology, two criteria replace exact truth as paramount: practical utility and level of certainty.

The level of certainty required to make an adequate judgment under the law differs depending on whether one is considering an administrative hearing, an inquest, or a criminal case. Although it seems obvious that much greater certainty about the nature of things is required when legislators set national and educational policy than when a district superintendent decides whether to continue a local program, the rhetoric in evaluation implies that the same high level of certainty is required of both cases. If we were to first determine the level of certainty desired in a specific case, we could then more easily choose



appropriate methods. Naturalistic descriptions give us greater certainty in our understanding of the nature of an educational process than randomized, controlled experiments do, but less certainty in our knowledge of the strength of a particular effect. . . .

[O]ur first concern should be the practical utility of our knowledge, not its ultimate truthfulness [Smith, 1978:17].

In studying the utilization of evaluation research (Patton, 1978) I found that decision makers and information users did not expect evaluation reports to produce "truth." Nor did they treat evaluation reports as containing "truth" in any fundamental sense. Rather, they viewed evaluation findings as additional information that they could and did combine with other information (political, experiential, other research, colleague opinions, and so on), all of which fed into a slow, evolutionary process of program development. Utilization of evaluation findings is not something that suddenly and concretely occurs at some distinct moment in time when decision makers are enlightened by truth. Rather, utilization is a diffuse and gradual process of reducing decision maker uncertainty within an existing social context. The purpose of evaluation research, then, is to provide relevant and useful information to decision makers, the criteria for usefulness and relevance being negotiated with decision makers and information users during the conceptual phase of the evaluation.

To say that the evaluator provides useful information that constitutes a perspective on the program being evaluated is not to say that the information provided should not be, to the full extent of the evaluator's capabilities, accurate, valid, reliable, and, yes, truthful within the ordinary, common-sense meaning of that term. The difference between truth and perspective is a matter of degree, the important point being that evaluation researchers consider how they will present their own work to others and how they themselves will regard what they do. House summarizes the issue with appropriate modesty for evaluators:

Expecting evaluation to provide compelling and necessary conclusions hopes for more than evaluation can deliver. Especially in a pluralistic society, evaluation cannot produce necessary propositions. But if it cannot produce the necessary, it can provide the credible, the plausible, and the probable. Its results are less than certain but still may be useful [House, 1977:5].



Perhaps the following Sufi story will provide students with some guidance in their deliberations about the difference between truth and perspective.

Mulla Nasrudin was on trial for his life. He was accused of no less a crime than treason. These charges had been brought by the sages who were ministers to the king charged with advising the king on matters of great importance. Nasrudin was charged with going from village to village inciting the people by saying: "The king's wise men do not speak truth. They do not even know what truth is. They are confused."

Nasrudin was brought before the king and the court. "How do you plead, guilty or not guilty?"

"I am both guilty and not guilty," replied Nasrudin.

"What, then, is your defense?"

Nasrudin turned and pointed to the nine wise men who were assembled in the court. "Have each Sage write an answer to the following question: 'What is water?'"

The king commanded the Sages to do as they were asked. The answers were handed to the king who read to the court what each Sage had written.

The first wrote: "Water is to remove thirst."

The second: "It is the essence of life."

The third: "Rain."

The fourth: "A clear, liquid substance."

The fifth: "A compound of hydrogen and oxygen."

The sixth: "Water was given to us by God to use in cleansing and purifying ourselves before prayer."

The seventh: "It is many different things—rivers, wells, ice, lakes, so it depends."

The eighth: "A marvelous mystery that defies definition."

The ninth: "The poor man's wine."

Nasrudin turned to the court and the king, "I am guilty of saying that the wise men are confused. I am not, however, guilty of treason because, as you see, the wise men are confused. How can they know if I have committed treason if they cannot even decide what water is? If the Sages cannot agree on the truth about water, something which they consume every day, how can one expect that they can know the truth about other things?"

The king ordered that Nasrudin be set free.

## THEORY

I once met with a principal investigator for a major national project which included a fieldwork component. He began our conversation, in response to my question about how things were going, by saying:



“The data analysis has come together really nicely, and we’ve got good descriptive information about the major patterns of activity in the program. Our problem is that we haven’t found a theory to pull it all together yet.”

Concern with the development and verification of theory seems to capture the hopes, ideals, and pretensions of social science. The issue concerns us here because so much of the writing on qualitative methods includes lengthy exhortations about the importance of theory. Denzin, one of the major writers on qualitative methods, asserts,

Research methods are of little use until they are seen in the light of theoretical perspectives. Substantive specialty is of little use or interest until it is firmly embedded within a theoretical framework and grounded upon sound research strategies. . . . What is needed is a common theoretical framework that can be consistently applied to all phases of the sociological act [Denzin, 1978:4].

Denzin goes on to suggest that the theoretical perspective of symbolic interactionism can provide such an overarching framework.

Bruyn’s classic work on “The Methodology of Participant Observation” (1963) is concerned throughout with the linkage between methods and theory. Douglas (1976) discusses at length the relationship between field methods and one’s adoption of either conflict theory or consensus theory in social science as a means of understanding the world. Pelto and Pelto (1978), in their review of anthropological methods, argue that “every piece of research has some relationship to theory, and everyone develops some explicit or implicit strategy or set of strategies that link day-to-day research activities to broad theoretical frameworks” (p. 251). They argue that it is impossible to conduct research without linkage to some theoretical system that suggests main features, principles of causality, and favored modes of observation. One of the most influential books on qualitative research is the Glaser and Strauss volume, *The Discovery of Grounded Theory* (1967), which is concerned entirely with the relationship between qualitative data and theory generation and verification. Where, then, does all this concern with theory leave evaluation researchers who use qualitative methods?

The feelings of many evaluation researchers about theory was expressed by Voltaire’s character Martin in *Candide*: “Let us work without theorizing, ‘tis the only way to make life endurable.”

Because evaluation researchers often ignore theoretical issues altogether, evaluators are accused of being technicians who simply



collect data without regard to the theoretical relevance or causal models suggested by their data. Such an approach is often an appropriate one, because in many cases what decision makers need and want is specific data relevant to narrow, technical issues that are helpful in monitoring or fine-tuning program operations. At other times they simply need good, solid descriptions of program activities and effects.

However, *evaluation research is by no means inherently non-theoretical*. To venture into the arena of causality is to undertake the task of theory generation and verification. The most common causal question in evaluation research is: Does the implemented program lead to the desired outcome? Or, What is the relationship between program activities and observed effects? Do the processes, activities, and treatments of the program cause or affect the behaviors, attitudes, skills, knowledge, and/or feelings of program participants?

My own view is that evaluation researchers need not take on the responsibility for either generating or verifying some broad theory of human behavior. Such theories, as presently developed in the social sciences, are generally abstract, jargon-laden, and esoteric. This does not mean, however, that evaluation researchers ought not become involved in the more concrete enterprise of thinking about what leads to what or what causes what in the programs they observed. In qualitative evaluation research theory construction is inductive, pragmatic, and highly concrete. Theoretical concerns are identified in and derived from the conceptual phase of the evaluation when decision makers and information users work with the evaluator to focus relevant evaluation questions. Given this perspective, there are three primary ways in which concern with theory becomes important in qualitative analysis.

## THEORETICAL PREDISPOSITIONS

First, evaluation researchers have a responsibility to reflect on, bring into consciousness, and make explicit whatever theoretical predispositions they may have with regard to the focus of a particular study. Such theoretical predispositions may concern certain substantive positions that the evaluator holds; for example, a belief that "labeling theory" best explains delinquency, a commitment to behavior modification as the best type of therapy for criminal offenders, or the belief that alcoholism is largely genetic in origin. During the writing of this book I was asked to intervene in an evaluation situation where midway through data collection the program staff of a short-term intervention program for learning-



disabled children found out that the evaluator had published articles and taught his graduate students that learning disabilities would only be remedied through long-term intervention. The evaluator's predisposition undermined his credibility with program staff.

Theoretical predispositions do not always involve commitment to specific substantive positions. The kinds of predispositions described in the previous paragraph are *relatively* easy to make explicit. The more insidious type of theoretical predisposition is the global or macro-level theory that constitutes a basic paradigm organizing the researcher's world view. The functionalist perspective in anthropology, the Marxian view in sociology, and behaviorism in psychology are paradigms of this order. The insidious nature of these theoretical perspectives depends on the extent to which a group of evaluation researchers have been so thoroughly socialized into a particular paradigmatic view that they are no longer aware that their perception has been shaped by theoretical blinders. Thomas Kuhn (1970) has described just how second-nature these perspectives can become.

Scientists work from models acquired through education and through subsequent exposure to the literature often without quite knowing or needing to know what characteristics have given these models the status of community paradigms. . . . That scientists do not usually ask or debate what makes a particular problem or solution legitimate tempts us to suppose that, at least intuitively, they know the answer. But it may only indicate that neither the question nor the answer is felt to be relevant to their research. Paradigms may be prior to, more binding, and more complete than any set of rules for research that could be unequivocally abstracted from them [Kuhn, 1970:46].

Douglas (1976) has argued that nearly everything one does in fieldwork can be affected by a basic predisposition to hold either a consensus or conflict view of society; fieldworkers trained in a functionalist/consensus tradition will approach their observations and interviewing quite differently from methodologists of a more conflict-oriented persuasion.

Evaluators have a responsibility to study themselves, to examine their own paradigmatic and theoretical predispositions, and to make those predispositions explicit. This will allow them to consider the extent to which their observations and analyses have been distorted by conscious or unconscious predispositions. Furthermore, by making such predispositions explicit, decision makers and information users can judge for themselves the extent to which some



subtle bias on the part of the evaluator has intruded in the data analysis.

## SPECULATIONS ON CAUSALITY

The second way in which a concern for theory can become important in qualitative analysis is in considering the relationships between program processes and observed outcomes, or other possible causal relationships that may help explain patterns in the data collected. Speculations on causal relationships are entirely appropriate—as long as they are clearly labeled as speculative.

Lofland (1971) is helpful in clarifying the role of causal speculation in qualitative analysis. He argues that the strong suit of the qualitative researcher is the ability “to provide an orderly description of rich, descriptive detail” (p. 59); the consideration of causes and consequences using qualitative data should be a “tentative, qualified, and subsidiary task” (p. 62). His modest proposal for thinking about the role of theory in qualitative methods primarily concerns how one talks about theoretical and causal findings in reporting the results of qualitative analysis.

It is perfectly appropriate that one be curious about causes, so long as one recognizes that whatever account or explanations he develops is *conjecture*. In more legitimacy-conferring terms, such conjectures are called *hypotheses* or *theories*. It is proper to devote a portion of one’s report to conjectured causes of variations so long as one clearly labels his conjectures, hypotheses or theories as being that [Lofland, 1971:62].

Such conjectures or speculations should not, however, be derived simply from theoretical predispositions of the type discussed earlier. *The cardinal principle of qualitative analysis is that causal relationships and theoretical statements be clearly emergent from and grounded in the phenomena studied. The theory emerges from the data; it is not imposed on the data.*

## DECISION MAKERS’ THEORIES OF ACTION

The third way in which a concern for theory can become important in qualitative analysis is in considering how to help decision makers and information users test *their own theories* about how programs operate. The causal model to be tested is the causal model on which program activities are based. Qualitative data can help decision makers and information users examine the degree to which their own



implementation ideals and program activities actually achieve desired outcomes through programmatic operation. Thus, both descriptively and analytically, qualitative data can be used to help evaluate a program's theory of action. Ways of helping decision makers and information users test their own programmatic theories of action are discussed in *Utilization-Focused Evaluation* (Patton, 1978:179-198).

This section has argued that evaluation research is by no means inherently nontheoretical. Indeed, theoretical issues can be important in three ways. First, evaluators have a responsibility to make as explicit as possible their own theoretical predispositions and to examine how those predispositions may have affected their observations and analyses. Second, evaluators doing qualitative analysis have a responsibility to report and explain whatever causes and consequences emerge during data analysis, clearly recognizing and stating that such theoretical linkages are speculative, and taking care to ground theoretical propositions in the empirical world, thus letting the theory emerge from the data. Third, evaluators can use qualitative data to help decision makers and information users reality-test their own theories of action about the linkages between program processes and program outcomes.

## GENERALIZATIONS

One final issue is left to be discussed before we get down to the basics, the nuts and bolts of data analysis, in the next chapter. To what extent is it the evaluator's responsibility to make generalizations from qualitative data? Generalizations are closely related to theory, the difference being that theory specifies the relationship among a set of variables while generalizations concern the extent to which whatever relationships are uncovered in a particular situation can be expected to hold true for every situation. The concern for generalizations again arises out of traditional social science emphases. The purpose of social science is generally held to be the making of empirical generalizations; social scientists are seldom interested in a particular situation, or particular study, for its intrinsic value. Their object is to study particular populations in order to generalize, and sampling procedures typically emphasize the importance of attention to "external validity" (Campbell and Stanley, 1966). In their major text, *Evaluation: A Systematic Approach*, Rossi et al. (1979) note the importance of deciding the extent to which evaluation findings



can be generalized (pp. 288-290). But what about the specific case of making generalizations from qualitative evaluation data?

Cronbach (1975), one of the major figures in educational measurement and evaluation, has given considerable attention to the issues of theory construction and generalizations. After reviewing 20-years of educational research, Cronbach concluded that social phenomena are too variable and too context-bound to lend themselves to generalization. He places particular emphasis on the importance of interpreting data in context rather than reducing the context to arrive at generalizations. Local conditions become primary. In interpreting local conditions and patterns he emphasizes providing information and developing concepts that will "help people use their heads instead of constructing generalizations and building theory."

Cronbach (1975) has also looked at generalizations outside of educational research, generalizations in natural sciences as well as the behavioral and social sciences. His conclusion:

Generalizations decay. At one time a conclusion describes the existing situation well, at a later time it accounts for rather little variance, and ultimately is valid only as history [Cronbach, 1975:122].

Cronbach's conclusion is not unlike that of Thomas Huxley, who wrote: "History warns us it is the customary fate of new truths to begin as heresies and to end as superstitions." In suggesting that generalizations have not stood up well in the sciences, Cronbach offers an alternative strategy that is excellent advice for the analyst of qualitative data.

Instead of making generalization the ruling consideration in our research, I suggest that we reverse our priorities. An observer collecting data in a particular situation is in a position to appraise a practice or proposition in that setting, observing effects in context. In trying to describe and account for what happened, he will give attention to whatever variables were controlled, but he will give equally careful attention to uncontrolled conditions, to personal characteristics, and to events that occurred during treatment and measurement. As he goes from situation to situation, his first task is to describe and interpret the effect anew in each locale, perhaps taking into account factors unique to that locale or series of events. . . . When we give proper weight to local conditions, any generalization is a working hypothesis, not a conclusion [Cronbach, 1975:124-125].

Evaluator Robert E. Stake (1978) has considered the issue of generalizations in contrast to "particularization," the notion that



knowledge lies in understanding particulars. He quotes William Blake on the subject.

To generalize is to be an idiot. To particularize is the lone distinction of merit. General knowledges are those that idiots possess.

Stake further comments:

Generalization may not be all that despicable, but particularization does deserve praise. To know particulars fleetingly, of course, is to know next to nothing. What becomes useful understanding is a full and thorough knowledge of the particular, recognizing it also in new and foreign contexts.

That knowledge is a form of generalization too, not scientific induction but *naturalistic generalization*, arrived at by recognizing the similarities of objects and issues in and out of context and by sensing the natural covariations of happenings. To generalize this way is to be both intuitive and empirical, and not idiotic [Stake, 1978:6].

Guba (1978) reviewed in depth three basic positions that might be taken in regard to the generalizability of naturalistic inquiry findings.

1. Generalizability is a chimera; it is impossible to generalize in a scientific sense at all. . . .
2. Generalizability continues to be important, and efforts should be made to meet normal scientific criteria that pertain to it. . . .
3. Generalizability is a fragile concept whose meaning is ambiguous and whose power is variable [Guba, 1978:68-70].

Having reviewed these three positions Guba proposes a resolution that recognizes the diminished value and changed meaning of generalizations.

The evaluator should do what he can to establish the generalizability of his findings. . . . Often naturalistic inquiry can establish at least the "limiting cases" relevant to a given situation. But in the spirit of naturalistic inquiry he should regard each possible generalization only as a working hypothesis, to be tested again in the next encounter and again in the encounter after that. *For the naturalistic inquiry evaluator, premature closure is a cardinal sin, and tolerance of ambiguity a virtue* [Guba, 1978:70; italics added].

House, in *The Logic of Evaluative Argument* (1977), also considers with care the question of the generalizability of evaluation



research findings. He approaches the issue in terms of audiences more than in terms of settings.

In evaluation, the social and psychological contexts become particularly relevant and the knowledge less certain. Under those conditions argumentation aimed at gaining the adherence and increasing the understanding of particular audiences, is more appropriate. Persuasion claims validity only for particular audiences and the intensity with which particular audiences accept the evaluation findings is a measure of this effectiveness. The evaluator does not aim at convincing a universal audience of all rational men with the necessity of his conclusions.

Persuasion is directly related to action. Even though evaluation information is less certain than scientific information addressed to a universal audience, persuasion is effective in promoting action because it focuses on a particular audience and musters information with which this audience is concerned [House, 1977:6].

The insights of House are particularly helpful because he not only places evaluation data in context, but he also places the evaluation report in context. The evaluation findings are most useful with regard to the particular setting from which those findings emerged, and the interpretation of findings is particular to those people who need to and expect to use the information that has been generated by evaluation research. This perspective makes it clear that the purpose of evaluation research is to provide information that is useful, information that permits action, and information that is relevant to the needs of decision makers and information users. Such a perspective combines the concern with truth, theory, and generalization into an overall perspective about how evaluation data will be received.

In summary, evaluation persuades rather than convinces, argues rather than demonstrates, is credible rather than certain, is variably accepted rather than compelling. This does not mean that it is mere oratory or entirely arbitrary. . . . *Once the burden of certainty is lifted, the possibilities for informed action are increased rather than decreased* (House, 1977:6-7; italics added).

## USEFUL INFORMATION FOR ACTION

There is a pragmatic bias running through the preceding section. Throughout I've been concerned with the practical, the concrete, and the achievable. This, to me, means that evaluators using qualitative methods provide perspective rather than truth, empirical assessment



of local decision makers' theories of action rather than generation and verification of universal theories, and context-bound information rather than generalizations. Evaluators can give up the burden of producing unassailable certainties and concentrate on the more immediate task of providing useful information to decision makers and information users. Thus, as the evaluator undertakes analysis and interpretation of qualitative data, it is helpful to keep in mind the admonition of Samuel Johnson: "As gold which he cannot spend will make no man rich, so knowledge which he cannot apply will make no man wise."

## A DOCUMENTOR'S PERSPECTIVE

by Beth Alberty

*The pages which follow provide a practical and concrete perspective on the points made in this chapter about truth, theory, and generalizations in evaluation research. This section is a reflective case study of the struggle experienced by one internal program evaluator in trying to figure out how to provide useful information to program staff from the voluminous qualitative data she collected. Beth was still putting together her reflections on her experiences as an internal evaluator-documentor when I met her at the meeting of the North Dakota Study Group on Evaluation in Grand Forks, North Dakota in February 1979. Later, when she sent me what she had written, I asked her if I could use it in this book to illustrate in concrete terms what it means to provide perspective through qualitative evaluation research. Beth begins by describing what she means by "documentation" and then shares her experiences as a novice in analyzing the data she collected, a process of moving from a mass of documentary material to a unified, holistic documentation. Beth's reflections on how she went about analyzing and interpreting the data she collected also provide an excellent introduction to the next and final chapter on how to analyze and interpret qualitative evaluation data.*

Internal documentation is currently being explored as an alternative to external program evaluation. For example, the Office of Education Teacher Center Program is encouraging the use of documentation by the teacher centers it funds as a way of meeting OE requirements for evaluation of center impact and effectiveness.

Documentation, as the word is commonly used, may refer to "slice of life" recordings in various media or to the marshalling of evidence in support of a position or point of view. We are familiar with "documentary" films; we require lawyers or journalists to "document" their cases. Both meanings contribute to my view of what documentation is, but they are far from describing it fully. Documentation, to my mind, is the interpretive reconstitution of a focal event, setting, project, or other phenomenon, based on



observation and on descriptive records set in the context of guiding purposes and commitments. Observations and records may be of many sorts—participants' observations and records in various forms, diaries, memoranda, minutes, interviews, questionnaires, photographs, film, tape recordings, work samples, and the like. The observations and records are designed to gather multiple points of view about multiple aspects of the phenomenon over time. Depending on audience and purposes, the observations and records can be juxtaposed, reworked, excerpted, and restated in a variety of ways, including by means of participant discussion, that result in analysis and illumination of what is being documented. It is with this restatement that the observations and records constitute a documentation, as distinct from a collection of documentary material.

I have always been a staff member of the situations I have documented, rather than a consultant or an employee of an evaluation organization. At first this was by accident, but now it is by conviction: My experience urges that the most meaningful evaluation in terms of a program's goals and commitments is one that is planned and carried out by the staff and that such an evaluation contributes to the program as well as to external needs for information. As a staff member, I participate in staff meetings and contribute to decisions. My relationships with other staff members are close and reciprocal. Sometimes I provide services or perform functions that directly fulfill the purposes of the program—for example, working with children or adults, answering visitors' questions, writing proposals and reports. Most of my time, however, is spent planning, collecting, reporting, and analyzing documentation. I should add that my positions have all been part-time; my experience suggests that this arrangement correctly acknowledges the important part that participants' own observations and record-keeping contribute to a documenting process, as well as the fact that documenting supports action. In any case, through all my engagements in a program, I am aware of a need to maintain as comprehensive a view as possible, certainly one larger than my immediate activity. This view enables me to make connections and find continuity between what I am observing, what has already happened, and what is aimed for.

## FIRST PERCEPTIONS

With these circumstances in mind, let me turn to the beginning plunge. As Carini (1975) points out, observing is the heart of documenting and it was into observing that I plunged, coming up delighted at the apparent ease and swiftness with which I could fish insight and ideas from the ceaseless ocean of activity around me. Indeed, the fact that observing (and record-keeping) does generate questions, insight, and matters for discussion is one of many reasons why records for any documentation should be gathered by those who actually work in the setting.

My observing took many forms, each offering a different way of releasing questions and ideas—interactive and noninteractive observations were



transcribed or discussed with other staff members, and thereby rethought; children's writing was typed out, the attention to every detail involving me in what the child was saying; notes of meetings and other events were rewritten for the record; and so on. Handling such detail with attention, I found, enabled me to see into the incident or piece of work in a way I hadn't on first look. Connections with other things I knew, with other observations I made, or questions I was puzzling over seemed to proliferate during these processes; new perceptions and new questions began to form.

I have heard others describe similarly their delighted discovery of the provocativeness of record-keeping processes. The teacher who begins to collect children's art, without perhaps even having a particular reason for the collecting, will, just by gathering the work together, begin to notice things about them that he or she had not seen before—how one child's work influences another's, how really different (or similar) are the trees they make, and so on. The in-school advisor or resource teacher who reviews all his or her contacts with teachers—as they are recorded or in a special meeting with his or her colleagues—may begin, for example, to see patterns of similar interest in the requests he or she is getting and thus become aware of new possibilities for relationships within the school.

My own delight in this apparently easy access to a first level of insight made me eager to collect more and more, and I also found the sheer bulk of what I could collect satisfying. As I collected more records, however, my enthusiasm gradually changed to alarm and frustration. There were *so many* things that could be observed and recorded, *so many* perspectives, such a *complicated* history! My feelings of wanting *more* changed to a feeling of needing to get *everything*. It wasn't enough for me to know how the program worked now—I felt I needed to know how it got started and how the present workings had evolved. It wasn't enough to know how the central part of the program worked—I felt I had to know about *all* its spinoff activities and from *all* points of view. I was quickly drawn into a fear of losing something significant, something I might need later on. Likewise, in my early observations of class sessions, I sought to write down everything I saw. I have had this experience of wanting to get everything in every setting in which I have documented, and I think it is not unique.

I was fortunate enough to be able to indulge these feelings and to learn from where they led me. It did become clear to me after a while that my early ambitions for documenting everything far exceeded my time and, indeed, the needs of the program. Nevertheless, there was a sense to them. Collecting so much was a way of getting to know a new setting, of orienting myself. And, not knowing the setting, I couldn't know what would turn out to be important in "reconstituting" it; also, the purpose of "reconstituting" it was sufficiently broad to include any number of possibilities from which I had not yet selected. In fact, I found that the first insights, the first connections that came from gathering the records were a significant part of the process of determining what would be important and what were the possibilities most suited to the purposes of the documentation. The process of gathering



everything at first turned out to be important and, I think, needs to be allowed for at the beginning of any documenting effort. Even though much of the material so gathered may remain apparently unused, as it was in my documenting, in fact it has served its purpose just in being collected. A similar process may be required even when the documentor is already familiar with the setting, since the new role entails a new perspective.

The first connections, the first patterns emerging from the accumulating records were thus a valuable aspect of the documenting process. There came a moment, however, when the data I had collected seemed more massive than was justified by any thought I'd had as a result of the collecting. I was ill at ease because the first patterns were still fairly unformed and were not automatically turning into a documentation in the full sense I gave earlier, even though I recognized them as part of the documentary data. Particularly, they did not function as "evaluation." Some further development was needed, but what? "What do I do with them now?" is a cry I have heard regularly since then from teachers and others who have been collecting records for a while.

I began with the relatively simple procedure of rereading everything I had gathered. Then I returned to rethink what my purposes were, and sought out my original resources on documentation. Rereading Carini's monograph and others in the North Dakota series, talking with the staff of The Prospect School and with my staff colleagues, I began to imagine a shape I could give to my records which would make a coherent representation of the program to an outside audience. At the same time I began to rethink how I could make what I had collected more useful to the staff I was working with in its work. Conceiving an audience was very important at this stage.

I will be returning to this moment of transition from initial collecting to rethinking later, to analyze the entry into interpretation that it entails. Descriptively, however, what occurred was that I began to see my observations and records as a body with its own configurations, interrelationships, and possibilities, rather than simply as excerpts of the larger program which related only to the program. Obviously, the observations and records continued to have meaning through their primary relationship to the setting in which they were made; but they also began to have meaning through their secondary relationships to each other.

These secondary relationships also emerge from observation as a process of reflecting. Here, however, the focus of observation is the setting as it appears *in and through* the observations and records that have accumulated, with all their representation of multiple perspectives and longitudinal dimensions. These observations *in and through records*—"thickened observations"—are of course confirmed and added to by continuing direct observation of the setting.

Beginning to see the records as a body and the setting through thickened observation is a process of integrating data. The process occurs gradually and requires a broad base of observation about many aspects of the program over some period of time. It then requires concentrated and systematic



efforts to find connections within the data and weave them into patterns, to notice changes in what is reported, and find the relationship of changes to what remains constant. As Carini has described, this process is supported by juxtaposing the observations and records in various ways, as well as by continual return to reobserve the original phenomenon. There is, in my opinion, no way to speed up the process of documenting. Reflectiveness takes time.

In retrospect I can identify my own approach to an integration of the data as the time when I began to give my opinions on long-range decisions and interpretations of daily events with the ease of any other staff member. Up to the moment of transition, I shared specific observations from the records and talked them over as a way of gathering yet more perspectives on what was happening. I was aware, however, that my opinions or interpretations were still personal. They did not yet represent the material I was collecting.

Thus, it may be that integration of the documentary material becomes apparent when the documentor begins to evince a broad perspective about what is being documented, a perspective that makes what has been gathered available to others without precluding their own perceptions. This perspective is not a fixed-point view of a finished picture, both the view and the picture constructed somehow by the documentor in private and then unveiled with a flourish. It is also not a personal opinion; nor does it arise from placing a predetermined interpretive structure or standard on the observations. The perspective results from the documentor's own current best integration of the many aspects of the phenomenon, of the teachers' or staff's aims, ideas, and current struggles, and of their historical development as these have been conveyed in the actions that have been observed and the records that have been collected.

As documentor, my perspective of a program or a classroom is like my perspective of a landscape. The longer I am in it, the sharper defined become its features, its hills and valleys, forests and fields, and the folds of distance; the more colorful and yet deeply shaded and nuanced in tone it appears; the more my memory of how it looks in other weather, under other skies, and in other seasons, and my knowledge of its living parts, its minute detail, and its history deepen my viewing and valuing of it at any moment. This landscape has constancy in its basic configurations, but is also always changing as circumstances move it and as my perceptions gather. The perspective the documentor offers to others must evoke the constancy, coherence, and integrity of the landscape, and its possibilities for changing its appearance. Without such a perspective, an organization or integration that is both personal and informed by all that has been gathered by myself and by others in the setting—others could not share what I have seen—could not locate familiar landmarks and reflect on them as they exhibit new relationships to one another and to less familiar aspects. All that material, all those observations and records, would be a lifeless and undoubtedly dusty pile.

The process of forming a perspective in which the data gathered are integrated into an organic configuration is obviously a process of inter-



pretation. I had begun documenting, however, without an articulated framework for interpretation or a format for representation of the body of records, like the theoretical framework a researcher brings to his or her data and the traditional format for reporting his or her work. Of course, there was a framework. Conceptions of artistic process, of learning and development, were inherent in G.A.M.E.'s work; but these were not explicit in its goals as a program to provide certain kinds of service. The plan of the documentation had called for certain results, but there was no specified format for presentation of results. Therefore, my entry into interpretation became a struggle with myself over what I was supposed to be doing. It was a long internal debate about my responsibilities and commitments.

When I began documenting at G.A.M.E., for example, I had priorities based on my experience and personal commitments. It seemed to me self-evidently important to provide art activities for children and to try and connect these to other areas of their learning. I knew that art was not something that could be "learned" or even experienced on a once-a-week basis, so I thought it was important to help teachers find various ways of integrating art and other activities into their classrooms. In coming to G.A.M.E. I had already made a personal estimate that what I was documenting was worthwhile and honest. I had found enough points of congruence between my priorities and the program. I could see how the various structures of the program specified ways of approaching the goals that seemed possible and that also enabled the elaboration of the goals.

This initial commitment was diffuse; I felt a kind of general enthusiasm and interest for the efforts I observed and a desire to explore and be helpful to the teachers. In retrospect, however, the commitment was sufficiently energizing to sustain me through the early phases of collecting observations and records, when I was not sure what these would lead to. Rather than restricting me, the commitment freed me to look openly at *everything* (as reflected in the early enthusiasm for collecting everything). Obviously, it is possible to begin documenting from many other positions of relative interest and investment, but I suspect that even if there is no particular involvement in program content on the part of the documentor, there must be at least some idea of being *helpful* to its staff. Otherwise, for example, the process of gathering data may be circumscribed.

At the point of beginning to "do something" with the observations and records, I was forced to specify the original commitment, to rethink my purposes and goals. Rereading the observations and records as a preliminary step in reworking to address different audiences, I found myself at first reading with an idea of "balancing" success and failure, an idea that constricted and trivialized the work I had observed and recorded. Thankfully, it was immediately evident from the data itself that such balance was not possible. If, during ten days of observation, a child's experience was intense one day and characterized by rowdy socializing the other nine, a simple weigh-off would not establish the success or failure of the child's experience. The idea was ludicrous. Similarly, the staff might be thorough in

---



its planning and follow-through on one day and disorganized on another day, but organization and planning were clearly not the totality of the experience for children.

Such weigh-offs implied an external, stereotyped audience awaiting some kind of quantitative proof, which I was supposed to provide in a disinterested way, like a traditional evaluator. The “balanced view” phase was also like my early record-gathering of everything. What I was documenting was still in fragments for me, and my approach was to the particulars, to every detail.

A second approach to interpreting, also brief, took a slightly broader view of the data, a view which acknowledged my original estimate of program value and attempted to specify it. Perceiving through the data the landscape-like configurations of program strength, I made assessments that included statements of past mistakes or inadequacies like minor “flaws” in the landscape (a few odd billboards and a garbage dump in one of Poussin’s dreams of classical Italy, for example) rather than debits on a balance sheet. Here again, the implication was of an external audience, expecting some absolute of accomplishment. The “flaws” could be “minor” only by reference to an implied *major* flaw—that of failing to carry out the program goals altogether.

The formulation of strength subsuming weakness could not withstand the vitality of the records I was reading. The reality the data portrayed became clearer as the inadequacy of my first formulations of how to interpret the documentary material was revealed. Similarly, the implications of external audience expectations were not justified by the actuality of my relationship to the program and staff. At G.A.M.E. my stated goal as documentor had been originally to set up record-keeping procedures that would preserve and make available to staff and to other interested persons aspects of the beginnings and workings of the program, and to collect and analyze some of the material as an assessment of what further possibilities for development actually existed. My goals had not been to evaluate in the sense of an external judgement of success or failure.

Thinking over what other approaches to interpretation were possible, I recalled that I had gathered documentary materials quite straightforwardly as a participant, whose engagement was initially through recognition of shared convictions and points of congruence with the program. Perhaps, I decided, I could share my viewpoint of the observations just as straightforwardly, as a participant with a particular point of view. In examining this possibility, I came to a view of interpreting observational data as a process of “rendering,” much as a piece of classical music is rendered by a performer. The interpretation follows a text closely—as a scientist might say, it sticks closely to the facts. But it also reflects the performer, specifically the performer’s particular manner of engagement in the enterprise shared by text and performer, the enterprise of music. The same relationship could exist, it seemed to me, between a body of observations and records gathered participatively and as documentor. The relationship would allow my personal experience and



viewpoint to enhance rather than distort the data. Indeed, I would become their voice. As Liam Hudson put it:

The interpenetration of observer and observed will then be seen not as an accident, peripheral to the social and behavioral sciences, but as a medium through which they work [quoted in Watkins, 1977:90].

Through this relationship I could make the observations available to staff and to other audiences in a way that was flexible and responsive to *their* needs, purposes, and standards. In so doing, of course, the framework of inherent conceptions underlying the work of the program would be incorporated. Thus, to interpret the observational data I had gathered, I had to reaffirm and clarify my relationship, my attachment to and participation in the program.

My initial engagement, with its strong coloring of prior interests and ideas, had never meant that I understood or was sympathetic with every goal or practice of every participant of the program all the time. In any joint enterprise, such as a school or program, there are diverse and multiple goals and practices. Part of the task of documenting is to describe and make these various understandings, points of view, and practices visible so that they can be reflectively considered by participants as the basis for planning. No participant agrees on all issues and points of practice. Part of being a participant is exploring differences and how these illuminate issues or contribute to practice. My participation allowed me to examine and extend the interests and ideas I came with as well as observing and recording those other people brought. In this process my engagement was deepened, enabling me to make assessments closer to the data than my first readings brought. These assessments are evaluation in its original sense of “drawing-value-from,” an interactive process of valuing, of giving weight and meaning.

In the context of renewed engagement and deepened participation, assessments of mistakes or inadequacies are construed as discrepancies between a particular practice and the intent behind it, between immediate and long-range purposes. The discrepancy is not a flaw in an otherwise perfect surface, but—like the discrepancy in a child’s understanding that stimulates new learning—is the occasion for growth. It is a sign of life and possibility. The burden of the discrepancy can lie either with the practice or with the intent, and that is the point for further examination. Assessment can also occur through the observation of and search for underlying themes of continuity between present and past intent and practice, and the points of change or transformation in continuity. Whereas discrepancy will usually be a more immediate trigger for evaluation, occasions for the consideration of continuity may tend to be longer-range—planning for the coming year, contemplating changes in staff and function, or commemorating an anniversary.



I have located the documentor as participant, internal to the program or setting, gathering and shaping data in ways that make them available to participants and potentially to an external audience. Returning to the image of a landscape, let me comment on the different forms availability assumes for these different audiences.

Participant access to the landscape through the documentor's perspective cannot be achieved through ponderous written descriptions and reports on what has been observed, but must be concentrated in interaction. Sometimes this may require the development of special or regular structures—a series of short-term meetings on a particular issue or problem; an occasional event that sums up and looks ahead; a regular meeting for another kind of planning. But many times the need is addressed in very slight forms, such as a comment in passing about something a child or adult user is doing, or about the appearance of a display, or the recounting of another staff member's observation. I do not mean that injecting documentation into the self-assessment process is a juggling act or some feat of manipulation; merely that the documentor must be aware that his or her role is to keep things open and that, while the observations and records are a resource for doing this, a sense of the whole they create is also essential. The landscape is, of course, changed by the new observations offered by fellow viewers.

The external audience places different requirements on the documentor who seeks to represent to it the documentary perspective. By external audience I refer to funding agencies, supervisors and school boards, institutional hierarchies, and researchers. Proposals, accounts, and reports to these audiences are generally required. They can be burdensome because they may not be organically related to the process of internal self-reflection and because the external audience has its own standards, purposes, and questions; it is unfamiliar with the setting and with the documentor, and it needs the time offered by written accounts to return and review the material. The external audience will need more history and formal description of the broad aspects than the internal audience, with commentary that indicates the significance of recent developments. This need can be met in the overall organization, arrangement, and introduction of documents which also convey the detail and vividness of daily activity.

To limit the report to conventional format and expectations would probably misrepresent the quality of thought, of relating, of self-assessment that goes into developing the work. If there is intent to use the occasion of a report for reflection—for example, by including staff in the development of the report—the reporting process can become meaningful internally while fulfilling the legitimate external demands for accounting. Naturally, such a comment engages the external audience in its own evaluative reflections *by evoking the phenomenon rather than reducing it*.

In closing, I return to what I see as the necessary engaged participation of the documentor in the setting being documented, not only for data-gathering but for interpretation. Whatever authenticity and power my perspective as documentor has had has come, I believe, from my commitment to the



development of the setting I was documenting and from the opportunities in it for me to pursue my own understanding, to assess and reassess my role, and to come to terms with issues as they arose.

We come to new settings with prior knowledge, experience, and ways of understanding, and our new perceptions and understandings build on these. We do not simply look at things as if we had never seen anything like them before. When we look at a cluster of light and dark greens with interstices of blue and some of deeper browns and purples, what we identify is a tree against the sky. Similarly, in a classroom, we do not think twice when we see, for example, a child scratching his head, yet the same phenomenon might be more strictly described as a particular combination of forms and movements. Our daily functioning depends on this kind of apparently obvious and mundane interpretation of the world. These interpretations are not simply personal opinion—though they certainly may be unique—nor are they made up. They are instead organizations of our perceptions as “tree” or “child scratching” and they correspond at many points with the phenomena so described.

It is these organizations of perception that convey to someone else what we have seen and that make objects available for discussion and reflection. Such organizations need not exclude our awareness that the tree is also a cluster of colors or that the child scratching his head is also a small human form raising its hand in a particular way. Indeed, we know that there could be many other ways to describe the same phenomena, including some that would be completely numerical—but not necessarily more accurate, more truthful, or more useful! After all, we organize our perceptions in the context of immediate purposes and relationships. The organizations must correspond to the context as well as to the phenomenon.

Facts do not organize themselves into concepts and theories just by being looked at; indeed, except within the framework of concepts and theories, there are no scientific facts but only chaos. There is an inescapable *a priori* element in all scientific work. Questions must be asked before answers can be given. The questions are all expressions of our interest in the world; they are at bottom valuations. Valuations are thus necessarily involved already at the stage when we observe facts and carry on theoretical analysis and not only at the stage when we draw political inferences from facts and valuations [Myrdal, 1969:9].

My experience suggests that the situation in documenting is essentially the same as what I have been describing with the tree and the child scratching and what Myrdal describes as the process of scientific research. Documentation is based on observation, which is always an individual response both to the phenomena observed and to the broad purposes of observation. In documentation observation occurs both at the primary level of seeing and recording phenomena and at secondary levels of reobserving the phenomena through a volume of records and directly, at later moments. Since docu-



mentation has as its purpose to offer these observations for reflection and evaluation in such a way as to keep alive and open the potential of the setting, it is essential that observations at both primary and secondary levels be interpreted by those who have made them. The usefulness of the observations to others depends on the documentor's rendering them as finely as he or she is able, with as many points of correspondence to both the phenomena and the context of interpretation as possible. *Such a rendering will be an interpretation that preserves the phenomena and so does not exclude but rather invites other perspectives.*

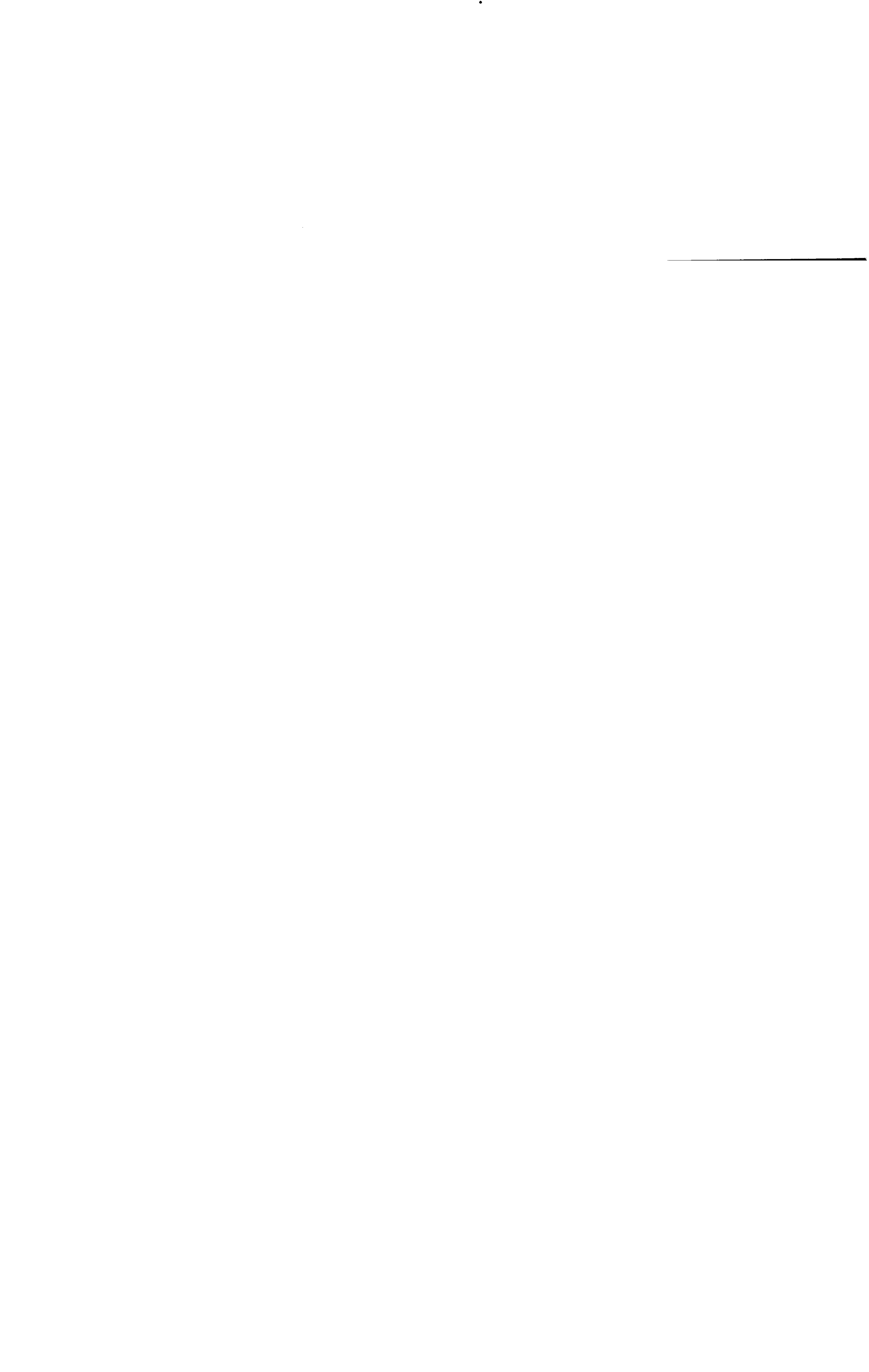
Of course, there is a role for the experienced observer from outside who can see a phenomenon freshly; who can suggest ways of obtaining new kinds of information about it, or, perhaps more importantly, point to the significance of already existing procedures or data; who can advise on technical problems that have arisen within a documentation; and who can even guide efforts to interpret and integrate documentary information. I am stressing, however, that the outside observer in these instances provides *support*, not judgment or the criteria for judgment.

The documentor's obligation to interpret his or her observations and those reflected in the records being collected becomes increasingly urgent, and the interpretations become increasingly significant, as all the observers in the setting become more knowledgeable about it and thus more capable of bringing range and depth to the interpretation. Speaking of the weight of her observations of the Manus over a period of some 40 years of great change, Margaret Mead clarifies the responsibility of the participant-observer to contribute to both the people studied and to a wider audience the rich individual interpretation of his or her own observations:

Uniqueness, now, in a study like this (of people who have come under the continuing influence of contemporary world culture), lies in the relationships between the fieldworker and the material. I still have the responsibility and incentives that come from the fact that because of my long acquaintance with this village I can perceive and record aspects of this people's life that no one else can. But even so, this knowledge has a new edge. This material will be valuable only if I myself can organize it. In traditional fieldwork, another anthropologist familiar with the area can take over one's notes and make them meaningful. But here it is my individual consciousness which provides the ground on which the lives of these people are figures [Mead, 1977:282-3].

In documenting it seems to me the contribution is all the greater, and all the more demanded, because what is studied is one's own setting and commitment.







## QUALITATIVE ANALYSIS AND INTERPRETATION

- The moment you cease observing, pack your bags, and leave the field you will get a remarkably clear insight about that one critical activity you should have observed . . . but didn't.
- The moment you turn off the tape recorder, say goodbye and leave the interview, it will become immediately clear to you what perfect question you should have asked to tie the whole thing together . . . but didn't.
- The moment you begin data analysis it will become perfectly clear to you that you're missing the most important pieces of information, and that without those pieces of information there is absolutely no hope of making any sense out of what you have.
- The complete analysis isn't.
- Evaluation reports finally make clear to decision makers what they had really wanted to know . . . but forgot to ask.

*From Halcolm's Laws of  
Evaluation Research  
à la Murphy*

### FOCUSING THE ANALYSIS

Focus in analyzing qualitative data collected from in-depth interviewing and fieldwork comes from the evaluation questions generated at the very beginning of the evaluation process: during the conceptual, question-focusing phase of the evaluation. So many times evaluators go through painstaking care, even agony, in the process of working with decision makers and information users to clearly conceptualize the evaluation and focus evaluation questions before data collection begins. Then, once the data are collected and analysis begins, they never look back over their notes to review and



renew their clarity on the central issues in the evaluation. It is not enough to count on remembering what the evaluation questions were. The early negotiations around the purpose of an evaluation usually involve important nuances. To reestablish those nuances for the purpose of helping focus the analysis it is important to review notes on decisions that were made during the conceptual part of the evaluation. (This assumes, of course, that the evaluator has treated the conceptual phase of the evaluation as a field experience and has kept detailed notes about the negotiations that went on and the decisions that were made.)

In addition, it is worth reopening discussions with decision makers and information users to make sure that the original focus of the evaluation is still relevant. This accomplishes two things. First, it allows the evaluator to make sure that the analysis will focus on needed information. Second, it prepares decision makers and information users for the evaluation results. At this point in time, the evaluator will also have a much better perspective on what kinds of questions it is possible to answer with the data that have been collected. It is possible to check out which questions take priority and to suggest new possibilities that have emerged as a result of the interviewing and/or fieldwork.

These discussions about the focus of the analysis will often coincide with the initial feedback that may take place in the field as described in Chapter Six. Social scientists have a tendency to want to conceal their results until they have carefully established what those results are and have polished their presentation to a fine point. It is important to remember that utilization of evaluation findings does not usually center on the final report—the final report is part of a total utilization process, but in many cases it is a minor part.

Evaluators who prefer to work diligently in the solitude of their offices until they can spring a final report on a waiting world may find that the world has passed them by. The reason is that evaluation feedback is most useful as part of a process of thinking about a program, rather than as a one-shot information input. Thus, evaluation surprises born of a sudden release of final reports are not likely to be particularly well received, nor are such reports likely to have great impact. Such surprises are more likely to increase rather than decrease uncertainty [Patton, 1978:264].

Sessions devoted to reestablishing the focus of the evaluation analysis and/or providing initial feedback need to be handled with considerable care. The evaluator will need to explain that analysis of



the data is a painstaking process requiring long hours of careful work, going over notes, organizing the data, looking for patterns, checking emergent patterns against the data, cross-validating data sources and findings, and making linkages among the various parts of the data and the emergent dimensions of the analysis. Thus, any discussion of findings at this point is extremely preliminary and can be directed only at the most general issues and the most striking, obvious results. If, in the course of conducting the more detailed and complete analysis of the data, the evaluator finds that statements made or feedback given during a preliminary session were inaccurate, it is important to let the decision makers and information users know about the discrepancy at once.

As noted in Chapter Six and as illustrated in Beth Alberty's reflections in the last chapter, there is typically not a precise point at which data collection ends and analysis begins. In the course of gathering data ideas about the analysis will occur to the people collecting the data. Those ideas constitute the beginning of analysis; They are part of the record of field notes. Whether one is doing in-depth interviewing or observations it is important to keep track of these analytical insights that occur during data collection. This overlapping of data collection and analysis improves both the quality of the data collected and the quality of the analysis so long as the evaluator is careful not to allow these initial interpretations to bias additional data collection. Indeed, instead of focusing additional data collection entirely on confirming initial field hypotheses, the evaluator should become particularly sensitive to looking for alternative explanations and other patterns that would invalidate initial insights. In any case, when data collection has ended and it is time to begin the final analysis, the evaluator has two primary sources to draw from in organizing the analysis: (1) the evaluation questions that were generated during the conceptual phase of the evaluation and clarified with decision makers prior to final analysis, and (2) analytic insights and interpretations that emerged during data collection.

## ORGANIZING THE DATA

The data generated by qualitative methods are voluminous. I have found no way of preparing students for the sheer massive volume of information with which they will find themselves confronted when data collection has ended. Sitting down to make sense out of pages of interviews and whole files of field notes can be overwhelming. Dealing with all those pieces of paper seems like an impossible task.



The first thing to do is make sure it's all there. Are the field notes complete? Are there any parts of detailed field notes that you put off to write later and never got to that need to be finished, even at this late date, before beginning the analysis? Are there any glaring holes in the data that can still be filled by collecting additional data before the analysis begins? Are interview transcriptions complete? Get a sense of the data; check out the quality of the information you have collected.

During the time I was writing this book I was teaching a seminar on evaluation research. Each student in the course was conducting some kind of evaluation, some using quantitative methods and some using qualitative methods. One student had conducted in-depth pre- and postinterviews with participants in a special program. Some thirty people had been interviewed with an hour to two hours of tape-recorded responses from each program participant. The transcription process took several weeks. When the transcriptions were complete she made copies of interviews from three participants and brought them to our seminar for assistance in doing the analysis. As I read the interviews I got a terrible sinking feeling in my stomach. While other students were going over the transcriptions, I pulled her aside and asked her what instructions she had given to the typist, and who the typist was. It was clear to me from reading just a few pages that she did not have verbatim transcriptions. The language in each interview was the same. The sentence structures were the same. The answers were grammatically correct. People in natural conversations simply do not talk that way. The grammar in natural conversations is atrocious. Sentences begin and then are interrupted by new sentences before the first sentences are completed. Without the knowledge of this student, and certainly without her request, the typist had decided to summarize the participants' responses because "so much of what they said was just rambling on and on about nothing." All of the interviews had to be transcribed again before analysis could begin.

Once the interviewer is certain that all the data are there, has checked out the quality of the data, and has filled in any missing gaps, *at least four complete copies should be made of all of the data.* If data collection has gone on over any long period of time, it is wise to make copies of the data as they are collected, being certain to put one copy in a safe place where it will not be disturbed, cannot be lost, and will not be destroyed. In any case, one complete copy of the data should be stored, preferably in a safety deposit box in a bank, for safekeeping. It is no exaggeration to say that these data are priceless. They are unique. The exact observations you have made, the exact

---



words people have spoken in interviews—these can never be recaptured in precisely the same way, even if new observations are undertaken and new interviews are conducted. They are precious material, and they should be treated as such.

Once a copy is put away for safekeeping, there remains one complete copy to use throughout the analysis, one copy for writing on, and one or more copies for cutting and pasting. A great deal of the work of qualitative analysis involves creative cutting and pasting of the data. It is usually best to have more than one copy for this purpose, for under no circumstances should one yield to the temptation to begin cutting and pasting the master copy. The master copy becomes a key resource for locating materials and maintaining the context for the raw data.

Once copies have been made of the data, the formal analysis can begin. The analysis of qualitative data is a creative process. It is also a process of intellectual rigor and a great deal of hard work. Because different people manage their creativity, intellectual endeavors, and hard work in different ways, there is no right way to go about organizing, analyzing, and interpreting qualitative data. Therefore, my description of how I work is not meant to be prescriptive. Each qualitative analyst must find his or her own process. After discussing how I approach getting into the data, I will describe some alternative ways of organizing and reporting qualitative data.

## GETTING STARTED ON CONTENT ANALYSIS

I begin by reading through all of my field notes or interviews and making comments in the margins or even attaching pieces of paper with staples or paper clips that contain my notions about what I can do with the different parts of the data. This is the beginning of organizing the data into topics and files. Coming up with topics is like constructing an index for a book or labels for a file system; look at what is there and give it a name, a label. The copy on which these topics and labels are written becomes the indexed copy of the field notes or interviews. The following are examples of topics used to organize my field notes from the evaluation of the wilderness education program described in the chapter on observation (Chapter Six).

Abbreviation: Part's Reacts Prog

Meaning: (Participants' reactions to the program)



Abbreviation: Parts Reacts Parts

Meaning: (Participants' reactions to other participants)

Abbreviation: Obs Parts Intera

Meaning: (Observations of participants' interactions)

Abbreviation: Obs S Role

Meaning: (Observations of staff role)

Abbreviation: Prog. Ideology

Meaning: (Statements of program ideology)

Abbreviation: Prog. Processes

Meaning: (Examples of program processes)

Abbreviation: Outcomes

Meaning: (Effects of program on participants)

Abbreviation: Sub-G.

Meaning: (Subgroup formations)

Abbreviation: Group Proc.

Meaning: (Group process)

The abbreviated topics are written directly on the relevant data passages, either in the margins or with an attached tab on the relevant page. The full labels in parentheses are the designations for separate files which serve to organize the data as the initial step in cutting and pasting. Many passages can serve several different purposes, patterns, or themes. That is the reason why multiple copies for cutting and pasting are necessary. Several readings of the data are necessary before they can be completely indexed.

In effect, this process of labeling the various kinds of data and establishing a data index is a first step in content analysis. The content of the data is being classified. A classification system is critical; without classification there is chaos. Simplifying the complexity of reality into some manageable classification scheme is the first step of analysis. As George Sand said in her *Nouvelles Lettres d'un Voyageur* (1869): "Classification is Ariadne's clue through the labyrinth of nature."

Where more than one person is working on the analysis, or where an evaluator has assistance in conducting the analysis, it is helpful to have more than one person classify the data. Each person codes the data into a classification scheme separately and then the results of the coding are compared and discussed. Important insights can emerge from the different ways in which two people look at the same set of data.



## COMPUTERIZED DATA PROCESSING

Sometimes a more elaborate classification analysis than a simple filing system is possible and desirable. This is particularly true of large projects where there is too much data for a single person to reasonably code; thus, a more formal classification scheme must be developed that can be used by trained coders who assist in organizing the data. In our study of the utilization of evaluation research which was the basis for *Utilization-Focused Evaluation* (Patton, 1978), lengthy interviews were conducted with 60 project officers, evaluators, and federal decision makers. The data were collected by trainees in the evaluation program at the University of Minnesota. Because the data had the potential for serving a number of different purposes in student work as well as in my own work, we decided to develop a comprehensive classification system that would provide easy access to the data by anyone who wanted to use it. Had only one investigator been intending to use the data, such an elaborate classification scheme would not have been necessary. However, to provide access to several students for different purposes, every paragraph in every interview was assigned a number consisting of a combination of the page number of that interview and the paragraph number on that page. Initial examinations of the interviews led to the development of a systematic and comprehensive classification scheme made up of 15 general categories within which were additional subcategories. Each general category and subcategory was given a code number. Every paragraph in every interview was then coded with as many numbers as necessary to describe the contents of that paragraph. One computer card was punched for each idea in each paragraph; the card contained the identification number of the interview, the general classification number of the content code, the subgroup classification number, the page number and paragraph number where the relevant passage was to be found, and a statement of no more than 50 letters describing the actual content of the passage in question. The classification scheme used to code the utilization of evaluation data appears in Appendix 9.1.

Every interview was coded twice by two independent coders. Each individual code, including redundancies, was keypunched and made part of a computer retrieval system that would permit a print-out of all passages on any subject included in the classification scheme, with brief descriptions of the content of those passages. The analyst could then use that content print-out to go directly to the full passages that were desired. In addition, the computer data-processing system



permitted easy cross-classification and cross-comparison of passages for more complex analyses.

Such an elaborate coding system is highly unusual. It is also very expensive and time-consuming to develop and implement such a comprehensive system. However, where data are going to be used by several people, or where data are going to be used over a long period of time, including additions to the data set over time, such a comprehensive and computerized system can be extremely useful and could actually save a great deal of time in the long run.

The more typical content analysis will depend on some physical sorting of the data into files that permit easy access to all the data that are relevant to a particular topic or theme. Once the data are organized in this way it is possible to begin describing, elaborating, and working with the data around each of the major topics or themes.

## **QUALITATIVE DESCRIPTION**

Evaluation reports based on qualitative methods will include a great deal of pure description of the program and the experiences of people in the program. The purpose of this description is to let the reader know what happened in the program, what it was like from the participants' point of view to be in the program, and what particular events or activities in the program were like. In reading through field notes and interviews the evaluator begins to look for those parts of the data that will be polished for presentation as pure description in the evaluation report. What is included by way of description will depend on what question the evaluation is attempting to answer. Often an entire activity will be reported in detail and depth because it represents a typical program experience. These descriptions are written in narrative form to provide a holistic picture of what has happened in the reported activity or event. Such a description is presented in Chapter Two; the activity observation presented in Chapter Two describes a parent education session in an early childhood family education program (see pp. 31-35).

## **CASE STUDIES**

### **GETTING STARTED ON CASE ANALYSIS**

The purpose of classifying qualitative data in preparation for content analysis is to facilitate the search for patterns and themes within a particular setting or across cases. This kind of qualitative

---



analysis is directed at the following kinds of programmatic and evaluation questions: What was the nature of the interactions among participants? What were participants' attitudes toward the program? What role did staff play in the program? What were the major activities in the program? What were the primary program processes?

Certain kinds of evaluation questions, however, are best answered through case analysis. Case analysis involves organizing the data by specific cases which permits in-depth study of these cases. Cases can be individuals, programs, institutions, or groups. The case study approach to qualitative analysis is a specific way of collecting data, organizing data, and analyzing data. The purpose is to gather comprehensive, systematic, and in-depth information about each case of interest. The starting point for case analysis, then, is making sure that the information for each case is as complete as possible.

Case data consists of all the information one has about the case. It includes all the interview data, the observational data, the documentary data, impressions and statements of others about the case, and data over time—in effect, all the information one has accumulated about the particular case or cases in question. These are the raw data for case analysis, and can amount to a large accumulation of information. At the individual level case data can include clinical records, statistical information about the person, background information, life history profiles, and diaries. At the program level case data can include program documents, program reports, interviews with program participants and staff, observations of the program, and program histories.

## THE CONTENT OF CASE STUDIES

Once the case data have been accumulated, the first task in case analysis is to write a *case record*. The case record pulls together and organizes the voluminous case data into a comprehensive, primary resource package. The case record includes all the major information that will be used in doing the case analysis and case study. Information is edited, redundancies are sorted out, parts are fitted together, and the case record is organized for ready access either chronologically and/or topically. The case record must be complete but manageable; it should include all the information needed for subsequent analysis, but it is organized at a level beyond that of the raw case data.

A case record should make no concessions to the reader in terms of interest or communication. It is a condensation of the case data



aspiring to the condition that no interpreter requires to appeal behind it to the data to sustain his interpretation. Of course, this criterion cannot be fully met: some case records will be better than others. The case record of a school attempts a portrayal through the organization of data alone, and a portrayal without theoretical aspirations [Stenhous, 1977:19].

The case record is used to construct a case study. The case study includes the information that would be communicated in the final report; it represents the data presentation in the report. The report may consist of several case studies which are then compared and contrasted, but the basic data of the study is the information provided about the cases. The case study is the descriptive, analytic, interpretive, and evaluative treatment of the more comprehensive descriptive data that is in the case record. Table 9.1 shows this sequence in moving from raw case data to the written case study.

*Table 9.1* The Process of Constructing Case Studies

---

Step one:	<i>Assemble the raw case data.</i> These data consist of all the information collected about the person or program for which a case study is to be written.
Step two:	<i>Construct a case record.</i> This is a condensation of the raw case data organizing, classifying, and editing the raw case data into a manageable and accessible package.
Step three:	<i>Write a case study narrative.</i> The case study is a readable, descriptive picture of a person or program making accessible to the reader all the information necessary to understand that person or program. The case study is presented either chronologically or thematically (sometimes both). The case study presents a holistic portrayal of a person or program.

---

The case study should take the reader into the case situation, a person's life, a group's life, or a program's life. Each case study in a report stands alone, allowing the reader to enter into the situation described on its own terms. At a later point in analysis it is possible to compare and contrast cases, but initially each case must be represented and understood as an idiosyncratic and unique phenomenon. The descriptions of the case should be holistic and comprehensive, given the focus of evaluation, and will include a myriad of dimensions, factors, variables, and categories woven together into an idiographic framework.

An example of a case study is presented in Appendix 9.2 at the end of this chapter; it is a case study of an individual prepared as part of an evaluation of an experience-based career education program. The data that went into the case study of each student in the program were



collected from multiple sources. Indeed, the approach used in gathering data for this case study is a particularly good example of how multiple sources of data can be brought together to construct a comprehensive picture of a particular case—in this instance, a case which illustrates a student's changing involvement in the program and changing attitudes and behaviors over time. The case data for each student in the study included:

- (a) observations of selected students at employer sites three times during the year;
- (b) interviews three times per year with the students' employer-instructors at the time of observation;
- (c) parent interviews once a year;
- (d) in-depth student interviews four times a year;
- (e) informal discussions with program staff; and
- (f) a review of student projects and other documents.

A total of 23 records was identified as secondary sources of data for each student and a set of guide questions was prepared for analyzing and reviewing each source. These records included employer evaluations of students, student products, test scores and staff evaluations [Fehrenbacher et al., 1976:7-8].

Information from all of these sources is brought together to produce a highly readable narrative that can be used by decision makers and information users to better understand what it was like to be in the program. The evaluation staff of the Northwest Regional Educational Laboratory went to great pains to carefully validate the information in the case studies they constructed. Different sources of information were used to cross-validate findings, patterns, and conclusions. Two evaluators reviewed the material in each case study to independently make judgments and interpretations about the content and meaning of the material in the case study. In addition, an external evaluator reviewed the raw data to check out any biases or unwarranted conclusions in the case study. Students were asked to read their own case studies and to comment on the accuracy of fact and interpretation in the study. Finally, to guarantee the readability of the case studies, a newspaper journalist was employed to help organize and edit the final versions of the case studies. Overall data-collection procedures and analytical strategies were reviewed by a national panel in order to eliminate design bias. Such a rigorous approach to the construction of case study narratives increases the confidence readers can have in that narrative and enhances the utility of the narrative by making the program experience accessible and



understandable to decision makers and information users. Both in its content and the process by which it was constructed, the case study example presented at the end of this chapter (Appendix 9.2) is exemplary of how qualitative case study data can be prepared and presented. The same process would apply to case study data at the group or program level. For an excellent example of program-level case studies in evaluation see Alkin et al. (1979).

How one compares and contrasts cases will depend on the purpose of the evaluation. The way in which cases were sampled will have an important bearing on how case studies are used in analysis. Critical cases, extreme cases, typical cases, and varied cases serve different evaluation purposes. (See Chapter Five for a discussion of these different case-sampling strategies and the purposes served by each.)

Once case studies have been organized and written the analytic strategies described in the remainder of this chapter can be used to further analyze and interpret the case study data.

## INDUCTIVE ANALYSIS

Inductive analysis means that the patterns, themes, and categories of analysis come from the data; they emerge out of the data rather than being imposed on them prior to data collection and analysis. The analyst looks for natural variation in the data. For evaluators, the study of natural variation will involve particular attention to variations in program processes and how participants respond to and are affected by programs. Two ways of representing the patterns emerge from analysis of the data. First, the analyst can use the categories developed and articulated in the program studied to organize presentation of particular themes. Second, the analyst may also become aware of categories or patterns for which the people in the program did not have labels or terms, and the analyst develops terms to describe these inductively generated categories. Each of these approaches to analysis is described below.

## INDIGENOUS TYPOLOGIES

Typologies are classification systems made up of categories that divide some aspect of the world into parts. Indigenous typologies have come to be called the “emic” approach to analysis in anthropology.

According to this view, cultural behavior should always be studied and categorized in terms of the inside view—the actors’ definition—of



human events. That is, the units of conceptualization in anthropological theories should be “discovered” by analyzing the cognitive processes of the people studied, rather than “imposed” from cross-cultural (hence, ethnocentric) classifications of behavior [Pelto and Pelto, 1978:54].

There is a strong emphasis coming out of the cross-cultural work of anthropologists on the importance of preserving and reporting the indigenous categories of people studied. Franz Boas was a major influence in this direction: “If it is our serious purpose to understand the thoughts of a people the whole analysis of experience must be based on their concepts, not ours” (Boas, 1943:314).

This kind of approach requires an analysis of the verbal categories used by participants and/or staff in a program to break up the complexity of reality into parts. It is a fundamental purpose of language to tell us what is important by giving it a name and therefore separating it from other things with other names. Once these labels have been identified from an analysis of what people in the program have said, the next step is to identify the attributes or characteristics that distinguish one thing from another. In describing this kind of analysis Charles Frake (1962) uses the example of a hamburger. Hamburgers can vary a great deal; there are many ways to prepare them or add to them, and they are still called hamburgers. However, when a piece of cheese is added to the meat, it is no longer a hamburger, it becomes a cheesburger. The task for the analyst is to discover what it is that separates “hamburger” from “cheesburger.” The purpose of this analysis is to discern and report “how people construe their world of experience from the way they talk about it” (Frake, 1962:74).

An example of this kind of analysis from evaluation comes from work we did in evaluating a program aimed at reducing the dropout rate among high school students. In observations and interviews at the targeted high school it became important to understand the ways in which teachers categorized students. With regard to problems of truancy, absenteeism, tardiness, and skipping class the teachers had come to label students as either “chronics” or “borderlines.” One teacher described the chronics as “the ones who are out there all the time and everything you do to get them in doesn’t work.” Another teacher said, “you can always pick them out, the chronics. They’re usually the same kids.” The borderlines, on the other hand, “skip a few classes, waiting for a response, and when it comes they shape up. They’re not so different from your typical junior high student, but when they see the chronics getting away with it they get more brazen



in their actions.” Another teacher said, “Borderlines are the ones who you see maybe two and three times a day, not constantly like the chronics.”

Not all teachers used precisely the same criteria to distinguish “chronics” from “borderlines,” but all teachers used these labels in talking about students. One of the tasks of qualitative analysis was to portray and understand the teachers’ views about dropouts as represented by this indigenous typology.

It became clear that in understanding the impact of the program on students, and the nature of program activities directed at reducing high school dropouts, it would be important to observe differences in the program between “borderlines” and “chronics.” It was difficult to get the teachers, in many cases, to even attempt to deal with the chronics.

It would have been impossible to understand fully the situations in that program as conceived by the teachers and experienced by the students without understanding this indigenous typology of “chronics” and “borderlines.” Moreover, this typology had important implications for how the program was organized and the extent to which different strategies were to be developed to deal with different kinds of students. These categories, then, became themes which were important throughout the data analysis and the final report.

Another example of an important typology that sheds light on program processes and outcomes was the differentiation among participants that emerged in the wilderness education program I evaluated. During the second year of the project one group of participants formed a subgroup and called themselves the “turtles.” They contrasted themselves to the “truckers.” On the surface, these labels were aimed at distinguishing different styles of hiking and backpacking, one slow and one fast. Beneath the surface, however, the terms came to represent different approaches to the wilderness and different styles of experience in relation to the wilderness and the program.

Every program gives rise to a special vocabulary which staff and participants can use to differentiate types of activities, different kinds of participants, different styles of participation, and different contributions to the program. These indigenous typologies are clues to the evaluator that the phenomena to which the labels refer are important to the people in the setting, and that to fully understand the setting it is necessary to understand those terms and their implications for the program.



## ANALYST-CONSTRUCTED TYPOLOGIES

While program staff and participants will have developed typologies for certain important features of the program, other important patterns may not have given rise to specific linguistic distinctions within the program. The second task of induction, then, is for the analyst to look for patterns, categories and themes for which a typology can be constructed to elucidate variations and contrasts in activities, participants, and/or staff. Lofland points out that such constructions must be done with considerable care to avoid creating things that are not really in the data. One major way of finding out whether or not such analyst-constructed typologies are accurate and useful is to present them to people in the program to find out if the constructions make sense.

At the farthest remove from participant-articulated designations, the analyst assumes the task of *constructing* patterns that appear to exist but remain unconceived in the phenomenology of the participants. It is this latter task of observer construction that is the most hazardous and most subject to the legitimate charge of imposing a world of meaning on the participants that better reflects the observer's world than the world under study. When the observer constructs participation patterns he is more likely to impute participation than to discern it. Nonetheless, it still seems worthwhile to attempt construction. . . . The best and most stringent test of observer constructions is their recognizability to the participants themselves. When participants themselves say, "yes, that is there, I'd simply never noticed it before," the observer can be reasonably confident that he has tapped into extant patterns of participation [Lofland, 1971:34].

A good example of this kind of evaluator-generated typology is to be found in the evaluation of the National Museum of Natural History, Smithsonian Institution, directed by Robert L. Wolfe and Barbara L. Tymetz (1978). They conducted a naturalistic inquiry of the "Ice Age Mammals and Emergence of Man" exhibit at the Museum. From their observations, they identified four different kinds of visitors to the exhibit.

[I]t is appropriate to describe our perceptions of the kind of visitors that we have observed in Hall 6 over the past several months. These descriptions are progressive in that each new category identifies a person more serious about the exhibit hall. We are unsure of how the readers of this report would describe the different kinds of visitors, but nevertheless, we described them as follows:



*The Commuter*

This is the person who merely uses the hall as a vehicle to get from the entry point to the exit point. . . .

*The Nomad*

This is a casual visitor—a person who is wandering through the hall, apparently open to become interested in *something*. The Nomad is not really sure why he or she is in the hall and not really sure that s/he is going to find anything interesting in this particular exhibit hall. Occasionally the Nomad stops, but it does not appear that the nomadic visitor finds any one thing in the hall more interesting than any other thing.

*The Cafeteria Type*

This is the interested visitor who *wants* to get interested in something, and so the entire museum and the hall itself is treated as a cafeteria. Thus, the person walks along, hoping to find something of interest, hoping to “put something on his or her tray” and stopping from time to time in the hall. While it appears that there is something in the hall that spontaneously sparks the person’s interest, we perceive this visitor has a predilection to becoming interested, and the exhibit provides the many things from which to choose.

*The V.I.P.—Very Interested Person*

This visitor comes into the hall with some prior interest in the content area. This person may not have come specifically to the hall, but once there, the hall serves to remind the V.I.P.’s that they were, in fact, interested in something in that hall beforehand. The V.I.P. goes through the hall much more carefully, much slower, much more critically—that is, they move from point to point, they stop, they examine aspects of the hall with a greater degree of scrutiny and care [Wolf and Tymetz, 1978:10-11].

This typology of types of visitors became important in the full evaluation because it permitted analysis of different kinds of museum experiences and different activities undertaken in the museum. Moreover, the authors recommended that when conducting interviews with museum visitors to get their reactions to exhibits that the interview data be differentially valued depending on the type of person who was being interviewed.

The typology above describes characteristics of participants. A different kind of typology was developed to describe different kinds of museum activities. In evaluating what people learn in a museum and how they learn, the evaluators found it important to develop a typology of different kinds of museum experiences. They called these “Museum Encounters of the First, Second, and Third Kind.”



*Museum Encounters of the First Kind:*

This encounter occurs in halls which use display cases as the primary approach to specimen presentation. Essentially, the visitor is a passive observer to the "objects of interest." Interaction is visual and may occur only at the awareness level. The visitor is probably not provoked to think or consider ideas beyond the visual display.

*Museum Encounters of the Second Kind:*

This encounter occurs in halls which employ a variety of approaches to engage the visitor's attention and/or learning. The visitor has several choices to become active in his/her participation. . . . The visitor is likely to perceive, question, compare, hypothesize, etc.

*Museum Encounters of the Third Kind:*

This encounter occurs in halls which invite high levels of visitor participation. Such an encounter invites the visitor to observe phenomena in process, to create, to question the experts, to contribute, etc. Interaction is personalized and within the control of the visitor [Wolf and Tymetz, 1978:39].

The primary purpose of typologies is to *describe*. These typologies can later be used to make interpretations about the nature of the program, but the first purpose is description based on an analysis of the patterns that appear in the data. But how does one recognize patterns in qualitative data and turn those patterns into meaningful categories?

## DEVELOPING CATEGORY SYSTEMS

Guba (1978) asserts that in focusing the analysis of qualitative data an evaluator must deal first with the problem of "convergence." The problem of convergence is figuring out what things fit together. This leads to a classification system for the data.

Guba suggests several steps for converting field notes and observations about issues and concerns into systematic categories of analysis. The evaluator-analyst begins by looking for "recurring regularities" in the data. These regularities represent patterns that can be sorted into categories. Categories should then be judged by two criteria: "internal homogeneity" and "external heterogeneity." The first criterion concerns the extent to which the data that belong in a certain category hold together or "'dovetail' in a meaningful way." The second criterion concerns the extent to which differences among categories are bold and clear. "The existence of a large number of unassignable or overlapping data items is good evidence of some basic fault in the category system" (Guba, 1978:53). The naturalistic evaluator then works back and forth between the data and the



classification system to verify the meaningfulness and accuracy of the categories and the placement of data in categories. When several different classification systems have been developed, some priorities must be established to determine which category systems are more important than others. Prioritizing is done according to the salience, credibility, uniqueness, heuristic value, feasibility, special interests, and materiality of the classification schemes. Finally, the category system or set of categories are tested for completeness.

1. The set should have internal and external plausibility, a property that might be termed "integratability." Viewed internally, the individual categories should appear to be consistent; viewed externally, the set of categories should seem to comprise a whole picture. . . .
2. The set should be reasonably inclusive of the data and information that do exist. This feature is partly tested by the absence of unassignable cases, but can be further tested by reference to the problem which the inquirer is investigating or by the mandate given the evaluator by his client/sponsor. If the set of categories did not appear to be sufficient, on logical grounds, to cover the facets of the problem or mandate, the set is probably incomplete.
3. The set should be reproducible by another competent judge. . . . The second observer ought to be able to verify that a) the categories make sense in view of the data which are available, and b) the data have been appropriately arranged in the category system. . . .

The category system auditor may be called upon to attest that the category system "fits" the data and that the data have been properly "fitted into" it.

4. The set should be credible to the persons who provided the information which the set is presumed to assimilate. . . . Who is in a better position to judge whether the categories appropriately reflect their issues and concerns than the people themselves? [Guba, 1978:56-57].

The second problem discussed by Guba is the problem of "divergence." By this he means that the evaluator must deal with how to "flesh out" the categories. He suggests that this is done by processes of extension (building on items of information already known), bridging (making connections among different items), and surfacing (proposing new information that ought to fit and then verifying its existence). The analyst brings closure to the process when sources of information have been exhausted, when sets of



categories have been saturated so that new sources lead to redundancy, when clear regularities have emerged that feel integrated, and when the analysis begins to “overextend” beyond the boundaries of the issues and concerns guiding the analysis.

The steps and procedures suggested by Guba for analyzing qualitative data are not mechanical or rigid. The process of data analysis is to a major extent “ ‘arty’ and intuitive.”

The task of converting field notes and observations about issues and concerns into systematic categories is a difficult one. No infallible procedure exists for performing it [Guba, 1978:53].

This effort at uncovering patterns, themes, and categories is a creative process that requires making carefully considered judgments about what is really significant and meaningful in the data. Since qualitative analysts do not have statistical tests to tell them when an observation or pattern is significant, they must rely on their own intelligence, experience, and judgment. This sometimes leads to the making of the qualitative analyst’s equivalent of Type I and Type II errors from statistics. The evaluator-analyst may decide that something is not significant when in fact it is; or, conversely, the analyst may attribute significance to something that is meaningless. A story illustrates this problem by making judgments about what is really significant.

Halcolm was approached by a woman who handed him something. Without hesitation Halcolm returned the object to the woman. The many young disciples who followed Halcolm to learn his wisdom began arguing among themselves about the special meaning of this interchange. A variety of interpretations were offered.

When Halcolm heard of the argument among his young followers he called them together and asked each one to report on the significance of what they had observed. When they finished he said, “The real purpose of the exchange was to enable me to show you that you are not yet sufficiently masters of observation to know when you have witnessed a meaningless interaction.”

## LOGICAL ANALYSIS

While working inductively the analyst is looking for emergent patterns in the data. These patterns, as noted in the preceding sections, can be represented as dimensions, categories, classification schemes, and categories. Once some dimensions have been



constructed, either using participant-generated constructions or evaluator-generated constructions, it is sometimes useful to cross-classify different dimensions to generate new insights about how the data can be organized and to look for patterns that may not have been immediately obvious in the initial, inductive analysis. Creating cross-classification matrices is an exercise in logic. This procedure involves creating potential categories by crossing one dimension or typology with another, and then working back and forth between the data and one's logical constructions, filling in the resulting matrix. This logical system will create a new typology all parts of which may or may not actually be represented by the data. Thus, the analyst moves back and forth between the logical construction and the actual data in the ongoing search for understanding through description.

In the high school dropout program described above the focus of activities in the program was aimed at increasing the capabilities of teachers in their efforts to reduce absenteeism, skipping classes, and tardiness. An external team of consultant/change agents worked with teachers in the school to help them develop approaches to the dropout problem. Observations of the program and interviews with the teachers gave rise to two dimensions, one aimed at describing teachers' beliefs about what kind of programmatic intervention was, from their experience, effective with dropouts. The inductive analysis of the data suggested that teachers' behaviors toward dropouts could be conceptualized along a continuum according to the extent to which teachers were willing to take direct responsibility for doing something about the problem. Thus, this dimension varied from taking responsibility to shifting responsibility to others.

The second dimension concerned the teachers' views about effective intervention strategies. The inductive analysis revealed three perspectives among the teachers. Some teachers believed that a rehabilitation effort was needed to help kids with their problems; some teachers preferred a maintenance or caretaking effort aimed at just keeping the school running, maintaining the system; and still other teachers favored finding some way of punishing students for their unacceptable and inappropriate behaviors, no longer letting them get away with the infractions they had been committing in the past. Figure 9.1 shows what happens when these two dimensions are crossed. Six cells are created, each of which represents a different kind of teacher role in response to the program.

The evaluator-analyst working with this data had been struggling in the inductive analysis to find the patterns that would express the different kinds of teacher roles manifested in the program. He had



Taking Responsibility ← → Shifting Responsibility to Others

Teachers' Beliefs About How to Intervene with Dropouts

Rehabilitation



Maintenance (Caretaking)



Punishment

<p><b>Counselor/Friend:</b> Help kids directly</p>	<p><b>Referral Agent:</b> Refer them to other helping agencies</p>
<p><b>Traffic Cop:</b></p> <p>Just keep them moving through the system</p>	<p><b>Ostrich:</b></p> <p>Ignore the situation and hope someone else does something</p>
<p><b>Old Fashioned School Master:</b></p> <p>Make them feel the consequences</p>	<p><b>Complainer:</b></p> <p>Somebody should remove the problem kids</p>

Figure 9.1 An Empirical Typology of Teacher Roles in Dealing with High School Dropouts



tried several constructions, but none of them quite seemed to work—they were not true to the data. When he described to me the other dimensions he had generated, I suggested that he cross them, as shown in Figure 9.1. When he did so he said that “the whole thing immediately fell into place.” Working back and forth between the matrix and the data, he generated a full descriptive analysis.

The description of teacher roles served several purposes. First, it gave teachers a mirror image of their own behaviors and attitudes. It could thus be used to help teachers make more explicit their own roles. Second, it could be used by the external team of consultants to more carefully gear their programmatic efforts toward different kinds of teachers who were acting out the different roles. It is relatively clear that an omnibus strategy for helping teachers establish a program that would reduce dropouts would not work in this school. Teachers manifesting different roles would need to be approached and worked with in different ways. By formally conceptualizing the different kinds of roles, the external team could more clearly conceptualize their program. Third, the description of teacher roles provided insights into the nature of the dropout problem. Having identified the various roles, the evaluator-analyst had a responsibility to report on which kinds of roles were more in evidence in this school and which kinds of roles were represented by only a few or even no teachers.

One must be extremely careful about this kind of logical analysis. It is easy for a matrix to begin to manipulate the data and to force them into categories created by the cross-classification rather than using the matrix as a system of logically generated sensitizing concepts to be tested out by the actual data. Sensitizing concepts provide direction and insight without rigorous imposition of the narrowly operationalized preordinate concepts that are mandatory in quantitative measurement. In the case of matrix analysis, the empty cell of a matrix (the cell created by crossing two dimensions for which no name or label immediately occurs) sensitizes the analyst to the possibility of a category of activity or behavior that has either been overlooked in the data or that is logically a possibility in the setting but has not been manifested. The latter case falls into those types of activities and behaviors which are important to note because they did *not* occur.

## COMMUNICATION THROUGH METAPHORS

The classification systems used and developed in the preceding examples have relied on metaphors as a way of communicating



connative meanings of the various categories. The “commuter”–“nomad”–“cafeteria type”–“V.I.P.” (very interested person) category system developed by Wolf and Tymetz (1978) to describe visitors to the National Museum of Natural History is an example of the use of metaphors to report the patterns in qualitative data. Likewise, their classification of museum environments and activities, Encounters of the First, Second, and Third Kind, refers to the popular science fiction movie, *Close Encounters of the Third Kind*. In the dropout study the teacher role descriptions rely on metaphors when depicting role types as “traffic cop,” “old-fashioned school master” and “the ostrich role.”

Metaphors can be powerful and clever ways of communicating findings. A great deal of meaning can be conveyed in a single phrase with a powerful metaphor. Moreover, developing and using metaphors can be fun, both for the analyst and for the reader. It is important, however, to make sure that the metaphor serves the data and not vice versa. The creative evaluator who finds a powerful metaphor may be tempted to manipulate the data to fit the metaphor. Moreover, because metaphors carry implicit connotations, it is important to make sure that the data fit the most prominent of those connotations so that what is communicated is what the evaluator-analyst wants to communicate. Finally, one must avoid reifying metaphors and acting as if the world were really the way the metaphor suggests it is.

The metaphor is chiefly a tool for revealing special properties of an object or event. Frequently, theorists forget this and make their metaphors a real entity in the empirical world. It is legitimate, for example, to say that a social system is like an organism, but this does not mean that a social system *is* an organism. When metaphors, or concepts, are reified, they lose their explanatory value and become tautologies. A careful line must be followed in the use of metaphors, so that they remain a powerful means of illumination [Denzin, 1978:46].

The analysis presented in the next section does not rely on metaphors. It does, however, rely on the sensitizing concepts of program processes and program outcomes basic to evaluation research. This analysis shows how a matrix can be created by cross-classifying program processes and program outcomes to organize the linkages between process and outcomes in qualitative analysis.



## A PROCESS/OUTCOMES MATRIX

The linkage between processes and outcomes is a fundamental issue in many program evaluations. As noted in Chapter Four, I suggested that an evaluation research design based on qualitative methods might be particularly appropriate under circumstances where either program processes or program impacts, or both, were largely unspecified, for whatever reasons. Sometimes the reason is because outcomes were meant to be individualized; sometimes the program is simply uncertain about what the outcomes will be; and in many programs neither processes nor impacts have been carefully articulated. Under such conditions one purpose of the evaluation may be to help articulate program processes, program impacts, and the linkages between the two. This task can be facilitated by constructing a process/outcomes matrix to organize the data.

Figure 9.2 is an abstraction of how such a matrix can be constructed. Major program processes or identified implementation components are listed along the left side. Types or levels of outcomes are listed across the top. The category systems for program processes and outcomes are developed from the data in the same way that other typologies are constructed (see previous sections). The cross-classification of any process with any outcome produces a cell in the matrix; for example, the first cell in Figure 9.2 is created by the intersection of process 1 with outcome a. The information that goes in cell 1-a (or any other cell in the matrix) describes linkages, patterns, themes, program content, or actual activities that help us understand the relationships between processes and outcomes. Such relationships may have been identified by participants themselves during interviews or discovered by the evaluator in analyzing the data. In either case, the process/outcomes matrix becomes a way of organizing, thinking about, and presenting the qualitative connections between program implementation dimensions and program impacts.

A couple of examples will help make the notion of the process-outcomes matrix more concrete. Suppose we have been evaluating a juvenile justice program which places delinquent youths in foster homes. We have visited several foster homes; observed what the home environments are like; and interviewed the juveniles, the foster home parents, and the probation officers. A regularly recurring theme in the interviews is the importance of the process of "letting kids learn to make their own decisions." A regularly recurring outcomes theme is "keeping the kids straight" (reduced recidivism). The blank cell in a matrix created by crossing the program process ("kids making their own decisions") with the program outcome



	a	b	c	d
1				
2	<b>LINKAGES EXPRESSED AS THEMES</b> <b>PATTERNS, QUOTATIONS, PROGRAM</b>			
3	<b>CONTENT OR ACTUAL</b> <b>ACTIVITIES</b>			
4				

Program Processes  
or  
Implementation Components

Figure 9.2 Matrix of Linkages Between Program Processes and Impacts



(“keeping kids straight”) creates a data analysis question: What actual decisions do juveniles make that are supposed to lead to reduced recidivism? We then carefully review our field notes and interview quotations looking for data that helps us understand how people in the program have answered this question based on their actual behaviors and practices. By describing what decisions juveniles actually make in the program the decision makers to whom our findings are reported can make their own judgments about the strength or weakness of this linkage between this program process and the desired outcome. Moreover, once the descriptive analysis is complete the evaluator is at liberty to offer interpretations and judgments about the nature and quality of this supposed process/outcome connection.

### RECOGNIZING PROCESSES, OUTCOMES, AND LINKAGES IN QUALITATIVE DATA

Because of the centrality of the sensitizing concepts “program process” and “program impact” in evaluation research it may be helpful to provide a more detailed description of how these concepts can be used in qualitative analysis. How does one recognize a program process? Learning to identify and label program processes is a critical evaluation skill. This sensitizing notion of “process” is a way of talking about the common action that cuts across program activities, observed interactions, and program content. The example I shall use to consider these issues is the wilderness education program I evaluated and discussed throughout the observations chapter (Chapter Six). That program, entitled the Southwest Field Training Project, involved using the wilderness as a medium for training professional educators in the philosophy, methods, and uses of experiential education by engaging those educators in their own experiential learning process.

In the Southwest Field Training Project participants go from their normal urban environments into the wilderness for a week; they spend at least one day and night completely alone in some wilderness spot “on solo”; at times, while backpacking, the group is asked to walk silently so as not to be distracted from the wilderness sounds and images by conversation; in group discussions participants are asked to talk about what they have observed about the wilderness and how they feel about being in the wilderness; and participants are asked to write about the wilderness environment in journals that they keep. What do these different activities have in common, and how



can that commonality be expressed? There are many different ways of talking about the underlying process:

- (a) experiencing the wilderness,
- (b) learning about the wilderness,
- (c) appreciating the wilderness,
- (d) immersion in the environment,
- (e) developing awareness of the environment,
- (f) becoming conscious of the wilderness, or
- (g) developing sensitivity to the environment.

Any of these phrases, each of which consists of some verb form (experiencing, learning, developing, and so on) and some noun form (wilderness, environment), captures some nuance of the process. The evaluator-analyst works back and forth between the data (field notes and interviews) and his or her conception of what it is that needs to be expressed to find the most fitting language to describe the process. What language do people in the program use to describe what those activities and experiences have in common? What language comes closest to capturing the essence of this particular process? What level of generality or specificity will be most useful in separating out this particular set of things from other things? How do program participants and staff react to the different terms that could be used to describe the process?

It is not unusual in any particular part of an analysis to go through several different phrases before finally coming up with the exact language that will go into a final report. In the Southwest Field Training Project we began with the phrase “Experiencing the Wilderness” and after several revisions finally described the process as “developing sensitivity to the environment” because the broader label permitted us to include discussions and activities that were aimed at helping participants understand how they were affected by and how they affected their normal institutional environments. Experiencing the wilderness became a specific sub-process that was part of the more global process of developing sensitivity to the environment. Program participants and staff played a major role in determining the final phrasing and description of this process. Other processes identified as important in the implementation of the program were:

- (a) encountering and managing stress;
- (b) sharing in group settings;
- (c) examining professional activities, needs, and commitments;
- (d) assuming responsibility for articulating personal needs;



- (e) exchanging professional ideas and resources; and
- (f) formally monitoring experiences, processes, changes, and impacts.

It is important to keep in mind as one struggles with finding the right language to communicate themes, patterns, and processes that there is no absolutely *right* way of stating what emerges from the analysis. There are only more and less *useful* ways of expressing what the data reveal.

Identifying and conceptualizing program outcomes and impacts can also be either an inductive or logical process. Inductively, the evaluator-analyst looks for changes in participants, expressions of change, program ideology about outcomes and impacts, and ways that people in the program make distinctions between “those who are getting *it*” and “those who aren’t getting *it*.” In highly individualized programs the statements about change that emerge from program participants and staff may be global. Such outcomes as “personal growth,” increased “awareness,” and “insight into self” are difficult to operationalize and standardize. That is precisely the reason qualitative methods may be used to evaluate such outcomes! The task for the evaluator-analyst, then, is to describe what actually happens to people in the program and what they say about what happens to them.

In many qualitative evaluation projects data on outcomes are gathered primarily or entirely through the use of in-depth interviews. Organizing and presenting such data involves the same classification processes discussed throughout this chapter. What must be organized and presented are quotations from interviewees. Sufficient quotive data should be presented to illuminate and support whatever analysis the evaluator provides in narrative form. Appendix 9.3 at the end of this chapter presents portions of a report describing the effects on participants of their experiences in the wilderness education program. The data presented are derived entirely from in-depth, open-ended interviews. For another example of how to organize, present, and interpret interview data see chapters two and four of *Utilization-Focused Evaluation* (Patton, 1978); these chapters discuss the results of interviews with decision makers and evaluators on the utilization of evaluation research in specific projects.

Where data on both program processes and participant outcomes are available, organizing that data may require some kind of logical scheme that permits the linkage of program processes to participant outcomes. One such logically derived scheme was used to organize the data in the Southwest Field Training Project. First, a classifi-



cation scheme that described different types of outcomes was conceptualized:

- (a) changes in knowledge,
- (b) changes in attitudes,
- (c) changes in feelings,
- (d) changes in behaviors, and
- (e) changes in skills.

With this scheme it was possible to organize the data to describe the content and major themes of the program so that the reader would know what kinds of changes were occurring and how those changes were perceived by participants to be related to specific program processes. *I emphasize that the process/outcomes matrix is merely an organizing tool; the data from participants themselves and from field observations provide the actual linkages between processes and outcomes.*

What was the relationship between the program process of “developing sensitivity to the environment” and these individual-level outcomes? Space permits only a few examples from the data.

*Skills:* “Are you kidding? I learned how to survive without the comforts of civilization. I learned how to read the terrain ahead and pace myself. I learned how to carry a heavy load. I learned how to stay dry when it’s raining. I learned how to tie a knot so that it doesn’t come apart when pressure is applied. You think those are metaphors for skills I need in my work? You’re damn right they are.”

*Attitudes:* “I think it’s important to pay attention to the space you’re in. I don’t want to just keep going through my life oblivious to what’s around me and how it affects me and how I affect it.”

*Feelings:* “Being out here, especially on solo, has given me confidence. I know I can handle a lot of things I didn’t think I could handle.”

*Behaviors:* “I use my senses in a different way out here. In the city you get so you don’t pay much attention to the noise and the sounds. But listening out here I’ve also begun to listen more back there. I touch more things too, just to experience the different textures.”

*Knowledge:* “I know about how this place was formed, its history, the rock formations, the effects of the fires on the vegetation, where the river comes from and where it goes.”

A different way of thinking about organizing data around outcomes was to think of different *levels* of impact: effects at the individual level, effects on the group, and effects on the institutions



from which participants came into the program. The staff hoped to have impacts at all of these levels. Thus, it was possible to organize the data by looking at what themes emerged when program processes were crossed with levels of impact. How did developing sensitivity to the environment affect individuals? How did the process of developing sensitivity to the environment affect the group? What was the effect of developing sensitivity to the environment on the institutions to which participants returned after their wilderness experiences? The process/outcomes matrix thus becomes a way of asking questions of the data, an additional source of focus in looking for themes and patterns in the hundreds of pages of field notes and interview transcriptions.

### CAUSES, CONSEQUENCES, AND RELATIONSHIPS

Thus far the data analysis has emphasized the tasks of organization and description. Even the process/outcomes matrix was aimed at providing a mechanism for organizing and describing the themes, patterns, activities, and content of the program, rather than at elucidating causal linkages between processes and outcomes. As the process/outcomes matrix demonstrates, however, there is often a fine line between description and causal interpretation. To the extent that one is *describing* the causal linkages suggested by and believed in by program participants and staff, the evaluator-analyst has not crossed the line from description into causal interpretation. Once the tasks of organization and description are complete it may then be appropriate to move on to consideration of causes, consequences, and relationships.

Chapter Eight, "The Purpose of Qualitative Analysis," discussed the tentative nature of causal theorizing in qualitative research. Statements about which things appear to lead to other things, which parts of the program produce certain effects, and how processes lead to outcomes are areas of speculation, conjecture, and hypothesizing. That's all right! When careful study of the data gives rise to ideas about causal linkages, there is no reason to deny decision makers and information users the benefit of those insights simply because they cannot be proven. What is important is that such statements be clearly qualified as what they are—speculation, conjecture, and hypothesizing. The evaluator who has studied the program, lived with the data from the field, and reflected at length about the patterns and themes that run through the data is in as good a position as anyone else at that point to speculate, make conjectures, and

---



formulate hypotheses. If decision makers and information users have asked for such information—and in my experience they virtually always welcome these kinds of speculative analyses—there is no reason not to share insights with them to help them think about their own causal presuppositions and hypotheses.

One of the biggest dangers for evaluators doing qualitative analysis is that when they begin to consider causes, consequences, and relationships they fall back on the linear assumptions of quantitative analysis and begin to specify isolated variables that are mechanically linked together out of context. In attempting to present a holistic picture of what the program is like and in struggling to understand the phenomenological nature of a particular set of activities and people in a specific context, simple statements of linear relationships may be more distorting than illuminating. It is the ongoing challenge, paradox, and dilemma of qualitative analysis that we must be constantly moving back and forth between the phenomenon of the program and our abstractions of that program, between the descriptions of what has occurred and our analysis of those descriptions, between the complexity of reality and our simplifications of those complexities, between the circularities and interdependencies of human activity and our need for linear, ordered statements of cause-effect.

Gregory Bateson traces at least part of the source of our struggle to the ways we have been taught to think about things. We are told that a “noun” is the “name of a person, place, or thing.” We are told that a “verb” is an “action word.” These kinds of definitions, Bateson argues, were the beginning of teaching us that “the way to define something is by what it supposedly *is* in itself—not by its relations to other things.”

Today all that should be changed. Children could be told that a noun is a word having a certain relationship to a predicate. A verb has a certain relationship to a noun, its subject, and so on. Relationship could now be used as a basis for definition, and any child could then see that there is something wrong with the sentence, “ ‘Go’ is a verb.” . . .

We *could* have been told something about the pattern which connects: that all communication necessitates *context*, and that without context there is no meaning [Bateson, 1978:13].

Without belaboring this point about the difference between linear causal analyses (x causes y) and a holistic perspective that describes the interdependence and relatedness of complex phenomena, I would



simply offer the reader a Sufi story. I suggest trying to analyze the data represented by the story in two ways. First, try to isolate specific variables that are important in the story; decide which are the independent and which the dependent variables; and then write a statement of the form: These things caused this thing. Then read the story again. For the second analysis try to distinguish among and label the different meanings of the situation expressed by the characters observed in the story; then write a statement of the form: These things and these things came together to create \_\_\_\_\_. Don't try to decide that one approach is right and the other is wrong; simply try to experience and understand the two approaches.

Walking one evening along a deserted road, Mulla Nasrudin saw a troop of horsemen coming towards him. His imagination started to work; he imagined himself captured and sold as a slave, or robbed by the oncoming horsemen, or impressed into the army. Fearing for his safety, Nasrudin bolted, climbed a wall into a graveyard, and lay down in an open tomb.

Puzzled at this strange behavior the men—honest travelers—pursued Nasrudin to see if they could help him. They found him stretched out in the grave, tense and quivering.

“What are you doing in that grave? We saw you run away and see that you are in a state of great anxiety and fear. Can we help you?”

Seeing the men up close Nasrudin realized that they were honest travelers who were genuinely interested in his welfare. He did not want to offend the travelers or embarrass himself by telling them how he had misperceived them. Nasrudin simply sat up in the grave and said, “You ask what I'm doing in this grave. If you must know, I can tell you only this: *I* am here because of *you*, and *you* are here because of *me*” [adapted from Shah, 1972:16].

## VALIDATION AND VERIFICATION

The preceding sections have focused on some strategies for making sense out of qualitative data. The remainder of this chapter will be concerned with the problem of deciding how much to trust the data analysis. There are two parts to the issue of trusting the data. First, the person or persons analyzing the data must make a determination of how much confidence to place in their own analysis. Second, the data analysis must be presented to others in such a way that they can verify and validate the findings of the analysis for themselves.

---



In the previous chapter I attempted to lift off the shoulders of evaluators the burden of uncovering and presenting Truth. In speculating on causes, consequences, and relationships in this chapter the emphasis has again been on the humble notion that, in the end, all we can provide is perspective. The perspective gained through careful qualitative analysis is not arbitrary, nor is it predetermined, but it does fall short of being Truth. This section discusses ways of validating and verifying the *perspective* that emerges through qualitative analysis.

When struggling with the process of analysis it is helpful to keep in mind that the basic purpose of qualitative analysis is to provide useful, meaningful, and credible answers to the evaluation questions of decision makers and information users. To be meaningful the answers provided must relate directly to the questions that have been asked; to be useful the answers must be understandable and clearly presented; and to be credible the evaluator must demonstrate that the perspective presented will hold up under careful scrutiny. The evaluator subjects the data to careful scrutiny by validating and verifying evaluation findings. There are several major strategies for validating and verifying the results of qualitative analysis.

## RIVAL EXPLANATIONS

Once the evaluator-analyst has described the patterns, linkages, and accompanying explanations that have emerged from the analysis, it is important to look for rival or competing themes and explanations. This can be done both inductively and logically. Inductively it involves looking for other ways of organizing the data that might lead to different findings. Logically it means thinking about other logical possibilities and then seeing if those possibilities can be supported by the data. When considering rival hypotheses and competing explanations the strategy to be employed is not one of attempting to disprove the alternatives; rather, the analyst looks for data that *support* alternative explanations. Failure to find strong supporting evidence for alternative explanations helps increase confidence in the original, principal explanation generated by the evaluator. It is likely that comparing alternative explanations or looking for data in support of alternative patterns will not lead to clear-cut “yes there is support” versus “no there is not support” kinds of conclusions. It is a matter of considering the *weight* of evidence and looking for the *best* fit between data and analysis. It is important to write down what alternative classification systems, themes, and explanations are considered and “tested” during data



analysis. Reporting on what alternative explanations were considered and how those alternatives were considered in the formal evaluation report lends considerable credibility to the final set of findings offered by the evaluator.

## NEGATIVE CASES

Closely related to the testing of alternative explanations is the search for negative cases. Where patterns and trends have been identified, our understanding of those patterns and trends is increased by considering the instances and cases that do not fit within the pattern. For example, in an employment training program where the large majority of participants complete the program and find satisfying employment, the most important analysis may be an examination of the program dropouts and the people who do not find satisfactory employment. In the Southwest Field Training Project involving wilderness education virtually all participants reported some kind of "personal growth" as a result of their participation in the project; however, the two people who reported "no change" provided particularly useful insights into how the program operated and affected participants.

While the evaluator can be exhorted to be as rigorous in attempting to support alternative hypotheses and to understand negative or deviant cases as that same evaluator was in building support from the data for the original explanation, readers of the evaluation findings will have to make their own decisions about the plausibility of alternate explanations and the reasons why deviant cases do not fit within dominant patterns.

[T]here are no guidelines specifying how and how long to search for negative cases or how to find alternative hypotheses given a specified body of qualitative data. Thus the analyst's attempt to convey his theory's credibility may still be unsuccessful because of possible bias in his search for negative cases or for reasonable alternative hypotheses [Glaser and Strauss, 1967:230].

I would also note that the section of the report that involves exploration of alternative explanations and consideration of why certain cases do not fall into the main pattern can be among the most interesting sections of a report to read. When well written, this section of a report reads something like a detective novel in which the evaluator (detective) looks for clues that lead in different directions and tries to sort out which direction makes the most sense given the

---



clues (data) that are available. Moreover, the tone of a report is different when the evaluator is willing to openly consider other possibilities than those finally recommended as most reasonable. Compare the approach of weighing alternatives to the report where all the data lead in a single-minded fashion, in a rising crescendo, toward an overwhelming presentation of a single point of view.

### TRIANGULATION: RECONCILING QUALITATIVE AND QUANTITATIVE DATA

The chapter on design talked about the potential importance of using different data-collection techniques and different evaluation research strategies to study the same program. It is in data analysis that the strategy of triangulation really pays off. There are basically two kinds of triangulation that contribute to verification and validation of qualitative analysis: (1) checking out the consistency of findings generated by different data-collection methods and (2) checking out the consistency of different data sources within the same method.

Triangulation of methods will most often revolve around comparing data collected through some kind of qualitative methods with data collected through some kind of quantitative methods. This is seldom a straightforward process because, given the argument in the early chapters of this book about the kinds of questions that particularly lend themselves to qualitative methods, it is highly likely that quantitative methods and qualitative methods will eventually answer different questions that do not easily come together to provide a single, well-integrated picture of the situation. Shapiro (1973) describes in detail her struggle to resolve basic differences between qualitative data and quantitative data in her study of Follow Through Classrooms; she eventually concluded that some of the conflicts between the two kinds of data were a result of measuring different things, although the ways in which different things were measured were not immediately apparent until she worked to sort out the conflicting findings. She began with greater trust in the data derived from quantitative methods and ended by believing that the most useful information came from the qualitative data.

An article by M.G. Trend (1978) of ABT Associates is required reading for anyone who is becoming involved in a project—particularly a team project—which will involve both the collection of qualitative and quantitative data and where different members of the team have responsibilities for different kinds of data. The Trend study involved an analysis of three HUD social experiments designed to



test the concept of using direct cash housing allowance payments to help low-income families obtain decent housing on the open market. The analysis of qualitative data from a participant observation study produced results that were at variance with those generated by analysis of quantitative data. The credibility of the qualitative data became a central issue in the analysis.

The difficulty lay in conflicting explanations or accounts, each based largely upon a different *kind* of data. The problems we faced involved not only the nature of observational *versus* statistical inferences, but two sets of preferences and biases within the entire research team. . . .

Though qualitative/quantitative tension is not the only problem which may arise in research, I suggest that it is a likely one. Few researchers are equally comfortable with both types of data, and the procedures for using the two together are not well developed. The tendency is to relegate one type of analysis or the other to a secondary role, according to the nature of the research and the predilections of the investigators. . . . Commonly, however, observational data are used for "generating hypotheses," or "describing *process*." Quantitative data are used to "analyze *outcomes*," or "verify hypotheses." I feel that this division of labor is rigid and limiting [Trend, 1978:352].

There is no magic in triangulation. The evaluator using different methods to investigate the same program should not expect that the findings generated by those different methods will automatically come together to produce some nicely integrated whole. Indeed, the evidence is that one ought to expect initial conflicts in findings from qualitative and quantitative data, and expect those findings to be received with varying degrees of credibility. In endorsing the notion of triangulation, Trend maintains that it is useful to bring a variety of data and methods to bear on the same problem. In doing so, however, Trend suggests that

we give different viewpoints the chance to arise, and postpone the immediate rejection of information or hypotheses that seem out of joint with the majority viewpoint. Observationally derived explanations are particularly vulnerable to dismissal without a fair trial [Trend, 1978:352-353].

### TRIANGULATION: COMPARING MULTIPLE QUALITATIVE DATA SOURCES

The second type of triangulation involves triangulating data sources. This means comparing and cross-checking consistency of

---



information derived at different times and by different means within qualitative methods. It means (1) comparing observational data with interview data; (2) comparing what people say in public with what they say in private; (3) checking for the consistency of what people in a situation say about this situation over time; and (4) comparing the perspectives of people from different points of view—staff views, client views, funder views, and views expressed by people outside the program, where those are available to the evaluator. It means validating information obtained through interviews by checking program documents and other written evidence that can corroborate what interview respondents report.

As with triangulation of methods, triangulation of data sources within qualitative methods will seldom lead to a single, totally consistent picture. It is best not to expect everything to turn out the same. The point is to study and understand when and why there are differences. The fact that observational data produce different results than interview data does not mean that either or both kinds of data are invalid, although that may be the case. More likely, it means that different kinds of data have captured different things and so the analyst attempts to understand the reasons for the differences. At the same time, consistency in overall patterns of data from different sources and reasonable explanations for differences in data from different sources contributes significantly to the overall credibility of the findings presented in the evaluation report.

### TRIANGULATION: MULTIPLE PERSPECTIVES FROM MULTIPLE OBSERVERS

A third kind of triangulation is “investigative triangulation which simply means that multiple as opposed to singular observers are employed” (Denzin, 1978:297). Triangulating observers or using several interviewers helps reduce the potential bias that comes from a single person and provides means of more directly assessing the reliability and validity of the data obtained. Douglas (1976) is particularly articulate in arguing the importance of team efforts in fieldwork.

Triangulating observers provides a check on bias in data collection. A related strategy is triangulating analysts—that is, having two or more persons independently analyze the same qualitative data set and then compare their findings. This was the approach we used in analyzing the interview data on the utilization of evaluation research reported in *Utilization-Focused Evaluation* (Patton, 1978). Recent



variations on the team approach in evaluation research include Michael Scriven's (1972a) strategy which involves having two separate teams for the evaluation, one which performs a goal-based evaluation (assessing the stated outcomes of the program) and a second team conducting a goal-free evaluation (see Chapter Four). A different team research strategy is the advocacy-adversary model suggested by Wolf (1975). The advocacy team gathers information that supports the proposition that the program is doing a good job and attaining its goals; the adversary team gathers information that leads to the conclusion that the program ought to be changed or terminated. Again, the advocacy-adversary model as presented in the literature typically involves two separate teams working on both data collection and analysis. A variation of this strategy would be to arbitrarily create advocacy and adversary teams only during the analysis stage so that both teams work with the same set of data but each team works toward marshalling that data to support different, opposite conclusions.

These different types of triangulation—methods triangulation, triangulation of data sources, and investigator triangulation—are all strategies for reducing systematic bias in the data. In each case the strategy involves checking findings against other sources. *Triangulation is a process by which the evaluator can guard against the accusation that a study's findings are simply an artifact of a single method, a single data source, or a single investigator's bias.*

#### DESIGN CHECKS: KEEPING METHODS AND DATA IN CONTEXT

One possible source of distortion in findings of qualitative evaluation reports is the nature of research design decisions used for data collection. Thus, it is important to consider the rival methodological hypothesis that the findings are due to distortions introduced by the sampling strategies used in the study. Three kinds of sampling errors can arise in qualitative research designs. There may be distortion in the situations that were sampled for observation (since it is seldom possible to observe all situations); there may be distortions introduced by the time periods during which observations took place—that is, problems of temporal sampling; and third, the findings may be distorted because of selectivity in the people who were sampled either for observations or interviews. In considering how sampling strategies may have affected evaluation findings the analyst returns to consideration of the reasons for having made initial design decisions (see Chapter Five). To the extent that those de-

---



sign decisions were based directly on the kinds of evaluation questions that were being asked, apparent distortions produced by sampling decisions may have been purposeful and deliberate given a calculated desire to study intensively only certain situations, certain time periods, or certain people. Under these conditions the problem is no longer one of distortion of the data actually collected, but is a question of the extent to which the findings can be generalized to other situations, other time periods, and other people. Thus, the evaluator-analyst must be careful to limit conclusions to those situations, time periods, persons, and contexts for which the data are applicable.

The importance of reporting both methods selected and resulting data in their proper contexts cannot be overemphasized. *Keeping things in context is a cardinal principle of qualitative analysis.* Mulla Nasrudin was once called upon to make this point to his monarch.

Although he was supposed to be a wise and holy man, Nasrudin was accused of being almost illiterate.

One day the ruler of his country decided to put this to the test.

“Write something for me, Nasrudin,” said he.

“I would willingly do so, but I have taken an oath never to write so much as a single letter again,” said Nasrudin.

“Well, write something in the way in which you used to write before you decided not to write, so that I can see what it was like.”

“I cannot do that, because every time you write something your writing changes slightly through practice. If I wrote now, it would be something written for now.”

“Then bring me an example of his writing, anyone who has one,” ordered the ruler.

Someone brought a terrible scrawl that the Mulla had once written to him.

“Is this your writing?” asked the monarch.

“No,” said Nasrudin. “Not only does writing change with time, but you are now showing a piece of writing done by me to demonstrate to someone how he should *not* write” [Shah, 1973:92].

## EVALUATOR EFFECTS

There are four ways in which the presence of the evaluator, or the fact that an evaluation is taking place, can distort the findings of a study:

- (1) reactions of program participants and staff to the presence of the evaluator;



- (2) changes in the evaluator (the measuring instrument) during the course of the evaluation—that is, instrumentation effects;
- (3) the predispositions or biases of the evaluator; and
- (4) evaluator incompetence.

The presence of an evaluator can certainly make a difference in how a program operates and in its outcomes. The fact that a study is being conducted may create a halo effect so that staff perform in an exemplary fashion and participants are motivated to “show off.” On the other hand, the presence of the evaluator may create so much tension and anxiety that performances are below par. Problems of reactivity are well documented in anthropological literature. That is one of the prime reasons why qualitative methodologists advocate long-term observations that permit an initial period during which evaluators (observers) and the people in the setting being observed get a chance to get used to each other. Denzin’s (1978) advice concerning the reactive effects of observers is, I think, applicable to the specific case of evaluator-observers:

It is axiomatic that observers must record what they perceive to be their own reactive effects. They may treat this reactivity as bad and attempt to avoid it (which is impossible), or they may accept the fact that they will have a reactive effect and attempt to use it to advantage. . . .

The reactive effect will be measured by daily field notes, perhaps by interviews in which the problem is pointedly inquired about, and also in daily observations [Denzin, 1978:200].

In brief, the evaluator has a responsibility to think about the problem, make a decision about how to handle it in the field, and then attempt to monitor evaluator/observer effects. My own view is that evaluator effects are considerably overrated, particularly by evaluators. There is more than a slight touch of self-importance in some concerns about reactivity. Lillian Weber, director of the Workshop Center for Open Education, City College School of Education, New York, once set me straight on this issue, and I pass her wisdom on to my colleagues. In doing observations of open classrooms, I was concerned that my presence, particularly the way kids flocked around me as soon as I entered the classroom, was distorting the situation to the point where it was impossible to do good observations. Lillian laughed and suggested to me that what I was experiencing was the way those classrooms actually were. She went on to note that this was common among visitors to schools; they were always con-



cerned that the teacher, knowing visitors were coming, whipped the kids into shape for those visitors. She suggested that under the best of circumstances a teacher might get kids to move out of habitual patterns into some model mode of behavior for as much as ten or fifteen minutes, but that, habitual patterns being what they were, kids would rapidly revert to normal behaviors and whatever artificiality might have been introduced by the presence of the visitor would likely become apparent.

*Evaluators should strive to neither overestimate nor underestimate their effects, but to take seriously their responsibility to describe and study what those effects are.*

The second concern about evaluator effects arises from the possibility that the evaluator changes during the course of the evaluation. One of the ways this sometimes happens in anthropological research is when the participant observers "go native" and absorb themselves in the local culture. The epitome of this in a shorter-term observation is the story of the observers who became converted to Christianity while observing a Billy Graham crusade (Lang and Lang, 1960). An evaluator observing a chemical dependency program may find that he or members of his family are chemically dependent and therefore change his entire approach to a study. Evaluators sometimes become personally involved with program participants or staff and therefore lose their sensitivity to the full range of events occurring in the setting.

Johnson (1975) and Glazer (1972) have reflected on how they and others have been changed by doing field research. The consensus of advice on how to deal with the problem of changes in the observer as a result of involvement in research is the same as advice about how to deal with the reactive effects created by the presence of observers.

It is central to the method of participant observation that changes will occur in the observer; the important point, of course, is to record these changes. Field notes, introspection, and conversations with informants and colleagues provide the major means of measuring this dimension, . . . for to be insensitive to shifts in one's own attitudes opens the way for placing naive interpretations on the complex set of events under analysis [Denzin, 1978:200].

The third concern about evaluator effects has to do with the extent to which the predispositions or biases of the evaluator may affect data analysis and interpretations. In effect, all of the procedures for validating and verifying data analysis that have been presented in this chapter are aimed at reducing distortions introduced by evalu-



ator predisposition. However, this issue remains difficult because, on the one hand, these procedures are aimed at substantiating the validity of the data and yet, on the other hand, I am arguing that the data inevitably represent perspective rather than truth. Consideration of this dilemma and apparent contradiction necessitates consideration of what it means to be “objective” and “subjective” in the conduct of research.

## OBJECTIVITY AND SUBJECTIVITY

The most frequent charge heard in attacks on qualitative methodology is that evaluation research conducted with qualitative methods is inevitably “subjective”—an aspersion connoting the very antithesis of scientific inquiry. Objectivity is traditionally considered the sine qua non of *the* scientific method. To be subjective means to be biased, unreliable, and irrational. Subjective data imply opinion rather than fact, intuition rather than logic, impression rather than confirmation. Social scientists are exhorted to eschew subjectivity and make sure that their work is “objective.” The conventional means for controlling subjectivity and maintaining objectivity are the methods of quantitative social science: distance from the setting and people being studied, formal operationalism and quantitative measurement, manipulation of isolated variables, and experimental designs. Close examination, however, reveals that these methodological approaches constitute a value-laden paradigm and that identification of objectivity as the major virtue of that paradigm is an ideological statement the function of which is to legitimize, preserve, and protect the dominance of a single evaluation methodology (Patton, 1978). The ways in which measurements are constructed in psychological tests, questionnaires, cost-benefit indicators, and routine management information system data are no less open to the intrusion of the evaluator’s biases than making observations in the field or asking questions in interviews. *Numbers do not protect against bias; they merely disguise it.*

Scriven has insisted that quantitative methods are no more synonymous with objectivity than qualitative methods are synonymous with subjectivity:

Errors like this are too simple to be explicit. They are inferred confusions in the ideological foundations of research, its interpretations, its application. . . . It is increasingly clear that the influence of ideology on methodology and of the latter on the training and behavior of researchers and on the identification and disburse-



ment of support is staggeringly powerful. Ideology is to research what Marx suggested the economic factor was to politics and what Freud took sex to be for psychology [Scriven, 1972a:94].

Scriven's lengthy discussion of objectivity and subjectivity in educational research is a major contribution in the struggle to detach the notions of objectivity and subjectivity from their traditionally narrow associations with quantitative and qualitative methodology, respectively. He presents a clear explanation of how objectivity has been confused with consensual validation of something by multiple observers. Yet, a little reflection will yield many instances where the majority of scientists (or other people) were factually wrong while one dissenting observer actually described things as they really were (Kuhn, 1970).

*Qualitative objectivity* has to do with the *quality* of the observations made by an evaluator. Scriven emphasizes the importance of being factual about observations rather than being distant from the phenomenon being studied. *Distance does not guarantee objectivity; it merely guarantees distance.* On the other hand, getting close enough to the situation observed to experience it firsthand means that evaluators can learn from their experiences, thereby generating personal insights. "For the social scientists to refuse to treat their own behavior as data from which one can learn is really tragic" (Scriven, 1972a:99).

Guba has considered the issues of objectivity and subjectivity in considerable depth. He notes that in all areas of social science the data collected should be reliable, factual, and confirmable. "There seems to be no intrinsic reason why the methods of a properly trained naturalistic inquirer should be any more doubtful a source of such data than the methods of an investigator using a more quantitative approach" (Guba, 1978:74-75). He suggests that the issue is more clearly stated by talking about "neutrality" of the evaluator rather than objectivity or subjectivity. The neutral evaluator is not predisposed toward certain findings on an a priori basis. Yet, this word change introduces its own problems. "Neutrality" may imply that the evaluator does not care about the findings. House suggests that lack of caring is hardly reassuring to decision makers and information users.

Even qualitative objectivity is insufficient for evaluation, for it carries the aura of neutrality. People being evaluated do not want a neutral evaluator, one who is unconcerned about the issues. A person on trial



would not choose a judge totally removed from his own social system. . . .

The evaluator must be seen as caring, as interested, as responsive to the relevant arguments. He must be impartial rather than simply objective.

The impartiality of the evaluator must be seen as that of an actor in events, one who is responsive to the appropriate arguments but in whom the contending forces are balanced rather than non-existent. The evaluator must be seen as not having previously decided in favor of one position or the other [House, 1977:45-46].

The politics of evaluation mean that evaluators must make their own peace with how they are going to describe what they do. The meaning and connotations of objectivity, subjectivity, neutrality, and impartiality will have to be worked out with particular decision makers and information users in specific evaluation settings. Essentially, these are all concerns about the extent to which the data can be trusted. All of the validating and verifying procedures discussed in this chapter are concerned with the issue of the trustworthiness of the data. For better or worse, the trustworthiness of the data is tied directly to the trustworthiness of the evaluator who collects and analyzes the data. Thus, the fourth and final issue concerning evaluator effects is the issue of competence.

Competence is demonstrated by using the verification and validation procedures necessary to establish the quality of analysis. Competence is demonstrated by building a "track record" of fairness and responsibility. Competence involves neither overpromising nor underproducing in evaluation research. Finally, the concern for evaluator competence means that decision makers and information users bear major responsibility in making sure that the evaluators they employ have the knowledge and skills they desire. It is not enough to shift the responsibility for the qualitative evaluation totally into the lap of the evaluator. Decision makers and information users share responsibility for the credibility of the evaluation by their selection of an evaluator and their identification of evaluation issues to be studied. This leads me to the final, and perhaps the most important, procedure for validating and verifying data analysis.

## SUBJECT AND AUDIENCE REACTIONS TO THE ANALYSIS

Evaluators can learn a great deal about the accuracy, fairness, and validity of their data analysis by having the people described in that



data analysis react to what is described. To the extent that participants in the study are unable to relate to the description and analysis in a qualitative evaluation report it is appropriate to question the credibility of the report. Alkin et al. (1979), in studying how evaluations are used, presented each case study to the people in the setting described and asked them for both verbal and written reactions. They then included these written reactions to the evaluation report as part of the data in the report. *The ultimate tests of the credibility of an evaluation report is the response of decision makers and information users to that report.* It is at that point that the evaluator's perspective is joined to the perspective of the people who must use the information. House suggests that the more "naturalistic" the evaluation, the more it relies on its audiences to reach their own conclusions, draw their own generalizations, and make their own interpretations. House is articulate and insightful on this critical point:

[U]nless an evaluation provides an explanation for a particular audience, and enhances the understanding of that audience by the content and form of the argument it presents, it is not an adequate evaluation for that audience, even though the facts on which it is based are verifiable by other procedures. One indicator of the explanatory power of evaluation data is the degree to which the audience is persuaded. Hence, an evaluation may be "true" in the conventional sense but not persuasive to a particular audience for whom it does not serve as an explanation. *In the fullest sense, then, an evaluation is dependent both on the person who makes the evaluative statement and on the person who receives it* [House, 1977:42; italics added].

## INTELLECTUAL RIGOR

The thread that runs through these procedures and techniques for verifying and validating qualitative data is their dependence on the intellectual rigor of the evaluator. There are no clear-cut rules about how to proceed. The task is to do one's best to make sense out of things. A qualitative analyst returns to the data over and over again to see if the constructs, categories, explanations, and interpretations make sense, if they really reflect the nature of the phenomena. Creativity, intellectual rigor, perseverance, insight—these are the intangibles that go beyond the routine application of scientific procedures. As Nobel prize-winning physicist Percy Bridgman put it: "There is no scientific method as such, but the vital feature of a scientist's procedure has been merely to do his utmost with his mind, *no holds barred*" (quoted in Mills, 1961:58; italics in original).



## PRESENTATION OF FINDINGS

I find in my own work that final reports frequently have less impact than the direct, face-to-face interactions I have with decision makers and information users to provide them with feedback about evaluation findings and to share with them the nature of the data. Final reports often serve an important dissemination function to audiences beyond immediate decision makers and information users, but they are not automatically and necessarily the primary source of information for those who are intended to actually *utilize* evaluation findings. I have done evaluations that involved no polished, final report. Such reports take a tremendous amount of time and energy, and certain utilization situations may not be well served by putting all those evaluation resources into the production of a polished final report that will sit on a shelf somewhere. Eyebrows may be raised when evaluators ask: "Is there any reason to produce a final, written report for this evaluation?" But it is a question worth asking, and, in my opinion, the burden of proof lies with the decision makers and information users to justify production of a full report.

Normally, of course, a full report will be produced. The contents, length, and nature of the report is partly a matter for negotiation between evaluators and decision makers. While individual style will and should affect what a final report looks like, there are some basic areas that should be covered in a comprehensive final evaluation report that presents qualitative data.

### A REPORT OUTLINE

The report outline suggested below presents major portions of a comprehensive final report. The exact order of the sections is less important than the fact that all of the different portions must be present.

- I. Purpose of the evaluation
    - A. Context of the evaluation
      1. How did the evaluation originate?
      2. Who is the evaluation for?
      3. How is the evaluation funded?
      4. How was the evaluator selected?
    - B. Evaluation Focus
      1. What questions are to be answered by the evaluation?
      2. Why these questions?
      3. What actions are anticipated or decisions to be made as a result of the evaluation?
-



- II. Methods decisions
  - A. Appropriateness of methods
    - 1. How do the methods being employed follow from the kinds of evaluation questions being asked?
    - 2. What are the strengths and weaknesses of the methods being used in light of the evaluation purpose?
  - B. What design and sampling decisions were made, for what reasons, and with what consequences?
    - 1. Situations to be sampled
    - 2. Time periods to be sampled
    - 3. People to be sampled
- III. Presentation of the data
  - A. Descriptive information about the program
    - 1. Program origins and history
    - 2. Program activities, processes, and goals
    - 3. Client and staff characteristics
  - B. Description of findings organized around evaluation questions, issues, and concerns generated by decision makers and information users
    - 1. Descriptive information based on observations and/or interviews. What happened? What was said?
    - 2. Provide whatever information is needed to take the reader into the situation being described and evaluated.
  - C. Analysis of the data
    - 1. Presentation of patterns, themes, tendencies, trends, and motifs that emerge from the data
    - 2. Presentation of categories, classification systems, and typologies
      - a. participant-generated typologies to explain their world
      - b. evaluator-generated typologies
  - D. Interpretations and explanations
    - 1. Linkages between categories and dimensions
    - 2. Relationships, things that appear to go together, interdependent parts
    - 3. Notions about causes and consequences, including hypotheses about the relationship between program processes and program outcomes
- IV. Validation and verification of findings
  - A. Details about actual implementation of methods and reporting on any departures from expected procedures. How was the study done? How were the data actually collected?
  - B. Credibility of the findings
    - 1. Discussion of rival hypotheses and alternative explanations
    - 2. Analysis of negative or deviant cases
    - 3. Triangulation
      - a. of methods
      - b. of sources
      - c. of investigators
    - 4. Evaluator effects—the evaluator's personal role and perspective



5. Transcriptions of any recorded reactions from participants or others who have examined the study
- V. Conclusions and recommendations (By request, this is sometimes the first section of the report; it is put up front so that decision makers can turn right to it.)
- A. What are the basic findings?
  - B. What are the implications of the findings?
  - C. What are the recommendations?
    1. Recommendations from program participants, staff, or others
    2. Recommendations of the evaluator

To facilitate reading a report it is often helpful to go through much of the entire sequence for each evaluation question—that is, to present the description, analysis, and interpretation of each focus in the evaluation all in one place. An evaluation report should be readable, understandable, and relatively free of academic jargon. It is the data that should impress the reader, not the academic training of the evaluator.

Even a comprehensive report will have to omit a great deal of information collected by the evaluator. Focus is essential. Evaluators who try to include everything risk losing their readers in the sheer volume of the presentation. Lofland (1971) calls the decisions that must be made about what material to leave out of a report “the agony of omitting.”

It can happen that an overall structure that organizes a great deal of material happens also to leave out some of one's most favorite material and small pieces of analysis. . . . [U]nless one decides to write a relatively disconnected report, he must face the hard truth that no overall analytic structure is likely to encompass every small piece of analysis and all the empirical material that one has on hand. . . .

The underlying philosophical point, perhaps, is that everything is related to everything else in a flowing, even organic fashion, making coherence and organization a difficult and problematic human task. But in order to have any kind of understanding, we humans require that some sort of order be imposed upon that flux. No order fits perfectly. All order is provisional and partial. Nonetheless, understanding requires order, provisional and partial as it may be. It is with that philosophical view that one can hopefully bring himself to accept the fact that he cannot write about everything that he has seen (or analyzed) and still write something with overall coherence or overall structure [Lofland, 1971:123].

The advice I find myself repeating most often to students when they are writing reports is: “FOCUS! FOCUS! FOCUS!” The agony of omitting on the part of the evaluator is matched only by the



readers' agony in having to read those things that were not omitted—but should have been.

## BALANCE BETWEEN DESCRIPTION AND ANALYSIS

One of the major decisions that has to be made about what to omit involves a corresponding decision about how much description to include. Description and quotation are the essential ingredients of qualitative inquiry. Sufficient description and direct quotations should be included to allow the reader to enter into the situation and thoughts of the people represented in the report. Description should stop short, however, of becoming trivial and mundane. The reader does not have to know absolutely everything that was done or said—again, the problem of focus.

Description is balanced by analysis and interpretation. Endless description becomes its own muddle. The purpose of analysis is to organize the description so that it is manageable. Description is balanced by analysis and leads into interpretation. An interesting and readable report provides sufficient description to allow the reader to understand the analysis, and sufficient analysis to allow the reader to understand the description.

Verification and validation information need not be relegated to a separate section. Parenthetical remarks throughout the text about information that has been validated is helpful to the reader at the point at which findings are presented. For example, if I am presenting a description of some program process and I then speculate on the relationship between that process and client outcomes, it is appropriate at that point to mention that (1) staff and clients agreed with this part of the analysis when they read it, (2) I experienced that linkage personally in my own participation in the program, and (3) this connection was independently arrived at by two analysts looking at the data separately.

It is also important that the evaluator *not* pretend that all findings are equally credible. The writer bears some responsibility to help the reader sort out the strengths and weaknesses of various parts of the description and analysis. Qualitative analysis does not have the parsimonious statistical significance tests of quantitative analysis. Statistical tests of significance are shorthand ways of telling the reader how seriously to take the findings. In qualitative analysis the analyst must make judgments that provide clues for the reader as to the writer's belief about variations in the credibility of different findings: when are patterns "clear"; when are they "strongly



supported by the data”; and when are the patterns “weak.” Readers will ultimately make their own decisions and judgments about these matters, but the evaluator’s opinions and speculations, after he or she has struggled with the data, deserve to be reported.

## THE EXECUTIVE SUMMARY

As one might suspect, qualitative reports tend to be relatively lengthy. This can be a major problem when busy decision makers do not have the time or will not take the time to read a lengthy report. A critical skill that evaluators must develop is the ability to produce an executive summary of one or two pages that presents the essential findings, conclusions, and reasons for confidence in the summary. The executive summary is a dissemination document, a political instrument, and cannot be—nor is it meant to be—a full and fair representation of the study. An executive summary or abstract should be written in plain language, should be highly focused, and should be directed at the essence of the evaluation. It is helpful to keep in mind when writing the executive summary that more people are likely to read the summary than any other document produced from the evaluation.

## UTILIZATION STRATEGIES

Every chapter in this book has emphasized the strategic nature of evaluation research. An evaluation strategy that includes a fundamental concern for the usefulness and actual utilization of evaluation data does not put off until the end of the study consideration of how to get decision makers and information users to pay attention to the results. The foundation for utilization is laid at the beginning of the process with that very first interaction between the evaluator and the people who are to use the evaluation findings. A strategy for enhancing utilization informs and frames every subsequent decision made about the evaluation.

The process of identifying and organizing decision makers and information users during the initial phase of the evaluation is guided by the strategic speculation that these are the people who will use the evaluation results. Defining the relevant evaluation issues, making clear the purpose of the study, and focusing the important evaluation questions are processes that require the evaluator to have developed a strategy about how the conceptualization of the evaluation will affect its impact months later. Selecting the appropriate methods to answer the questions posed by decision makers requires familiarity

---



with a variety of evaluation methods strategies. Nor are data collection, data analysis, and data interpretation routine processes. Strategic options are available at each step in the evaluation process, strategic options that affect the utilization of evaluation findings. Thus the theme of the first chapter: Strategic evaluation researchers must possess a large repertoire of skills, competences, and methods—including qualitative methods—in order to function within the utilization-focused framework of “a paradigm of choices.”

Still, there are no guarantees that evaluation results will be used. The strategies described in this book are aimed at increasing the probability that evaluation findings will be meaningful, understandable, relevant, and useful. That’s what strategies are—ways of increasing the likelihood that the desired end will be achieved. Despite the best planning and the most careful strategic attention to utilization criteria throughout the evaluation process, the Sisyphean task of enhancing the utilization of evaluation findings remains one of the primary occupational hazards to the mental health of evaluators. There will always be the feedback session in which, having spent a year of one’s life gathering data, pouring over it, and writing a rigorous and conscientious final report, some decision maker says: “Well, now, I know that you put a lot of work into this. I’m anxious to hear all about what you’ve learned. I’ve got about ten minutes before my next appointment.”

Weep not. Use that ten minute period well! *Be prepared* to make it count.

## A LOOK TO THE FUTURE

It is my hope that this book will make a contribution to the ongoing debate about evaluation methods. Evaluators argue among themselves about many things. There are disagreements about the appropriate purposes of evaluation research, appropriate models, appropriate methods, appropriate types of data analysis, and appropriate utilization expectations, to name a few. These disagreements represent real options, and the field of evaluation research is richer for these options.

What must be avoided is bringing premature closure to these debates. From time to time (and, it seems, with increasing frequency) regulations surface in various federal and state agencies that prescribe universal, standardized evaluation measures and methods for all programs funded by those agencies. I oppose all such regulations in the belief that local program processes are too diverse and client outcomes are too complex to be fairly represented



nationwide, or even statewide, by some narrow set of prescribed measures and methods—whether the mandate be for quantitative *or* qualitative approaches.

So far evaluators have avoided being certified, licensed, inspected, and otherwise regulated. Methods are not yet entirely handed down from on high by officials, bureaucrats, and other “authorities.” But always there is someone around who would like to put an end to both evaluator discretion and indiscretion.

Halcolm will tell you this:

The establishment of an orthodox evaluation methodology is no different from the establishment of a state religion. Officially telling you what methods to use is only one step removed from officially telling you what results to find. At that point utilization of findings will cease to be an issue—for there will be nothing to use, only orders to follow.

*From Halcolm: Political Treatise on Evaluation*

When methods decisions are based on some universal mandate rather than situational merit, evaluation research offers no challenge, requires no subtlety, presents no risk, and allows for no accomplishment. My belief that the current state of evaluation practice does offer challenge, does require subtlety, does present risk, and does allow for accomplishment is what sustained the writing of this book on qualitative evaluation methods.

---



**Appendix 9.1**  
**CODEBOOK FOR CONTENT ANALYSIS OF INTERVIEWS WITH**  
**DECISION MAKERS AND EVALUATORS ABOUT**  
**THEIR UTILIZATION OF EVALUATION RESEARCH**

*Program Evaluated*

0101	nature or kind of program
0102	program relationship to government hierarchy
0103	funding (source, amount, determination of, etc.)
0104	purpose of program, initiation
0105	history of program (duration, changes, termination, etc.)
0106	program effectiveness
0107	other, miscellaneous

*Evaluator Role in Specific Study*

0201	evaluator's role in initiation and planning stage
0203	evaluator's role in data-collection stage
0204	evaluator's role in final report and dissemination
0205	relationship of evaluator to program (internal/external)
0206	evaluator's organization (type, size, staff, etc.)
0207	opinions/feelings about role in specific study
0208	evaluator's background
0209	comments on evaluator, evaluator process
0210	other, miscellaneous

*Decision Maker's Role in Specific Study*

0301	decision maker's role in initiation and planning stage
0302	decision maker's role in data-collection stage
0303	decision maker's role in final report and dissemination
0304	relationship of decision maker to program
0305	relationship of decision maker to other people or units in government
0306	comments on decision maker and decision-making process (opinions, feelings, facts, knowledge, etc.)
0307	other, miscellaneous

*Project Officer Role in Specific Study*

0401	project officer's role in initiation and planning stage
0402	project officer's role in data-collection stage
0403	project officer's role in final report and dissemination
0404	relationship of project officer to program
0405	comments on project officer (opinions, feelings, facts, knowledge, etc.)
0406	other, miscellaneous



*Interactions*

- Interactions about or during:
- 0501 comments on the process of selecting the evaluator (bidding, RFP)
  - 0502 interactions during or about initiation of study
  - 0503 interactions during or about design of study
  - 0504 interactions during or about data collection
  - 0505 interactions during or about final report/findings
  - 0506 interactions during or about dissemination

*Planning and Initiation Process of This Study (how and who started)*

- 0601 initiator
- 0602 interested groups or individuals
- 0603 circumstances surrounding initiation
- 0604 planning for dissemination at initiation phase
- 0605 other, miscellaneous

*Purpose of Study (why)*

- 0701 description of purpose
- 0702 changes in purpose
- 0703 other, miscellaneous

*Political Context*

- 0801 description of political context
- 0802 effects on study
- 0803 other, miscellaneous

*Expectations for Utilization*

- 0901 description of expectations
- 0902 holders of expectations
- 0903 effect of expectations on study (design, methodology, final report, etc.)
- 0904 relationship of expectations to specific decisions
- 0905 reasons for lack of expectations
- 0906 people mentioned as not having expectations
- 0907 effect of lack of expectations on study (design, methodology, final report)
- 0908 other comments on expectations/lack of expectations

*Data Collection, Analysis, Methodology*

- 1001 methodological quality
- 1002 methodological appropriateness
- 1003 factors affecting data collection and methodology (positive factors, problems, etc.)
- 1004 other, miscellaneous



*Findings, Final Report*

- 1101 description of findings/recommendations
- 1102 reception of findings/recommendations
- 1103 comments on final report (forms, problems, quality)
- 1104 comments and description of dissemination (forms, problems, extent of)
- 1105 other, miscellaneous

*Impact of Specific Study*

- 1201 description of impacts on program
- 1202 description of nonprogram impacts
- 1203 impact of specific recommendations
- 1204 other, miscellaneous

*Factors and Effect on Utilization*

- 1301 lateness
- 1302 methodological quality
- 1303 methodological appropriateness
- 1304 positive/negative findings
- 1305 surprise findings
- 1306 central/peripheral objectives
- 1307 point in life of program
- 1308 presence/absence of other studies
- 1309 political factors
- 1310 interaction with evaluators
- 1311 resources
- 1312 most important factor
- 1313 other, miscellaneous

*General Comments*

- 1401 opinion of government use of evaluations in general
- 1402 effect of this study on opinion
- 1403 comments on evaluations/evaluators in general
- 1404 comments on decision maker/decision-making process in general
- 1405 comments on project officer and P.O. process in general
- 1406 comments on utilization/impact of evaluations in general
- 1407 evaluations in general
- 1408 comments on methodology in general
- 1409 factors affecting utilization in general
- 1410 other, miscellaneous



*Background Information of Interviewee*

- 1501 years in Federal Government/Evaluation Research
- 1502 academic discipline
- 1503 academic degree
- 1504 other, miscellaneous



## Appendix 9.2

### MIKE: AN ILLUSTRATIVE CASE STUDY<sup>1</sup>

BY FEHRENBACHER ET AL. (1976) (USED BY PERMISSION OF NORTHWEST REGIONAL EDUCATIONAL LABORATORY)

*Background:* Sitting in a classroom at Metro City High School was difficult for Mike in 1972-73. In some classes he was way behind. In math he was always the first to finish a test. "I loved math and could always finish a test in about ten minutes, but I wasn't doing well in my other classes," Mike explained.<sup>2</sup>

He first heard about Experience-Based Career Education (EBCE) when he was a sophomore. "I really only went to the assembly to get out of one of the classes I didn't like," Mike confessed.

But after listening to the EBCE explanation, Mike was quickly sold on the idea. He not only liked the notion of learning on the job, but also thought the program might allow him to work at his own speed. The notion of no grades and no teachers also appealed to him.

Mike took some descriptive materials home to his parents and they joined him for an evening session at the EBCE learning center to find out more about the program. Now after two years in the program, Mike is a senior and his parents want his younger brother to get into the program.

Early EBCE testing sessions last year verified the inconsistency of Mike's experiences in school. While his reading and language scores were well below the average scored by a randomly selected group of juniors at his school, he showed above average abilities in study skills and demonstrated superior ability in math.<sup>3</sup>

On a less tangible level, EBCE staff members early last school year described Mike as being hyperactive, submissive, lacking in self-confidence and unconcerned about his health and physical appearance when he started the EBCE program. He was also judged to have severe writing deficiencies. Consequently, Mike's EBCE learning manager devised a learning plan that would build his communications skills (in both writing and interpersonal relations) while encouraging him to explore several career possibilities. Mike's job experiences and projects were designed to capitalize on his existing interests and to broaden them.

*First Year EBCE Experiences.* A typical day for Mike started at 8:00 a.m., just as in any other high school, but the hours in between varied

1. All names and references are fictitious in this illustrative case study although the narrative is based on an actual case history.
2. Unless otherwise indicated, statements in quotation marks refer to comments made by the student during an interview with the evaluator. Pseudonyms were used to replace students' actual names.
3. When the terms "above average" or "below average" are used in this section it will mean that a student's score was greater than one standard deviation above or below the EBCE group mean for that variable, meaning that less than 18 percent of his peers scored above or below that level.



considerably. When he first arrived at the EBCE learning center, Mike said he usually spent some time “fooling around” with the computer before he worked on projects underway at the center.

On his original application, Mike indicated his career preference would be computer operator. This led to an opportunity in the EBCE program to further explore that area and to learn more about the job. During April and May, Mike’s second learning level experience took place in the computer department of City Bank Services. He broke up his time there each day into morning and afternoon blocks, often arriving before his employer instructor did for the morning period. Mike usually spent that time going through computer workbooks. When his employer instructor arrived they went over flow charts together and worked on computer language.

Mike returned to the high school for lunch and a German class he selected as a project. EBCE students seldom take classes at the high school but Mike had a special interest in German since his grandparents speak the language.

Following German class, Mike returned to the learning center for an hour of work on other learning activities and then went to City Bank. “I often stayed there until 5:00 p.m.,” Mike said, even though high school hours ended at three.

Mike’s activities and interests widened after that first year in the EBCE program but his goal of becoming a computer programmer was reinforced by the learning level experience at City Bank. The start of a new hobby—collection of computer materials—also occurred during the time he spent at City Bank. “My employer instructor gave me some books to read that actually started the collection,” Mike said.

Mike’s interests in animals also was enhanced by his EBCE experience. Mike has always liked animals and his family has owned a horse since he was 12 years old. By picking blueberries Mike was able to save enough to buy his own colt two years ago. One of Mike’s favorite projects during the year related to his horse. The project was designed to help Mike with Basic Skills and to improve his critical thinking skills. Mike read about breeds of horses and how to train them. He then joined a 4-H group with hopes of training his horse for show.

Several months later, Mike again focused on animals for another EBCE project. This time he used the local zoo as a resource, interviewing the zoo manager and doing a thorough study of the Alaskan Brown Bear. Mike also joined an Explorer Scouting Club of volunteers to help at the zoo on a regular basis. “I really like working with the bears,” Mike reflected. “They were really playful. Did you know when they rub their hair against the bars it sounds like a violin?” Evaluation of the zoo project, one of the last Mike completed during the year, showed much improvement. The learning manager commented to Mike, “You are getting your projects done faster, and I think you are taking more time than you did at first to do a better job.”

Mike got off to a slow start in the area of Life Skills development. Like some of his peers, he went through a period described by one of the learning managers as “freedom shock.” When removed from the more rigid structure



normally experienced in a typical school setting, Mike tended to avoid his responsibility to the more “academic” side of his learning program. At first, Mike seldom followed up on commitments and often did not let the staff know what he was doing. By the end of the year, he had improved remarkably in both of these behavior areas.

Through the weekly writing required in maintaining his journal, Mike demonstrated a significant improvement in written communications both in terms of presenting ideas and feelings and in the mechanics of writing. Mike also noted an interesting change in his behavior. “I used to watch a lot of TV and never did any reading,” Mike said at the beginning of the following year. “I read two books last year and have completed eight more this summer. Now I go to the book instead of the television,” he added. Mike’s favorite reading materials are science fiction.

Mike also observed a difference in his attitude about homework. “After going to school for six hours I wouldn’t sit down and do homework. But in the EBCE program I wasn’t sitting in a classroom, so I didn’t mind going home with some more work on my journal or projects.”

Mike’s personal development was also undergoing change. Much of this change was attributed to one of his employer instructors, an elementary school teacher, who told him how important it is in the work world to wash and wear clean clothes. Both she and the project staff gave Mike much positive reinforcement when his dress improved. That same employer also told Mike that she was really interested in what he had to say and therefore wanted him to speak slower so he could be understood.

Mike’s school attendance improved while in the EBCE program. During the year, Mike missed only six days. This was better than the average absence for others in the program, which was found to be 12.3 days missed during the year, and much improved over his high school attendance.

Like a number of other EBCE students in his class, Mike went out on exploration level experiences but completed relatively few other program requirements during the first three months of the school year. By April, however, he was simultaneously working on eight different projects and pursuing a learning level experience at City Bank. By the time Mike completed his junior year he had finished nine of the required thirteen competencies, explored nine business sites, completed two learning levels and carried through on eleven projects. Two other projects were dropped during the year and one is uncompleted but could be finished in the coming year.

On a more specific level, Mike’s competencies included transacting business on a credit basis, maintaining a checking account, designing a comprehensive insurance program, filing taxes, budgeting, developing physical fitness, learning to cope with emergency situations, studying public agencies and operating an automobile.

Mike did not achieve the same level of success on all of his job sites. However, his performance consistently improved throughout the year. Mike criticized the exploration packages when he started them in the first months



of the program and, although he couldn't pinpoint how, said they could be better. His own reliance on the questions provided in the package was noted by the EBCE staff with a comment that he rarely followed up on any cues provided by the person he interviewed. The packets reflected Mike's disinterest in the exploration portion of EBCE work. They showed little effort and a certain sameness of remarks about his impressions at the various sites.

Mike explored career possibilities at an automobile dealer, an audiovisual repair shop, a supermarket, an air control manufacturer, an elementary school, a housing development corporation, a city public works, a junior high school and a bank services company.

Mike's first learning level experience was at the elementary school. At the end of three and one-half months the two teachers serving as his employer instructors indicated concern about attendance, punctuality, initiative in learning and amount of supervision needed to see that Mike's time was used constructively. Mike did show significant improvement in appropriate dress, personal grooming and quality of work on assignments.

Reports from the second learning level experience—at the computer department of the bank services company—showed a marked improvement. The employer instructor there rated Mike satisfactory in all aspects and by the time of the final evaluation gave excellent ratings in ten categories—attendance/punctuality, adhering to time schedules, understanding and accepting responsibility, observing employer rules, showing interest and enthusiasm, poise and self-confidence, using initiative in seeking opportunities to learn, using employer site learning resources, beginning assigned tasks promptly and completing tasks assigned.

During the latter part of the school year, Mike worked on several projects at once. He worked on a project on basic electricity and took a course on "Beginning Guitar" for project credit.

To improve his communications skills, Mike also worked on an intergroup relations project. This project grew out of an awareness by the staff that Mike liked other students but seemed to lack social interaction with his peers and the staff. Reports at the beginning of the year indicated that he appeared dependent and submissive and was an immature conversationalist. In response to these observations, Mike's learning manager negotiated project objectives and activities with him that would help improve his communications skills and help him solve some of his interpersonal problems. At the end of the year Mike noted a positive change related to his communications skills. "I can now speak up in groups," he said.

Mike's unfinished project related to his own experience and interests. He had moved to the Portland area from Canada ten years previously and frequently returns to see relatives. The project was on immigration laws and regulations in the functional citizenship area. At the same time, it will help Mike improve his grammar and spelling. Since students have the option of completing a project started during their junior year when they are a senior, Mike had a chance to finish the project this year.



Of the year Mike said, "It turned out even better than I thought." Things he liked best about the new experience in EBCE were working at his own speed, going to a job and having more freedom.

At the end of the year, Mike's tests showed significant increases in both reading and language skills. In the math and study skill area where he was already above average only slight increases were indicated.

Tests on attitudes, given both at the beginning and the end of the year, indicated positive gains in self-reliance, understanding of roles in society, tolerance for people with differences in background and ideas than his, and openness to change.

Aspirations did not change for Mike. He still wants to go into computer programming after finishing college. "When I started the year I really didn't know too much about computers. I feel now that I know a lot and want even more to make it my career."

*(The description of Mike's second year in EBCE are omitted. We pick up the case study after the second year description.)*

*Mike's Views of EBCE.* Mike felt that his EBCE experiences, especially the learning levels, had improved all of his basic skills. He felt that he had the freedom to do the kinds of things he wanted to do while at employer sites. These experiences, according to Mike, have strengthened his vocational choice in the field he wanted to enter and have caused him to look at educational and training requirements plus some other alternatives. For instance, Mike tried to enter the military, figuring it would be a good source of training in the field of computers, but was unable to because of a medical problem.

By going directly to job sites Mike has gotten a feel for the "real world" of work. He said his work at computer repair-oriented sites furthered his conceptions of the patience necessary when dealing with customers and fine degree of precision needed in the repair of equipment. He also discovered how a customer engineer takes a problem, evaluates it and solves it.

When asked about his work values Mike replied, "I figure if I get the right job, I'd work at it and try to do my best . . . in fact, I'm sure that even though I didn't like the job I'd still do more than I was asked to. . . . I'd work as hard as I could." Although he has always been a responsible person, he feels that his experiences in EBCE have made him more trustworthy. Mike also feels that he is now treated more like an adult because of his own attitudes. In fact, he feels he understands himself a lot more now.

Mike's future plans concern trying to get a job in computer programming at an automobile dealership or computer services company. He had previously done some computer work at the automobile dealership in relationship to a project in Explorer Scouts. He also wants more training in computer programming and has discussed these plans with the student coordinator and an EBCE secretary. His attitude towards learning is that it may not be fun but it is important.



When asked in which areas he made less growth than he had hoped to, Mike responded, "I really made a lot of growth in all areas." He credits the EBCE program for this, finding it more helpful than high school. It gives you the opportunity to "get out and meet more people and get to be able to communicate better with people out in the community."

Most of Mike's experiences at the high school were not too personally rewarding. He did start a geometry class there this year, but had to drop it as he had started late and could not catch up. Although he got along all right with the staff at the high school in the past he felt the teachers there had a "barrier between them and the students." The EBCE staff "treat you on a more individual type circumstance . . . have the time to talk to you." In EBCE you can "work at your own speed . . . don't have to be in the classroom."

Mike recommends the program to most of his friends, although some of his friends had already dropped out of school. He stated, "I would have paid to come into EBCE, I think it's really that good of a program. . . . In fact, I've learned more in these two years in EBCE than I have in the last four years at the high school." He did not even ask for reimbursement for travel expenses because he said he liked the program so much.

*The Views of His Parents.* When Mike first told his parents about the program they were concerned about what was going to be involved and whether it was a good program and educational. When interviewed in March, they felt that EBCE has helped Mike to be more mature and know where he is going.

Mike's parents said they were well informed by the EBCE staff in all areas. Mike tended to talk to them about his activities in EBCE, while the only thing he ever talked about at the high school was photography. Mike's career plans have not really changed since he entered EBCE and his parents have not tried to influence him, but EBCE has helped him to rule out mechanic and truck driving as possible careers.

Since beginning the EBCE program his parents have found Mike to be more mature, dependable and enthusiastic. He also became more reflective and concerned about the future. His writing improved and he read more.

There are no areas where his parents felt that EBCE did not help him and they rated the EBCE program highly in all areas.

*Test Progress Measures on Mike.* Although Mike showed a great improvement in almost all areas of the *Comprehensive Test of Basic Skills* during the first year of participation, his scores decline considerably during the second year. Especially significant were the declines in Mike's arithmetic applications and study skills scores.

Mike's attitudinal scores all showed a positive gain over the two-year total period, but also tended to decline during the second year of participation. On the semantic differential, Mike scored significantly below the EBCE mean at FY 75 posttest on the *community resources, adults, learning and work* scales.



Mike showed continued growth over the two-year period on the *work, self-reliance, communication, role, and trust* scales of the *Psychosocial Maturity Scale*. He was significantly above the EBCE posttest means on the *work, role, and social commitment* scales and below average on only the *openness to change* scale. The *openness to change* score also showed a significant decline over the year.

The staff rated Mike on seven student behaviors. At the beginning of the year he was significantly above the EBCE mean on “applies knowledge of his/her own aptitudes, interests, and abilities to potential career interests” and below the mean on “understands another person’s message and feelings.” At posttest time he was still below the EBCE mean on this latter behavior as well as on “demonstrates willingness to apply Basic Skills to work tasks and to vocational interests.”

Over the course of the two years in the EBCE program Mike’s scores on the *Self-Directed Search* (SDS) showed little change in pattern, although the number of interests and competencies did expand. Overall, realistic (R) occupations decreased and enterprising (E) occupations increased as his code changed from RCI (where C is conventional and I is investigative occupations) at pretest FY 74 to ICR at pretest FY 75 (a classification which includes computer operators and equipment repairers) to CEI at posttest FY 75. However, the I was only one point stronger than the R and the CER classification includes data processing workers. Thus, Mike’s SDS codes appeared very representative of his desired occupational future.

*Evaluators’ Reflections.* Mike’s dramatic declines in attitudes and basic skill scores reflect behavior changes which occurred during the second half of his second year of the program and were detected by a number of people. In February at a student staffing meeting his learning manager reported of Mike that “no progress is seen in this zone with projects . . . still elusive . . . coasting right now . . . may end up in trouble.” The prescription was to “watch him—make him produce . . . find out where he is.” However, at the end of the next to last zone in mid-May the report was still “the elusive butterfly! (Mike) needs to get himself in high gear to get everything completed *on time!!!*” Since the posttesting was completed before this time, Mike probably coasted through the posttesting as well.

Other data suggesting his lack of concern and involvement during the second half of his senior year was attendance. Although he missed only two days the first half of the year, he missed thirteen days during the second half.

Mike showed a definite change in some of his personality characteristics over the two years he spent in the EBCE program. In the beginning of the program he was totally lacking in social skills and self-confidence. By the time he graduated, he had made great strides in his social skills (although there was still much room for improvement). However, his self-confidence had grown to the point of overconfidence. Indeed the employer instructor on his last learning level spent a good deal of time trying to get Mike to make a realistic appraisal of his own capabilities.



When interviewed after graduation, Mike was working six evenings a week at a restaurant where he worked part-time for the last year. He hopes to work there for about a year, working his way up to cook, and then go to a business college for a year to study computers.



### Appendix 9.3

## EXCERPTS FROM AN ILLUSTRATIVE INTERVIEW ANALYSIS: REFLECTIONS ON OUTCOMES FROM PARTICIPANTS IN A WILDERNESS EDUCATION PROGRAM

by Jeanne Campbell and  
Michael Patton

Experiences affect people in different ways. This experiential education truism means that the individual outcomes, impacts, and changes that result from participation in some set of activities are seldom predictable with any certainty. Moreover, the meaning and meaningfulness of such changes as do occur are likely to be highly specific to particular people in particular circumstances. While the individualized nature of learning is a fundamental tenet of experiential education, it is still important to stand back from those individual experiences in order to look at the patterns of change that cut across the specifics of person and circumstances. One of the purposes of the evaluation of the Learninghouse Southwest Field Training Project was to do just that—to document the experiences of individuals and then to look for the patterns that help provide an overview of the project and its impacts.

A major method for accomplishing this kind of reflective evaluation was the conduct of follow-up interviews with the eleven project participants. The first interviews were conducted at the end of October 1977, three weeks following the first field conference in the Gila Wilderness of New Mexico. The second interviews were conducted during the third week of February, three weeks after the wilderness experience in the Kofa Mountains of Arizona. The third and final interviews were conducted in early May following the San Juan River conference in southern Utah. All interviews were conducted by telephone. The average interview took twenty minutes with a range from fifteen to thirty-five minutes. Interviews were tape-recorded and transcribed for analysis.

The interviews focus on three central issues: (1) How has your participation in the Learninghouse Project affected you personally? (2) How has your participation in the project affected you professionally? (3) How has your participation in the Learninghouse Project affected your institution? In the pages which follow, participant responses to these questions are presented and analyzed. The major purpose of the analysis was to organize participant responses in such a way that overall patterns would become clear. The emphasis throughout is on letting participants speak for themselves. The challenge for the evaluators was to present participant responses in a cogent fashion that integrates the great variety of experiences and impacts recorded during the interviews.

### PERSONAL CHANGE

*“How has your participation in the Learninghouse Project affected you personally? What has been the impact of the project on you as a person?”*



Questions about personal change generated more reactions from participants than subsequent questions about professional and institutional change. There is an intensity to these responses about individual change that makes it clear just how significant these experiences were in stimulating personal growth and development. Participants attempted throughout the interviews to indicate that they felt differently about themselves as persons because of their Learninghouse experiences. While such personal changes are often difficult to articulate, the interviews reflect a variety of personal impacts.

### CONFIDENCE: A SENSE OF SELF

During the three weeks in the wilderness, participants encountered a number of opportunities to test themselves. Can I carry a full pack day after day, uphill and downhill? Can I make it up that mountain? Do I have anything to contribute to the group? As participants encountered and managed stress, they learned things about themselves. The result was often an increase in personal confidence and a greater sense of self.

It's really hard to say that LH did one thing or another. I think increased self-confidence has helped me do some things that I was thinking about doing. And I think that came, self-confidence came about largely because of the field experiences. I, right after we got back, I had my annual merit evaluation meeting with my boss, and at that I requested that I get a, have a change in title or a different title, and another title really is what it amounts to, and that I be given the chance for some other responsibilities that are outside the area that I work in. I want to get some individual counseling experience, and up to this point I have been kind of hesitant to ask for that, but I feel like I have a better sense of what I need to do for myself and that I have a right to ask for it at least [Cliff, post-Kofas].

I guess something that has been important to me in the last couple of trips and will be important in the next one is just the outdoor piece of it. Doing things that perhaps I'd not been willing to attempt before whatever reason. And finding I'm better at it than expected. Before I was afraid [Charlene, post-Kofas].

The interviews indicate that increased confidence came not only from physical accomplishments but also—and especially—from interpersonal accomplishments.

After the Kofas I achieved several things that I've been working on for two years. Basically, the central struggle of the last two years of my life has been to no longer try to please people. No matter what my own feelings and needs are I try to please you. And in the past I had done whatever another person wanted me to do in spite of my own feelings



and needs. And to have arrived at a point where I could tend to my own feelings and take care of what I needed to do for me is by far the most important victory I've won . . . a major one.

In the Kofas, I amazed myself that I didn't more than temporarily buy into how . . . I was being described . . . when I didn't recognize myself yet. And that's new for me. In the past I'd accept others' criticisms of me as if they were indeed describing me . . . and get sucked into that. And I felt that was an achievement for me to hold onto my sense of myself in the face of criticisms has long been one of my monsters I've been struggling with, so to hold onto me is, especially as I did, was definitely an achievement [Billie, post-Kofas].

I've been paying a lot of attention to not looking for validation from other people. Just sticking with whatever kinds of feelings I have and not trying to go outside of myself . . . and lay myself on a platter for approval. I think the project did have a lot to do with that, especially this second trip in the Kofas [Greg, post-Kofas].

I would say the most important thing that happened to me was being able to talk to other people quite honestly about, I think really about their problems more than mine. That's very interesting in that I think that I had, I think I had an effect upon Billy and Charlene both. As a result of that it gave me a lot more confidence and positive feelings. Do you follow that? Where rather than saying I had this problem and I talked to somebody and they solved it for me, it was more my helping other people to feel good about themselves that made me feel more adequate and better than myself [Rod, post-Gila].

Another element of confidence concerns the extent to which one believes in one's own ideas—a kind of intellectual confidence.

I think if I take the whole project into consideration, I think that I've gained a lot of confidence myself in some of the ideas that I have tried to use, both personally and let's say professionally. Especially in my teaching aspects, especially teaching at a woman's college where I think one of our roles is not only to teach women subject matter, but also to teach them to be more assertive. I think that's a greater component of our mission than normally would have it at most colleges. I think that a lot of the ideas that I had about personal growth and about my own interactions with people were maybe reinforced by the LH experience, so that I felt more confident about them, and as a result they have come out more in my dealings with people. I would say specifically in respect to a sort of a more humanistic approach to things [Rod, post-Kofas].

Increased confidence for participants was often an outcome of learning that they could do something new and difficult. At other times, however,



increased confidence emerged as a result of finding new ways to handle old and difficult situations, for example, learning how to recognize and manage stress.

A change I've noticed most recently and most strongly is the ability to recognize stress. And also the ability to recognize that I can do a task without needing to make it stressful which is something I didn't know I did. So what I find I wind up doing, for example, is when I've had a number of things happen during the day and I begin to feel myself keying up I find myself very willing to say both to close friends and to people I don't know very well, I can't deal with this that you're bringing me. Can we talk about it tomorrow? This is an issue that really needs a lot of time and a lot of attention. I don't want to deal with it today, can we talk later . . . etc. So I'm finding myself really able to do that. And I'm absolutely delighted about it.

*(Whereas before you just piled it on?)*

Exactly. I'd pile it and pile it until I wouldn't understand why I was going in circles [Charlene, post-Kofas].

## PERSONAL CHANGE—OVERVIEW

The personal outcomes cited by Learninghouse participants are all difficult to measure. What we have in the interviews are personal perceptions about personal change. The evidence, in total, indicates that participants felt differently and, in many cases, behaved differently as a result of their project participation. Different participants were affected in different ways and to varying extents. One participant reported virtually no personal effects from the experiences.

And as far as the effect it had on me personally, which was the original question, okay, to be honest with you, to a large degree it had very little effect, and that's not a dig on the program, because at some point in people's lives I think things start to have smaller effect, but they still have effect. So I think that for me, what it did have an effect on was tolerance. Because there were a lot of things that occurred on the trip that I didn't agree with. And still don't agree, but I don't find myself to be viciously in disagreement any longer, just plainly in disagreement. So it was kind of like before, I didn't want to listen to the disagreement, or I wanted to listen to it but resolve it. Now, you know, there's a third option, that I can listen to it, continue to disagree with it and not mind continuing to listen to it [Cory, post-San Juan].

The more common reaction, however, was surprise at just how much personal change occurred.



My expected outcome was increase the number of contacts in the Southwest, and *everyone* of my expected outcomes were professional. That you know, much more talk about potential innovations in education and directions to go, and you know, field based education what that's about, and I didn't expect at all, which may not be realistic on my part, but at least I didn't expect at all—the personal impact [Charlene, post-Gila].

For others the year's participation in Learninghouse was among the most important learning experiences of a lifetime, precisely because the project embraced personal as well as professional growth.

I've been involved in institutions and in projects as an educator, let's say, for 20 years. I mean I started out teaching in high school, going to the NSF institutions during the summertime and I've gone to a lot of Chatauqua things and a lot of conferences, you know, of various natures. And I really think that this project has by far the greatest . . . has had by far the greatest impact on me. And I think that the reason is that in all the projects that I've had in the past . . . they've been all very specifically oriented towards one subject or toward one . . . more of a, I guess, more of a science, more of a subject matter orientation to them. Whereas this having a process orientation has a longer effect. I mean a lot of the things I learn in these instances is out of date by now and you keep up with the literature, for example, and all that and maybe that stimulates you to keep up . . . but in reality as far as a growth thing on my part, I think on the part of other participants, I think that this has been phenomenal. And I just think that this is the kind of thing that we should be looking towards funding on any level, federal, or any level [Rod, post-San Juan].

We come now to a transition point in this report. Having reported participants' perceptions about personal change, we want to report the professional outcomes of the Learninghouse project. The problem is that in the context of a holistic experience like the Southwest Field Training Project, the personal-professional distinction becomes arbitrary. A major theme running throughout discussions during the conferences was the importance of reducing the personal-professional schism, the desirability of living an integrated life and being an integrated self. This theme is reflected in the interviews, as many participants had difficulty responding separately to questions about personal versus professional change.

### **PERSONAL/PROFESSIONAL CHANGE**

Analytically, there is at least a connotative difference between personal and professional change. For evaluation purposes, we tried to distinguish



one from the other as follows: personal changes concern the thoughts, feelings, behaviors, intentions, and knowledge people have about themselves; professional changes concern the skills, competences, ideas, techniques, and processes people use in their work. There is, however, a middle ground. How does one categorize changes in thoughts, feelings, and intentions about competences, skills, and processes? There are changes in the person that affect that person's work. This section is a tribute to the complexity of human beings in defying the neat categories of social scientists and evaluators. This section reports changes that, for lack of a better nomenclature, we have called simply personal/professional impacts.

The most central and most common impact in this regard concerned changes in personal perspective that affected fundamental notions about and approaches to the world of work. The wilderness experiences and accompanying group processes permitted and/or forced many participants to stand back and take a look at themselves in relation to their work. The result was a changed perspective. The following three passages are from interviews conducted after the first field conference in the Gila, a time when the contrasts provided by the first wilderness experience seemed to be felt most intensely.

The trip came at a real opportune time. I've been on this new job about 4-5 weeks and was really getting pretty thoroughly mired in it, kind of overwhelmed by it, and so it came after a particularly hellish week, so in that sense it was just a critical, really helpful time to get away. To feel that I had, to remember that I had some choices, both in terms of whether I stayed here or went elsewhere, get some perspective of what it was I actually wanted to accomplish in higher education rather than just surviving to keep my sanity. And it gave me some, it renewed some of my ability to think of doing what I wanted to do here at the University, or trying to, that there were things that were important for me to do rather than just handling the stuff that poured across my desk [Henry, post-Gila].

I think it's helped make me become more creative, and just, and that's kind of tied in with the whole idea of the theory of experiential education. And the way we approached it on these trips. And so for instance I'm talking with my wife the other night, after I got Laura's paper that she'd given in Colorado, and I said you oughta read this because you can go out and teach history and you know, experientially. Then I gave her an idea of how I would teach frontier history for instance, and I don't know beans about frontier history. But it was an idea which, then she told another friend about it, and this friend says oh, you can get a grant for that. You know. So that was just a real vivid example, and I feel like, it's, I've been able to apply, or be creative in a number of different situations, I think just because I give myself a certain freedom, I don't know, I can't quite pinpoint what brought it about, but I just, feel more creative in my work [Cliff, post-San Juan].



You know my biggest problem is I've been trying to save the world, and what I'm doing is pulling back. Because, perhaps the way I've been going about it has been wrong or whatever, but at least my motives are clearer and I know much more directly what I need and what I don't need and so I'm more open but less, yeah, as I said, I've been in a let's save the world kind of thing, now I feel more realistic and honest [Charlene, post-Gila].

I've been thinking about myself and my relationship to men and my boss, and especially to ideas about fear and risk . . . I decided that I needed to become a little more visible at the department. After the October experience, I just said I was a bit more ready to become visible at the department level. And I volunteered then to work on developing a department training policy and develop the plan and went down to the department and talked to the assistant about it and put myself in a consulting role while another person was assigned the actual job of doing it. And I think that I was ready to make that decision and act on it after I first of all got clear that I was working on male-female relationships. My department has a man, again, not a terribly easy one to know, so it's a risk for me to go talk with him and yet I did it. I was relatively comfortable and felt very good and very pleased with myself that I had done that and I think that's also connected [Billie, post-Kofas].

The connection between personal changes and professional activities was an important theme throughout the Learninghouse Project. The passages reported in this section illustrate how that connection took hold in the minds and lives of project participants. As we turn now to more explicit professional impacts, it is helpful to keep in mind the somewhat artificial and arbitrary nature of the personal-professional distinction.

*(Omitted are sections on changed professional knowledge about experiential education; use of journals; group facilitation skills; individual professional skills; and personal insights regarding work and professional life; and the specific projects participants undertook professionally. Also omitted are sections on institutional impacts. We pick up the report in the concluding section.)*

## FINAL REFLECTIONS

Personal change . . . professional change . . . institutional change . . . Evaluation categories aim at making sense out of an enormously complex reality. The reflections by participants throughout the interviews make it clear that most of them came away from the Learninghouse program feeling



changes in themselves. Something had touched them. Sometimes it meant a change in perspective that would show up in completely unexpected ways.

For one thing, I just finished the purchase of my house. First of all, that's a new experience for me. I've never done it before. I've never owned a home and never even wanted to. It seemed odd to me that my desire to "settle down" or make this type of commitment to a *place* occurred just right after the Gila trip. Just sort of one of those things that I woke up and went, "Wow, I want to stay here. I like this place. I want to buy it." And I had never in my life lived in a house or a place that I felt that way about. I thought that was kind of strange. And I do see that as a function of personal growth and stability. At least some kind of stability.

Other areas of personal growth: one has been, and this kind of crosses over I think into the professional areas, and that would be an ability to gain perspective. Certainly the trips I think . . . incredibly valuable for gaining perspective on what's happening in my home situation, my personal life, my professional life . . . the whole thing. And it has allowed me to focus on some priority types of things for me. And deal with some issues that I've been kind of dragging on for years and years and not really wanting to face up with them or deal with them. And I have been able to move on and move through those kinds of things in the last 6 or 9 months or so to a much greater extent than ever before [Tom, post-San Juan].

Other participants came away from the wilderness experiences with a more concrete orientation that they could apply to work, play, and life.

The thing that I realized as I was trying to make some connections between the river and raft trip, was that in some ways I can see the parallels of my life being kind of like our raft trip was, and the rapids, or the thrill ride, and they're a lot of fun, but it's nice to get out of them for a while and dry off. It's nice sometimes to be able to just drift along and not worry about things. But a lot of it also is just hard work. A lot of times I wish I could get out of it and go a different way, and that's been kind of a nice thing for me to think about and kind of a viewpoint to have whenever I see things in a lull or in a real high speed pace, that I can say, "Okay, I'm going to be in this for a while, but I'm going to come out of it and go into something else". And so that's kind of a metaphor that I use as somewhat of a philosophy or point of view that's helpful as I go from day to day [Cliff, post-San Juan].

A common theme that emerged as participants reflected on their year's involvement with Learninghouse was a new awareness of options, alternatives, and possibilities.



I would say that if I have one overall comment, the effect of the first week overall, is to renew my sense of the broader possibilities in my job and in my life. Opens things to me. I realize that I have a choice to be here and be myself. And since I have a choice, there are responsibilities. Which is a good feeling [Henry, post-Gila].

I guess to me what sticks out overall is that the experience was an opportunity for me to step out of the rest of my life and focus on it and evaluate it, both my personal life and my work, professional life aspect [Michael, post-San Juan].

As participants stood back and examined themselves and their work they seemed to discover a clarity that had previously been missing. Perspective, awareness, clarity . . . stuff of which personal/professional/institutional change is made.

I think I had a real opportunity to explore some issues of my own worth with a group of people who were willing to allow me to explore those. And it may have come later, but it happened then. On the Learninghouse, through the Learninghouse . . . and I think it speeded up the process of growing for me in that way, accepting my own worth, my own ideas about education, about what I was doing, and in terms of being a teacher it really aided my discussions of people and my interactions. It really gave me a lot of focus on what I was doing. I think I would've muddled around a long time with some issues that I was able to, I think, gain some clarity on pretty quickly by talking to people who were sharing their experience and were working towards the same goals, self-directed learning, and experiential education [Greg, post-San Juan].

I think what happened is that for me it served as a catalyst for some personal changes, you know, the personal, institutional, they're all wound up, bound up together. I think I was really wrestling with jobs and career and so on. For me the whole project was a catalyst, a kind of permission to look at things that I hadn't looked at before. One of the realizations, one of the insights that I had in the process was, kind of neat on my part, to become concrete, specific in my actions in my life, no matter whether that was writing that I was doing, or if it was in my job, or whatever it was. But to really pay attention to that. I think that's one of the things that happened to me [Peter, post-San Juan].

These statements from interviews do not represent a final assessment of the impacts of the Learninghouse Southwest Field Training Project. Several participants resisted the request to make summary statements about the effects and outcomes of their participation in the program because they didn't want to force premature closure.



*(Can you summarize the overall significance of participation in the project?)*

I do want to make a summary, and I don't again. . . . It feels like the words aren't easy and for me being very much a words person, that's unusual. It's not necessarily that the impact hasn't been in the cognitive areas. There have been some. But what they've been, where the impact has been absolutely overwhelming is in the affective areas. Appreciation of other people, appreciation of this kind of education. Though I work in it, I haven't *done* it before! A real valuing of people, the profession, of my colleagues in a sense that I never had before. . . .

The impact feels like it's been dramatic, and I'm not sure that I can say exactly how. I'm my whole . . . it all can be summarized perhaps by saying I'm much more in control. In a good kind of sense. In accepting risk and being willing to take it; accepting challenge and being willing to push myself on that; accepting and understanding more about working at the edge of my capabilities . . . what that means to me. Recognizing very comfortably what I can do and feeling good about that confidence, and recognizing that what I haven't yet done, and feeling okay about trying it. The whole perception of confidence has changed [Charlene, post-San Juan].

The Learninghouse program was many things—the wilderness, a model of experiential education, stress, professional development—but most of all, the project was the people who participated. In response after response participants talked about the importance of the people to everything that happened. Because of the dominance of that motif throughout the interviews, we want to end this report with that highly personal emphasis.

I said before I think that to know some people, that meant a lot to me, people who were also caring. And people who were also involved, very involved in some issues, philosophical and educational, that were pretty basic not only to education, but to living. Knowing these people has been really important to me. It's given me a kind of continuity and something to hold onto in the midst of a really frustrating, really difficult situation where I didn't have people where I could get much feedback from, or that I could share much thinking about, talking about, and working with. It's just kind of basic issues. That kind of continuity is real important to just my feelings, important to myself. Feeling like I have someplace to go. . . .

Sometimes I feel funny about placing so much emphasis on the people. . . . But the people have really meant a lot to me as far as putting things together for myself. Being able to have my hands in something that might, that really offers me a way to go [Greg, post-San Juan].



## REFERENCES

- Alkin, Marvin C. Evaluation: Who needs it? Who cares? *Studies in Educational Evaluation*, 1975, 1, 201-212.(a)
- Alkin, Marvin C. Framing the decision context. *AERA Cassette Series in Evaluation*. Washington, DC: American Educational Research Association, 1975. (b)
- Alkin, Marvin C., Daillak, Richard, & White, Peter. *Using evaluations: Does evaluation make a difference?* Beverly Hills, CA: Sage, 1979.
- Anderson, Richard B. The effectiveness of follow through: What have we learned? Presented at the annual meeting of the American Educational Research Association, New York City, April 5, 1977.
- Bandler, Richard, & Grinder, John. *Patterns of the hypnotic techniques of Milton H. Erickson, M.D., vol. 1*. Cupertino, CA: Meta Publications, 1975.
- Barker, Roger G. *Ecological psychology*. Stanford: Stanford University Press, 1968.
- Bateson, Gregory. The pattern which connects. *CoEvolution Quarterly*, 1978, Summer, 5-15.
- Becker, Howard, & Geer, Blanche. Participant observation and interviewing: A comparison. In W.J. Filstead (Ed.), *Qualitative methodology*. Chicago: Markham, 1970.
- Bloch, Arthur. *Murphy's law and other reasons things go wrong*. Los Angeles, CA: Price/Stern/Sloan, 1979.
- Blumer, Herbert. *Symbolic interactionism*. Englewood Cliffs, NJ: Prentice-Hall, 1969.
- Blumer, Herbert. Methodological principles of empirical science. In N.K. Denzin (Ed.), *Sociological methods: A sourcebook*. New York: McGraw-Hill, 1978.
- Boas, Franz. Recent anthropology. *Science*, 1943, 98, 311-314, 334-337.
- Bogdan, Robert, & Taylor, Stephen J. *Introduction to qualitative methods*. New York: John Wiley, 1975.
- Bruyn, Severyn. The methodology of participant observation. *Human Organization*, 1963, 21, 224-235.
- Bruyn, Severyn. *The human perspective in sociology: The methodology of participant observation*. Englewood Cliffs, NJ: Prentice-Hall, 1966.
- Bussis, Anne, Chittenden, Edward A., & Amarel, Marianne. *Methodolgy in educational evaluation and research*. Princeton, NJ: Educational Testing Service, 1973. (mimeo)
- Campbell, Donald T. Qualitative knowing in action research. Presented at the 1974 Annual Meeting of the American Psychological Association, New Orleans, 1974.
- Campbell, Donald T., & Stanley, Julius C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1966.



- Carini, Patricia F. *Observation and description: An alternative methodology for the investigation of human phenomena*. North Dakota Study Group on Evaluation Monograph Series. Grand Forks: University of North Dakota, 1975.
- Castenada, Carlos. *Journey to Ixtlan*. New York: Pocket Books, 1973.
- Cervantes Saavedra, Miguel de. *Don Quixote*. New York: Signet Classics, 1964.
- Conrad, Joseph. *Heart of darkness*. New York: Dell, 1960.
- Cronbach, Lee J. Beyond the two disciplines of scientific psychology. *American Psychologist*, 1975, 30, 116-127.
- Denzin, Norman K. The logic of naturalistic inquiry. In N.K. Denzin (Ed.), *Sociological methods: A sourcebook*. New York: McGraw-Hill, 1978. (a)
- Denzin, Norman K. *The research act*. New York: McGraw-Hill, 1978. (b)
- Douglas, Jack D. *Investigative social research: Individual and team field research*. Beverly Hills, CA: Sage, 1976.
- Edwards, Ward, Guttentag, Marcia, & Snapper, Kurt. A decision-theoretic approach to evaluation research. In E.L. Struening and M. Guttentag (Eds.), *Handbook of evaluation research, vol. 1*. Beverly Hills, CA: Sage, 1975.
- Elmore, Richard F. Follow through planned variation. In W. Williams and R. Elmore (Eds.), *Social program implementation*. New York: Academic Press, 1976.
- Fehrenbacher, Harry L., Owens, Thomas R., & Haenn, Joseph F. *The use of student case study methodology in program evaluation*. Research Evaluation Development Paper Series No. 10. Portland, OR: Northwest Regional Educational Laboratory, 1976.
- Feiman, Sharon. *Teacher curriculum work center: A descriptive study*. North Dakota Study Group on Evaluation Monograph Studies. Grand Forks: University of North Dakota, 1975.
- Feiman, Sharon. Evaluating teacher centers. *School Review*, 1977, 8, 395-411.
- Filstead, William J. (Ed.). *Qualitative methodology*. Chicago: Markham, 1970.
- Filstead, William J. Qualitative methods and program evaluation: some issues and observations. Presented at the Evaluation Research Society annual meeting, Washington, D.C., 1978
- Frake, Charles. The ethnographic study of cognitive systems. In T. Gladwin and W.H. Sturtevant (Eds.), *Anthropology and human behavior*. Washington, DC: Anthropology Society of Washington, 1962.
- Garfinkle, Harold. *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall, 1967.
- Gephart, William J. On truth. *CEDR Quarterly*, 1978, 11, 2-6.
- Glaser, Barney G., & Strauss, Anselm L. *Discovery of grounded theory: Strategies for qualitative research*. Chicago: AVC, 1967.
- Glaser, Edward M., & Backer, Thomas E. A clinical approach to program evaluation. *Evaluation*, 1972, Fall, 54-59.
- Glazer, Myron. *The research adventure: Promise and problems of fieldwork*. New York: Random House, 1972.
- Guba, Egon G. *Toward a methodology of naturalistic inquiry in educational evaluation*. CSE Monograph Series in Evaluation No. 8. Los Angeles: Center for the Study of Evaluation, University of California, Los Angeles, 1978.
- Hage, Jerald. *Techniques and problems of theory construction in sociology*. New York: Wiley Interscience, 1972.
- House, Ernest R. *The logic of evaluative argument*. CSE Monograph Series in Evaluation No. 7. Los Angeles: Center for the Study of Evaluation, University of California, Los Angeles, 1977.
-



- House, Ernest R. Assumptions underlying evaluation models. *Educational Researcher*, 1978, 7, 4-12.
- Hyman, Herbert H. *Interviewing in social research*. Chicago: University of Chicago Press, 1954.
- Johnson, John M. *Doing field research*. Beverly Hills, CA: Sage, 1975.
- Junker, Buford H. *Field work: An introduction to the social sciences*. Chicago: University of Chicago Press, 1960.
- Kartunnen, Lauri. Remarks on presuppositions. Presented at the Texas Conference on Performances, Conversational Implicature and Presuppositions, March 1973. (mimeo)
- Katzer, Jeffrey, Cook, Kenneth H., & Crouch, Wayne W. *Evaluating information: A guide for users of social science research*. Reading, MA: Addison-Wesley, 1978.
- Kourilsky, Marilyn. An adversary model for educational evaluation. *Evaluation Comment*, 1974, 4.
- Kuhn, Thomas. *The structure of scientific revolutions*. Chicago: University of Chicago Press, 1970.
- Lang, K., & Lang, G.E. Decisions for Christ: Billy Graham in New York City. In M. Stein, A.J. Vidich, and D.M. White (Eds.), *Identity and anxiety*. New York: Free Press, 1960.
- Levine, Murry. Scientific method and the adversary. *American Psychologist*, 1974, September, 666-677.
- Lofland, John. *Analyzing social settings*. Belmont, CA: Wadsworth, 1971.
- McLaughlin, Milbrey. Implementation as mutual adaptation. In W. Williams and R.F. Elmore (Eds.), *Social program implementation*. New York: Academic Press, 1976.
- Mead, Margaret. *Letters from the field, 1925-1975*. New York: Harper & Row, 1977.
- Mills, C. Wright. *The sociological imagination*. New York: Oxford University Press, 1961.
- Moos, Rudolf. *Evaluating correctional and community settings*. New York: Wiley Interscience, 1975.
- Morris, Lynn Lyons, & Fitz-Gibbon, Carol Taylor. *Evaluator's handbook*. Beverly Hills, CA: Sage, 1978. (a)
- Morris, Lynn Lyons, & Fitz-Gibbon, Carol Taylor. *How to deal with goals and objectives*. Beverly Hills, CA: Sage, 1978. (b)
- Morris, Lynn Lyons, & Fitz-Gibbon, Carol Taylor. *How to design a program evaluation*. Beverly Hills, CA: Sage, 1978. (c)
- Myrdal, Gunnar. *Objectivity in social research*. New York: Random House/Pantheon, 1969.
- NWREL (Northwest Regional Educational Laboratory). *3-on-2 evaluation report, 1976-77, vols. I, II, and III*. Portland, OR: Northwest Regional Educational Laboratory, 1977.
- Olson, Ruth Anne. *A value perspective on evaluation*. Marcy Open School, Minneapolis Public Schools, 1974. (mimeo)
- Owens, Thomas. *Education evaluation by adversary proceeding*. In E. House (Ed.), *School evaluation: The politics and process*. Berkeley, CA: McCutchan, 1973.
- Owens, Thomas, Haenn, Joseph F., & Fehrenbacher, Harry L. *The use of multiple strategies in the evaluation of an experience-based career education program*. Research Evaluation Development Paper Series No. 9. Portland, OR: Northwest Regional Educational Laboratory, 1976.



- Parlett, Malcolm & Hamilton, David. Evaluation as illumination: A new approach to the study of innovatory programs. In G.V. Glass (Ed.), *Evaluation studies review annual, vol. 1*. Beverly Hills, CA: Sage, 1976.
- Partnow, Elaine. *The quotable woman, 1800-on*. Garden City, NY: Anchor Books, 1978.
- Patton, Michael Quinn. *Alternative evaluation research paradigm*. North Dakota Study Group on Evaluation Monograph Series. Grand Forks: University of North Dakota, 1975.
- Patton, Michael Quinn. *Utilization-focused evaluation*. Beverly Hills, CA: Sage, 1978.
- Payne, Stanley L. *The art of asking questions*. Princeton, NJ: Princeton University Press, 1951.
- Peers, John, & Bennett, Gordon *1001 logical laws . . . for all walks of life*. New York: Doubleday, 1979.
- Pelto, Pertti J., & Pelto, Gretel H. *Anthropological research: The structure of inquiry*. Cambridge: Cambridge University Press, 1978.
- Perrone, Vito. *The abuses of standardized testing*. Bloomington, IN: Phi Delta Kappa Educational Foundation, 1977.
- Perrone, Vito, Patton, Michael Quinn, & French, Barbara. *Does accountability count without teacher support?* Minneapolis: Center for Social Research, University of Minnesota, 1976.
- Polanyi, Michael, & Prosch, Harry. *Meaning*. Chicago: University of Chicago Press, 1975.
- Powdermaker, Hortense. *Stranger and friend*. New York: W.W. Norton, 1966.
- Provus, Malcolm. *Discrepancy evaluation for educational program improvement and assessment*. Berkeley, CA: McCutchan, 1971.
- Richardson, Stephen A., Dohrenwend, Barbara Snell, & Klein, David. *Interviewing, its forms and functions*. New York: Basic Books, 1965.
- Rossi, Peter H., & Williams, W. (Eds.). *Evaluating social programs: Theory, practice, and politics*. New York: Seminar Press, 1972.
- Rossi, Peter H., Freeman, Howard E., & Wright, Sonia R. *Evaluation: A systematic approach*. Beverly Hills, CA: Sage, 1979.
- Sanders, J., & Cunningham, D. A structure for formative evaluation in product development. *Review of Educational Research*, 1973, 43, 217-236.
- Sanders, J., & Cunningham, D. *Techniques and procedures for formative evaluation*. Research Evaluation Development Paper Series No. 2. Portland, OR: Northwest Regional Educational Laboratory, 1974.
- Scott, Myrtle, & Eklund, Susan J. Ecological methods in the study of administrative behavior. Presented at the 1979 American Educational Research Meetings, San Francisco, 1979.
- Schriren, Michael. The methodology of evaluation. In R.W. Tyler, R.M. Gagne, and M. Scriven (Eds.), *Perspectives of curriculum evaluation*. AERA Monograph Series on Curriculum Evaluation No. 1. Chicago: Rand McNally, 1967.
- Scriven, Michael. Prose and cons about goal-free evaluation. *Evaluation Comment*, 1972, 3, 1-7. (a)
- Scriven, Michael. Objectivity and subjectivity in educational research. In L.G. Thomas (Ed.), *Philosophical redirection of educational research: The seventy-first yearbook of the National Society for the Study of Education*. Chicago: University of Chicago Press, 1972. (b)



- Shah, Idries. *The exploits of the incomparable Mullah Nasrudin*. New York: Dutton, 1972.
- Shah, Idries. *The subtleties of the inimitable Mulla Nasrudin*. New York: Dutton, 1973.
- Shapiro, Edna. Educational evaluation: Rethinking the criteria of competence. *School Review*, 1973, November, 523-549.
- Shils, Edward A. Social inquiry and the autonomy of the individual. In D. Lerner (Ed.), *The human meaning of the social sciences*. Cleveland, OH: Meridian, 1959.
- Smith, Nick L. Truth, complementarity, utility, and certainty. *CEDR Quarterly*, 1978, 11, 16-17.
- Stake, Robert E. *Evaluating the arts in education: A responsive approach*. Columbus, OH: Charles E. Merrill, 1975.
- Stake, Robert E. The case study method in a social inquiry. *Educational Researcher*, 1978, 7, 5-8.
- Stenhouse, Lawrence. *Case study as a basis for research in a theoretical contemporary history of education*. East Anglia, England: Centre for Applied Research in Education, University of East Anglia, 1977.
- Strike, Kenneth. Explaining and understanding: The impact of science on our concept of man. In L.G. Thomas (Ed.), *Philosophical redirection of educational research: The seventy-first yearbook of the National Society for the Study of Education*. Chicago: University of Chicago Press, 1972.
- Thompson, Mark. *Evaluation for decision in social programmes*. Lexington, MA: D.C. Heath, 1975.
- Trend, M.G. On the reconciliation of qualitative and quantitative analyses: A case study. *Human Organization*, 1978, 37, 345-354.
- Trow, Martin. Comment on "Participant observation and interviewing: A comparison." In W.J. Filstead (Ed.), *Qualitative methodology*. Chicago: Markham, 1970.
- Tucker, Eugene. The follow through planned variation experiment: What is the payoff? Presented at the annual meeting of the American Educational Research Association, New York City, April 5, 1977.
- Watkins, Mary M. *Waking dreams*. New York: Harper & Row, 1977.
- Wax, Rosalie H. *Doing fieldwork: Warnings and advice*. Chicago: University of Chicago Press, 1971.
- Webb, Eugene J., Campbell, Donald T., Schwartz, Richard, & Sechrest, Lee. *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago: Rand McNally, 1966.
- Weiss, Carol (Ed.). *Evaluating action programs*. Boston: Allyn & Bacon, 1972. (a)
- Weiss, Carol (Ed.). *Evaluation research: Methods of assessing program effectiveness*. Englewood Cliffs, NJ: Prentice-Hall, 1972. (b)
- Willems, Edwin P., & Rausch, Harold L. *Naturalistic viewpoints in psychological research*. New York: Holt, Rinehart & Winston, 1969.
- Williams, Walter. Implementation analysis and assessment. In W. Williams and R.F. Elmore (Eds.), *Social program implementation*. New York: Academic Press, 1976.
- Williams, Walter, & Elmore, Richard F. (Eds.). *Social program implementation*. New York: Academic Press, 1976.
- Wirth, Louis. Preface. In K. Mannheim, *Ideology and utopia*. New York: Harcourt Brace Jovanovich, 1949.



- Wolf, Robert L. Trial by jury: A new evaluation method. *Phi Delta Kappan*, 1975, November.
- Wolf, Robert L., & Tymitz, Barbara. Whatever happened to the giant wombat: An investigation of the impact of the Ice Age Mammals and Emergence of Man Exhibit, National Museum of Natural History, Smithsonian Institutes. (mimeo)
- Worthen, Blaine R. Competencies for educational research and evaluation. *Educational Researcher*, 1975, 4, 13-16.
- Worthen, Blaine R. Eclecticism and evaluation models: Snapshots of an elephant's anatomy. Presented at the annual meeting of the American Educational Research Association, New York City, April 5, 1977.
- Worthen, Blaine R., & Sanders, James R. (Eds.). *Educational evaluation: Theory and practice*. Worthenton, OH: Charles A. Jones, 1973.



# INDEX

- accreditation model, 50-53
- active-reactive-adaptive evaluators, 18, 58-59
- adversary model, 50-53, 332
- advocacy-adversary model, 50-53, 332
- analysis, types of:
  - analyst-constructed typologies, 309-311
  - case studies, 302-306, 351-358
  - category systems, 311-313
  - causal, 324-326
  - computerized, 301-302
  - content, 41, 109-115, 299-326, 347-350
  - description, 30-31, 36, 57, 63-64, 78, 88, 124, 139-140, 159-164, 302, 324, 343-344
  - emic, 306-307
  - indigenous typologies, 306-311
  - inductive, 40-41, 43, 46, 50, 53, 55, 80-83, 125, 276, 306-309, 327
  - linkages, 320-324
  - logical, 313-316, 327
  - matrix, 318-320
  - metaphors, 316-317
- appropriateness of methods, 13, 15, 17, 20, 50, 116-117, 126, 272-273
- art criticism, 50-52
- audiences, 282, 291, 338-339
  
- background questions, 209-210
- behavioral objectives, 50-51
- behavioral questions, 207, 210
- boundary problem, 95
- breaking routines, 79-80, 89, 125
  
- case studies, 64-65, 76, 78, 88, 100-107, 134, 302-306, 351-358
  
- causality, 276, 278, 320-326
- Change Agent Study, 69-70
- checklist for using qualitative methods, 88-89
- choices, paradigm of, 17-19, 345
- clarity of questions, 225-231
- closeness to settings studied, 36, 43-44, 47, 124-125, 163, 337
- comparing programs, 66-68
- competency of evaluators, 338
- conceptualizing evaluation, 15
- confidentiality, 107, 133, 153
- connoisseurship, 52
- content analysis, 41, 109-115, 299-326, 347-350
- content of interviews, 207-211
- context, 9, 40, 45, 125, 251, 280-283, 322-323, 325
- controlling interviews, 243-246
- convergence in analysis, 311-312
- conversational interviews, 197-200, 206
- cost benefit or effectiveness, 18
- covert observations, 130-134, 138, 151
- creativity, 75, 159-160, 299, 313, 339
  
- decision-making model, 50, 57-58
- deinstitutionalization, 74, 77
- demographic questions, 209-210
- description, 30-31, 36, 57, 63-64, 78, 88, 124, 139-140, 159-160, 161-164, 302, 324, 343-344
- dichotomous questions, 213-219
- dimensions of fieldwork, 138
- discipline, 251
- discovery mode, 36, 46-47, 125, 185
- Discrepancy Evaluation, 72
- divergence in analysis, 312-313
- diversity, 66-68



- documentation, process of, 171, 283-293
- documents, 152-154, 157-159
- duration of observations, 134-135, 138
- early childhood education, 31-35
- ecological psychology, 44
- emic approach to analysis, 306-307
- entry into the field, 306-307
- ethnomethodology, 44
- Evaluation Research Society, 51
- evaluator effects, 333-336
- executive summary, 344
- Experience-Based Career Education, 64-65, 351-358
- experience questions, 207, 210
- experimental designs, 17, 18, 42, 109-115
- experts, 52
- feedback, 46, 71-73, 166, 172, 186-188, 291, 296-297, 340-344, 345
- feeling questions, 207, 210
- field notes, 160-165, 168, 183, 246-251
- fieldwork, 124, 157-159, 169-189
  - ending, 184-189
  - entry into, 169-179
  - personal experience of, 193-194
  - routinization of, 179-184
  - rules of, 192-193
  - stages of, 169-189
- focusing analysis, 295-297, 342-344
- focusing evaluation questions, 96-97, 99, 135-136, 138, 295-297
- Follow Through, 70-71, 76-77, 329
- formative evaluation, 71-93, 88, 133, 159, 166, 187-188, 289
- generalizations, 132, 134, 279-282
- generating new insights, 79-80, 89
- goal-free evaluation, 50, 55-57, 89, 332
- goals, 56-57, 89, 290
- grounded theory, 18, 41, 80-83, 89, 275
- group interviews, 201
- Halcolm's Evaluation Wisdom:
  - Beatitudes, 17, 253-254
  - Biography, 195-196
  - Koans, 265
  - Laws, 15, 89-90
- Methodological Chronicles, 121
- Oracles, 39
- Parables, 21, 37, 267
- Proverbs, 265
- Political Treatise, 346
- Sufi stories, 48, 116-117, 154, 252-253, 267, 274, 313
- Universal Encyclopedic Dictionary, 49
- Heisenberg Uncertainty Principle, 189
- holistic-inductive paradigm, 43-44, 46, 54, 109-110
- holistic perspective, 40, 43, 46-47, 50, 53, 87, 125, 134, 135, 304, 325-326
- human environment, observations of, 141-142
- hypothetico-deductive paradigm, 19, 41, 109-115
- ideology, 336-337
- illuminative evaluation, 55
- impartiality, 338
- implementation evaluation, 68-71, 88
- individualized outcomes, 62-64, 65, 88, 106
- inductive analysis, 40-41, 43, 46, 50, 53, 55, 80-83, 125, 276, 306-309
- infiltrating the setting, 170, 172
- informal interactions, observations of, 144-147
- inner perspective, 44
- input measures, 50
- insight, 15, 159-160, 163-164, 284-285, 297, 339
- interpretations, 36, 159-164, 184, 251, 267-268, 282, 286-293, 324-326, 335, 343-344
- interview analysis, a sample, 359-368
- interview notes, 246-251
- interview quality control, 251
- interview questions, types of, 207-211
- interviews, types of, 18, 197-206
  - informal, conversational, 197-200, 206
  - interview guide, 197, 200-202, 206
  - standardized open-ended, 197, 202-206
- interviews, wording questions in, 211-246



- announcement formats, 234-238
- attention-getting, 237
- clarity of, 225-231
- controlling, 243-246
- dichotomous, 213-219
- follow-up, 238-240
- neutral, 231-233
- open-ended, 211-219
- prefactory, 234-238
- presuppositions, 219-222
- probes, 238-240
- recognition, 240-243
- role-playing, 233-234
- simulation, 233-234
- singular, 222-224
- support, 240-243
- "why?" questions, 228-231
- introspection, 44-45, 125-126, 127, 287, 333-336
- investigation paradigm, 131
- judgments, 52, 57, 73, 155, 159, 163, 171, 293, 313, 343
- Kalamazoo Accountability Evaluation, 23-30
- key informants, 182-183
- knowledgeable questions, 208, 210
- language, 147-149, 225-227, 307-308
- leading questions, 232
- legislative monitoring, 77-78, 89, 171, 225-227
- logical analysis, 313-316
- logical positivism, 45, 269-271
- matrix analysis, 318-320
- methodological mixes, 108-115
- meta-evaluation, 93-94
- metaphor, 52, 82, 316-317
- mixed designs, 112-115
- morals, 17, 130-131, 134
- multiple methods, 18, 157-159, 329-332
- naturalistic behaviorism, 44
- naturalistic inquiry, 41-43, 46, 50, 52, 53, 80-83, 106, 109-115, 184-185, 189, 339
- negative cases, 328
- neutrality in analysis, 337-338
- neutral questions, 231-233
- nonverbal communication, 149-150
- norm-referenced tests, 51
- North Dakota Study Group on Evaluation, 10, 12, 283
- objectives, 50-51, 56-57
- objectivity, 56, 336-338
- observatons, focus of, 137-156
  - documents, 152-154
  - human, social environment, 141-142
  - informal interactions, 144-147
  - language, 147-149
  - nonevents, 154-156
  - nonverbal communication, 149-150
  - participant behaviors, 142-144
  - program activities, 142-144
  - program setting, 139-141
  - unobtrusive measures, 150-152
  - unplanned activities, 144-147
- observations, personal nature of, 292-293
- observations, types of, 127-138
  - brief, 134-135, 138
  - broad, 135-136, 138
  - complete, 132
  - covert, 130-134, 138
  - long duration, 134-135, 138
  - narrow, 135-136, 138
  - onlooker, 127-130, 138
  - overt, 130-134, 138
  - participant, 127-130, 138
- open education, 62-63, 85-86, 98, 226
- open-ended questions, 211-219
- Operation Reach-Out, 111-114
- opinion questions, 207, 210
- outcomes evaluation, 62-64, 65, 80-81, 318-324, 330, 359-368
- output measures, 50
- paradigms, 17-20, 336, 345
- parent education program, 31-35
- participant behaviors, observation of, 142-144
- participant observation, 18, 30, 127-130, 138, 191, 289-291, 335
- part-time observation, 176-179
- personalizing evaluation, 83-84, 86, 89
- perspective, 191-192, 196-197, 269-274, 282, 283-293, 327, 336-339



- phenomenology, 44-45, 51, 198-199, 325
- political considerations, 18, 84-87, 336-337
- positivism, 45, 269-271
- practice exercises,  
    analysis, 326  
    interviewing, 209-210, 214-216  
    observation, 139
- practice of evaluation, 58-59
- predispositions, 276-278
- preparation for observation, 122-124
- presentation of findings, 340-344
- presupposition questions, 219-222
- probes, 238-240
- process evaluation, 60-62, 65, 88, 318-324, 330
- professional review, 51
- program activities, observation of, 142-144
- program setting, 139-141
- purposeful sampling, 100-107
- quality, 73-74, 88
- quality control, 251
- quantitative data, 18, 19, 22, 336
- quantitative instrumentation, lack of, 75-76
- questionnaires, 9, 23-29
- quotations, 30-31, 36, 163, 246-251, 343, 359-368
- random sampling, 100, 105
- rapport, 231
- reactivity, 333-336
- reciprocal relationships, 172
- relevance, 18
- repertoire of methods, 17-18, 20, 48
- reporting results, 340-344
- responsive evaluation, 54
- revolutions, 79
- rigor, 18, 251, 299, 339-340
- rival explanations, 327-328
- role of the evaluator-observer, 138
- routinization of fieldwork, 179-184
- rules of observational research, 192-193
- sampling, 98-107, 209-210, 332-333
- secondary data, 18
- self-esteem, 75
- sensitizing concepts, 137-138, 316-317, 320
- sensory questions, 208-210
- sequencing of questions, 222-224
- singular questions, 222-224
- social ecology, 141-142
- Southwest Field Training Project, 320-324, 328, 359-368
- standardized interviews, 202-206
- standardized tests, 9, 17, 18, 22, 51, 75-77, 189
- state of the art considerations, 75-76, 89
- statistical significance, 74, 112, 313, 343
- strategic considerations, 17, 20, 39-40, 43-44, 46-47, 49, 60, 87-89, 100-106, 109-110, 114-116, 205, 344
- subjectivity, 45, 51, 54, 336-338
- Sufi stories (Shah), 227, 326, 333
- symbolic interaction, 44
- systems analysis, 50
- tape recording, 247-251
- teacher centers, 67
- technology of fieldwork, 165-168
- theory, 18, 41, 80-83, 185-186, 274-279, 292
- thickened observations, 286
- time frame of questions, 209-210
- time sampling, 106-107, 134-135
- training for observation, 122-124, 128
- trade-offs, 95-98, 136
- transaction model of evaluation, 50, 54-55
- triangulation, 18, 51, 108-109, 157-159, 329-332
- trust, 170, 172
- truth, 132, 133, 268-274
- typologies, 306-313
- unanticipated outcomes, 56, 89
- units of analysis, 99-100, 143
- unobtrusive measures, 18, 76-77, 88, 150-152
- unplanned activities, observation of, 144-147
- useful evaluations, 12-13, 15, 17-18, 83-84, 89, 94-95, 107, 134, 271-274, 282-283, 344-345
- utilization-focused evaluation, 10, 17, 59-60, 69, 82, 84-87, 132, 279, 301, 322, 331, 345



- validation of data, 326-339
- validity, 18, 75, 88, 94, 122-123, 130, 282
- value of observational data, 124-126
- value questions, 207, 210
- variations in design, 111-115
- variations in observational data, 124-126
- variations in qualitative interviewing, 197-206
- verification, 46-47, 80-83, 184-186, 275-276, 297, 326-339
- verstehen*, 44-45
- “why?” questions, 228-231



## ABOUT THE AUTHOR

MICHAEL QUINN PATTON is Director of the Minnesota Center for Social Research and Adjunct Professor in the Hubert H. Humphrey Institute of Public Affairs, both at the University of Minnesota, where he was named outstanding teacher of the year in 1976. He holds M.S. and Ph.D. degrees in sociology from the University of Wisconsin, and is the author of numerous articles, reports, and conference papers in the field of evaluation research. Dr. Patton regularly conducts evaluation workshops and has also served as an evaluation consultant to many educational and human services projects in the United States and abroad. Dr. Patton is the author of *Utilization-Focused Evaluation* (Sage, 1978).