Qualitative Data Analysis

An Expanded Sourcebook

Qualitative Data Analysis

Second Edition

Matthew B. Miles
A. Michael Huberman

SAGE Publications
International Educational and Professional Publisher
Thousand Oaks London New Delhi

For information address:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
6 Bonhill Street
London EC2A 4PU
United Kingdom

SAGE Publications India Pvt. Ltd.
M-32 Market
Greater Kailash I
New Delhi 110 048 India

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Miles, Matthew B.
Qualitative data analysis : an expanded sourcebook / Matthew B. Miles, A. Michael Huberman. — 2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-8039-4653-8 (cl) — ISBN 0-8039-5540-5 (pb)
1. Social sciences—Research. 2. Education—Research.
I. Huberman, A. M. II. Title.
H62.M437 1994
300′.723—dc20 93-41204
CIP

04 05 17

Sage Production Editor: Rebecca Holland

Contents

List of Figures, Tables, and Boxes

Figures

Tables

Boxes

Acknowledgments

The first edition of this book grew out of our experience in two linked research projects. One, beginning in 1978, was the field study component of the Study of Dissemination Efforts Supporting School Improvement (Department of Education Contract 300-78-0527), led by David P. Crandall of The Network, Inc. We are indebted to him for his steady encouragement and support and that of Ann Bezdek Weinheimer, project officer from the Office of Planning, Budgeting and Evaluation.

In the field study itself, Beverly Loy Taylor and Jo Ann Goldberg were strong colleagues; their fieldwork and case study analysis, along with ours, led to volume four of the DESSI final report, *People, Policies, and Practices: Examining the Chain of School Improvement,* later published as *Innovation Up Close* (Plenum, 1984).

The second project, The Realities of School Improvement Programs: Analysis of Qualitative Data (NIE grant G-81-001-8), gave us the opportunity to develop our methodological ideas further and to write the first edition of this book. Rolf Lehming, of the Program on Dissemination and Improvement of Practice, was our project officer; we valued his sustained interest and advice.

The ideas in the first edition—and indeed in this one—do not necessarily reflect the views or policies of the Department of Education. But we remain grateful for its sponsorship of these studies.

In the last 10 years, many people have contributed to our understanding of qualitative data analysis and to the development of the second edition. We have experimented in the company of colleagues with studies that expanded, tested, and refined the methods described in the first edition. We are indebted to Ann Lieberman, Ellen Saxl, Myrna Cooper, Vernay Mitchell, and Sharon Piety-Jacobs, who joined Miles in a study (1983-1985) of school "change agents"; to the late Eleanor Farrar, Karen Seashore Louis, Sheila Rosenblum, and Tony Cipollone in a study with Miles (1985-1989) of urban high school reform; to Per Dalin, Adriaan Verspoor, Ray Chesterfield, Hallvard Kuløy, Tekle Ayano, Mumtaz Jahan, and Carlos Rojas, whom we assisted in a World Bank study (1988-1992) of educational reform in Bangladesh, Ethiopia, and Colombia; to Marie-Madeleine Grounauer and Gianreto Pini, Huberman's associates in a teachers' life cycle study (1982-1986); and to Monica Gather-Thurler and Erwin Beck, associates in Huberman's study of research use (1984-1988).

As always, the process of teaching from the book taught us a great deal. There are too many participants to list, but we were fortunate to have led an extended series of seminars at the universities of Nijmegen and Utrecht (strong thanks to Rein van der Vegt) and at many other universities as well: Geneva, Zürich, Paris, Dijon, Leuven, Göteborg, Montreal, Toronto, Queen's, Utah, Monash, Melbourne, and Adelaide.

During 1990-1991 we sent an informal survey to a wide range of people engaged in qualitative research, asking for collegial advice and examples of their work. Our warm thanks to the 126 researchers who responded; they provided a wide range of ideas, papers, advice, and cautions that were immensely helpful. Many of these colleagues are quoted or cited in this book.

Because computer technology advanced dramatically during this period, we ourselves did the word processing. But we must acknowledge the help of Judith Woolcock, who computer-scanned and cleaned up the text of the first edition for use in our revision. Ali Yildirim provided strong administrative and secretarial support, prepared tables and figures with great patience and skill, and produced the reference list. Sadi Seferoglu aided with graphic production and with library search. Michael Moon's production work on the indexes was outstanding. Our warm appreciation to Tina Hill for her excellent typesetting and to Rebecca Holland for her thoughtful production editing.

For permission to reproduce previously published (and some unpublished) material as exhibits, we are grateful to Russell Bernard, Robert Bogdan, Kathleen Carley, Thomas Carney, Mark Chesler, Dick Corbett, David Crandall, Kathleen Eisenhardt, David Flinders, Christina Gladwin, Randy Hodson, Diane Kell, Klaus Krippendorff, Anton Kuzel, Dean Lee, Kenneth Leithwood, Merry Merryfield, John Ogbu, John Owen, Michael Patton, Marc Pillet, Laurel Richardson, G. Mark Schoepfle, Nick Smith, Suzanne Stiegelbauer, Barry Turner, Karl Weick, Oswald Werner, Bruce Wilson, and Harry Wolcott.

Grants supporting the extensive retrieval and synthesis work for this edition came to us from the John D. and Catherine T. MacArthur Foundation, where Peter Gerber provided thoughtful support, and from Sage Publications. Sara Miller McCune and David McCune of Sage took a keen interest in the project. We are grateful for the active, intelligent guidance that our editor, Mitch Allen, provided throughout the work.

We owe a very special debt to Carolyn Riehl. Her ability to locate and extract interesting ideas—both substantive and methodological—from a wide range of qualitative studies is remarkable. She was a strong third colleague during our extended period of retrieval and ordering.

The Appendix, on computer programs, draws centrally on the work of Eben Weitzman. His gift for combining technical clarity and a good sense of user friendliness is indispensable for qualitative researchers faced with the burgeoning growth of analysis software.

Drafts of this edition were reviewed by many people: Our warm thanks for the thoughtful advice of Martha Ann Carey, Rick Ginsberg, David Marsh, Joseph Maxwell, Betty Miles, Renata Tesch, Harry Wolcott, and an anonymous Sage reviewer.

As with the first edition, we are grateful to each other for the energy, stimulation, and productivity of our work in a demanding enterprise. We'll soon be at our 15th year of collaboration. The only disappointing feature this time was that the work occupied just 2 years. Still, there's always the next common venture . . .

M. B. M.
A. M. H.

sion from Pergamon Press Ltd, Headington Hill Hall, Oxford OX3 OBW, UK.

Figure 4.3, from "Some Practical Aspects of Qualitative Data Anlysis: One Way of Organizing the Cognitive Processes Associated With the Generation of Grounded Theory" by B. A. Turner, 1981, *Quality and Quantity, 15*, pp. 225-247. Copyright © 1981 by Kluwer Academic Publishers. Reprinted by permission.

Box 4.5, from *Testing, Reform, and Rebellion* by D. Corbett and B. Wilson, 1991, Norwood, NJ: Ablex. Copyright © 1991 by Ablex Publishing Corporation. Reprinted by permission.

Figure 4.6, from *What's In Good Field Notes?* by R. Bogdan and L. Easterday, n.d., Syracuse, NY: Syracuse University, Center on Human Policy. Reprinted by permission of the author.

Box 4.6, from *Constructing Scenes and Dialogue to Display Findings in Case Study Reporting* by M. M. Merryfield, 1990, Columbus: Ohio State University, College of Education. Reprinted by permission of the author.

Figure 4.7, from "Coding Choices for Textual Analysis: A Comparison of Content Analysis and Map Analysis" by K. Carley, 1993, *Sociological Methodology, 23*, pp. 75-126. Copyright © 1993 by the American Sociological Association. Reprinted by permission.

Figure 4.8, from *La mise en pratique de la recherche* by A. M. Huberman and M. Gather-Thurler, 1991, Bern, Switzerland: P. Lang. Copyright © 1991 by Peter Lang A. G. Reprinted by permission.

Figure 4.15, from *Professionals' Views of the "Dangers" of Self-Help Groups* (CRSO Paper 345), 1987, Ann Arbor, MI: Center for Research on Social Organization. Reprinted by permission of the author.

Figure 5.1, from *Collaborative Inquiry Methodology* by T. F. Carney, 1990, Windsor, Ontario, Canada: University of Windsor, Division for Instructional Development. Reprinted by permission of the author.

Figures 5.6 and 5.7, from *Mapping Restructuring: Final Technical Report* by N. Khattri and M. B. Miles, 1993, New York: Center for Policy Research. Reprinted by permission from Center for Policy Research.

Box 5.1, from "The Consequences of Poetic Representation: Writing the Other, Rewriting the Self" by L. Richardson, 1992, in *Investigating Subjectivity: Research on Lived Experience* by C. Ellis and M. G. Flaherty (Eds.), Newbury Park, CA: Sage Publications, Inc. Copyright © 1992 by Sage Publications, Inc. Reprinted by permission.

Box 5.3, from "Through the Eye of the Beholder: On the Use of Qualitative Methods in Data Analysis" (R & D Report 3137) by S. Stiegelbauer, M. Goldstein, and L. L. Huling, 1982, Austin: University of Texas, R & D Center for Teacher Education. Reprinted by permission of the the author.

Box 5.5, from "How to Record Activities" by O. Werner, 1992, *Cultural Anthropology Methods Newsletter, 4*, pp. 1-3. Reprinted by permission of CAM and author.

Box 5.6, from *Systematic Fieldwork: Vol. 2. Ethnographic Analysis and Data Management* by O. Werner and G. M. Schoepfle, 1987, Newbury Park, CA: Sage Publications, Inc. Copyright © 1987 by Sage Publications, Inc. Reprinted by permission of Sage Publications, Inc., and the authors.

Box 5.8, from *Chester High School: Working Against the Odds* by M. B. Miles and S. Rosenblum, 1987, New York: Center for Policy Research. Reprinted by permission of Center for Policy Research.

Box 5.9 and Table 5.13, from *Research Methods in Anthropology* (2nd Edition) by H. R. Bernard, 1994, Newbury Park, CA: Sage Publications, Inc. Copyright © 1994 by Sage Publications, Inc. Reprinted by permission.

Box 6.3, from *The Social Psychology of Organizing* by K. Weick, 1979, Reading, MA: Addison-Wesley. Reprinted with permission of McGraw-Hill.

Box 6.4, from "Federal Intervention in the Mid-1980's: Charting the Influence of the Participation and Equity Program at the Institutional Level" by J. Owen and R. Hartley, 1987, *Australian Educational Researcher, 15*, pp. 71-85. Reprinted by permission of the author.

Box 6.5, from *How the School Improvement Strategies of Transformational Leaders Foster Teacher Development* (paper presented at the Sixth Annual Conference of the Department of Educational Administration and Centre for Leadership Development, OISE, Toronto) by K. Leithwood, D. Jantzi, and B. Dart, 1991. Reprinted by permission of the authors.

Figure 7.3, from "The Professional Life Cycle of Teachers" by A. M. Huberman, 1989, *Teachers College Record, 9*, pp. 31-57. Reprinted by permission of Teachers College Press, Columbia University and the author.

Figures 7.4, 7.5, 7.6, from *Ethnographic Decision Tree Modeling* by C. H. Gladwin, 1989, Newbury Park, CA: Sage Publications, Inc. Reprinted by permission.

Box 7.4, from "Making Fast Strategic Decisions in High-Velocity Environments" by K. M. Eisenhardt, 1989, *Academy of Management Journal, 32*, pp. 543-576. Reprinted by permission of Academy of Management Journal, College of Business Administration, Texas A & M University.

Table 7.10, from "Building Theories From Case Study Research" by K. M. Eisenhardt, 1989, *Academy of Management Journal, 14*, pp. 532-550. Reprinted by permission of Academy of Management Journal, College of Business Administration, Texas A & M University.

Table 8.7, from *Parents and Federal Education Programs: Vol. 7. Methodologies Employed in the Study of Parental Involvement*, by D. Lee, W. Kessling, and R. Melaragno, 1981, Santa Monica, CA: System Development Corporation.

Figure 10.1, from "Clustering" by K. Krippendorff in *Multivariate Techniques in Human Communication Research*, 1980, New York: Academic Press. Reprinted by permission of the author.

Box 10.2, from "The Active Worker: Compliance and Autonomy at the Workplace" by R. Hodson, 1991, *Journal of Contemporary Ethnography, 20*, pp. 47-78. Reprinted by permission of Sage Publications, Inc.

Table 11.1, from "In Search of Ethical Guidance: Constructing a Basis for Dialogue" by D. J. Flinders, 1992, *Qualitative Studies in Education, 5*, pp. 101-116. Copyright © 1992. Reprinted by permission of Taylor and Francis Ltd.

1

Introduction

We wrote this book to address a critical need faced by researchers in all fields of the human sciences. Put simply: How can we draw valid meaning from qualitative data? What methods of analysis can we use that are practical, communicable, and non-self-deluding—in short, will get us knowledge that we and others can rely on?

A. The General Problem

Qualitative data, usually in the form of words rather than numbers, have always been the staple of some fields in the social sciences, notably anthropology, history, and political science. In the past decade, however, more researchers in basic disciplines and applied fields (psychology, sociology, linguistics, public administration, organizational studies, business studies, health care, urban planning, educational research, family studies, program evaluation, and policy analysis) have shifted to a more qualitative paradigm. As L. M. Smith (1992b) observed, the terms *ethnography, field methods, qualitative inquiry, participant observation, case study, naturalistic methods,* and *responsive evaluation* have become practically synonymous.

Qualitative data are sexy. They are a source of well-grounded, rich descriptions and explanations of processes in identifiable local contexts. With qualitative data one can preserve chronological flow, see precisely which events led to which consequences, and derive fruitful explanations. Then, too, good qualitative data are more likely to lead to serendipitous findings and to new integrations; they help researchers to get beyond initial conceptions and to generate or revise conceptual frameworks. Finally, the findings from qualitative studies have a quality of "undeniability." Words, especially organized into incidents or stories, have a concrete, vivid, meaningful flavor that often proves far more convincing to a reader—another researcher, a policymaker, a practitioner—than pages of summarized numbers.

The expansion of qualitative inquiry since the first edition of this book (Miles & Huberman, 1984) has been phenomenal. The base of books, articles, and papers we collected for this second edition has more than tripled over that for the first. Debate on underlying epistemological issues has continued vigorously (Guba, 1990). There are full-fledged handbooks (Denzin & Lincoln, 1994; LeCompte, Millroy, & Preissle, 1992), Sage's Qualitative Research Methods series (more than 24 volumes), new journals (*Qualitative Studies in Education, Qualitative Health Research*), newsletters (*Cultural Anthropology Methods*), annual forums (Ethnography in Education Research Forum; Qualitative Research in Education Conference), computer bulletin boards (QUIL), software meetings (International Conferences on Computers and Qualitative Methodology), and special qualitative interest groups in most major professional associations.

Yet, in the flurry of this activity, we should be mindful of some pervasive issues that have not gone away. These issues include the labor-intensiveness (and extensiveness over months or years) of data collection, frequent data overload, the distinct possibility of researcher bias, the time demands of processing and coding data, the adequacy of sampling when only a few cases can be managed, the generalizability of findings, the credibility and quality of conclusions, and their utility in the world of policy and action.

Seen in traditional terms, the reliability and validity of qualitatively derived findings can be seriously in doubt (Dawson, 1979, 1982; Ginsberg, 1990; Kirk & Miller, 1986; Kvale, 1989a; LeCompte & Goetz, 1982). Although standards for the "goodness" of qualitative findings may well differ from traditional ones, as Lincoln and Guba (1985, 1990), Wolcott (1992), and others have emphasized, the general problem remains. (We discuss this further in Chapter 10.)

Although many researchers, from graduate students struggling with dissertations to experienced researchers, work alone on their projects and often focus on single cases, qualitative work is becoming more complex. Increasingly we see "multisite, multimethod" studies (A. G. Smith & Louis, 1982) that may combine qualitative and quantitative inquiry (Rossman & Wilson, 1984), carried out by a research team working with comparable data collection and analysis methods (Herriott & Firestone, 1983; Yin, 1984).

Beyond these issues, a deep, dark question about qualitative studies remains. As one of us has written:

> The most serious and central difficulty in the use of qualitative data is that methods of analysis are not well formulated. For quantitative data, there are clear conventions the researcher can use. But the analyst faced with a bank of qualitative data has very few guidelines for protection against self-delusion, let alone the presentation of unreliable or invalid conclusions to scientific or policy-making audiences. How can we be sure that an "earthy," "undeniable," "serendipitous" finding is not, in fact, *wrong?* (Miles, 1979b, p. 591)

Since 1979, it must be said, the shared craft of qualitative analysis has advanced. For example, matrix and network displays are no longer rare. And although phenomenology has been called "a method without techniques," its practitioners have begun to explicate their procedures (Kvale, 1988; Melnick & Beaudry, 1990; Pearsol, 1985). Grounded theory methods are described more concretely (Strauss & Corbin, 1990), and the motto of the *Cultural Anthropology Methods Newsletter* is "Methods belong to all of us." Sieber's (1976) finding that less than 5-10% of the pages in seven well-respected field methods texts were devoted to analysis is by now outdated: More recent texts have addressed analysis problems far more seriously (Bernard, 1988; Bogdan & Biklen, 1992; Goetz & LeCompte, 1984; Merriam, 1988; Patton, 1990; Smith & Manning, 1982; Spradley, 1979; Werner & Schoepfle, 1987a, 1987b).

Still the problem of confidence in findings has not gone away. Clearly, we still lack a bank of explicit methods to draw on. We need to keep working at sensible canons for qualitative data analysis, in the sense of shared ground rules for drawing conclusions and verifying their sturdiness.

Some qualitative researchers still consider analysis to be an art form and insist on intuitive approaches to it. We are left with the researcher's telling us of classifications and patterns drawn from the welter of field data, in ways that are irreducible or even incommunicable. We do not really see how the researcher got from 3,600 pages of field notes to the final conclusions, as sprinkled with vivid illustrations as they may be.

Other researchers have hesitated to focus on analysis issues on the grounds that unequivocal determination of the validity of findings is impossible (Becker, 1958; Bruyn, 1966; Lofland, 1971; Wolcott, 1992). More profoundly, for some phenomenologically oriented, interpretivist, and constructivist researchers, there is no unambiguous social reality "out there" to be accounted for, so there is little need to evolve methodological canons to help explicate its laws (see Dreitzel, 1970). In this view, social processes are ephemeral, fluid phenomena with no existence independent of social actors' ways of construing and describing them.

At times it seems as if the competing, often polemical arguments of different schools of thought about how qualitative research should be done properly use more energy than the actual research does. We confess to liking George Homans' (1949) remark: "People who write about methodology often forget that it is a matter of strategy, not of morals" (p. 330).

This book is written in the belief that, as qualitative researchers, we need to keep sharing our craft—that is, the explicit, systematic methods we use to draw conclusions and to test them carefully. We need methods that are credible, dependable, and replicable in *qualitative* terms. This is the need our book addresses.

B. The Nature of This Book

This is a practical sourcebook for all researchers who make use of qualitative data. It aims to share the current state of the craft, drawing on our experience and that of many colleagues in the design, testing, and use of qualitative data analysis methods. Strong emphasis is placed on data displays—matrices and networks—that go beyond ordinary narrative text. Each method of data display and analysis is described and illustrated in detail, with practical suggestions for the user for adaptation and use.

Audiences

The book is for *practicing researchers* in all fields whose work, whether basic or applied, involves the struggle with actual qualitative data analysis issues.

An important subset of that audience is the *beginning researcher*—a graduate student or junior staff member—who is working with qualitative data. We have encountered many students launched on qualitative dissertations or research projects who feel overwhelmed and undertrained. With them in mind, we keep the language accessible and supportive, and we offer suggestions for using the book in qualitative methods courses.

A third audience is *staff specialists and managers,* who rely on qualitative information as a routine part of their work and who need practical methods for making the best use of it.

Many examples used in the book are drawn from educational research, both ours and others'. We also include examples from other domains—health care, public health, anthropology, sociology, psychology, business studies, political science, public administration, program evaluation, library science, organizational studies, criminology, communication, computer science, family studies, policy research—to underline that the methods are generic, not field-limited.

Some of the methods reported here grew out of multiple-case studies of organizations, carried out by a research team. But do not despair if (a) you are working alone, (b) your study has just one case, or (c) you are focusing at the individual or small group level. There are many relevant examples for you, along with targeted advice.

Approach

This is a sourcebook. We do not consider it to be a comprehensive handbook. Rather, we have tried to bring together a serviceable set of resources, to encourage their use, and, above all, to stimulate their further development, testing, and refinement.

The resources came via two routes. First, we retrieved and synthesized a great deal of work published during the last decade. Second, we drew a snowball sample of qualitative researchers and sent them an informal survey, asking them about methods of qualitative data analysis they had been using, and inviting them to send specific examples, exhibits, and suggestions. Many of the ideas from these 126 colleagues have been included.

This book is about *doing analysis.* We cover questions of research design and data collection only as they bear on analysis, and only glancingly address matters such as access to field sites and trust-building with informants. Others have dealt with these issues repeatedly and well, and we cite their work along the way.

We have taken as concrete and direct an approach as possible, staying close to the reader's elbow and aiming to be a knowledgeable guide through uneven territory. Although in each chapter we aim to provide a coherent intellectual frame for specific methods, we always emphasize hands-on work with actual data. For each of the methods outlined, we give specific illustrations, with enough detail so that the reader can see how things work, can try the method, and, most important, can revise the method in future work.

We also tend to be pragmatic. Although we outline our epistemological views in section D below, we believe, perhaps less naively than the reader might think at first, that any method that works—that will produce clear, verifiable, credible meanings from a set of qualitative data—is grist for our mill, regardless of its antecedents.

These methods are manageable and straightforward. They do not necessarily require prolonged training or a specialized vocabulary. We can add that the experience of inventing analysis methods and of using/adapting those of others has been a happy, productive one.

The strongest message of this book is not that these particular methods should be applied scrupulously, but that the creation, testing, and revision of simple, practical, and effective analysis methods remain the highest priority for qualitative researchers.

The spirit of that quest is well summed up by a thoughtful remark on the first edition by a European sociologist (W. Hutmacher, personal communication, 1983):

> I think you've worked out a comprehensive solution to many of the methodological problems we *have* to resolve, that we resolve poorly and, as a result, that we often cover up when reporting out to our peers. But yours isn't the only solution nor the last one we'll come upon. We have to admit that we're *all* casting about, you included.

The quest continues. In our survey of qualitative researchers, we asked about issues that were unclear or puzzling. One researcher replied: "Everything is unclear and puzzling. . . . Improved methodology, however, raises confidence to a much more significant plane and provides a more certain base (though not an absolute one) for action."

This book was written to share the "casting about," experimentation, dialogue, and learning that good qualitative analysis requires. We remain convinced that concrete, sharable methods do indeed "belong to all of us." In the past decade, we've found that refining and developing analysis methods on new projects had clear payoff; our confidence in findings was greater, and credibility for our research, practice, and policy audiences was enhanced. We hope our experience will be helpful to our colleagues, as theirs has been helpful to us.

C. Our Orientation

It is good medicine, we think, for researchers to make their preferences clear. To know how a researcher construes the shape of the social world and aims to give us a credible account of it is to know our conversational partner. If a critical realist, a critical theorist, and a social phenomenologist are competing for our attention, we need to know where each is coming from. Each will have diverse views of what is real, what can be known, and how these social facts can be faithfully rendered.

At the time of the first edition of this book, we thought of ourselves as "realists" (Huberman & Miles, 1985). We still do. But "realism" has come to mean many things. We see ourselves in the lineage of "transcendental realism" (Bhaskar, 1978, 1989; Harré & Secord, 1973; Manicas & Secord, 1982). That means we think that social phenomena exist not only in the mind but also in the objective world—and that some lawful and reasonably stable relationships are to be found among them. The lawfulness comes from the regularities and sequences that link together phenomena. From these patterns we can derive constructs that underlie individual and social life. The fact that most of those constructs are invisible to the human eye does not make them invalid. After all, we all are surrounded by lawful physical mechanisms of which we're, at most, remotely aware.

Human relationships and societies have peculiarities that make a realist approach to understanding them more complex—but not impossible. Unlike researchers in physics, we must contend with institutions, structures, practices, and conventions that people reproduce and transform. Human meanings and intentions are worked out within the frameworks of these social structures—structures that are invisible but nonetheless real. In other words, social phenomena, such as language, decisions, conflicts, and hierarchies, exist objectively in the world and exert strong influences over human activities because people construe them in common ways. Things that are believed become real and can be inquired into.[1]

We agree with interpretivists who point out that knowledge is a social and historical product and that "facts" come to us laden with theory. We affirm the existence and importance of the subjective, the phenomenological, the meaning-making at the center of social life. Our aim is to register and "transcend" these processes by building theories to account for a real world that is both bounded and perceptually laden, and to test these theories in our various disciplines.

Our tests do not use "covering laws" or the deductive logic of classical positivism. Rather, our explanations flow from an account of how differing structures produced the events we observed. We aim to account for events, rather than simply to document their sequence. We look for an individual or a social process, a mechanism, a structure at the core of events that can be captured to provide a *causal description* of the forces at work.

Transcendental realism calls both for causal explanation and for the evidence to show that each entity or event is an instance of that explanation. So we need not only an explanatory structure but also a grasp of the particular configuration at hand. That is one reason why we have tilted toward more inductive methods of study.

As Erickson (1977) put it, social facts are embedded in social action, just as social meaning is constituted by what people *do* in everyday life. These meanings are most often discovered

> by hanging around and watching people carefully and asking them why they do what they do. . . . [Given] this orientation toward social meaning as embedded in the concrete, particular doings of people, qualitative researchers are reluctant to see attributes of the doing abstracted from the scene of social action and counted out of context. (p. 58)

Our aim here has been to be explicit about our biases, not to persuade anyone of their superior virtue or even their reasonableness. Like Howe (1988), we are wary of abstract epistemological arguments that do not connect operationally with the actual research practices used to gain knowledge.

At the working level, it seems hard to find researchers encamped in one fixed place along a stereotyped continuum between "relativism" and "postpositivism." Now scores of postpositivists are using naturalistic and phenomenological approaches. At the same time, an increasing number of interpretively oriented ethnographers are using predesigned conceptual frames and instruments, especially when dealing with multiple cases. Few postpositivists will dispute the validity and importance of subjective meanings, and few phenomenologists still practice pure hermeneutics. Critical theorists use symbolic interactionism to discover social determinisms.[2] In epistemological debates it is tempting to operate at the poles. But in the actual practice of empirical research, we believe that all of us—realists, interpretivists, critical theorists—are closer to the center, with multiple overlaps.

Furthermore, the lines between epistemologies have become blurred. Current perspectives such as pragmatism and critical theory have qualities of both interpretivism and postpositivism. Approaches like ours, which do away with correspondence theory (direct, objective knowledge of forms) and include phenomenological meaning, are hard to situate. Some researchers (e.g., Pitman & Maxwell, 1992) have argued that realist and interpretivist methods are both attempts to build coherent arguments that relate theoretical claims to independently measured facts. Others (Lee, 1991) have aimed to show that each perspective adds a meaningful layer without necessarily contradicting the others: a subjective understanding, an interpretive understanding (as rendered by the researcher), a positivist understanding (theoretical propositions according to rules of formal logic).

The paradigms for conducting social research seem to be shifting beneath our feet, and an increasing number of researchers now see the world with more pragmatic, ecumenical eyes. Our view is that sharing more about our craft is essential, and that it is possible to develop practical standards—workable across different perspectives—for judging the goodness of conclusions. Even if we happen to be dubious about postpositivist canons, we are still accountable for the rationality and trustworthiness of our methods.

We may face a risk of formalization when we dissect and reassemble the analytic procedures used by qualitative researchers, but not a large one. For the time being, we seem to be in a lively, partially explored realm far removed from canonical sterility. To us it seems clear that research is actually more a craft than a slavish adherence to methodological rules. No study conforms exactly to a standard methodology; each one calls for the researcher to bend the methodology to the peculiarities of the setting (cf. Mishler, 1990). At the least, we need to find out what qualitative researchers actually *do* when they assemble and analyze data from the field.[3]

Readers looking at the methods in this sourcebook will find them to be orderly ones, with a good degree of formalization. Many colleagues prefer intuitive, relaxed voyages through their data, and we wish them well. We have opted for thoroughness and explicitness, not just because it suits us, but because vague descriptions are of little practical use to others. Note, however, that some techniques in this book call for metaphorical thinking, figurative representations, even free associations. And the overall structure of the text allows for some techniques to be used and others to be left aside. We advise you to look behind any apparent formalism and seek out what will be useful in your own work.

D. Varieties of Qualitative Research

Qualitative research may be conducted in dozens of ways, many with long traditions behind them. To do them all justice is impossible here. For our purposes the question is, What do some different varieties of qualitative research have to say about *analysis?* Can we see some common practices, some themes?

First, we might look at some thoughtful efforts to array the whole range of qualitative study approaches. Wolcott (1992) shows a literal "tree" (Figure 1.1) of nearly two dozen strategies organized according to preferred styles of collecting data. His classification turns around *methods.*

Tesch (1990), whose tree is computer-generated (Figure 1.2), sorts 27 types of qualitative research according to three major substantive questions: What are the characteristics of language itself? Can we discover regularities in human experience? Can we comprehend the meaning of a text or action? These are broad families of research *purposes.*

Jacob's (1987) taxonomy sorts five major qualitative research traditions (ecological psychology, holistic ethnography, ethnography of communication, cognitive anthropology, and symbolic interactionism) by using dimensions including "assumptions about human nature and society," the "focus" (the content examined, at what social system level), and "methodology" (research design, data collection, and qualitative/quantitative analysis).

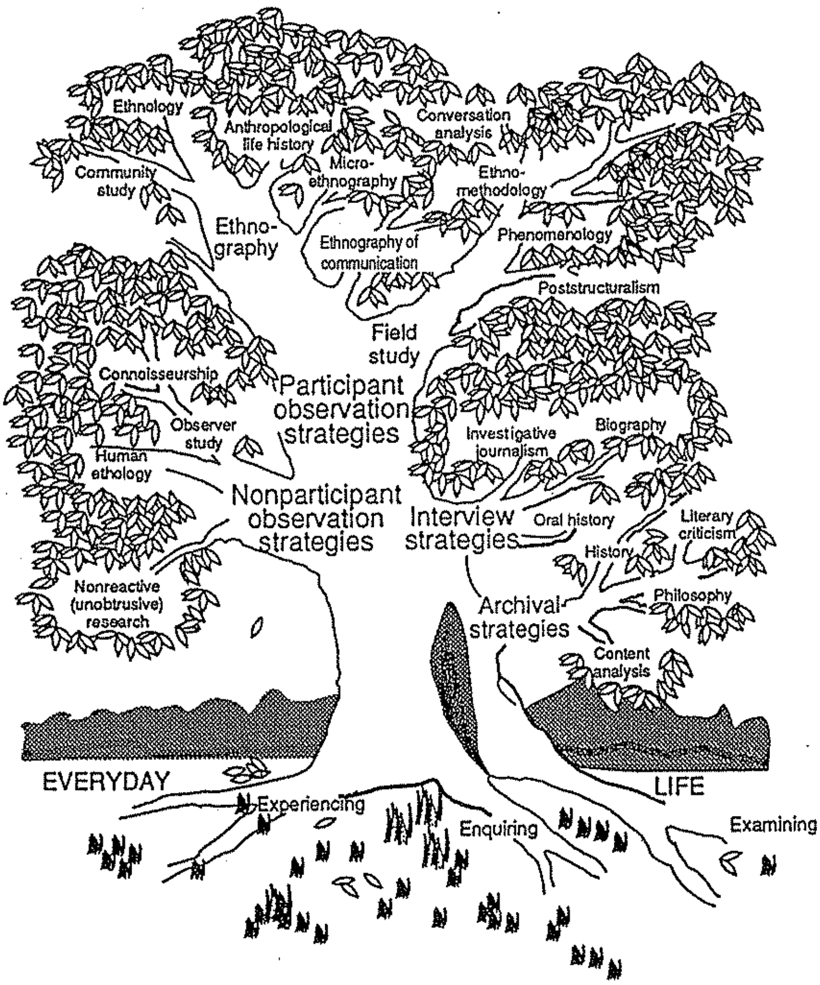
The reader interested in these sorting efforts will enjoy exploring them more fully. But as comprehensive and clarifying as these catalogs and taxonomies may be, they turn out to be basically incommensurate, both in the way the different qualitative strands are defined and in the criteria used to distinguish them. The mind boggles in trying to get from one to another.

Also, such taxonomies can go out of date quickly. For example, as the more "interpretivist" stream gains currency, we have qualitative research being conducted in history, literature, and journalism, while historians are using videos, tape recorders, interviews, and statistics to supplement traditional data sources.

Recurring Features of Qualitative Research

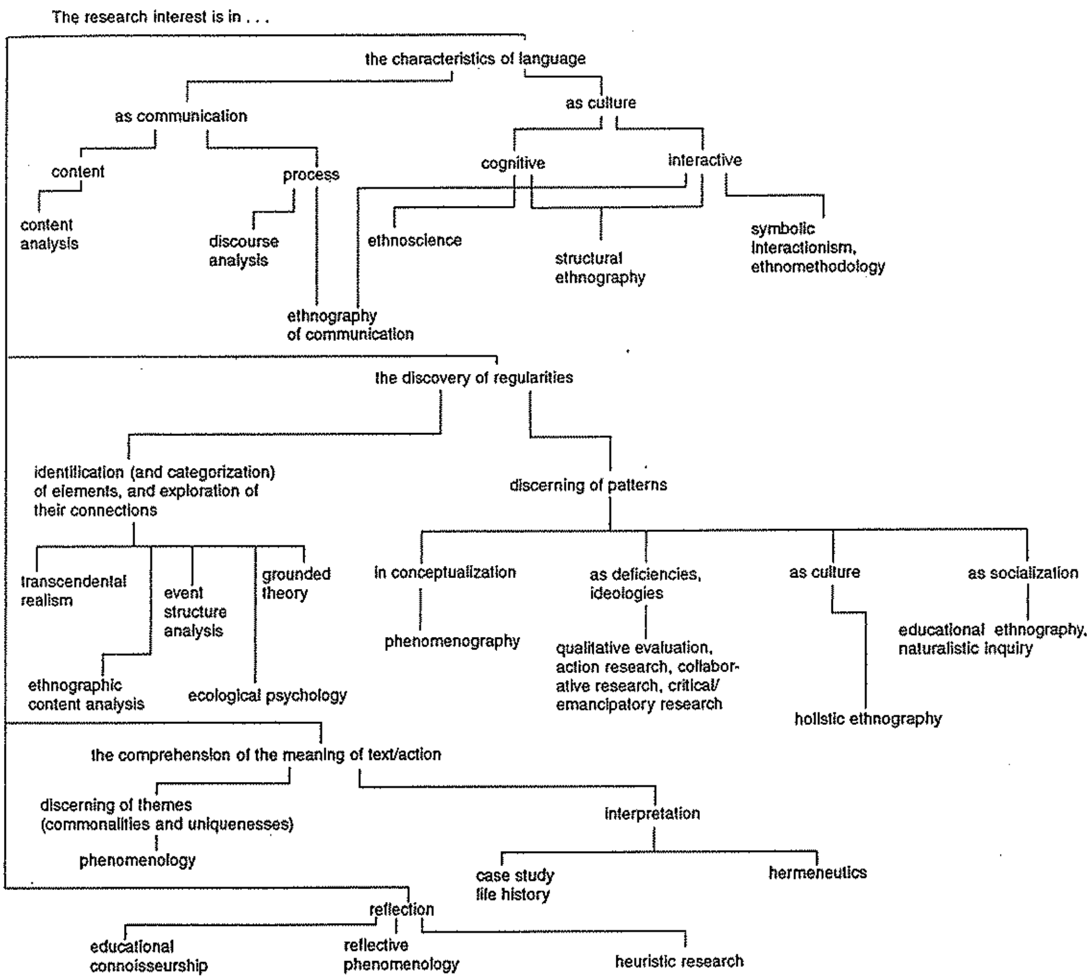
Do any features occur in most species of qualitative inquiry? Let's make a try at a list, with the understanding that some exemplars will be left dangling. We concur with Wolcott's (1982) emphasis on the "naturalist" nature of most qualitative research—even though that term, too, has undergone a sea change. By combining some of his descriptors with several of ours, we can suggest some recurring features of "naturalist" research:

Figure 1.1
Qualitative Strategies in Educational Research (Wolcott, 1992)



- Qualitative research is conducted through an intense and/or prolonged contact with a "field" or life situation. These situations are typically "banal" or normal ones, reflective of the everyday life of individuals, groups, societies, and organizations.
- The researcher's role is to gain a "holistic" (systemic, encompassing, integrated) overview of the context under study: its logic, its arrangements, its explicit and implicit rules.
- The researcher attempts to capture data on the perceptions of local actors "from the inside," through a process of deep attentiveness, of empathetic understanding (*Verstehen*), and of suspending or "bracketing" preconceptions about the topics under discussion.
- Reading through these materials, the researcher may isolate certain themes and expressions that can be reviewed with informants, but that should be maintained in their original forms throughout the study.

Figure 1.2
Graphic Overview of Qualitative Research Types (Tesch, 1990)



- A main task is to explicate the ways people in particular settings come to understand, account for, take action, and otherwise manage their day-to-day situations.
- Many interpretations of this material are possible, but some are more compelling for theoretical reasons or on grounds of internal consistency.
- Relatively little standardized instrumentation is used at the outset. The researcher is essentially the main "measurement device" in the study.
- Most analysis is done with words. The words can be assembled, subclustered, broken into semiotic segments. They can be organized to permit the researcher to contrast, compare, analyze, and bestow patterns upon them.

These may be a "core" of recurring features for naturalistic studies, but they are configured and used differently in any particular research tradition. Let's have an illustrative look at three of these: interpretivism, social anthropology, and collaborative social research. We emphasize the *analytic* differences among them.

Three Approaches to Qualitative Data Analysis

Interpretivism. This line of inquiry has a long intellectual history. Dilthey's (1911/1977) thesis that human discourse and action could not be analyzed with the methods of natural and physical science was the defining conceptual perspective. Human activity was seen as "text"—as a collection of symbols expressing layers of meaning.

How would one interpret such a text? For Dilthey and the *phenomenologists,* the way led through "deep understanding," an empathy or indwelling with the subject of one's inquiries. For the *social interactionists,* interpretation comes via the understanding of group actions and interactions. In both cases there is an inevitable "interpretation" of meanings made both by the social actors and by the researcher.

Phenomenologists often work with interview transcripts, but they are careful, often dubious, about condensing this material. They do not, for example, use coding, but assume that through continued readings of the source material and through vigilance over one's presuppositions, one can reach the *"Lebenswelt"* of the informant, capturing the "essence" of an account—what is constant in a person's life across its manifold variations. This approach does not lead to covering laws, but rather to a "practical understanding" of meanings and actions.

Interpretivists of all types also insist that researchers are no more "detached" from their objects of study than are their informants. Researchers, they argue, have their own understandings, their own convictions, their own conceptual orientations; they, too, are members of a particular culture at a specific historical moment. Also they will be undeniably affected by what they hear and observe in the field, often in unnoticed ways. An interview will be a "co-elaborated" act on the part of both parties, not a gathering of information by one party. Note the analytic problem: If researchers use few preestablished instruments, it will be difficult to separate out "external" information from what they themselves have contributed when decoding and encoding the words of their informants.

Qualitative researchers in *semiotics,* in *deconstructivism,* in *aesthetic criticism,* in *ethnomethodology,* and in *hermeneutics* often have pursued this general line of inquiry—each, of course, with special emphases and variations.

Social anthropology. The primary methodology in this field, ethnography, stays close to the naturalist profile we just described: extended contact with a given community, concern for mundane, day-to-day events, as well as for unusual ones, direct or indirect participation in local activities, with particular care given to the *description* of local particularities; focus on individuals' perspectives and interpretations of their world; and relatively little prestructured instrumentation, but often a wider use of audio- and videotapes, film, and structured observation than in other research traditions.

Analytically, some points are notable. First, ethnographic methods tend toward the descriptive. The analysis task is to reach across multiple data sources (recordings, artifacts, diaries) and to condense them, with somewhat less concern for the conceptual or theoretical meaning of these observations. Of course, in deciding what to leave in, what to highlight, what to report first and last, what to interconnect, and what main ideas are important, analytic choices are being made continuously.

Social anthropologists are interested in the behavioral regularities in everyday situations: language use, artifacts, rituals, relationships. These regularities often are expressed as "patterns" or "language" or "rules," and they are meant to provide the inferential keys to the culture or society under study. As Van Maanen (1979) put it, the prime analytic task is to "uncover and explicate the ways in which people in particular (work) settings come to understand, account for, take action and otherwise manage their day-to-day situation." This "uncovering" and "explicating" is typically based on successive observations and interviews, which are reviewed analytically to guide the next move in the field.

Finally many social anthropologists are concerned with the genesis or refinement of theory. They may begin with a conceptual framework and take it out to the field for testing, refinement, or qualification. Much cross-cultural theory in socialization, parenting, and kinship has resulted from field research in a variety of settings.

Researchers in *life history, grounded theory, ecological psychology, narrative studies,* and in a wide range of *applied* studies (education, health care, family studies, program evaluation) often take this general line. This perspective has informed much of our own work, though we, along with *case study* analysts (e.g., Yin, 1991), have gravitated to more fully codified research questions, more standardized data collection procedures, and more systematic devices for analysis.

Collaborative social research. In this genre collective action is undertaken in a social setting. The protagonists may seek out a set of researchers comfortable with the action and willing to accompany the process in real time (see Schensul & Schensul, 1992). This accompaniment takes one of two typical forms: "reflexivity," where the research remains in an asking or questioning stance; or dialectics, where researchers and local actors may have opposing interpretations of the data.

As a general strategy for institutional change in this vein, *action research* has been practiced since the 1920s (Whyte, 1991). The researchers, with local help, design the outlines of a "field experiment" (e.g., changing the offerings in an institutional cafeteria, redesigning the operations and staffing of a ship). The data are collated and given to the "activists," both as feedback and to craft the next stage of operations. Note that this approach incorporates some of the features of naturalistic studies: participant observation, sensitivity to participants' concerns, focus on descriptive data in the initial phases, nonstandardized instrumentation, a holistic perspective, the search for underlying themes or patterns.

These points also hold for *collaborative action research* (Oja & Smulyan, 1989), where the researchers join closely with the participants from the outset. The aim is to transform the social environment through a process of critical inquiry—to act on the world, rather than being acted on. This approach is found in fields such as *critical ethnography* (Thomas, 1993) and *action science* (Argyris, Putnam, & Smith, 1985). The analytic tasks emphasize the use of action-related constructs, seen in a melioristic frame, and intellectual "emancipation" through unpacking taken-for-granted views and detecting invisible but oppressive structures.[4]

Analytic Methods: Some Common Features

Given these diverse approaches, can we see features that recur during any style of qualitative analysis? On the face of it, there may be some irreconcilable couples—for example, the quest for lawful relationships (social anthropology) versus the search for "essences" that may not transcend individuals, and lend themselves to multiple compelling interpretations (phenomenology).

Still, some analytic practices may be used across different qualitative research types. Here is a fairly classic set of analytic moves arranged in sequence:

- Affixing codes to a set of field notes drawn from observations or interviews
- Noting reflections or other remarks in the margins
- Sorting and sifting through these materials to identify similar phrases, relationships between variables, patterns, themes, distinct differences between subgroups, and common sequences
- Isolating these patterns and processes, commonalities and differences, and taking them out to the field in the next wave of data collection
- Gradually elaborating a small set of generalizations that cover the consistencies discerned in the database
- Confronting those generalizations with a formalized body of knowledge in the form of constructs or theories

We'll return to recurring features such as these, while acknowledging the desirable diversity of analytic approaches now in use.[5] Next, however, we need to take a step back to ask, What kind of data are we actually faced with in qualitative studies?

E. The Nature of Qualitative Data

General Nature

In some senses, all data are qualitative; they refer to essences of people, objects, and situations (Berg, 1989). We have a "raw" experience, which is then converted into words ("His face is flushed." . . . "He is angry.") or into numbers ("Six voted yes, four no." . . . "The thermometer reads 74 degrees.").

In this book we focus on data in the form of words—that is, language in the form of extended text. (Qualitative data also can appear as still or moving images, but we do not deal with these forms.)[6]

The words are based on *observation, interviews, or documents* (or as Wolcott [1992] puts it, "watching, asking, or examining"). These data collection activities typically are carried out in close proximity to a local setting for a sustained period of time.

Finally, such data are not usually immediately accessible for analysis, but require some *processing*. Raw field notes need to be corrected, edited, typed up; tape recordings need to be transcribed and corrected.

Some Underlying Issues

But it is not so simple. The words we attach to fieldwork experiences are inevitably framed by our implicit concepts. In fact, as Counelis (1991) suggests, a written description ("data") of someone clenching a fist and grimacing as "angry" is a conceptual substitute for the direct experience of one's own feelings and perceptions.

The processing of field notes is itself problematic. As Atkinson (1992) points out, they are really texts constructed by the field-worker on the basis of observation and participation: "What may be generated as 'data' is affected by what the ethnographer can treat as 'writable' and 'readable.' " Similarly, transcription of tapes can be done in many ways that will produce rather different texts.

Furthermore, what we consider as a descriptive, first-order "fact" (the number of arrests in a precinct, to use Van Maanen's [1983b] example) rapidly ramifies out into the interpretations and explanations that the people being studied have ("The brown-nosing lieutenant is cracking down"), *and* into the researcher's second-order conception of "what's going on"—the interpretations of the interpretations (positing the patrol officer's turf as a central idea).

The influence of the researcher's *values* is not minor (e.g., what one thinks about the fairness of arrests).

To put it another way, qualitative data are not so much about "behavior" as they are about *actions* (which carry with them intentions and meanings and lead to consequences). Some actions are relatively straightforward; others involve "impression management"—how people want others, including the researcher, to see them.

Furthermore, those actions always occur in specific situations within a social and historical context, which deeply influences how they are interpreted by both insiders and the researcher as outsider.

Thus the apparent simplicity of qualitative "data" masks a good deal of complexity, requiring plenty of care and self-awareness on the part of the researcher.

Strengths of Qualitative Data

What is important about well-collected qualitative data? One major feature is that they focus on *naturally occurring, ordinary events in natural settings,* so that we have a strong handle on what "real life" is like.

That confidence is buttressed by *local groundedness,* the fact that the data were collected in close proximity to a specific situation, rather than through the mail or over the phone. The emphasis is on a specific *case,* a focused and bounded phenomenon embedded in its context. The influences of the local context are not stripped away, but are taken into account. The possibility for understanding latent, underlying, or nonobvious issues is strong.

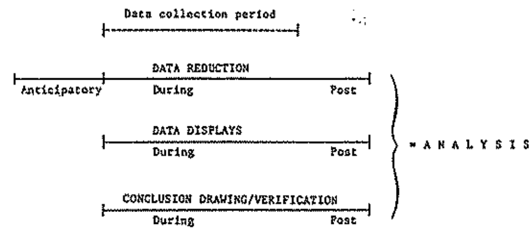
Another feature of qualitative data is their *richness and holism,* with strong potential for revealing complexity; such data provide "thick descriptions" that are vivid, nested in a real context, and have a ring of truth that has strong impact on the reader.

Furthermore, the fact that such data are typically collected over a *sustained period* makes them powerful for studying any process (including history); we can go far beyond "snapshots" of "what?" or "how many?" to just how and why things happen as they do—and even *assess causality* as it actually plays out in a particular setting. And the inherent *flexibility* of qualitative studies (data collection times and methods can be varied as a study proceeds) gives further confidence that we've really understood what has been going on.

Qualitative data, with their emphasis on people's "lived experience," are fundamentally well suited for locating the *meanings* people place on the events, processes, and structures of their lives: their "perceptions, assumptions, prejudgments, presuppositions" (van Manen, 1977) and for connecting these meanings to the *social world* around them.

We make three other claims for the power of qualitative data, to which we return during later chapters. They often

Figure 1.3
Components of Data Analysis: Flow Model



have been advocated as the best strategy for discovery, exploring a new area, *developing hypotheses.* In addition we underline their strong potential for *testing hypotheses,* seeing whether specific predictions hold up. Finally, qualitative data are useful when one needs to supplement, validate, explain, illuminate, or reinterpret *quantitative* data gathered from the same setting.

The strengths of qualitative data rest very centrally on the competence with which their analysis is carried out. What do we mean by analysis?

F. Our View of Qualitative Analysis

Our general view of qualitative analysis is outlined in Figure 1.3. We define *analysis* as consisting of three concurrent flows of activity: data reduction, data display, and conclusion drawing/verification. We explore each of these themes in more depth as we proceed through the book. For now, we make only some overall comments.

Data Reduction

Data reduction refers to the process of selecting, focusing, simplifying, abstracting, and transforming the data that appear in written-up field notes or transcriptions. As we see it, data reduction occurs continuously throughout the life of any qualitatively oriented project. Even before the data are actually collected (see Figure 1.1), anticipatory data reduction is occurring as the researcher decides (often without full awareness) which conceptual framework, which cases, which research questions, and which data collection approaches to choose. As data collection proceeds, further episodes of data reduction occur (writing summaries, coding, teasing out themes, making clusters, making partitions, writing memos). The data reduction/transforming process continues after fieldwork, until a final report is completed.

Data reduction is not something separate from analysis. It is *part* of analysis. The researcher's decisions—which data chunks to code and which to pull out, which patterns best summarize a number of chunks, which evolving story to tell—*are all analytic choices.* Data reduction is a form of analysis that sharpens, sorts, focuses, discards, and organizes data in such a way that "final" conclusions can be drawn and verified. As Tesch (1990), points out, it also can be seen as "data condensation."

By "data reduction" we do *not* necessarily mean quantification. Qualitative data can be reduced and transformed in many ways: through selection, through summary or paraphrase, through being subsumed in a larger pattern, and so on. Occasionally it may be helpful to convert the data into primitive quantities (e.g., the analyst decides that the case being looked at has a "high" or "moderate" degree of administrative centralization), but this is not always wise. Even when it does look like a good analytical strategy, our advice is to keep the numbers, and the words you used to derive the numbers, together in your ensuing analysis. It is important not to strip the data at hand from the context in which they occur.

Data Display

The second major flow of analysis activity is data display. Generically, a *display* is an organized, compressed assembly of information that permits conclusion drawing and action. In daily life, displays vary from gasoline gauges to newspapers to computer screens to factor analysis printouts. Looking at displays helps us to understand what is happening and to do something—either analyze further or take action—based on that understanding.

The most frequent form of display for qualitative data in the past has been *extended text.* As we note later, text (in the form, say, of 3,600 pages of field notes) is terribly cumbersome. It is dispersed, sequential rather than simultaneous, poorly structured, and extremely bulky. Using only extended text, a researcher may find it easy to jump to hasty, partial, unfounded conclusions. Humans are not very powerful as processors of large amounts of information; our cognitive tendency is to reduce complex information into selective and simplified gestalts or easily understood configurations. Or we drastically overweight vivid information, such as the exciting event that jumps out of page 124 of the field notes after a long, "boring" passage. Pages 109 through 123 may suddenly have been collapsed, and the criteria for weighting and selecting may never be questioned. Extended text can overload humans' information-processing capabilities (Faust, 1982) and preys on their tendencies to find simplifying patterns.

In the course of our work, we have become convinced that better displays are a major avenue to valid qualitative analysis. The displays discussed in this book include many types of matrices, graphs, charts, and networks. All are designed to assemble organized information into an immediately accessible, compact form so that the analyst can see what is happening and either draw justified conclusions or move on to the next step of analysis the display suggests may be useful.

As with data reduction, the creation and use of displays is not separate from analysis, it is a *part* of analysis. Designing a display—deciding on the rows and columns of a matrix for qualitative data and deciding which data, in which form, should be entered in the cells—are analytic activities. (Note that designing displays also has clear *data reduction* implications.)

The dictum "You are what you eat" might be transposed to "You know what you display." In this book we advocate more systematic, powerful displays and urge a more inventive, self-conscious, iterative stance toward their generation and use.

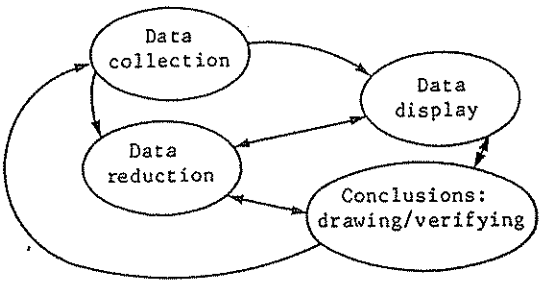
Conclusion Drawing and Verification

The third stream of analysis activity is conclusion drawing and verification. From the start of data collection, the qualitative analyst is beginning to decide what things mean—is noting regularities, patterns, explanations, possible configurations, causal flows, and propositions. The competent researcher holds these conclusions lightly, maintaining openness and skepticism, but the conclusions are still there, inchoate and vague at first, then increasingly explicit and grounded, to use the classic term of Glaser and Strauss (1967). "Final" conclusions may not appear until data collection is over, depending on the size of the corpus of field notes; the coding, storage, and retrieval methods used; the sophistication of the researcher; and the demands of the funding agency, but they often have been prefigured from the beginning, even when a researcher claims to have been proceeding "inductively."

Conclusion drawing, in our view, is only half of a Gemini configuration. Conclusions are also *verified* as the analyst proceeds. Verification may be as brief as a fleeting second thought crossing the analyst's mind during writing, with a short excursion back to the field notes, or it may be thorough and elaborate, with lengthy argumentation and review among colleagues to develop "intersubjective consensus," or with extensive efforts to replicate a finding in another data set. The meanings emerging from the data have to be *tested* for their plausibility, their sturdiness, their "confirmability"—that is, their *validity.* Otherwise we are left with interesting stories about what happened, of unknown truth and utility.

We have presented these three streams—data reduction, data display, and conclusion drawing/verification—as interwoven before, during, and after data collection in paral-

Figure 1.4
Components of Data Analysis: Interactive Model



lel form, to make up the general domain called "analysis." The three streams can also be represented as shown in Figure 1.4. In this view the three types of analysis activity and the activity of data collection itself form an interactive, cyclical process. The researcher steadily moves among these four "nodes" during data collection and then shuttles among reduction, display, and conclusion drawing/verification for the remainder of the study.

The coding of data, for example *(data reduction)*, leads to new ideas on what should go into a matrix *(data display)*. Entering the data requires further data reduction. As the matrix fills up, preliminary *conclusions* are drawn, but they lead to the decision, for example, to add another column to the matrix to *test* the conclusion.

In this view, qualitative data analysis is a continuous, iterative enterprise. Issues of data reduction, of display, and of conclusion drawing/verification come into figure successively as analysis episodes follow each other. But the other two issues are always part of the ground.

Such a process is actually no more complex, conceptually speaking, than the analysis modes quantitative researchers use. Like their qualitative brethren, they must be preoccupied with data reduction (computing means, standard deviations, indexes), with display (correlation tables, regression printouts), and with conclusion drawing/verification (significance levels, experimental/control differences). But their activities are carried out through well-defined, familiar methods, are guided by canons, and are usually more sequential than iterative or cyclical. Qualitative researchers, on the other hand, are in a more fluid—and a more pioneering—position.

Thus, as we've suggested, qualitative analysis needs to be well documented as a process—mainly to help us learn. Purposes of "auditing" aside, we need to understand more clearly just what is going on when we analyze data, to reflect, refine our methods, and make them more generally usable by others. See Chapter 10, section D for more.

G. Using This Book

Overview

This book is organized roughly according to the chronology of qualitative research projects, from initial design to final reports. For a quick overview of that sequence, see Chapter 13. A run through the Table of Contents will also help.

Format of Specific Methods

We've designed this sourcebook to be as practical as possible. Each method is described in this format:

Name of method.

Analysis problem. The problem, need, or difficulty faced by a qualitative data analyst, for which the method proposed is a useful solution.

Brief description. What the method is and how it works.

Illustration. In more detail, a "minicase" showing how the method is developed and used. Usually this section has a variety of subheadings, such as "Building the Display," "Entering the Data," and "Analyzing the Data."

Variations. Alternative approaches using the same general principle. Relevant work of other researchers is cited.

Advice. Summarizing comments about use of the method, and tips for using it well.

Time required. Approximate estimates (contingent on subject matter, researcher's skill, research questions being asked, number of cases, etc.).

The text also includes supplementary methods, described in a briefer format, that can be used with or instead of the principal method being discussed.

Suggestions for Users

Ideas about what a reader should "do" with any particular book are often presumptuous, mistaken, or both. As someone has pointed out, a book is essentially a random-access display that users activate merely by turning their eyes toward it. Authors have no control over what readers end up doing. Nevertheless we give some advice to different types of users, based on our own and others' experience with the first edition.

Experienced researchers. This is a sourcebook. Colleagues have told us they have used it in several ways.

1. Browsing. The book contains a wide range of material, so simply exploring it in an unstructured way can be fruitful.

2. Problem solving. Anyone opening the book comes to it with more or less specifically defined problems in doing qualitative data analysis. The index has been designed to be "problem-sensitive" to permit easy access to appropriate sections of the book. The Table of Contents can also be used in this way.

3. A to Z. Some readers prefer to go through a book sequentially, from start to finish. We have organized the book so it makes sense that way. Some A-to-Z readers have told us that they found it useful to jump forward from the first part of Chapter 5 (within-case descriptive analysis) to Chapter 9 (matrix displays) and Chapter 10 (tactics) and then return to Chapter 6 (within-case explanatory analysis) before going on to cross-case analysis (Chapters 7 and 8) and the remainder of the book.

4. Operational use. For readers conducting an ongoing qualitative research project, either alone or with colleagues, it's useful to read particular sections focusing on upcoming analysis tasks (e.g., the formation of research questions, coding, time-ordered displays), then to discuss them with available colleagues, and then to plan next steps in the project, revising the methods outlined here or developing new ones. Chapters 2 and 3 are particularly helpful during proposal development and in the start-up and design phases of a project.

5. Research consulting. The book can be used by people with an advisory or consulting role in the start-up and ongoing life of research projects. Assuming good problem identification, a research consultant can work with the client in either a problem-solving or a direct training mode (see below) to aid with thoughtful project design, and coping with early problems.

Teachers of research methods. Some colleagues have used the book as a main text, others as a supplementary one. In either case our very strong advice is to engage students in active design, data collection, and analysis. Several of our readers also said that teacher and students need to engage in analytic work *together* as colleagues. The book is not designed to be helpful in the type of methods course that is "about" qualitative research and provides no direct experience in doing it. Actual data are needed.

A typical teaching sequence is:

Chapter 2 (pre-data collection work) (with most emphasis on conceptual framework, research questions, and sampling)
Chapter 3 (technical and managerial design issues)
Chapter 4 (analysis during data collection)
Chapters 5 and 6 (within-case analysis)
Chapter 10 (conclusion-drawing and verification tactics)
Chapters 7 and 8 (cross-case analysis)
Chapter 9 (matrix building and use)
Chapter 11 (ethics)
Chapter 12 (reports)
Chapter 13 (overview, review)

A colleague told us that he has begun a course by discussing section F of Chapter 1 ("Our View of Qualitative Data Analysis") and then asking students to draw conclusions from an existing data set, such as displays from our book *Innovation Up Close* (Huberman & Miles, 1984), first working with ideas in Chapters 5 and 6 and then turning to Chapters 7 and 8. After that, he supports students in doing quick studies of their own, analyzing the data they produce.

We've found that the book can also be used in intensive time-block teaching. For a 3-day training workshop (maximum size: 25 participants), the most profitable sequence begins with careful attention to each of the first four sections of Chapter 2, then goes on to work on coding in Chapter 4, and then focuses in more depth on a few display modes, one by one, in Chapters 5 through 8. These might include partially ordered, time-ordered, and conceptually ordered displays in Chapter 5; causal networks in Chapter 6; then similarly ordered meta-matrices and case-ordered predictor-outcome displays in Chapters 7 and 8.

Repeated practice in using various tactics in Chapter 10, as suggested there, is also useful. Our experience is that a rapid overview of the tactics is best done *after* people have had direct experience with a few analysis episodes.

For each topic, we have used a learning approach like this, carried out by individuals or working pairs, who stay together throughout the workshop:

1. Introductory lecture and/or reading to clarify the main conceptual points of the section.
2. A brief learning task (e.g., drawing a conceptual framework, designing a section of a coding scheme, designing a matrix, drawing an event-state network, interpreting a filled-out matrix, writing an analysis). Time here is usually 30 minutes or less.
3. Comparing the products of individuals or pairs by using an overhead projector or newsprint; drawing generalizations, giving advice.
4. If step 2 has not involved actual data from the participants' current or projected research, there should be a period for application to their own work. Consultative help should be available from the workshop or course leader.

The same general principles apply when the book is being used in a semester-long course, although the coverage will be deeper and more leisurely. Interim exercises focusing on actual research tasks, critiqued in class, are particularly productive. Active, reflective self-documentation through personal logs or journals is also profitable (see also Chapter 10, section D).[7]

Students and other novice researchers. We give some direct advice here, keeping in mind that you will often be working alone, usually on a single case, and may be feeling worried about the quality of your study—dissertation or not.

1. This book helps you on *analysis.* Use other, introductory books to help with the basics (e.g., Bogdan & Biklen, 1992; Glesne & Peshkin, 1992; LeCompte & Preissle, with Tesch, 1993; Lofland & Lofland, 1984).
2. Protect yourself from getting overwhelmed by looking at the big picture. Look at the overview (Figure 13.1) in Chapter 13; read Chapter 1, section F; scan the ideas on displays in Chapter 9. Remember that the whole book is organized in rough time sequence.
3. Scan the "Advice" sections for each chapter.
4. Learn by doing. Use your own study (whether it is in the planning stage or under way) as a vehicle and apply to it the suggestions in each chapter. When reading this book, write out your own conclusions from the display examples and compare with those of friends or colleagues.
5. Compensate for the problem of having to work alone by finding someone to be a critical friend to react to your work as you go.
6. Keep an informal log or journal of what you are running up against. This tactic will help your learning and will be useful when you write up your study.
7. Don't worry about the jargonlike names of particular displays; the issue is what a display can do for you.
8. The time estimates supplied are very rough. Don't be alarmed if you take a lot less or much more than the estimate.
9. The biggest enemy of your learning is the gnawing worry that you're not "doing it right." Dissertation work tends to encourage that. But any given analytic problem can be approached in many useful ways. Creativity, inventing your way out of a problem, is definitely the better stance. Some of the more interesting methods you see in the book were created by students and other beginning researchers.

Staff specialists and managers. The workshop formats described above have been used with organizational consultants interested in improving their use of qualitative data in diagnosing organizations. It's also useful to have participants work in pairs or small teams through a complete consulting case study, analyzing a bank of data collected from clients and generating action interventions. Most of the suggestions for experienced researchers can be helpful for this audience as well.

Notes

1. We like the way Phillips (1990) makes this point. He notes that researchers are fully able to investigate beliefs and other forms of cognition:

> We can inquire into the beliefs of a society, how they came about, what their effects are, and what the status is of the evidence that is offered in support of the truth of the beliefs. And we can get these matters right or wrong—we can describe these beliefs correctly or incorrectly, or we can be right or make mistakes about their origins or their effects. It simply does not follow from the fact of the social construction of reality that scientific inquiry becomes impossible or that we have to become relativists. (p. 42)

2. If these various epistemological labels seem like arcane shorthand to you, we suggest the thoughtful, articulate debates presented in the collection by Guba (1990).

3. Mishler (1990) has a pragmatic, articulate view of the research scientist as more of a craftsperson than a logician: "Competence depends on apprenticeship training, continued practice and experience-based, contextual knowledge of the specific methods applicable to a phenomenon of interest, rather than on an abstract 'logic of discovery' and application of formal 'rules' " (p. 435).

4. The reader may have noted that we have left out some of the more linguistically oriented approaches to qualitative analysis (e.g., *ethnoscience, discourse analysis, content analysis*), along with some more recent approaches straddling epistemological borders (*phenomenography*). We suggest readings on these at several points.

5. The analytic sequence depicted here is probably closest to ethnographic methods, as extended by work in grounded theory. It moves from one inductive inference to another by selectively collecting data, comparing and contrasting this material in the quest for patterns or regularities, seeking out more data to support or qualify these emerging clusters, and then gradually drawing inferences from the links between other new data segments and the cumulative set of conceptualizations.

For a thoughtful and practical look at this general approach, see Wolcott (1994), who distinguishes three major operations: *description* (rendering "what's going on here," including respondents' words), *analysis* (showing how things work, through systematically identifying key factors and relationships), and *interpretation* (making sense of meanings in context—"What's to be made of it all?"). All three are needed, with the balance depending on the study.

Our own approach is sometimes more deductive; it may begin with an orienting set of relationships or constructs and derive from them a provisional coding system, which then is applied to initial field notes; condensed data in systematic displays are used to aid with conclusion drawing, enriched and tested by a new cycle of data from the next round.

Much phenomenological research stops at the moment of generalization and makes little attempt to connect to sets of constructs or covering laws. Some phenomenological methods, such as the "hermeneutical circle" used for the interpretation of texts, center more on interpretation than on gaining firm empirical knowledge of social or natural facts. Still the aim is to construct a coherent, internally consistent argument—one with theoretical referents from a series of "empirical facts" in the form of texts, perceptions, and social acts. This process often involves multiple readings and condensations in the search for regularities and "essences"; with this empathy and familiarity (sometimes supported by further dialogue with the informant), farther-ranging interpretations, often based on additional social facts, may be undertaken.

The analytic challenge for all qualitative researchers is finding coherent descriptions and explanations that still include all of the gaps, inconsistencies, and contradictions inherent in personal and social life. The risk is forcing the logic, the order, and the plausibility that constitute theory making on the uneven, sometimes random, nature of social life. Yet without theory we can be left with banal, unilluminating descriptions.

6. We have found some treatments of images as data especially useful. Harper (1989) suggests that such data can be used in several ways: a scientific, descriptive mode; a narrative, storytelling mode; a "reflexive" mode where people respond to pictures of themselves and their environments; or a phenomenological mode, deepening the personal meanings of the researcher.

Ball and Smith (1992) discuss still photographs and remind us that they are not automatically more "realistic" than words, are very much subject to interpretation and captioning, depend on the context, and can be staged or faked.

7. Treatments of teaching qualitative methods we have found especially useful include Strauss's (1988) thoughtful reflections and his associated 1987 book; the special issue of *Anthropology and Education Quarterly* (Vol. 14, no. 3, 1983, which includes helpful articles by Bogdan, Hall, and Lofland and Lofland); Ely, Anzul, Friedman, Garner, and Steinmetz (1991), which has many vivid, detailed examples of student work; and Webb and Glesne (1992), who surveyed 75 teachers of qualitative methods and summarized typical problems students encounter.

Certain themes recur: Field-based courses are essential (the only way to learn to do analysis is to do it); strong mutual support among students is critical, as is a collaborative student-teacher relationship; course size needs to be small (fewer than 15), and the course should last for at least a semester; and self-reflection through journals and mutual critique is very important.

2

Focusing and Bounding the Collection of Data

THE SUBSTANTIVE START

Contrary to what you might have heard, qualitative research designs do exist. Some are more deliberate than others. At the proposal stage and in the early planning and start-up stages, many design decisions are being made—some explicitly and precisely, some implicitly, some unknowingly, and still others by default. The qualitative researcher is beginning to focus on the study's issues, the cases to be studied, the data to be collected, and how these data will be managed and analyzed.

This book is about *analysis*. Why are we talking about design? As Figure 1.3 suggests, study design decisions can, in a real sense, be seen as analytic—a sort of *anticipatory data reduction*—because they constrain later analysis by ruling out certain variables and relationships and attending to others. Design decisions also permit and support later analysis; they prefigure your analytic moves.

Some design decisions are mainly conceptual: the conceptual framework and research questions, sampling, case definition, instrumentation, and the nature of the data to be collected. Others (discussed in Chapter 3), though they appear in the guise of "management" issues, are equally focusing and bounding: how data will be stored, managed, and processed; what computer software may be used to support the work; and which agreements are made with the people being studied.

We cannot deal thoroughly here with qualitative research design; see the detailed, helpful suggestions made by Marshall and Rossman (1989). In this chapter we discuss the analytic issues that arise as a study is bounded, focused, and organized. We provide specific examples, but want to emphasize that these issues must be dealt with uniquely in any particular study. They may be approached loosely or tightly; in either case, initial design decisions nearly always lead to redesign. Qualitative research designs are not copyable patterns or panaceas that eliminate the need for building, revising, and "choreographing" your analytic work (Preissle, 1991).

Tight Versus Loose: Some Trade-offs

Prior to fieldwork, how much shape should a qualitative research design have? Should there be a preexistent conceptual framework? A set of research questions? Some predesigned devices for collecting data? Does such prior bounding of the study blind the researcher to important features in the case, or cause misreading of local informants' perceptions? Does lack of bounding and focusing lead to indiscriminate data collection and data overload? These are recurrent questions in qualitative analysis, and they have started up lively debate. Let's try to order the terms of the debate and to explain our own position.

Any researcher, no matter how unstructured or inductive, comes to fieldwork with *some* orienting ideas. A sociologist may focus on families or organizations (rather than, say, on rock formations or anthills) and, within that focus, will look for data marked by conceptual tags (roles, relationships, routines, norms). If that researcher looks at closets or lunchrooms, it is not with the eyes of an architect or a cook, but with an interest in what the room and its contents have to say about the patterns shared by people using it. A psychologist would orient differently toward the same phenomena, "seeing" motivation, anxiety, communication, and cognition.

The conventional image of field research is one that keeps prestructured designs to a minimum. Many social anthropologists and social phenomenologists consider social processes to be too complex, too relative, too elusive, or too exotic to be approached with explicit conceptual frames or standard instruments. They prefer a more loosely structured, emergent, inductively "grounded" approach to gathering data: The conceptual framework should emerge from the field in the course of the study; the important research questions will come clear only gradually; meaningful settings and actors cannot be selected prior to fieldwork; instruments, if any, should be derived from the properties of the setting and its actors' views of them.

We go along with this vision—up to a point. Highly inductive, loosely designed studies make good sense when experienced researchers have plenty of time and are exploring exotic cultures, understudied phenomena, or very complex social phenomena. But if you're new to qualitative studies and are looking at a better understood phenomenon within a familiar culture or subculture, a loose, inductive design may be a waste of time. Months of fieldwork and voluminous case studies may yield only a few banalities. As Wolcott (1982) puts it, there is merit in openmindedness and willingness to enter a research setting looking for questions as well as answers, but it is "impossible to embark upon research without some idea of what one is looking for and foolish not to make that quest explicit" (p. 157).

Tighter designs are a wise course, we think, for researchers working with well-delineated constructs. In fact, we should remember that qualitative research can be outright "confirmatory"—that is, can seek to test or further explicate a conceptualization. Tighter designs also provide clarity and focus for beginning researchers worried about diffuseness and overload.

So a case can be made for tight, prestructured qualitative designs and for loose, emergent ones. Much qualitative research lies between these two extremes. Something is known conceptually about the phenomenon, but not enough to house a theory. The researcher has an idea of the parts of the phenomenon that are not well understood and knows where to look for these things—in which settings, among which actors. And the researcher usually has some initial ideas about how to gather the information. At the outset, then, we usually have at least a rudimentary conceptual framework, a set of general research questions, some notions about sampling, and some initial data-gathering devices.

How prestructured should a qualitative research design be? Enough to reach the ground, as Abraham Lincoln said when asked about the proper length of a man's legs. It depends on the time available, how much already is known about the phenomena under study, the instruments already available, and the analysis that will be made.

Our stance lies off center, toward the structured end. To our earlier epistemological reasons, we should add a few that are more mundane. First, the looser the initial design, the less selective the collection of data; *everything* looks important at the outset if you are waiting for the key constructs or regularities to emerge from the case, and that wait can be a long one. The researcher, submerged in data, will need months to sort it out. You may have that kind of time if you're doing a dissertation or are funded by a long-term grant, but most projects are time constrained.

Second, fieldwork may well involve multiple-case research, rather than single-case studies. If different fieldworkers are operating inductively, with no common framework or instrumentation, they are bound to end up with the double dilemma of data overload and lack of comparability across cases.[1]

Then, too, we should not forget why we are out in the field in the first place: to describe and analyze a pattern of relationships. That task requires a set of analytic categories (cf. Mishler, 1990). Starting with them (deductively) or getting gradually to them (inductively) are both possible. In the life of a conceptualization, we need both approaches—and may well need them from several field researchers—to pull a mass of facts and findings into a wide-ranging, coherent set of generalizations.

Finally, as researchers, we do have background knowledge. We see and decipher details, complexities, and subtleties that would elude a less knowledgeable observer. We know some questions to ask, which incidents to attend to closely, and how our theoretical interests are embodied in the field. Not to "lead" with your conceptual strength can be simply self-defeating.

Clearly, trade-offs are involved here. In multiple-case research, for example, the looser the initial framework, the more each researcher can be receptive to local idiosyncrasies—but cross-case comparability will be hard to get, and the costs and the information load will be colossal.

Tightly coordinated designs face the opposite dilemma: They yield more economical, comparable, and potentially generalizable findings, but they are less case-sensitive and may entail bending data out of contextual shape to answer a cross-case analytic question. The solution may well lie in avoiding the extremes.

With this backdrop, let's look more closely at the aspects of a study design involving decisions about focusing and bounding the collection of qualitative data in the field. In this chapter we focus on conceptual aspects, including developing a conceptual framework, formulating research questions, defining the case, sampling, and instrumentation. We turn to management issues in Chapter 3.

A. Building a Conceptual Framework

Rationale

Theory building relies on a few general constructs that subsume a mountain of particulars. Categories such as "social climate," "cultural scene," and "role conflict" are the labels we put on intellectual "bins" containing many discrete events and behaviors. Any researcher, no matter how inductive in approach, knows which bins are likely to be in play in the study and what is likely to be in them. Bins come from theory and experience and (often) from the general objectives of the study envisioned. Setting out bins, naming them, and getting clearer about their interrelationships lead you to a conceptual framework.

Doing that exercise also forces you to be selective—to decide which variables are most important, which relationships are likely to be most meaningful, and, as a consequence, what information should be collected and analyzed—at least at the outset. If multiple researchers are involved, the framework helps them study the same phenomenon in ways that will permit an eventual cross-case analysis.

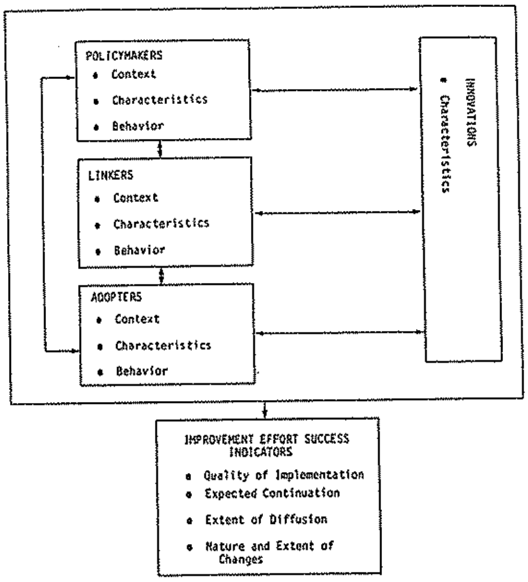
Brief Description

A conceptual framework explains, either graphically or in narrative form, the main things to be studied—the key factors, constructs or variables—and the presumed relationships among them. Frameworks can be rudimentary or elaborate, theory-driven or commonsensical, descriptive or causal.

Illustrations

Let's look at a few examples. First, Figure 2.1 presents a rudimentary, mostly descriptive framework from a large-scale contract research study (The Network, Inc., 1979).

Figure 2.1
Conceptual Framework for a Study of the Dissemination of Educational Innovations (The Network, Inc., 1979)

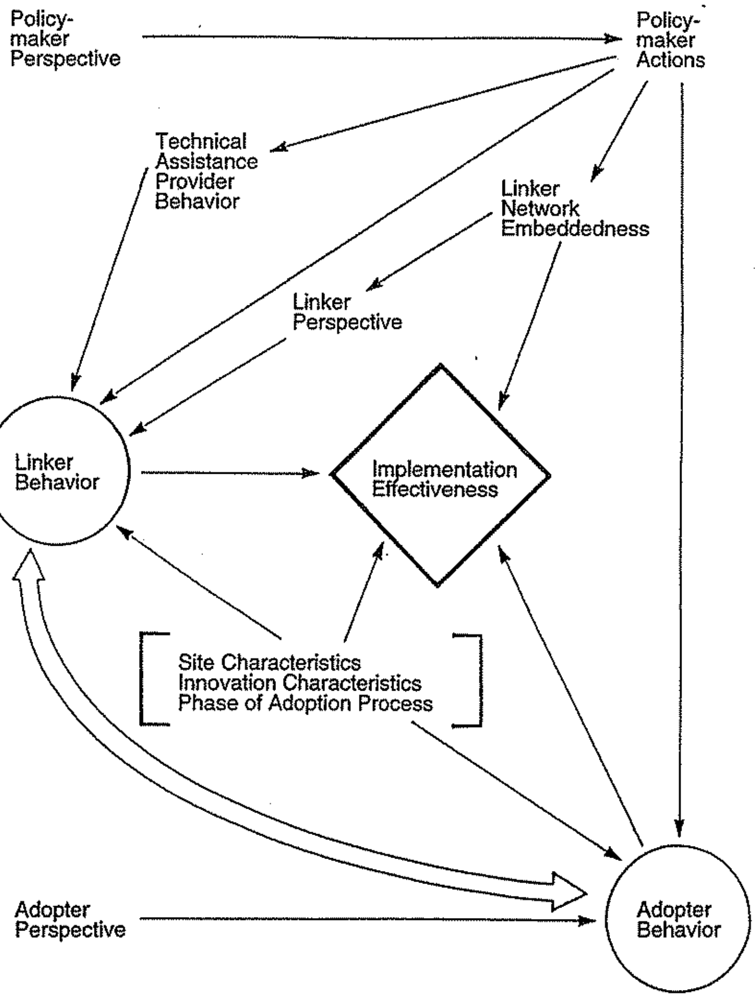


The study's general objectives were to examine several programs aimed at "school improvement" through the dissemination of exemplary innovations, to understand the reasons for implementation success, and to make policy recommendations.

Here we see an example of the bins approach. The framework is mostly a visual catalogue of *roles* to be studied (policymakers, linkers, adopters) and, within each role, where these people work and what they do (context, characteristics, behavior). A second major aspect of the study is the *innovations,* notably their characteristics. A third aspect is the *outcomes* of the innovations (improvement effort success indicators).

What does this framework do for the researcher? First, it specifies who and what will and will not be studied. For example, it looks as if the people who developed the innovations will not be studied. It also appears that the study will focus on four types of *successful* outcomes. Second, the framework assumes some *relationships,* as indicated by the arrows. Some of these relationships are purely logical—for instance, the idea that adopters and the innovations will influence one another—but the arrows also mirror empirical findings.

Figure 2.2
Second Conceptual Framework for a Study of the Dissemination of Educational Innovations
(The Network, Inc., 1979)

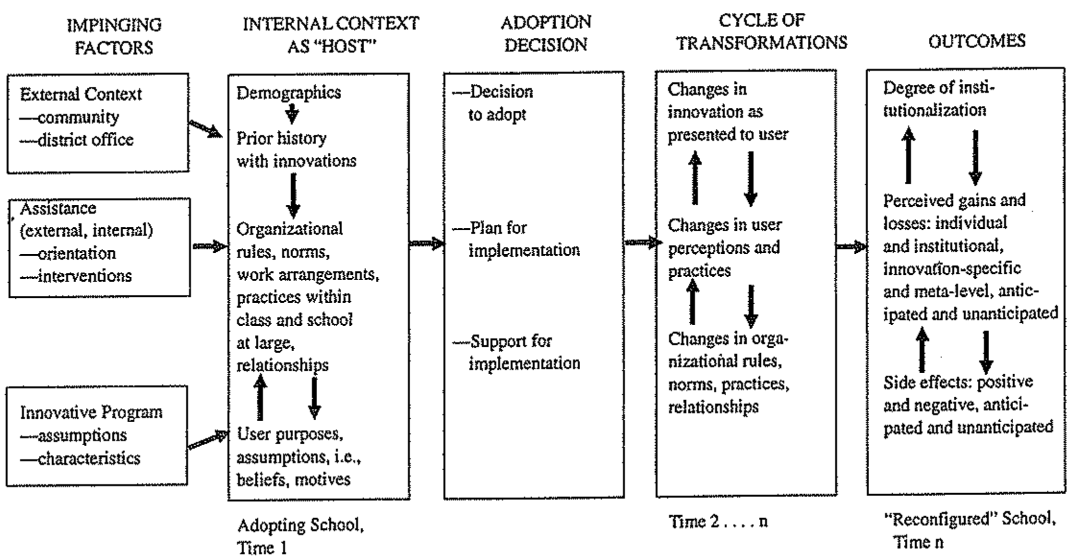


We see here the focusing and bounding function of a conceptual framework. Some, not all, actors are going to be studied, along with some, not all, aspects of their activity. Only some relationships will be explored, certain kinds of outcomes measured, and certain analyses made—at least at the outset.

Now for a slightly more complex, more inferential conceptual frame using some of the same variables (Figure 2.2). It comes from the same study. It is a refinement of the first illustration, with heavier bets being made on the interrelationships. For example, "policymakers" are hypothesized to influence "linkers" through the provision of technical assistance and through interventions in the linkers' network.

There are few two-way arrows in this cut. The researcher is deciding to collect information *selectively,* at least on the first go, to test some hypotheses. Similarly, it looks as if the study will focus more heavily on "linker behavior," "adopter behavior," and "implementation effectiveness"—that is, on variables coming later in the

Figure 2.3
Conceptual Framework for a Multicase "School Improvement" Field Study, Initial Version
(Huberman & Miles, 1984)



causal chain indicated by the arrows. "Linker perspective," for example, will be studied only as a presumed consequence of "network embeddedness" and as a predictor of "linker behavior."

On our continuum from exploratory to confirmatory designs, the first illustration is closer to the exploratory end and the second to the confirmatory. Let's have a look at one about midway along the continuum (Figure 2.3). This framework is of particular interest in that it lays out the study from which we draw many of our subsequent exhibits (Huberman & Miles, 1983b, 1984).[2]

Once again, we have the bins, labeled as *events* (e.g., "prior history with innovations"), *settings* (e.g., "community, district office . . . adopting school"), *processes* (e.g., "assistance," "changes in user perceptions and practices"), and *theoretical constructs* (e.g., "organizational rules"). Some of the outcomes are hypothesized (e.g., "degree of institutionalization"), but most are open-ended ("perceived gains and losses"). The directional arrows follow time flow, but some bets still are being made (e.g., that most assistance comes early and that reciprocal changes will occur among the innovation, its users, and the organization).

But the contents of each bin are less predetermined than in Figure 2.1. Each researcher in the study will have to find out what the "characteristics" of the innovations are at the field site and how these factors will affect "implementation effectiveness." This is still a very general brief.

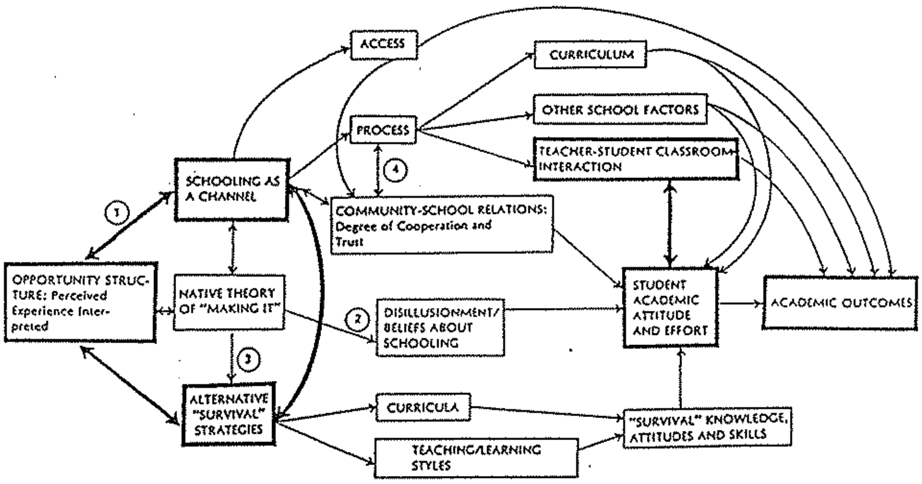
It is also a brief that can change en route, as this conceptual framework did. As qualitative researchers collect data, they revise their frameworks—make them more precise, replace empirically feeble bins with more meaningful ones, and reconstrue relationships. Conceptual frameworks are simply the current version of the researcher's map of the territory being investigated. As the explorer's knowledge of the terrain improves, the map becomes correspondingly more differentiated and integrated, and researchers in a multiple-case study can coordinate their data collection even more closely.

Variations

Here is an ethnographic framework for the study of minority children's school experience (Box 2.1).

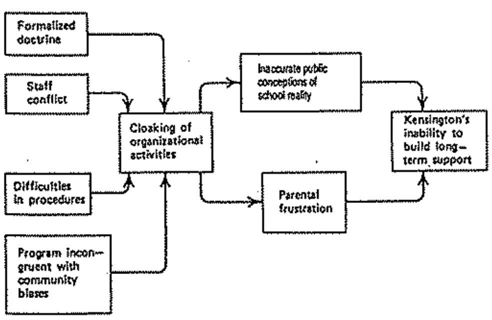
The rectangles with bold outline are the bins with the highest research priority. The numbers by "arrows" indi-

Box 2.1
Conceptual Framework for an Ethnographic Study (Ogbu, 1981)



cate the relationships to be examined initially; some show one-way influence, others two-way. There are purely descriptive bins (curriculum, teaching/learning styles) and more conceptual labels (opportunity structure, "survival" knowledge and strategies). Ensuing, inevitable changes can be mapped onto this frame, or they can call for a recasting.

Box 2.2
Emergent Conceptual Framework: The Antecedents and Consequences of Cloaking Organizational Activities (Smith & Keith, 1971)



Conceptual frameworks can also evolve and develop out of fieldwork itself. L. M. Smith and Keith (1971) were among the first to do this graphically. Box 2.2 shows an example drawn from their study of the creation of a new school. The researchers noticed something they called "cloaking of organizational activities"—keeping internal functioning protected from external view. Why did this happen, and what were the consequences? Smith and Keith believed that the school's "formalized doctrine"—its philosophy and setup, along with procedural difficulties, staff conflict, and poor program fit with community biases—led to the cloaking. The cloaking, in turn, led to inaccurate public perceptions and parental frustration; Kensington School could not build long-term support for itself.

Although the cloaking idea came from prior organizational research, the way it patterned at Kensington plus the other associated variables essentially were derived inductively. Smith and Keith used many such emergent conceptual frameworks to explicate their understanding.

Advice

Here are some suggestions that summarize and extend what has been said in this section.

1. Conceptual frameworks are best done graphically, rather than in text. Having to get the entire framework on a single page obliges you to specify the bins that hold the discrete phenomena, to map likely relationships, to divide variables that are conceptually or functionally distinct, and to work with all of the information at once. Try your hand at it, especially if you are a beginning researcher.

2. Expect to do several iterations, right from the outset. There are probably as many ways of representing the main variables as there are variables to represent, but some—typically later cuts—are more elegant and parsimonious than others.

3. If your study has more than one researcher, have each field researcher do a cut at a framework early on and then compare the several versions. This procedure will show, literally, where everyone's head is. It usually leads to an explication of contentious or foggy areas that otherwise would have surfaced later on, with far more loss of time and data.

4. Avoid the no-risk framework—that is, one that defines variables at a very global level and has two-directional arrows everywhere. This avoidance amounts essentially to making no focusing and bounding decisions, and is little better than the strategy of going indiscriminately into the field to see what the site has to "tell." However, you can *begin* with such an omnibus framework—Figure 2.1 is close to the no-risk framework—as a way of getting to a more selective and specific one.

5. Prior theorizing and empirical research are, of course, important inputs. It helps to lay out your own orienting frame and then map onto it the variables and relationships from the literature available, to see where the overlaps, contradictions, refinements, and qualifications are.

Time Required

Taken singly, each iteration of a conceptual framework does not take long. If you've already done much thinking about the study and are on top of the literature, an initial cut might take 45 minutes to an hour. If you doubt this estimate, we should mention that, in training workshops using these materials, participants always have been able to develop preliminary conceptual frameworks for their qualitative studies in under 30 minutes. The issue is usually not developing a giant scheme *de novo,* but making explicit what is already in your mind. Furthermore, working briskly seems to help cut away the superfluous—even to foster synthesis and creativity.

Successive cuts of a framework are more differentiated and have to resolve the problems of the prior one, so they take longer—an hour or two. If you are new to the field, the first revision may take 2 hours or so. But it is enjoyable work.

B. Formulating Research Questions

Rationale

It is a direct step from conceptual framework to research questions. If I have a bin labeled "policymaker," as in Figure 2.1, and within it a subhead called "behavior," I am implicitly asking myself some questions about policymakers' behaviors. If I have a two-way arrow from the policymaker bin to an "innovation" bin, as again in Figure 2.1, my question has to do with how policymakers behave in relation to the introduction of innovations and, reciprocally, how different kinds of innovations affect policymakers' behaviors.

If my conceptual framework is more constrained, so are my questions. In Figure 2.2, my interest in "policymaker actions" is more focused. I want to ask, How do policymakers' actions affect adopter behavior, linkers' "embeddedness," linker behavior, and the behavior of different technical "assisters"? Here I am down to specific variables in a bin and specific relationships between bins. Naturally the nature and quality of the relationships will take some pondering before the research questions come clear.

What do these questions do for me? First, they make my theoretical assumptions even more explicit. Second, they tell me what I want to know most or first; I will start by channeling energy in those directions. My collection of data will be more focused and limited.

I am also beginning to make some implicit sampling decisions. I will look only at *some* actors in *some* contexts dealing with *some* issues. The questions also begin to point me toward data-gathering devices—observations, interviews, document collection, or even questionnaires.

Finally, the rough boundaries of my analysis have been set, at least provisionally. If I ask, How do policymaker actions affect adopter behavior? I will be looking at data on adopter behaviors, policymaker actions, and their influences on each other, *not* at what else affects adopter behaviors. The research questions begin to operationalize the conceptual framework.

This is clearly a deductive model. We begin with some orienting constructs, extract the questions, and then start to line up the questions with an appropriate sampling frame and methodology. Inductivists could argue that we might have the wrong concepts, the wrong questions, and the wrong methodology, and only would find how wrongheaded we were when equivocal or shallow findings appeared. Maybe. But inductivists, too, are operating with research questions, conceptual frameworks, and sampling matrices—though their choices are more implicit and the links between framework and procedures less linear. Nevertheless these choices will serve to bound and focus their study.

Take, for example, the problem of understanding police work—a domain that remains somewhat arcane and obscure despite (or perhaps because of) TV shows. It deserves an inductive approach. But which facets of police work will be studied? You can't look at them all. And where will you study them? You can't look everywhere. And when? If you "delay" that decision for a few months until you've spent some time, say, at the precinct station, that simply reflects two tacit sampling decisions (start at the precinct house, then recheck after a while).

Suppose the implicit research question was, How do arrests and bookings work? That choice immediately excludes many other issues and leads to *sampling and instrumentation* choices (e.g., using observation, rather than official documents; selection of different kinds of suspects, crimes, styles of apprehending suspects, types of officers). These sampling and instrumentation decisions, often inexplicit, are actually delimiting the settings, actors, processes, and events to be studied. In sum, the research questions, implicit or explicit, constrain the possible types of analyses.

Research questions and conceptual frameworks—either implicit/emerging or prespecified—affect each other. Let's illustrate.

Manning (1977), in his field study of police work, wanted to study arrests. He found that officers often made "discretionary" arrests, bending laws and regulations to fit their own, often private, purposes. Arrests were often justified after the fact, depending on the situation, and the definition of a "crime" turned out to be extremely fluid. Each arrest included implicit "negotiations" that brought informal and formal rules into play.

These findings, though, depend on Manning's conceptual framework. Readers of Schutz and Goffman will recognize immediately the language of social phenomenology: situational actions, negotiated rule meanings, and the like. A Marxist analysis of the same research questions in the same settings would probably show us how the police serve class-bound interests by stabilizing existing social arrangements. A mainstream social psychologist might focus on autonomy and social influence as the core issues.

In other words, people, including research people, have preferred "bins" and relational "arrows" as they construe and carve up social phenomena. They use these explicitly or implicitly to decide which questions are most important and how they should get the answers. We believe that better research happens when you make your framework—and associated choices of research questions, cases, sampling, and instrumentation—explicit, rather than claiming inductive "purity."

Brief Description

The formulation of research questions may precede or follow the development of a conceptual framework. The questions represent the facets of an empirical domain that the researcher most wants to explore. Research questions may be general or particular, descriptive or explanatory. They may be formulated at the outset or later on, and may be refined or reformulated in the course of fieldwork.

Questions probably fall into a finite number of types, many of which postulate some form of relationship. Box 2.3 shows the types of specific questions used in an evaluation of reading instruction for special education children using computer-assisted materials, and their more general form.

Illustration

Our school improvement study shows how a conceptual framework hooks up with the formulation of research questions. Look back at Figure 2.3, where the main variable sets of the study were laid out. Look at the third column, entitled "Adoption Decision." Its component parts are listed inside the bin: the decision to adopt, the plan for implementation, and support for implementation. The task, then, is to decide what you want to find out about these topics. The procedure we used was to cluster specific research questions under more general ones, as shown in Figure 2.4.

Notice the choices being made within each topical area. For example, in the first two areas, the main things we want to know about the "decision to adopt" are who was involved, how the decision was actually made, and how important this project was relative to others. All of the questions seem to be functional, rather than theoretical or descriptive—they have to do with getting something done.

We can also see that, conceptually, more is afoot than we saw in the conceptual framework. The "requisite conditions" in the last clump of questions indicate that the researchers have some a priori notions about which factors make for the greatest preparedness.

When such a research question gets operationalized, an attempt will be made to determine whether these conditions were present or absent at the various field sites, and whether that made any difference in the execution of the project. This is a good example of how research questions feed directly into data collection. Of course, field researchers will be attentive to other, as yet undreamed-of,

Box 2.3

Types of Research Questions: Example (N. L. Smith, 1987, p. 311)

Sample Questions	*General Form*
Causal-Research	
Do children read better as a result of this program?	Does X cause Y?
Do children read better in this program as compared with other programs?	Does X cause more of Y than Z causes of Y?
Noncausal-Research	
What is the daily experience of the children participating in this program?	What is X?
Are the remedial centers located in the areas of primary need?	Is X located where Y is lowest?
Noncausal-Policy	
What do we mean by "special education children" and "remediation"?	What does "Y" mean?
Is this program receiving support from state and local officials for political rather than educational reasons?	Why does S support X?
Noncausal-Evaluation	
What are the characteristics of the best CAI materials being used?	What makes W good?
How do the various minority groups view this program and judge its quality?	Does T value X?
Noncausal-Management	
What is the cost-effectiveness of the program compared with other programs?	Is X more cost-effective than Z?
How can we maximize the scheduling of classes at the centers with minimum expense?	How are U maximized and V minimized simultaneously?

Figure 2.4

General and Specific Research Questions Relating to the Adoption Decision (School Improvement Study)

How was the adoption decision made?
- Who was involved (e.g., principal, users, central office people, school board, outside agencies)?
- How was the decision made (top-down, persuasive, consultative, collegial-participative, or delegated styles)?

How much priority and centrality did the new program have at the time of the adoption decision?
- How much support and commitment was there from administrators?
- How important was it for teachers, seen in relation to their routine, "ordinary" activities, and any other innovations that were being contemplated or attempted?
- Realistically, how large did it loom in the scheme of things?
- Was it a one-time event or one of a series?

What were the components of the original plan for implementation?
- These might have included front-end training, monitoring and debugging/troubleshooting unexpected problems, and ongoing support.
- How precise and elaborate was this plan?
- Were people satisfied with it at the time?
- Did it deal with all of the problems anticipated?

Were the requisite conditions for implementation ensured before it began?
- These might have included commitment, understanding, materials and equipment, skills, time allocation, and organizational backup.
- Were any important conditions seen as missing? Which were most missing?

"requisite conditions," and the idea of "requisite conditions" may not be retained throughout the study.

Advice

1. Even if you are in a highly inductive mode, it is a good idea to start with some general research questions. They allow you to get clear about what, in the general domain, is of most interest. They make the implicit explicit without necessarily freezing or limiting your vision.

2. If you are foggy about your priorities or about the ways they can be framed, begin with a foggy research question and then try to defog it. Most research questions do not come out right on the first cut, no matter how experienced the researcher or how clear the domain of study.

3. Formulating more than a dozen or so general questions is looking for trouble. You can easily lose the forest for the trees and fragment the collection of data. Having a large number of research questions makes it harder to see emergent links across different parts of the database and to integrate findings.

As we saw in Figure 2.4, a solution to research question proliferation is the use of major questions, each with subquestions, for clarity and specificity. It also helps to consider whether there is a key question, the "thing you really want to know."

4. It is sometimes easier to generate a conceptual framework *after* you've made a list of research questions. You look at the list for common themes, common constructs, implicit or explicit relationships, and so on, and then begin to map out the underlying framework joining these pieces. Some researchers operate best in this mode.

5. In a multiple-case study, be sure all field-workers understand each question and see its importance. Multiple-case studies have to be more explicit, so several researchers can be aligned as they collect information in the field. Unclear questions or different understandings make for non-comparable data across cases.

6. Once the list of research questions is generated and honed, look it over to be sure each question is, in fact, researchable. You can always think of trenchant questions that you or your informants have no real means of answering, nor you of measuring.

7. Keep the research questions in hand and review them during fieldwork. This closeness will focus data collection; you will think twice before noting down what informants have for lunch or where they park their cars. Unless something has an obvious, direct, or potentially important link to a research question, it should not fatten your field notes. If a datum initially ignored does turn out to be important, you will know it. The beauty of qualitative field research is that there is (nearly) always a second chance.

Time Required

Formulating the questions is an iterative process; the second version is sharper and leaner than the first, and the third cut gets the final few bugs out. Most time should be spent on the general questions because the range and quality of specific ones will depend on how good the overarching question is.

Assuming that you feel in good touch with the topics of your study (either through experience or literature review), drafting and iterating a set of six or seven general research questions should take at least 2 or 3 hours. The questions should not be done in one sitting. A question that looks spellbinding usually loses some of its appeal when you look again a few hours later. Specific subquestions should come more quickly but might add another hour or two. The time will vary with the researcher's experience, the nature of the study, and the complexity and explicitness of the conceptual framework, but these estimates are reasonable, in our experience.

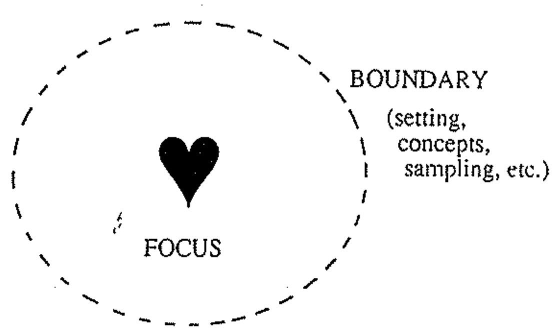
C. Defining the Case: Bounding the Territory

Rationale and Brief Description

Qualitative researchers often struggle with the questions of "what my case is" and "where my case leaves off." Abstractly, we can define a *case* as a phenomenon of some sort occurring in a bounded context. The case is, in effect, your unit of analysis. Studies may be of just one case or of several.

Figure 2.5 shows this graphically: There is a focus, or "heart," of the study, and a somewhat indeterminate boundary defines the edge of the case: what will not be studied.

Figure 2.5
The Case as the Unit of Analysis



Illustrations

What are some examples of cases? Sometimes the "phenomenon" may be an *individual* in a defined context:

A patient undergoing cardiovascular bypass surgery, before, during, and 6 months after surgery, in the context of his or her family and the hospital setting (Taylor, MacLean, Pallister, & White, 1988)

Note that the "heart" here is the patient. The boundary defines family and hospital as the context. The researchers will not, for example, interview the patient's colleagues at work or visit the restaurants where he or she dines. The bounding is also by time: No information will be gathered later than 6 months after hospitalization.

We can also expect that the boundary will be defined further by *sampling* operations, to which we'll come in a minute. For example, these researchers will not be interviewing the patient's children, only the spouse. But they will be sampling diet, exercise, and blood count data, as well as interview data from the patient's "lived experience."

Other examples of *individuals* as the case are:

A young person's work experience in his or her first "real" job (Borman, 1991): for example, Miriam's work as a bookkeeper in River City Bank

An uncommonly talented mechanic, in his shop amid the context of his friends, neighbors, and customers (Harper, 1987)

The researcher's grandfather, seen in his family and work context throughout his life course via his diaries (Abramson, 1992)

A case may also be defined by a *role:*

The role of "tramp" as it relates to police and rehabilitation staff (Spradley, 1979)

The role of a school principal in his specific school/ community setting, as studied by Wolcott (1973)

A teacher moving through "life cycles" during a teaching career (Huberman, 1993)

Or a *small group:*

An informal group of black men in a poor inner-city neighborhood (Liebow, 1967)

The architect, builder, and clients involved in the construction of a new house (Kidder, 1985)

Or an *organization:*

An inner-city high school engaged in a serious change effort to improve itself over a period of 4 years (Louis & Miles, 1990)

A Silicon Valley electronics firm competing in a fast-moving, turbulent market (Eisenhardt, 1989)

A nursery school and elementary school merging their staffs and facilities in the context of a national reform program in the Netherlands (van der Vegt & Knip, 1990)

Or a *community or "settlement":*

The Italian neighborhood in Boston's North End (Whyte, 1943)

Or a *nation:*

In Bolivia, seen over many years of history, the focus might be on causes of peasant revolts (Ragin, 1987)

These examples stress the nature and size of the *social unit,* but cases can be defined in other ways. As Werner and Schoepfle (1987a) usefully point out, a case can be located *spatially*—for example, the study of a nude beach reported in Douglas (1976).

In addition, they note, a case can be defined *temporally:* events or processes occurring over a specified period. See, for example, the case defined as an *episode or encounter:* Giorgi's (1975) study of what happened when a father gave a treasured chess set to his young son. A case also may be defined as an *event,* such as a school staff meeting; or as a *period of time,* as in the classic study *Midwest and Its Children* (Barker & Wright, 1971), which includes "One Boy's Day," a record of what Raymond did from the time he got out of bed until he reentered it; or as a sustained *process* (the adoption, implementation, and institutionalization of an innovative program by a school in its district, as in the study we already described [Huberman & Miles, 1984]).

So far we have discussed the "case" as if it were monolithic. In fact, as Yin (1984) points out, cases may have subcases "embedded" within them. A case study of a school may contain cases of specific classrooms; a case study of a hospital ward may have cases of specific doctor-patient relationships within it.

Single cases are the stuff of much qualitative research and can be very vivid and illuminating, especially if they are chosen to be "critical," extreme or unique, or "revelatory," as Yin (1984) suggests.

We argue in this book, with much recent practice to support us, that multiple cases offer the researcher an even deeper understanding of processes and outcomes of cases, the chance to test (not just develop) hypotheses, and a good picture of locally grounded causality. The question of just which cases to include in a sample is discussed below.

A comment on notation. We sometimes prefer—and use here and there in this book—the word *site* because it reminds us that a "case" always occurs in a specified social and physical *setting;* we cannot study individual cases devoid of their context in the way that a quantitative researcher often does.

Advice

1. Start intuitively. Think of the focus, or "heart," and build outward. Think of what you will *not* be studying as a way to firm up the boundary. Admit that the boundary is never quite as solid as a rationalist might hope.

2. Define the case as early as you can during a study. Given a starting conceptual framework and research questions, it pays to get a bit stern about what you are defining as a case; that will help clarify further both the framework and the questions.

3. Remember that sampling operations will define the case(s) further.

4. Attend to several dimensions of the case: its *conceptual* nature, its *social size,* its *physical* location, and its *temporal* extent.

Time Required

If a starting conceptual framework and research questions are reasonably clear, a first cut at case definition usually takes no more than a few minutes; discussion among members of a research team (or with interested colleagues) may occupy an hour or two as clarity emerges during successive iterations of the definition of the "case."

D. Sampling: Bounding the Collection of Data

Rationale

Sampling is crucial for later analysis. As much as you might want to, you cannot study everyone everywhere doing everything. Your choices—whom to look at or talk with, where, when, about what, and why—all place limits on the conclusions you can draw, and on how confident you and others feel about them.

Sampling may look easy. Much qualitative research examines a single "case," some phenomenon embedded in a single social setting. But settings have subsettings (schools have classrooms, classrooms have cliques, cliques have individuals), so deciding where to look is not easy. Within any case, social phenomena proliferate (science lessons, teacher questioning techniques, student unruliness, use of innovations); they, too, must be sampled. And the questions of multiple-case sampling add another layer of complexity. How to manage it all? We discuss some general principles and suggest useful references for detailed help.

Key features of qualitative sampling. Qualitative researchers usually work with *small* samples of people, nested in their context and studied in-depth—unlike quantitative researchers, who aim for larger numbers of context-stripped cases and seek statistical significance.

Qualitative samples tend to be *purposive,* rather than random (Kuzel, 1992; Morse, 1989). That tendency is partly because the initial definition of the universe is more limited (e.g., arrest-making in an urban precinct), and partly because social processes have a logic and a coherence that random sampling can reduce to uninterpretable sawdust. Furthermore, with small numbers of cases, random sampling can deal you a decidedly biased hand.

Samples in qualitative studies are usually not wholly prespecified, but can evolve once fieldwork begins. Initial choices of informants lead you to similar and different ones; observing one class of events invites comparison with another; and understanding one key relationship in the setting reveals facets to be studied in others. This is conceptually-driven sequential sampling.

Sampling in qualitative research involves two actions that sometimes pull in different directions. First, you need to set *boundaries:* to define aspects of your case(s) that you can study within the limits of your time and means, that connect directly to your research questions, and that probably will include examples of what you want to study. Second, at the same time, you need to create a *frame* to help you uncover, confirm, or qualify the basic processes or constructs that undergird your study.

Qualitative sampling is often decidedly *theory-driven,* either "up front" or progressively, as in a grounded theory mode. Suppose that you were studying how "role models" socialize children, and that you could only manage to look at four kindergarten classes. At first, that number seems very limited. But if you chose teachers according to relevant theory, you might pick them according to gender, sternness/nurturance, and socializing versus academic emphasis. And you would sample *within* each class for certain processes, such as deliberate and implicit modeling or application of sanctions. You might find also, as you went, that certain events, such as show-and-tell time or being read to by the teacher, were unusually rich with socialization actions, and then you would sample more carefully for these.

Sampling like this, both within and across cases, puts flesh on the bones of general constructs and their relationships. We can see generic processes; our generalizations are not to "all kindergartens," but to existing or new theories of how role modeling works. As Firestone (1993) sug-

Figure 2.6
Typology of Sampling Strategies in Qualitative Inquiry
(Kuzel, 1992; Patton, 1990)

Type of Sampling	*Purpose*
Maximum variation	Documents diverse variations and identifies important common patterns
Homogeneous	Focuses, reduces, simplifies, facilitates group interviewing
Critical case	Permits logical generalization and maximum application of information to other cases
Theory based	Finding examples of a theoretical construct and thereby elaborate and examine it
Confirming and disconfirming cases	Elaborating initial analysis, seeking exceptions, looking for variation
Snowball or chain	Identifies cases of interest from people who know people who know what cases are information-rich
Extreme or deviant case	Learning from highly unusual manifestations of the phenomenon of interest
Typical case	Highlights what is normal or average
Intensity	Information-rich cases that manifest the phenomenon intensely, but not extremely
Politically important cases	Attracts desired attention or avoids attracting undesired attention
Random purposeful	Adds credibility to sample when potential purposeful sample is too large
Stratified purposeful	Illustrates subgroups; facilitates comparisons
Criterion	All cases that meet some criterion; useful for quality assurance
Opportunistic	Following new leads; taking advantage of the unexpected
Combination or mixed	Triangulation, flexibility, meets multiple interests and needs
Convenience	Saves time, money, and effort, but at the expense of information and credibility

gests, the most useful generalizations from qualitative studies are *analytic,* not "sample-to-population."

General sampling strategies. Erickson (1986) suggests a generic, funneling sampling sequence, working from the outside in to the core of a setting. For example, in studying schools, he would begin with the school community (census data, walk around the neighborhood) and then enter the school and the classroom, staying several days to get a sense of the frequency and occurrence of different events. From there, the focus would tighten: specific events, times, and locations. Periodically, however, Erickson would "follow lines of influence . . . into the surrounding environment" to test the typicality of what was found in a given classroom, and to get a better fix on external influences and determinants.

In Figure 2.6 we list a range of sampling strategies, most usable either within a complex case or across cases. They can be designed ahead of time, or be evolved during early data collection. (Note: The word *case* in Figure 2.6 may refer either to "cases" taken as a whole or to "informants" within a single case setting.) How do such strategies affect analysis?

Maximum variation, for example, involves looking for outlier cases to see whether main patterns still hold. The *critical case* is the instance that "proves" or exemplifies the main findings. Searching deliberately for *confirming and disconfirming cases, extreme or deviant cases,* and *typical cases* serves to increase confidence in conclusions. Some of the strategies benefit inductive, theory-building analysis (e.g., *opportunistic, snowball or chain,* and *intensity*).

The remaining entries are fairly self-evident, except for *politically important cases.* These are "salient" informants who may need to be included (or excluded) because they connect with politically sensitive issues anticipated in the analysis.

Other strategies can be used for selection of informants prior to data collection. For example, Goetz and Lecompte (1984, cited in Merriam, 1988) offer some possibilities not mentioned by Patton and Kuzel (sources for Figure 2.6): *comprehensive sampling*—examining every case, instance, or element in a given population; *quota selection*—identifying the major subgroups and then taking an arbitrary number from each; *reputational case selection*—instances chosen on the recommendation of an "expert" or "key informant"; and *comparable case selection*—selecting individuals, sites, and groups on the same relevant characteristics over time (a replication strategy). Most of these strategies will increase confidence in analytic findings on the grounds of representativeness.

See also Guba and Lincoln (1989), who advocate *maximum variation* sampling, a deliberate hunt for negative instances or variations. This process may take the form of questions to informants, such as, "Whom do you know who sees things differently?" or "Where can I find patients who *don't* keep appointments?"

Johnson (1990), in his comprehensive treatment of selecting informants, also suggests *dimensional* sampling: The researcher lays out the dimensions on which variability is sought, then takes representative, "well-informed" informants for each contrasting dimension. The aim is to find people who are more knowledgeable, reliable, and accurate in reporting events that are usual, frequent, or patterned. (This strategy has risks: Such informants may assume greater uniformity than actually exists (Pelto & Pelto, 1975; Poggie, 1972).

The sampling strategies we've been discussing can be applied both within or across cases. Let's turn to some of the core issues in each of these domains.

Within-case sampling. Quantitative researchers usually think of cases as individual persons, draw a "sample" of persons, and then collect comparable "data points" from each. By contrast, a qualitative "case" may range widely in definition from individuals to roles, groups, organizations, programs, and cultures. But even when the case is an individual, the qualitative researcher has many within-case sampling decisions: Which *activities, processes, events, times, locations,* and *role partners* will I sample?

In our cardiovascular bypass patient example, we might want to sample diet and exercise activities; the processes of understanding, taking in, and acting on medical advice; events such as admission and discharge interviews; time periods including prehospitalization, hospitalization, and posthospitalization (once every 2 weeks); locations including recovery room, ward, and the patient's home; and role partners including the patient's physician, ward nurses, dietitian, and spouse.

Within-case sampling is almost always *nested*—for example, studying children within classrooms within schools within neighborhoods, with regular movement up and down that ladder.

A second major point is that such sampling must be *theoretically* driven—whether the theory is prespecified or emerges as you go, as in Glaser and Strauss's (1967) "theoretical sampling." Choices of informants, episodes, and interactions are being driven by a conceptual question, not by a concern for "representativeness." To get to the construct, we need to see different instances of it, at different moments, in different places, with different people. The prime concern is with the *conditions* under which the construct or theory operates, not with the generalization of the findings to other settings.

The third point is that within-case sampling has an *iterative* or "rolling" quality, working in progressive "waves" as the study progresses. Sampling is investigative; we are cerebral detectives, ferreting out answers to our research questions. We observe, talk to people, and pick up artifacts and documents. That leads us to new samples of informants and observations, new documents. At each step along the evidential trail, we are making sampling decisions to clarify the main patterns, see contrasts, identify exceptions or discrepant instances, and uncover negative instances—where the pattern does not hold. Our analytic conclusions depend deeply on the within-case sampling choices we made.

So within-case sampling helps us see a local configuration in some depth. What can adding cases do for us, and how do we create a sample of cases?

Multiple-case sampling. Multiple-case sampling adds *confidence* to findings. By looking at a range of similar and contrasting cases, we can understand a single-case finding, grounding it by specifying *how* and *where* and, if possible, *why* it carries on as it does. We can strengthen the precision, the validity, and the stability of the findings. We are following a *replication* strategy (Yin, 1991). If a finding holds in one setting and, given its profile, also holds in a comparable setting but does not in a contrasting case, the finding is more robust. The "multiple comparison groups" used in grounded theory work play a similar role.

With multiple-case studies, does the issue of *generalizability* change? Essentially, no. We are generalizing from one case to the next on the basis of a match to the underlying theory, not to a larger universe. The choice of cases usually is made on *conceptual* grounds, not on representative grounds. The cases often are arrayed on a continuum (e.g., highly gifted to underachieving pupils), with few exemplars of each, or they are contrasted (e.g., assertive and passive adolescents). Other, unique properties may be added (e.g., some assertive adolescents are from cities, some from rural areas). If you look closely at the cells of such a sampling frame, each is essentially unique. Because case study researchers examine intact settings in such loving detail, they know all too well that each setting has a few properties it shares with *many* others, some properties it shares with *some* others, and some properties it shares with *no* others. Nevertheless, the multiple-case sampling gives us confidence that our emerging theory is generic, because we have seen it work out—and not work out—in predictable ways.

Multiple-case sampling, although it may have iterative aspects, normally has to be thought through carefully. An explicit *sampling frame* is needed. It will be guided by the research questions and conceptual framework—either prespecified or emergent. Once again, random sampling will not help.

How many cases should a multiple-case study have? This question is not answerable on statistical grounds, of course. We have to deal with the issue conceptually: How many cases, in what kind of sampling frame, would give us confidence in our analytic generalizations? It also depends on how rich and complex the within-case sampling is. With high complexity, a study with more than 15 cases or so can become unwieldy. There are too many data to scan visually and too many permutations to account for. And the problems of practical and intellectual coordination among multiple researchers get very large, once you are over a staff of five or six people. Still, we've seen multiple-case studies in the 20s and 30s. The price is usually thinner data. And at some point you say: Why not do a survey?

Brief Description

Sampling involves decisions not only about which people to observe or interview, but also about settings, events, and social processes. Multiple-case studies also demand clear choices about which types of cases to include. Qualitative studies call for continuous refocusing and redrawing of study parameters during fieldwork, but some initial selection still is required. A conceptual framework and research questions can help set the foci and boundaries for sampling decisions.

Illustrations

Let's look briefly at two ends of the qualitative continuum: the unfettered, exploratory single-case study and the more constrained and focused multiple-case study. Suppose we wanted to deal with "police work," as in Manning's (1977) research. This example has no orienting conceptual frame, but rather a general perspective on social processes, notably how people literally "make sense" of their habitual surroundings. Part of this sense-making activity is developing and interpreting rules about legitimate and illegitimate behavior.

The decision to explore this perspective by studying the arrest and booking of suspects in a single precinct is a good example of a sampling choice. You could ask, How are laws interpreted by people enforcing them in face-to-face situations? and then select police officers as a sample of such "people" (rather than judges or fire inspectors). Or you can move right away from the general domain to sampling *events* and *processes* and ask, How do police officers interpret laws when arresting and booking suspects?

However you proceed, the sampling parameters are partially set by the framework and the research question: police work, rule interpreting, arrests, and booking. There is still room for choices within each dimension of the study, but the universe is now far more bounded and focused. To get a sense of a minimal set of initial sampling choices within this universe, let's array some options:

Sampling Parameters	*Possible Choices*
settings:	precinct station, squad car, scene of the crime, suspect's residence or hangout
actors:	police officers with different characteristics (e.g., rank, seniority, experience, race, beliefs, education) and suspects (age, race, beliefs, education, type of offense)
events:	arrests, bookings, possibly pursuits of suspects, and post hoc justifications of booking to other actors
processes:	making the arrest, doing the booking, relating to suspects, interpreting laws, justifying laws, generally negotiating law enforcement within the precinct

The researcher may have to touch most or all of these bases to get the research question well answered. The first base usually is the setting—say, the precinct station. From there, several options emerge:

1. Start with the precinct station, one kind of police officer, all bookings during the working day, and all instances of the social interactions around "legal" and "illegal" behavior that occur.
2. Start with the precinct station and all types of officers, bookings, and justifications for the booking.
3. Start with one officer and follow the officer through several episodes of arrests, pursuits, bookings, and justifications for them.
4. Start with a booking at the precinct station and then reconstitute the prior events.

Many permutations are possible, but a selection process is invariably at work. An ethnographer setting out to "hang around" a precinct is continuously making sampling decisions about what to observe, whom to talk with, what to ask, what to write down, whether to stay in one room or another. And these choices, in turn, are determined by the questions being asked, and the perspective—implicit or explicit—that determines why these questions, and not others, are being asked.

Questions of practicality also face us. There is a finite amount of time, with variable access to different actors and events, and an abundance of logistical problems. Very seldom does a start-up sampling frame survive the lovely imperfection and intractability of the field. It must be shifted and reframed.

Being selective calls for some restraint in the classes of data you go after. Here we might suggest some guidelines. For example, useful data would (a) identify *new leads* of importance, (b) *extend* the area of information, (c) *relate* or bridge already existing elements, (d) reinforce main *trends,* (e) account for *other information* already in hand, (f) exemplify or provide more evidence for an *important theme,* and (g) *qualify or refute* existing information.

Finally, sampling means just that: taking a smaller chunk of a larger universe. If I begin with a well-developed conceptualization, I can focus on one precinct station and one kind of police officer making one kind of booking; if my conceptualization stands up, I can make statements about bookings that may apply to other officers and other precincts. But to test and ramify those claims and to establish their *analytic generality,* I have to move on to several other precinct stations with similar and contrasting characteristics. Here again, the main goal is to strengthen the conceptual validity of the study, but the procedure also helps determine the conditions under which the findings hold.

To illustrate a multiple-case study, more predesigned, let's look at the sample of cases for the school improvement study (Table 2.1). There are 12 cases and 8 sampling dimensions, which means that each of the 12 cases is a unique configuration, but one sharing some dimensions with one or more sites—for example, its current status (expanding, ongoing, inactive), its longevity, its demographics, and its program type.

Table 2.2 brings out the dimensions of comparison and contrast more sharply. The case uniqueness means that our findings will be modestly representative of school improvement programs studied across the country. At the same time, the dimensions of comparison and contrast noted above will provide a test of the ideas in the conceptual framework. That test would let us make some general statements about the core processes and determinants at work.

We cannot, however, necessarily support generalization of the findings to other settings at a less abstract level. More specific findings (e.g., major transformations of innovations that occur in sites of certain types) may not hold where local circumstances are different—or even in some sites that share some of our sample's characteristics.

For example, the question of whether the outcomes found for the three rural NDN (externally developed) projects in this sample obtain for the full population of rural NDN projects is not an answer that this sample can provide. Our findings can, however, help define the parameters for a follow-up survey or a series of new case studies focused on this subpopulation. Many of the core processes, incidents, interactions, and outcomes will be found elsewhere; when they are, the general findings will be buttressed by a raft of particulars that give more shape and body to the emerging conceptualization.[3]

Incidentally, this 12-case sample was nested in a larger sample of cases from which survey data were collected; field study findings were expected to illustrate and possibly confirm trends in that larger population. Some sampling dimensions reflect this nesting (e.g., focusing only on NDN [externally developed] and IV-C [locally developed] programs, each an important part of the larger study; or the break by region and setting). Other dimensions (year of start-up, project status) follow more from the conceptual framework, which hypothesized specific kinds of changes over time; still other dimensions (program type, name, content) are generated from what was known about properties of the innovation being tried.

Note the focusing and bounding decisions that have been made. A far larger survey study is concentrating on all program types in all geographical areas. The field study we are discussing here will be looking only at the school improvement process, two types of programs, and, within these, at 12 projects. Each aspect acts like a decision tree that takes us into increasingly particularistic domains of study. So what we may eventually have to say about the "school improvement" process will be both highly general (through a conceptual test and revision of the framework) and highly particular (about how the overarching findings play out in specific programs in specific settings). To the extent that analyses converge across the multiple cases and with the survey, we can make some strong claims for the viability of our findings.

Each case is an ideal type. But it is far easier to move from a limited but well-delineated small sample to a larger one and to make some educated guesses about what we are likely to find, than to go directly from a single case to a larger set.

The prime interest of a multiple-case study is conceptual. We have some notions of how complex programs and practices are implemented in school settings, and we want to test them to see how they vary under different conditions. Sampling for variability on the year the project began provides a test of the process of change. Stratifying levels of growth (e.g., "dwindling," "expanding") for different projects in different settings provides a conceptual test of "success" in different local configurations. Noting covariations within the 12-case sample allows us to replicate some findings across cases and, through the contrasts observed, to distinguish between cases on dimensions that

Table 2.1
Characteristics of Field Study Sample

		SITE CONTEXT		ASPECTS OF THE INNOVATION				
SITE	PROGRAM SPONSORSHIP#	U.S. REGION	SETTING	YEAR PROJECT BEGAN	STATUS (as initially assessed)	PROGRAM TYPE	PROGRAM NAME OR INITIALS[x]	PROGRAM CONTENT
ASTORIA	(E)	Southeast	Small city	1978	Expanding	Add-on	EPSF	Early Childhood
BANESTOWN	(E)	Southeast	Rural	1979	Expanding	Pull-out	SCORE-ON	Reading/Math
BURTON	(E)	Midwest	Suburban	1979	Expanding	Add-on	IPLE	Law and Government
CALSTON	(E)	Midwest	Center city	1978	Ongoing	Drop-in	Matteson 4D	Reading
CARSON	(L)	Plains	Rural	1977	Expanding	Add-on	IPA	Individualized educational planning*
DUN HOLLOW	(L)	Northeast	Urban sprawl	1977	Dwindling	Add-on	Eskimo Studies	Social studies*
LIDO	(E)	Northeast	Rural	1976	Dwindling	Add-on	KARE	Environment
MASEPA	(E)	Plains	Rural	1978	Ongoing	Drop-in	ECRI	Language Arts*
PERRY-PARKDALE	(E)	Midwest	Suburban	1977	Ongoing	Sub-system	EBCE	Career education
PLUMMET	(L)	Southwest	Center city	1976	Ongoing	Sub-system	Bentley Center	Alternative school
PROVILLE	(L)	Southwest	Urban sprawl	1977	Dwindling	Pull-out	CEP	Vocational education
TINDALE	(L)	Midwest	Urban sprawl	1976	Ongoing	Drop-in	Tindale Reading Model	Reading

\# (E) = externally developed innovation
(L) = locally-developed innovation

x (I) program names are pseudonyms, to avoid identifying specific sites.

* Program is used in this site with a comprehensive sample of learners, rather than with low-achieving or marginal populations.

Table 2.2
Final Sampling Frame for Field Study

	National Diffusion Network Projects (NDN)			Title IV-C Projects (IV-C)		
	1979	1978	Earlier	1978	1977	Earlier
EXPANDING	SCORE-ON Banestown Rural Pull-out, in-school	EPSF Astoria Town/suburb Add-on, in-school			IPA Carson Rural Add-on, in-school	
ONGOING	IPLE Burton Suburb Drop-in, in-school/field	ECRI Masepa Rural Drop-in, in-school Matteson 4-D Calston Metro urban Drop-in, in-school	EBCE Perry-Parkdale Suburb Subsystem, in-school/field			Tindale Reading Tindale Urban sprawl Subsystem, in-school Bentley Center Plummet Urban Subsystem, in-school
INACTIVE, DWINDLING			KARE Lido Rural Add-on, field	Eskimo Curr. Dun Hollow Urban sprawl Add-on, in-school	CEP Proville Urban sprawl Pull-out, in-school	

are conceptually meaningful. This effect is much more powerful than a series of individual case studies over several years.

So we have done pretty well with a 12-case sample: careful illustration, strong construct validity, legitimate claims to external validity with respect to the core findings, and spadework for replication studies with well-identified subsets of settings and actors—all without control groups, random sampling, systematic treatments, or other procedures required in experimental and correlational research.

Note that this sampling matrix does not resolve some of the important *within*-case focusing and bounding decisions to be made in multiple-case studies. For example, if one case researcher observes only administrators and another only teachers, the comparability of the two cases is minimal. For these decisions, we usually need the conceptual framework or the research questions. Using them, we can agree on sampling parameters and comparable choices for initial fieldwork—settings, actors, events, processes.

Cross-case comparison is impossible if researchers operate in radically different settings, use no coherent sampling frame, or, worst of all, if they focus on different processes. These processes in cases flow out of the interplay of actors, events, and settings. They are usually at the core of the conceptual framework, and serve as the glue holding the research questions together. Key processes can be identified at the outset or gradually—often via pattern codes, reflective remarks, memos, and interim summaries, as we'll see in Chapter 4. Being explicit about processes and collecting comparable data on them will not only avoid costly distractions but also foster comparability and give you easier access to the core underlying constructs as you get deeper into data collection.

Advice

1. If you're new to qualitative research, rest assured that there is never "enough" time to do any study. So taking the tack, "I'll start somewhere and take it from there," is asking for trouble. It is probably a good idea to start with a fallback sample of informants and subsettings: the things you have to cover in light of what you know at that point. That sample will change later, but less than you may think.

2. Just thinking in sampling-frame terms is good for your study's health. If you are talking with one kind of informant, you need to consider why this kind of informant is important and, from there, who else should be interviewed or observed. This is also a good exercise for controlling bias.

3. In complex cases, remember that you are sampling people to get at characteristics of settings, events, and processes. Conceptually, the people themselves are secondary. This means watching out for an overreliance on talk, or on observation of informants, while you may neglect sampling for key events, interactions in different settings, and

episodes embodying the emerging patterns in the study. Remember also to line up these sampling parameters with the research questions as you go: Are my choices doing a representative, time-efficient job of answering them? Finally, the sampling choices at the start of the study may not be the most pertinent or data-rich ones. A systematic review can sharpen early and late choices.

4. In qualitative research, as well as in survey research, there is a danger of sampling too narrowly. In fact, points 2 and 3 above tug in that direction: Go to the meatiest, most study-relevant sources. But it is also important to work a bit at the peripheries—to talk with people who are not central to the phenomenon but are neighbors to it, to people no longer actively involved, to dissidents and renegades and eccentrics. Spending a day in the adjoining village, school, neighborhood, or clinics is also worth the time, even if you don't see the sense at that point.

There are rewards for peripheral sampling. First, you may learn a lot. Second, you will obtain contrasting and comparative information that may help you understand the phenomenon at hand by "de-centering" you from a particular way of viewing your other cases. As we all know, traveling abroad gives us insights into our own culture.

5. Spend some time on whether your sampling frame is feasible. Be sure the time is there, the resources are there, the requisite access to people and places is ensured, and the conditions are right for doing a careful job. Plan to study a bit less, rather than more, and "bank" the extra time. If you are done, the time is yours for a wider or deeper pass at the field. If not, you will need this time to complete your more modest inquiry under good conditions.

6. Three kinds of instances have great payoff. The first is the apparently "typical" or "representative" instance. If you can find it, try to find another one. The second is the "negative" or "disconfirming" instance; it gives you both the limits of your conclusions and the point of greatest variation. The third is the "exceptional" or "discrepant" instance. This instance will allow you to qualify your findings and to specify the variations or contingencies in the main patterns observed. Going deliberately after negative and atypical instances is also healthy in itself; it may force you to clarify your concepts, and it may tell you that you indeed have sampled too narrowly. (More on this in Chapter 10, section B.)

7. Apply some criteria to your first and later sampling plans. Here's a checklist, summarizing what we've said, and naming some underlying issues:

- Is the sampling *relevant* to your conceptual frame and research questions?
- Will the phenomena you are interested in *appear?* In principle, *can* they appear?
- Does your plan enhance *generalizability* of your findings, either through conceptual power or representativeness?
- Can *believable* descriptions and explanations be produced, ones that are true to real life?
- Is the sampling plan *feasible,* in terms of time, money, access to people, and your own work style?
- Is the sampling plan *ethical,* in terms of such issues as informed consent, potential benefits and risks, and the relationship with informants (see also Chapter 11)?

Time Required

It's difficult to provide workable guidelines here. Within-case sampling decisions tend to get made gradually, over time. Even so, a first cut at within-case choices should not take more than 2 or 3 hours in studies with a clear conceptual framework and research questions. At least one iteration before the first site visit is a good idea. For more exploratory or "grounded" studies, a few hours might suffice to decide where to start and, once at that starting place, which tools to deploy to collect information.

Sampling cases in a multiple-case study is nearly always a demanding experience. There are usually so many contending dimensions and so many alternative combinations of those dimensions that it is easy for researchers to lose intellectual control, get overwhelmed with the multiple possibilities, chew up a great deal of time, and finally say, "There's no rational way to do this." Setting up the possibilities in matrix form definitely helps.

In our experience you can usually expect to spend 3 or 4 hours for a first cut when a dozen or so cases are involved; two or three additional sessions, involving colleagues, are typically required. The work isn't simple, and cannot be hurried. After all, you are making some long-term commitments to the places where your basic data will be collected.

E. Instrumentation

Rationale

We've been emphasizing that conceptual frameworks, research questions, and sampling plans have a focusing and bounding role within a study. They give some direction to the researcher, before and during fieldwork, by clarifying *what* he or she wants to find out from *whom* and *why.*

Knowing what you want to find out, at least initially, leads inexorably to the question of how you will get that information. That question, in turn, constrains the analyses you can do. If I want to find out how suspects are arrested and booked, I may decide to *interview* people associated with this activity (police officers, suspects, attorneys), *observe* bookings, and collect arrest-relevant *documents* (e.g., regulations, transcripts). I also may take pictures of bookings or record them on tape (cf. Van Maanen, 1979). But how much of this instrumentation has to be designed prior to going out to the field? And how much structure should such instrumentation have?

Note that the term *instrumentation* may mean little more than some shorthand devices for observing and recording events—devices that come from initial conceptualizations loose enough to be reconfigured readily as the data suggest revisions. But note, too, that even when the instrumentation is an open-ended interview or observation, some technical choices must be made: Will notes be taken? Of what sort? Will the transaction be tape-recorded? Listened to afterward? Transcribed? How will notes be written up?

Furthermore, as Kvale (1988) points out, during an "open-ended" interview much interpretation occurs along the way. The person describing his or her "life world" discovers new relationships and patterns during the interview; the researcher who occasionally "summarizes" or "reflects" what has been heard is, in fact, condensing and interpreting the flow of meaning. As Kvale suggests, the "data" are not being "collected," but rather "co-authored." A warning here: The same things are happening even when the interview question is much more structured and focused. So let's not delude ourselves about total "control" and "precision" in our instrumentation—while remembering that attention to design can make a real difference in data quality and the analyses you can carry out.[4]

How much preplanning and structuring of instrumentation is desirable? There are several possible answers: "little" (hardly any prior instrumentation) to "a lot" (of prior instrumentation, well structured) to "it depends" (on the nature of the study). Each view has supporting arguments; let's review them in capsule form.

Arguments for little prior instrumentation.

1. Predesigned and structured instruments blind the researcher to the site. If the most important phenomena or underlying constructs at work in the field are not in the instruments, they will be overlooked or misrepresented.

2. Prior instrumentation is usually context-stripped; it lusts for universality, uniformity, and comparability. But qualitative research lives and breathes through seeing the context; it is the particularities that produce the generalities, not the reverse.

3. Many qualitative studies involve single cases, with few people involved. Who needs questionnaires, observation schedules, or tests—whose usual function is to yield economical, comparable, and parametric distributions for large samples?

4. The lion's share of fieldwork consists of taking notes, recording events (conversations, meetings), and picking up things (documents, products, artifacts). *Instrumentation* is a misnomer. Some orienting questions, some headings for observations, and a rough and ready document analysis form are all you need at the start—perhaps all you will ever need in the course of the study.

Arguments for a lot of prior instrumentation.

1. If you know what you are after, there is no reason not to plan in advance how to collect the information.

2. If interview schedules or observation schedules are not focused, too much superfluous information will be collected. An overload of data will compromise the efficiency and power of the analysis.

3. Using the same instruments as in prior studies is the only way we can converse across studies. Otherwise the work will be noncomparable, except in a very global way. We need common instruments to build theory, to improve explanations or predictions, and to make recommendations about practice.

4. A biased or uninformed researcher will ask partial questions, take selective notes, make unreliable observations, and skew information. The data will be invalid and unreliable. Using validated instruments well is the best guarantee of dependable and meaningful findings.

Arguments for "it depends."

1. If you are running an *exploratory*, largely descriptive study, you do not really know the parameters or dynamics of a social setting. So heavy initial instrumentation or closed-ended devices are inappropriate. If, however, you are doing a *confirmatory* study, with relatively focused research questions and a well-bounded sample of persons, events, and processes, then well-structured instrument designs are the logical choice. Within a given study, there can be both exploratory and confirmatory aspects that call for differential front-end structure, or there can be exploratory and confirmatory *times*, with exploration often called for at the outset and confirmation near the end.[5]

2. A *single-case* study calls for less front-end preparation than does a *multiple-case* study. The latter is looking forward to cross-case comparison, which requires some standardization of instruments so that findings can be laid side by side in the course of analysis. Similarly, a *freestanding* study has fewer constraints than a *multimethod*

Figure 2.7
Prior Instrumentation: Key Decision Factors

Little Prior Instrumentation	*"It Depends"*	*A Lot of Prior Instrumentation*
Rich context description needed		Context less crucial
Concepts inductively grounded in local meanings		Concepts defined ahead by researcher
Exploratory, inductive		Confirmatory, theory-driven
Descriptive intent		Explanatory intent
"Basic" research emphasis		Applied, evaluation or policy emphasis
Single case		Multiple cases
Comparability not too important		Comparability important
Simple, manageable, single-level case		Complex, multilevel, overloading case
Generalizing not a concern		Generalizability/representativeness important
Need to avoid researcher impact		Researcher impact of less concern
Qualitative only, free-standing study		Multimethod study, quantitative included

study (e.g., a field study tied to a survey, an idea we discuss further in Chapter 3). A *"basic"* study often needs less advance organizing than an *applied, evaluation,* or *policy* study. In the latter cases, the focus is tighter and the instrumentation more closely keyed to the variables of interest.

3. Much depends on the *case definition* and levels of analysis expected. A researcher studying classroom climate in an elementary school might choose to look intensively in 3 of the building's 35 classrooms, and so probably would be right to start with a looser, orienting set of instruments. If, however, an attempt is made to say something about how classroom climate issues are embedded in the working culture of the building as a whole, a more standardized, validated instrument—a questionnaire or a group interview schedule—may also be required.

In Figure 2.7 is a summary of some main issues in deciding on the appropriate amount of front-end instrumentation.

We think there is wisdom in all three stances toward front-end instrumentation and its degree of structure. The first stance (little prior instrumentation) puts the emphasis on certain types of validity: *construct* (Are the concepts well grounded?), *descriptive/contextual* (Is the account complete and thorough?), *interpretive* (Does the account connect with the "lived experience" of people in the case?), and *natural* (Is the setting mostly undisturbed by my presence?).[6]

The second stance (a lot of preinstrumentation) emphasizes *internal validity* (Am I getting a comparably measured response from different people?) and *generalizability* (Is this case a good instance of many others?), along with sheer *manageability* of data collection.

The third stance is both contingent and ecumenical, considering it unhelpful to reach for absolute answers in relative instances. Figure out first what kind of study you are doing and what kind of instruments you are likely to need at different moments within that study, and then go to work on the ones needed at the outset. But in all cases, as we have argued, the amount and type of instrumentation should be a function of your conceptual focus, research questions, and sampling criteria. If not, the tail is likely to be wagging the dog, and later analysis will suffer.

Brief Description

Instrumentation comprises specific methods for collecting data: They may be focused on qualitative or quantitatively organized information, and may be loosely to tightly structured.

Illustration

How can front-end instrument design be driven in different ways by a study's scope and focus? We give an ecumenical example, showing a mix of predesigned and open-ended instrumentation that follows the implications of a conceptual framework and research questions without locking in too tightly. Back to the school improvement study.

It is worth looking again at the *conceptual framework* (Figure 2.3) to recall the main variables and their flow over time. Remember also that (a) this is a multiple-case (N = 12) study and that (b) the phenomenon under study is moderately well, but not fully, understood from prior empirical research. Both of these points suggest that some front-end instruments are likely to be called for.

One important *research question* in the study was:

> In which ways did people redefine, reorganize, or reinvent the new program in order to use it successfully?

Looking back at Figure 2.3, we can see that the question derives from the fourth bin, "Cycle of Transformations," and that within that bin, from the first variable cluster, "changes in innovation as presented to user." Previous em-

Figure 2.8
Excerpts From Interview Guide, School Improvement Study

33. Probably you have a certain idea of how ________ looks to you now, but keep thinking back to how it first looked to you then, just before students came. How did it seem to you then?
    Probes:
    —Clearly connected, differentiated vs. unconnected, confusing
    —Clear how to start vs. awesome, difficult
    —Complex (many parts) vs. simple and straightforward
    —Prescriptive and rigid vs. flexible and manipulatable

34. What parts of aspects seemed ready to use, things you thought would work out OK?

35. What parts or aspects seemed not worked out, not ready for use?

36. Could you describe what you actually did during that week or so before you started using ________ with students?
    Probes:
    —Reading
    —Preparing materials
    —Planning
    —Talking (with whom, about what)
    —Training

....

40. Did you make any changes in the standard format for the program before you started using it with students? What kind of changes with things you thought might not work, things you didn't like, things you couldn't do in this school?
    Probes:
    —Things dropped
    —Things added, created
    —Things revised

pirical research and cognitive and social-psychological theory both led us to the idea that people will adapt or reinvent practices while using them.

The *sampling* decisions are straightforward. The question addresses teachers in particular, and to get the answer, we will have to observe or interview them or, ideally, do both. We should sample events such as the teacher's first encounter with the innovation, and processes such as assessing its strong and weak points and making changes in it to fit one's practice.

Let's look at the interview component of the *instrumentation.* We developed a semistructured interview guide. Each field researcher was closely familiar with the guide, but had latitude to use a personally congenial way of asking and sequencing the questions, and to segment them appropriately for different respondents.

The guide was designed after fieldwork had begun. An initial wave of site visits had been conducted to get a sense of the context, the actors, and how the school improvement process seemed to be working locally. From that knowledge, we went for deeper and broader understanding.

Now to dip into the guide near the point where the research question will be explored (Figure 2.8).

The interviewer begins by taking the informant back to the time just before he or she was to use the innovation with students, asking for detailed context—what was happening, who colleagues were, and what feelings they had.

Questions 33 through 36 move forward through time, asking how the innovation looked, its ready or unready parts, and what the teacher was doing to prepare for its use. Question 40 comes directly to the research question, assessing pre-use changes made in the innovation. The probes can be handled in various ways: as aids to help the interviewer flesh out the question, as prompts for items the informant may have overlooked, or as subquestions derived from previous research.

Later in the interview, the same question recurs as the interviewer evokes the teacher's retrospective views of early and later use, then moves into the present ("What changes are you making now?") and the future ("What changes are you considering?").

So all field researchers are addressing the same general question and are addressing it in similar ways (chronologically, as a process of progressive revision), although the question wording and sequencing will vary from one researcher to the next. If the response opens other doors, the interviewer will probably go through them, coming back later to the "transformation" question. If the response is uncertain or looks equivocal when the researcher reviews the field notes, the question will have to be asked again—perhaps differently—during the next site visit.

Advice

1. We have concentrated here on general principles of designing appropriate instrumentation, not on detailed technical help. For the latter we recommend treatments such as Spradley (1979), Weller and Romney (1988), and Mishler (1986) on interviewing. Judd, Smith, and Kidder (1991), R. B. Smith and Manning (1982), Werner and Schoepfle (1987a, 1987b), Goetz and LeCompte (1984), and Brandt (1981) are helpful on other methods as well, including questionnaires, observation, and document analysis. Good overviews are provided by Marshall and Rossman (1989), Fetterman (1989), Bogdan and Taylor (1975), and Denzin (1978).

2. Simply *thinking* in instrument design terms from the outset strengthens data collection as you go. If you regularly ask, Given that research question, how can I get an answer? it will sharpen *sampling* decisions (I have to ob-

serve/interview this class of people, these events, those processes), help clarify *concepts*, and help set *priorities* for actual data collection. You also will learn the skills of redesigning instrumentation as new questions, new subsamples, and new lines of inquiry develop.

Not thinking in instrument design terms can, in fact, lead to self-delusion: You feel "sensitive" to the site but actually may be stuck in reactive, seat-of-the-pants interviewing. That tactic usually yields flaccid data.

3. People and settings in field studies can be observed more than once. Not everything is riding on the single interview or observation. In qualitative research there is nearly always a second chance. So front-end instrumentation can be revised—in fact, should be revised. You learn how to ask a question in the site's terms and to look with new eyes at something that began to emerge during the first visit. Instrumentation can be modified steadily to explore new leads, address a revised research question, or interview a new class of informant.

4. In qualitative research, issues of instrument validity and reliability ride largely on the skills of the researcher. Essentially a *person*—more or less fallibly—is observing, interviewing, and recording, while modifying the observation, interviewing, and recording devices from one field trip to the next. Thus you need to ask, about yourself and your colleagues, How valid and reliable is this person likely to be as an information-gathering instrument?

To us, some markers of a good qualitative researcher-as-instrument are:

- some familiarity with the phenomenon and the setting under study
- strong conceptual interests
- a multidisciplinary approach, as opposed to a narrow grounding or focus in a single discipline
- good "investigative" skills, including doggedness, the ability to draw people out, and the ability to ward off premature closure

In some sociological or anthropological textbooks, lack of familiarity with the phenomenon and setting, and a single-disciplinary grounding are considered assets. But although unfamiliarity with the phenomenon or setting allows for a fertile "decentering," it also can lead to relatively naive, easily misled, easily distracted fieldwork, along with the collection of far too much data.

The problem is how to get beyond the superficial or the merely salient, becoming "empirically literate." You can understand little more than your own evolving mental map allows. A naive, undifferentiated map will translate into global, superficial data and interpretations—and usually into self-induced or informant-induced bias as well. You have to be knowledgeable to collect good information (Markus, 1977). As Giorgi (1986) puts it, "educated looking" is necessary.

Inexperience and single-discipline focus can lead to a second danger: plastering a ready-made explanation on phenomena that could be construed in more interesting ways. Thus presumably "grounded" theorizing can turn out to be conceptual heavy-handedness, without the researcher's even being aware of it. (Ginsberg, 1990, even suggests that the researcher's "counter-transference"—like the psychoanalyst's unacknowledged feelings toward the patient—is at work during data collection and must be surfaced and "tamed" through discussions with peers, careful retrospective analysis, and "audits." Transcending personal biases and limitations is not easy.)

On balance, we believe that a knowledgeable practitioner with conceptual interests and more than one disciplinary perspective is often a better research "instrument" in a qualitative study: more refined, more bias resistant, more economical, quicker to home in on the core processes that hold the case together, and more ecumenical in the search for conceptual meaning.[7]

Time Required

It is not possible to specify time requirements for instrumentation. So much depends on the modes of data collection involved, on how prestructured you choose to be, on the nature of the research questions, and on the complexity of the sample. More structured instrumentation, done with diverse samples, with a confirmatory emphasis, will take substantially more time to develop.

Summary Comments

We've looked at substantive moves that serve to focus and bound the collection of data—reducing it in advance, in effect. These moves include systematic *conceptual frameworks* organizing variables and their relationships, *research questions* that further define the objects of inquiry, defining the "heart" and boundaries of a study through *case definition*, planning for within-case and multiple-case *sampling*, and creating *instrumentation*. All of these moves serve both to constrain and support analysis. All can be done inductively and developmentally; all can be done in advance of data collection. Designs may be tight or loose. Such choices depend on not only your preferred research style but also the study's topic and goals, available theory, and the researcher's familiarity with the settings being studied.

In the next chapter, we pursue other, more technical issues of focusing and bounding a study during its design.

Notes

1. There is a paradox here (R. E. Herriott, personal communication, 1983): We may have more confidence in a finding across multiple cases when it was not anticipated or commonly measured, but simply "jumps out" of all the cases. Of course, researchers should stay open to such findings—although it cannot be guaranteed they will appear.

2. For a display showing the school settings and the kinds of innovations involved, see Tables 2.1 and 2.2 in this chapter, section D. The study's 12 case reports are available from the principal investigator (D. P. Crandall, The Network, Inc., 300 Brickstone Square, Suite 900, Andover, MA 01810). We recommend the Carson and Masepa cases for readers interested in comprehensive treatment, and the Lido case for a condensed treatment. The cross-case analysis appears in Huberman and Miles (1984).

3. Such support is a variant of the famous "degrees of freedom" argument made by Campbell (1975), who displaced the unit of analysis from the case to each of the particulars making up the case, with each particular becoming a test of the hypothesis or theory being tested. The greater the number of such particulars and the greater their overlap, the more confidence in the findings and their potential generalizability.

4. A fascinating example of interview design comes from Leary's (1988) description of the "cognitive interview" used in police work. A witness to a crime is asked to re-create the atmosphere of the scene (sounds, smells, lighting, furniture) and then to tell the story of what happened. The interviewer does not interrupt until the end of the story and then asks follow-up questions, starting with the end and working backward. Compared with conventional questioning, this design yields 35% more facts and increases the number of "critical facts" elicited as well.

5. A good example is Freeman, Klein, Riedl, and Musa's (1991) study of the strategies followed by computer programmers. In a first interview, the programmer told a story (e.g., of a debugging incident). The second interview was a joint editing of the story. The third interview was more conceptual, probing for decision points, options, and criteria; the fourth highlighted pitfalls and further decision points. Later interviews served to confirm an emerging theory of programmers' strategies.

6. These distinctions among types of validity are drawn from Warner (1991). See also the discussion of validity in Chapter 10, section C.

7. Extra ammunition on this point is supplied by Freeman's (1983) revisionist attack on Margaret Mead's (1928) pioneering study of Samoan culture. Freeman considers Mead's findings to be invalid as a result of her unfamiliarity with the language, her lack of systematic prior study of Samoan society, and her residence in an expatriate, rather than in a Samoan, household. For instance, well-meant teasing by adolescent informants may have led to her thesis of adolescent free love in Samoa. In addition, Freeman argues, because Mead was untutored in the setting, she had recourse to a preferred conceptual framework (cultural determinism). This weakened her findings still further.

Some epistemologists (e.g., see Campbell, 1975) have also made a strong case for fieldwork conducted by "an alert social scientist who has thorough local acquaintance." Whyte (1984) also wants the sociologist, like the physician, to have "first, intimate, habitual, intuitive familiarity with things; secondly, systematic knowledge of things; and thirdly, an effective way of thinking about things" (p. 282).

3

Focusing and Bounding the Collection of Data

FURTHER DESIGN ISSUES

Once the substantive issues are more or less clearly seen during study planning, some design issues remain. They are more mundane, but often critical for supporting successful analysis. We discuss here how qualitative and quantitative data may be linked in the study, and a series of topics on study management, including the use of computers, data management, staffing/time planning, and agreements made with study participants.

A. Linking Qualitative and Quantitative Data

Rationale

The late Fred Kerlinger, quantitative researcher par excellence, once said to one of us, "There's no such thing as qualitative data. Everything is either 1 or 0." Against this view, we have Berg's (1989) equally fervent dictum that all data are basically qualitative: To a raw experience, we may attach either words or numbers. Or as Campbell (1974) remarks, all research ultimately has a qualitative grounding.

Fierce battles have been fought on this topic. Quantitative studies have been linked with positivism and damned as incommensurable with naturalistic, phenomenological studies (J. K. Smith, 1983; J. K. Smith & Heshusius, 1986). Qualitative researchers have complained that they are disparaged as The Other, losing out against the powerful, prestigious establishment that takes quantitative methods for granted (Reinharz, 1993). Researchers are stereotyped as number crunchers or "navel-gazers"; Gherardi and Turner's (1987) thoughtful analysis of the conflict is charmingly titled *Real Men Don't Collect Soft Data.*[1]

But at bottom, we have to face the fact that numbers and words are *both* needed if we are to understand the world. As Kaplan (1964) puts it, "Quantities are *of* qualities, and a measured quality *has* just the magnitude expressed in its measure" (p. 207).

Gherardi and Turner (1987) suggest that the issue is one of knowing when it is useful to count and when it is "difficult or inappropriate to count at all," when data are nonstandardized and we have no clear rules for saying what is variation and what is error.

Weinstein and Tamur (1978) are also persuasive. They see quantification not as an end in itself, but rather as "a means of making available techniques which add power and sensitivity to individual judgment when one attempts to detect and describe *patterning* in a set of observations. . . . Why throw away anything helpful?" (p. 140).

We believe that the quantitative-qualitative argument is essentially unproductive. Like Reichardt and Cook (1979) and Miller and Fredericks (1991), we see no reason to tie the distinction to epistemological preferences. For example, Howe's analyses (1985, 1988) show that quantitative and qualitative methods are "inextricably intertwined," not only at the level of specific data sets but also at the levels of study design and analysis.

In a deeper sense, as Salomon (1991) points out, the issue is not quantitative-qualitative at all, but whether we are taking an "analytic" approach to understanding a few controlled variables, or a "systemic" approach to understanding the interaction of variables in a complex environment.[2]

The question, then, is not whether the two sorts of data and associated methods can be linked during study design, but whether it should be done, how it will be done, and for what purposes.

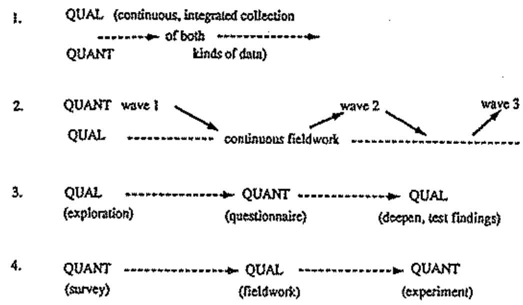
Why link qualitative and quantitative data? Rossman and Wilson (1984, 1991) suggest three broad reasons: (a) to enable confirmation or corroboration of each other via triangulation; (b) to elaborate or develop analysis, providing richer detail; and (c) to initiate new lines of thinking through attention to surprises or paradoxes, "turning ideas around," providing fresh insight.

Greene, Caracelli, and Graham (1989), after reviewing 57 mixed-method evaluation studies, extended this list. They proposed that such studies can help sequentially (results of the first method inform the second's sampling, instrumentation, etc.) and can expand the scope and breadth of a study by using different methods in different components.

Similarly Firestone (1987) suggests that, on the one hand, quantitative studies "persuade" the reader through de-emphasizing individual judgment and stressing the use of established procedures, leading to more precise and generalizable results. On the other hand, qualitative research persuades through rich depiction and strategic comparison across cases, thereby overcoming the "abstraction inherent in quantitative studies."

Sieber (1973) offers a detailed list of reasons to combine methods, with many examples. To illustrate: Quantitative data can help with the qualitative side of a study during *design* by finding a representative sample and locating deviant cases. It can help during *data collection* by supplying background data, getting overlooked information, and helping avoid "elite bias" (talking only to high-status respondents). During *analysis* quantitative data can help by showing the generality of specific observations, correcting the "holistic fallacy" (monolithic judgments about a case), and verifying or casting new light on qualitative findings.

Figure 3.1
Illustrative Designs Linking Qualitative and Quantitative Data



Looked at the other way, qualitative data can help the quantitative side of a study during *design* by aiding with conceptual development and instrumentation. They can help during *data collection* by making access and data collection easier. During *analysis* they can help by validating, interpreting, clarifying, and illustrating quantitative findings, as well as through strengthening and revising theory.

To benefit in these ways, study designers need, as Greene et al. (1989) propose, to consider some key issues: the complementary differences across methods and the purposes for using them, the phenomena being studied, the implicit paradigms used, and what will happen concretely—Are the quantitative and qualitative sides of equal status? Are they interactive or separate? How are they sequenced?

If you believe, as we do, that linkage is a good idea, it can lead to overall designs like those shown in Figure 3.1.

In Design 1, fieldwork involves steady, integrated collection of both quantitative and qualitative data, as needed to understand the case at hand.

Design 2 shows a multiwave survey, conducted in parallel with continuous fieldwork. The first survey wave may draw attention to things the field-worker should look for, the next fieldwork findings may lead to revisions in wave 2 of the survey, and so on.

Design 3 alternates the two kinds of data collection, beginning with exploratory fieldwork, leading to the development of quantitative instrumentation, such as a questionnaire. The questionnaire findings can be further deepened and tested systematically with the next round of qualitative work.

Design 4 shows another alternating style: An initial survey helps point the field-worker to phenomena of importance; the field-worker moves to develop a close-up, strong conceptual understanding of how things work; and a quantitative experiment is designed to test some resulting, perhaps competing, hypotheses.[3]

These designs underline an important point. Both types of data can be productive for descriptive, reconnoitering, exploratory, inductive, opening-up purposes. And both can be productive for explanatory, confirmatory, hypothesis-testing purposes. We agree with Madey (1978): "The methods are not always interchangeable. Each method can, however, be strengthened by using intrinsic qualities of the other" (p. 7).

For example, the careful measurement, generalizable samples, experimental control, and statistical tools of good quantitative studies are precious assets. When they are combined with the up-close, deep, credible understanding of complex real-world contexts that characterize good qualitative studies, we have a very powerful mix. As Jick (1979) notes, qualitative methods can be "the glue that cements the interpretation of multimethod results." (See also Krathwohl, 1991, for a thoughtful comparative analysis.)

Brief Description

We see qualitative-quantitative linkage at three levels. The first is the "quantizing" level, where qualitative information can be either counted directly (say, the number of times a doctor interrupts a patient during an interview) or converted into ranks or scales (in this teacher's classroom, "moderate" mastery of a specific innovation has occurred).[4]

The second level is that of linkage between distinct data types, where qualitative information (say, from an open-ended interview) is compared to numerical data (say, from a questionnaire the same person filled out).

The third is that of overall study design, such as the multimethod approaches sketched in Figure 3.1, or more complex ones; all may involve combinations of case study, survey, experiments, and unobtrusive-measure studies.[5]

Illustrations

The quantizing level. Here are some brief examples. Morse (1989) describes what she calls "appurtenant counting." She cites a study of teenage mothers and the frequency of their use of the word *stuff,* as in "We spent a lot of time together and stuff." The word occurred more than 100 times in 16 open-ended interviews, suggesting a "half-child mode of speech" being used by the women as they coped with adult responsibilities.

During our school improvement study, we gradually became aware of the importance of job mobility and looked at people in our sites who had changed jobs (moving up, out, sideways, or down). It proved helpful to know how many people had moved (75, for 12 sites), how many of these had moved because of their experience with the innovation (83%), and how many were actually upward moves (35%). Here, too, we made a site-level judgment (see Table 7.16).

We also converted some interview data into rating scales: the degree of pressure teachers had felt to adopt an innovation, their satisfaction with the assistance they had received, or the "roughness" or "smoothness" of the implementation. Because our cases were schools, not individuals, this conversion involved examining interviews from different people, checking degree of agreement, and arriving at a site-level rating. Three- to five-point scales seemed easiest and most reliable (e.g., high, moderate/high, moderate, moderate/low, low; or smooth, mostly smooth, mixed, rough, very rough). In the displays and analysis, we aimed to keep "numbers" like these closely associated with the words from which we drew the judgment, and to keep the words associated with the context.

Linking data types. Kell (1990) studied the effects of computers on classroom teaching and learning. The research team repeatedly visited classrooms in six school districts over a school year, doing interviews with teachers and administrators and making systematic classroom observations. They also asked teachers to complete two different standardized questionnaires at three times during the year, covering their theoretical orientation to reading and their concerns about adoption and use of innovations. Thus the two data types were closely linked, and they permitted Kell to study changes occurring over the whole school year.

Stringfield and Teddlie (1990, 1991) used achievement data to select 16 schools, half of which were low-achieving and half high-achieving "outliers." Then field-workers who, like the school personnel, did not know whether they were involved in a "high" or "low" site, collected qualitative information through interviews and "high-inference" observations, along with questionnaires and a quantitative measure of classroom behavior. Through both sorts of data, the field-workers correctly identified all highs and lows and correctly predicted that four low schools would later improve. The themes identified cast much light on what makes for a more and less "effective" school.[6]

Multimethod designs. Wilson and Bolland (1992) studied organizations that provide care for the elderly, wanting to test competing explanatory models. They collected numerical data on referrals and goal-oriented actions for 49 organizations and then built a nominated sample of community leaders (Who sets the community agenda for the elderly?). The referral and goal-oriented data were analyzed to determine which organizations were most central in the network of organizations; the leader data were examined to see which leaders were most central in their network. Semistructured interviews were conducted with central leaders. Combined analysis supported one of the theoretical models; this could not have been accomplished with either data set alone.

Maxwell, Bashook, and Sandlow (1986) combined ethnographic and experimental/control group methods in a study of how peer review committees influenced the learning of participating physicians. After exploratory fieldwork, they compared high- and low-influence committees to find recurring patterns and then developed a "model program" to be carried out in the original committees and in added control committees. Both qualitative and quantitative outcome data were examined, showing moderate to strong differences as predicted and illuminating some unexpected ones as well.[7]

Advice

1. Don't fall into a "default" mode that sees qualitative data as the only way of proceeding. Talk over your assumptions with a colleague to be sure you are not in an unquestioning, partisan frame of mind.

2. Consider whether your study can benefit from a quantitative aspect or component. Think *purposes* and think ahead: In the light of my research questions and the audiences for my study report, will qualitative information be enough, or should it be complemented by a numerical data set of some kind?

3. Remember that, during study design, you are also setting up the social system of the project. If you are a lone researcher, can you manage both sorts of information? What social or technical supports may you need? On this issue, as with many others, we advise finding and using a "friendly stranger" who can advise you, critique products, and provide a supportive, different perspective.

If you are part of a research team, what will be the consequences of a specific division of labor? In our experience, for example, keeping qualitative and quantitative researchers separate can feed invidious comparisons, stereotyping, and sterile arguments about "which data do you trust?" Also, postponing one type of data collection until late in the study means that the quality and robustness of data collection may suffer.

4. Look carefully at the costs you will incur by including a quantitative component. Brewer and Hunter (1989) have useful suggestions.

5. When sampling for a quantitative component, remember that you can fall back on a wide range of well-developed sampling methods.

B. Management Issues Bearing on Analysis

How a qualitative study is managed from Day 1 strongly influences the kinds of analyses that can be done, and how easily. Qualitative studies, especially those done by the lone researcher or the novice graduate student, are notorious for their vulnerability to poor study management. As Kvale (1988) points out in his entertaining 16-page analysis of the naive question, "How shall I find a method to analyze the 1,000 pages of interview transcripts I have collected?" the first answer is, "Never conduct interview research in such a way that you arrive at a situation where you [have to] ask such a question" (p. 90).

We make no pretense of being exhaustive on issues of study management, but want to point to a series of issues and associated design decisions with strong implications for analysis. We illustrate along the way with brief examples and give advice, largely from our own experience and the lore of colleagues. As it happens, remarkably little has been written specifically about the management of qualitative studies.[8]

We deal with four topics in turn: the use of computers, data management, staffing/time planning, and agreements made with study participants. Warning: You, like the zoologist's son, might learn more about penguins than you really wanted to know. You may want to skim or jump forward to other chapters on analysis as such, and return to these issues later.

Computer Use

The first edition of this book had less than half a page on computers as a means for data storage and retrieval. In fewer than 10 years, the development of new software has been phenomenal; on the order of two dozen programs well suited to the needs of qualitative researchers are now available. In our survey, three-quarters of the respondents reported using computer software for entering data, coding, search and retrieval, making displays, or building concepts.[9]

By now it is largely taken for granted that you need a good word processor to do qualitative research. Handwritten or dictated field notes, along with tape recordings, must be converted into analyzable text, which then needs to be reduced, displayed, and used to draw and verify conclu-

Figure 3.2
Uses of Computer Software in Qualitative Studies

1. Making notes in the field
2. Writing up or transcribing field notes
3. Editing: correcting, extending or revising field notes
4. Coding: attaching key words or tags to segments of text to permit later retrieval
5. Storage: keeping text in an organized database
6. Search and retrieval: locating relevant segments of text and making them available for inspection
7. Data "linking": connecting relevant data segments with each other, forming categories, clusters or networks of information
8. Memoing: writing reflective commentaries on some aspect of the data, as a basis for deeper analysis
9. Content analysis: counting frequencies, sequence or locations of words and phrases
10. Data display: placing selected or reduced data in a condensed, organized format, such as a matrix or network, for inspection
11. Conclusion drawing and verification: aiding the analyst to interpret displayed data and to test or confirm findings
12. Theory building: developing systematic, conceptually coherent explanations of findings; testing hypotheses
13. Graphic mapping: creating diagrams that depict findings or theories
14. Preparing interim and final reports

sions. Ordinary typing and retyping is far too slow and costly.

But it's also fair to say that the researcher who does not use software beyond a word processor will be hampered in comparison with those who do. As Tesch (1989) points out, computer-aided analysis can reduce analysis time, cut out much drudgery, make procedures more systematic and explicit, ensure completeness and refinement, and permit flexibility and revision in analysis procedures. Ragin and Becker (1989) add that the microcomputer is especially useful for "case-oriented" researchers, those interested in "interconnected arguments about interrelated events," rather than in sheer statistical "variation." Computing, they suggest, can move studies beyond the "handicraft production" that has characterized much qualitative research.

Issues and decisions. The basic issue is, What could I use computer software for in my project? In Figure 3.2 we summarize some typical uses. It looks a bit overwhelming. Do I really have to think about all this? The short answer is yes.

A longer, more humane answer addresses the next level of decision: What specific software (if any) should I be using? In the Appendix we provide suggestions on how to choose particular programs that are appropriate for your level of computer sophistication, for the project at hand, and for the sort of analysis you are contemplating.[10]

The question is whether I can get along effectively with, say, my familiar word processor, which may cover items 1, 2, 3, 4, 6, 8, and 14 but may be slow, difficult, and less effective for the other, more complex tasks listed in Figure 3.2. In most studies it is more practical to find and use additional analysis software than to rely only on your word processor. Coding will go faster, better, and more precisely with programs such as HyperQual, The Ethnograph, or Kwalitan. Storing and retrieving relevant chunks of text will be much quicker and easier with programs such as Sonar Professional, Orbis, askSam, the Text Collector, or Metamorph. Associations and links within your data can be made more easily with programs such as ATLAS/ti, HyperQual, or HyperRESEARCH. Data can be displayed more easily through programs such as AQUAD, ATLAS/ti, and Orbis. Conceptual and graphic mapping can be facilitated by programs such as Inspiration, MetaDesign, MECA, and SemNet. Theories can be formulated and tested more easily with programs such as AQUAD, ATLAS/ti, QCA, NUDIST, and HyperRESEARCH. These programs and others are listed in the Appendix and are described in detail by Tesch (1990) and by Weitzman and Miles (1993).

Advice.

1. The question, What software is best? cannot be answered meaningfully (cf. Bernard, 1991). You have to get specific about the kind of database you are building in your project and about the kind of analysis you will be doing. See the Appendix for more.

2. Work at a level that is comfortable for you, technically speaking, and extend your understanding with the help of friends whose computer skills are stronger than yours. As we explain in the Appendix, friends are crucial "enablers," supplying help of many kinds. You learn from your friends, and they learn from helping you.

3. Although tutorials supplied with programs can help, the best way to learn a new program is to use it on real tasks, with support from friends.

4. Allow time in your study design for learning. More generally (cf. Wolfe, 1992), you usually are faced with trade-offs among (a) your time to learn the program, (b) your time to set up the program to do what you want, and (c) the time you will take to do the actual work by using the program. If you do less on (a) and (b), (c) usually will be slower.

5. No single program can do everything well. You may end up with several programs, each with specific strengths, rather than with one all-purpose program.

6. Remember that software developers are restless. Programs obsolesce and go off the market; programs are refined, sometimes fussily so, sometimes beautifully; new programs appear. Stay tuned, and think critically about what you need as your projects and research career evolve.

Data Management

The idea of data management is familiar to quantitative researchers; they have been taught to think in terms of systematic data sets, codebooks, data cleaning, documentation of defined variables, records of analyses carried out, and so on. Yet, as Freedland and Carney (1992) found, 20% or more of quantitative studies, including those on clinical drug trials (!), have serious deficiencies along these lines and could not be precisely reconstructed by the researcher, let alone be replicated by others. Similarly, data can easily be miscoded, mislabeled, mislinked, and mislaid (Wolfe, 1992) without careful data management plans.

Data management is just as—perhaps even more—important for qualitative researchers. Typically, large amounts of data come from several cases or sites, and sometimes include numerical data sets as well. How to keep track of it all? How to permit easy, flexible, reliable use of the data? How to make them available to several members of a research team? How to do all of this at different points in time over a project's life? How to set things up so that the study could, in principle at least, be verified by someone else or be replicated? As both Wolfe (1992) and Levine (1985) point out, data management and data analysis are integrally related. There is no firm boundary between them.

Issues and decisions. The main issues are ensuring (a) high-quality, accessible data, (b) documentation of just what analyses have been carried out, and (c) retention of data and associated analyses after the study is complete. Many questions are at hand, but they boil down to these: What do I need to anticipate in the way of data management needs? and How thorough do I need to be?

Drawing on ideas from information and library science, Levine (1985) outlined five useful general principles for a storage and retrieval system for qualitative data. Here they are, with examples:

1. *Formatting:* how written-up fieldwork notes are laid out, physically embodied, and structured. For example, notes might be headed with the name of the researcher, the site, the person(s) involved, and the date—and be set up in paragraphs, with lines numbered sequentially.

Fieldwork notes might be put into loose-leaf notebooks or file folders, on file cards or edge-punched McBee cards, and/or in computer files.

Notes also can be sorted into a structure of file types, including those for specific persons, for events, or for topics relevant to the study at hand. Levine suggests that the overall file structure is a map of the data for the researcher's use.

2. *Cross-referral:* information in one file showing where information in another can be found. For example, a file on a specific nurse's work might indicate particular patients she had worked with, along with the locations in each patient file of specific interactions she had had with that patient. Metamorph is an example of a "relational database manager" that can accomplish such a task easily.

3. *Indexing:* a generic term for what usually is called "coding." It includes (a) defining clear categories (codes), (b) organizing these into a more or less explicit structure, embodied in a "thesaurus" or codebook, and (c) pairing of the codes with appropriate places in the database. (We have more to say about coding in the next chapter.) The coding system can evolve during the research or be prestructured, be mostly descriptive or analytic/explanatory, and be relatively exhaustive/comprehensive or precise and focused (cf. Levine's [1985] discussion).

4. *Abstracting:* a condensed summary of longer material (e.g., the complete field notes for a classroom observation, the school newspaper for October 17), which is linked clearly in the file structure to the longer material.

5. *Pagination:* using unique numbers/letters as locators of specific material in field notes. For example, B J K 1 22 locates for Brookside Hospital the first interview with Dr. Jameson by researcher Kennedy, page 22.

Although Levine's (1985) suggestions are cast in terms of "manual data manipulation," they are also key for computerized systems. In our view all of these functions can be accomplished more easily and simply with computer software. BUT NOTE: You will still need a well-organized physical place for storage and retrieval of raw field notes, tapes, edited hard copies, memos, and so on.

What, in summary, needs to be included in the data that are "managed" in a project—and retained after the project? Figure 3.3 presents an ideal wish list.

Do I have to set all this up, and use it all, and keep it all? No.

Will I be better off if I think about these possibilities and their relative importance during design of the study? Yes.

Is it easier to do it with the right software? Yes, two or three times easier. Some programs automatically generate and store material of this sort as you go along.

Advice.

1. Think ahead about these issues if you can, and make some preliminary decisions: What will my data files look

Figure 3.3
What to Store, Retrieve From, and Retain

1. Raw material: field notes, tapes, site documents
2. Partially processed data: write-ups, transcriptions. Initial version, and subsequent corrected, "cleaned", "commented-on" versions
3. Coded data: write-ups with specific codes attached
4. The coding scheme or thesaurus, in its successive iterations
5. Memos or other analytic material: the researcher's reflections on the conceptual meaning of the data
6. Search and retrieval records: information showing which coded chunks or data segments the researcher looked for during analysis, and the retrieved material; records of links made among segments
7. Data displays: matrices or networks used to display retrieved information, along with the associated analytic text. Revised versions of these
8. Analysis episodes: documentation of what you did, step by step, to assemble the displays and write the analytic text
9. Report text: successive drafts of what is written on the design, methods, and findings of the study
10. General chronological log or documentation of data collection and analysis work
11. Index of all of the above material

like? How will they be organized? How can I get the information from them that I need?

2. Set up a first-cut physical filing system early in the project and think early and actively about the software you will need.

3. Expect to revise and extend your data management system as you go.

4. Remember that "research data sets are fragile assets" (Freedland & Carney, 1992). If they are lost, erased, vandalized, or damaged, you are out of luck. MAKE BACKUPS.

5. Save the more obsessive aspects of data management for:

- complex, multicase, multiresearcher, multimethod projects that get chaotic in a hurry if they are not well planned and managed
- high-stakes projects where much is riding on the results and the world is waiting for well-defended answers
- projects you believe are crucial or even definitive
- projects for which you expect or are seeking an external audit

Staffing and Time Planning

Qualitative studies are chronically called "labor intensive," but the details are rarely provided. Here, too, we rely on our experience and that of colleagues.

Issues and decisions. Who will do the work, and how much time will it take? These innocent questions mask more complexity than you often expect.

We've said that lone researchers, new or experienced, should find a critical friend, partner, or colleague who can supply alternative perspectives, support, and protection from bias and self-delusion. In studies that have more than one staff member, there will always be diversity of experience, background, and skills.

That complexity means that the social system of the project needs attention; it will not develop automatically as you wish. It is crucial to build strong relationships with partners or within larger staffs. We've found it useful, for example, to devote plenty of early time to work on core issues (the conceptual framework, research questions, sampling) and to more general "maintenance" issues, such as hopes for and worries about the project, reports of "what else is happening in my life," and procedural ground rules for joint work.

One other rule of thumb we've found useful: Avoid sharp senior-junior divisions of labor, such as having juniors do the fieldwork and the seniors do the analysis and writing. Senior people need to be directly involved in data collection in order to have a concrete feel for what the field setting is like. Junior people will function poorly as "hired hands." Both need to be actively and mutually engaged in thinking about the project and the meaning of emerging findings. You cannot be a "principal investigator" in a field study without spending time in the field.

There comes a time in the planning of every project when we say, "Let's do the arithmetic." On newsprint or a chalkboard, we begin assembling information, starting with the number of expected days of fieldwork and adding multiples of those days for writing up field notes, coding, completing displays, doing cross-case analyses, and writing interim and final reports. Days are added for the issues reviewed in Chapter 2: conceptual framework and research questions, sampling, and instrumentation. Days are needed also for partner or staff meetings, which can be time-expensive (e.g., a four-person staff meeting every other week for 2 hours uses 26 person-days a year).

Then comes the crunch: We have X person-days in the budget, and $X + 34$, or $X + 12$, or $X + 73$ required if we are to deal with our sample of cases in the depth intended. (Coming out with X minus a number is not typical.) The first arithmetic pass usually leads to redefining within-case and/or cross-case sampling, and the depth and detail of analysis being planned.

The multipliers we use look something like this. For each day of fieldwork, expect to spend:

- 2 or 3 days processing field notes (writing them up, correcting, etc.) If tapes are being transcribed, the multiple may

Figure 3.4
The "Arithmetic" for a Project

Task	*Days*
Instrument revision	2
New instrument development	2
Entry to one new site	2
Added literature abstracting	15
Site contact: 1.5 days × 3 visits × 3 sites	13.5
Interview write-ups and coding (3x)	40.5
Case analyses: 6 people × 3 sites × 1.5 days	27
Cross-case analyses: 3 sites × 2 days	6
Site write-up: 3 × 2 days	6
Weekly meetings: 2 hours × 2 staff × 35 weeks	20
Reports:	
Interim	6
Final	25
Journal articles (2)	20
Total	185

run from 4 to 8, depending on the fineness of detail and the transcriber's familiarity with the content

- 1 or 2 days coding (depending on the fineness and complexity of the evolving scheme)
- 1 or 2 days completing displays and writing (depending on the number and type of displays)

These are multipliers for a single case. In a multiple-case study, you (a) multiply by the number of cases, (b) consider what cross-case analyses will be carried out and what the within-case and cross-case reports will look like, and (c) make a total-days estimation.

Here is an illustration from a project one of us is conducting with an associate (Figure 3.4). It was to involve three in-depth "cognitive mapping" interviews over the course of a school year with six people in each of three schools that were engaged in restructuring efforts. (Note: Preliminary conceptual work and instrumentation and pilot work in two sites already had been completed.)

In this case only 116 days of staff time were available. That led, in short order, to decisions to limit the study to two schools rather than three, to meet only every other week, to eliminate one of the journal articles, and to cut the literature abstracting by half. We did not want to jeopardize the "in-depth" aspect of the study by reducing field contact or by doing briefer analyses.

Advice.

1. Take time deliberately for start-up of the relationship between you and your partner, or the functioning of the staff group. Research teams are not built in a day.

2. At the start, people rarely have all of the skills they need for a study. Allow for learning time on such issues as fluency with new software, reliable use of a coding scheme, or drawing conclusions from data displays.

3. Do the "arithmetic" early, before your plans for within-case and cross-case sampling are firm. Repeat the arithmetic after a first round of fieldwork, write-ups, and analysis.

4. The estimates shown are for qualitative data. We will not make estimates for quantitative data, except to note that instrument development is typically much longer and analysis is shorter. If a quantitative component is part of the study, extra days should be added to its own "arithmetic" for linking/interplay with qualitative findings.

5. In a study of any complexity, it will pay to make a matrix of individual staffers by types of work days, to be sure you are working in the realm of the possible.

6. Time plans usually suffer from excessive optimism. Get a cynical skeptic to critique them.

Agreements With Study Participants

Usually study participants and researchers need to reach some explicit agreements about shared expectations.[11] Our intent here is to examine the analysis-related aspects of agreements made with those whose daily lives are being examined.

Issues and decisions. The main issue is, What explicit expectations do we want to build with study participants that will maintain and improve the quality of conclusions?

We might think first of a "meta-agreement": Are we contemplating an equal-status *"narrative inquiry"* model, in which researchers and participants are each "telling their stories" (Connelly & Clandinin, 1990)? Or are we heading for a *collaborative action research* model, in which researchers join forces with people facing a problem to help them study and resolve it (Schensul & Schensul, 1992)? Or will it be a more *traditional* model that differentiates expert researchers from researched-upon "informants"? The first two models imply more shared control over the design and conduct of the project than does the last.

Whatever the basic relationship implied in such a meta-agreement, some matters need to be clarified with participants at the outset. (Still, remember that these matters may not be fully understood on both sides until the study un-

Figure 3.5
Questions for Agreement With Study Participants

1. How much time and effort will be involved?
2. What kind of data collection is involved (e.g., observation, interviewing, journal writing, life histories)?
3. Is participation voluntary?
4. Who will design and steer the study?
5. Will material from participants be treated confidentially?
6. Will participants' anonymity be maintained?
7. Who will produce descriptive and explanatory products?
8. Will participants review and critique interim and final products?
9. What benefits will accrue to participants—both informants and researchers?

folds.) Figure 3.5 presents issues that might be considered as the elements of an initial set of expectations.[12]

Many different agreements can be reached. We comment on the implications of some for doing analysis. A fuller discussion appears in Chapter 11, which deals with ethical issues as they play out during the course of a study.

Data collection agreements (#2) that include active involvement of participants, such as journal writing, are threatened when such participation is coerced, even gently (#3). Such agreements also tend to move the study in the direction of shared study design and steering (#4).

Vagueness about confidentiality (#5), as when a researcher voluntarily or involuntarily passes on a participant's comments to another, often has "drying up" or distorting effects on subsequent data collection; relationships may get strained, and subsequent analyses may be biased.

The same goes for anonymity (#6). An individual, group, or organization not assured in advance of nonidentifiability in study reports may provide biased data (self-censored, defensive, rosy) if it is believed that an accurate, identifiable account would jeopardize some interest. In any case, anonymity of individuals is difficult or impossible to assure when a case study of a group or organization is read by its members. (An alternative agreement is to use real names from the outset; under these circumstances, individuals only provide information they regard as public or nondamaging. This, too, has its conclusion-narrowing aspects.)

A typical agreement is that researchers will produce the products of the study (#7). This agreement rests on the traditional assumption that well-trained researchers will get good data and will draw well-founded conclusions. In the "narrative inquiry" model, however, expertise resides in participants as much as in researchers. The collaborative action research model implies that participant expertise is developed through the researcher's facilitation during the process. In either case, issues of "goodness of analysis" are just as pressing as in the traditional model (see Chapter 10, section C).

Study products are sometimes fed back to participants (#8) as a way of providing "member checks" on the accuracy of descriptions, explanations, and interpretations. Agreements can vary: Does an individual see material about his or her case before others do? Can an individual or group censor or veto material, or even block publication? Or is the agreement only that errors of fact will be corrected and that alternative interpretations will be included in footnotes? Such agreements can improve both the quality of data and the quality of final conclusions, but they also can result in truncated or distorted conclusions if someone has been given, and exercises, the right of censorship.

Researchers usually benefit (#9) from their studies through insights, recognition, promotion, new grants, and consulting. That is why they keep on researching. Participants' benefits are often posed quite vaguely at the start: "the chance to reflect," "clarifying ideas," "learning what others are doing." Advance agreements that mention assistance (as in collaborative research), consultation or training, joint authorship, or shared royalties as expected benefits may improve the quality of the data and the conclusions. If the benefits do not materialize, data and conclusion quality may suffer.

Advice.

1. Be clear in your mind what you want your agreement with participants to be like. Commit it to paper as a vehicle for discussion with them while negotiating entry and access. Once the agreement is clear, a brochure with a summary of the study and its ground rules is helpful for a study that has multiple participants, most of whom will not have been party to initial discussions.

2. Don't assume that an official agreement with a site is understood by everyone you meet. Check.

3. Incorporate in data collection plans an explicit procedure for logging participants' understanding (or misunderstanding) of agreements, including any threats to data or conclusion quality that you see.

Summary Comments

Added design issues that make a big difference in analysis include how qualitative data may be linked with quantitative information from the same setting, and a series of nuts-and-bolts management issues. Careful decisions about which computer software to use for what purposes need to be made. A systematic plan for data management—storage and retrieval of everything from raw data to final study reports—is equally important. Building good colleague and staff relationships is essential, as is initial and recurring time planning. Finally, as with all of these design decisions, the agreements made with study participants about their participation, privacy, access to study reports, and benefits can make a large difference in the quality of analysis that is possible.

This concludes our discussion of the focusing and bounding activities that occur during study design. We turn next to the analysis that takes place during the early stages of a project.

Notes

1. Have a look at this extended quote from Gherardi and Turner (1987):

> The message . . . is that quantitative work is courageous, hard biting, hard work. Collecting hard data means making hard decisions, taking no nonsense, hardening one's heart to weaklings, building on a hard core of material, using hard words to press on to hard won results which often carry with them promises of hard cash for future research and career prospects. By contrast soft data [are] weak, unstable, impressible, squashy and sensual. The softies, weaklings or ninnys who carry it out have too much of a soft spot for counter-argument for them to be taken seriously, they reveal the soft underbelly of the social science enterprise, are likely to soft-soap those who listen to them. They are too soft-hearted, pitying, and maybe even foolish to be taken seriously, so that it is only right that they should be employed on soft money. (p. 5)

Gherardi and Turner underline the ridiculousness of such language. For a thoughtful treatment of the entire qualitative-quantitative topic, we recommend Bryman (1988).

2. Salomon's (1991) analytic/systemic distinction is similar to Ragin's (1987) distinction between "variable-oriented" and "case-oriented" studies, to Maxwell and Miller's (1992) definition of "paradigmatic" and "syntagmatic" studies, and to Mohr's (1982) discussion of "variance" theories that focus on variables and their relationships and "process" theories that focus on events and the processes that connect them (see also Maxwell, 1992a). We explore this topic early in Chapter 7.

3. These qualitative/quantitative designs are deliberately simplified. See Patton (1990) for many other examples, including mixtures of experimental, content-analytic, statistical, and naturalistic components.

4. There is, of course, a long and well-developed tradition of dealing quantitatively with qualitative data: content analysis. The issue is one of counting the frequency and sequencing of particular words, phrases, or concepts. For an incisive, thorough review, see Carley (1990). Carley also demonstrates that texts with exactly the same number of specific concepts can easily have quite different meanings, once the relations among concepts are taken into account. Thus a "network" analysis is required.

Note also the tradition of using quantitative information in anthropological fieldwork. Helpful recent compilations of methods are by Werner and Schoepfle (1987a, 1987b) and Bernard (1988).

5. Among the best treatments of multimethod studies are Fielding and Fielding (1986), Louis (1982), Firestone and Herriott (1983), and Brewer and Hunter (1989).

6. In this and many other multicase studies, it is possible to rank cases and to benefit from the availability of nonparametric statistical techniques (Gibbons, 1992a, 1992b) for contrasting cases.

7. For additional exhibits of linking data types and multimethod designs, see Bryman (1988), who provides many specific examples.

Caracelli and Greene (1993) have very usefully defined and illustrated strategies for analysis in multimethod designs. These include transforming quantitative data into qualitative and vice versa, developing typologies in one data type to apply to the other, looking at extreme cases with both types of data, and consolidation or merging of both data types into a common data set.

8. See also Chapter 1 in Werner and Schoepfle (1987b) for detailed suggestions on the mechanics of qualitative data management.

9. Tesch believes (1989) that diffusion of computer programs to researchers was slowed by negative attitudes toward computers, epistemological reservations, lack of information, or insufficiently powerful and relevant programs. She may be right; it is true that nearly half of the software users we heard from were using a particular program, The Ethnograph. But it is likely that all of these forces have weakened in the years since her report, and that use of computing in qualitative research will continue to accelerate.

10. The reader is referred to the thoughtfully educative treatment in Tesch (1990). She explains carefully what software can do for the qualitative researcher, describes six programs in detail, and discusses the features of some two dozen others. Weitzman and Miles (1994) provide close-up information on 2 programs, comparing their features and suggesting guidelines for choice.

11. Many "agreements" are tacit. We don't discuss here such topics as gaining entry, building rapport, maintaining trust, developing an equal-status relationship, avoiding co-optation, retaining critical distance, ensuring information flow, and the like. For useful discussions of these issues, see Lofland and Lofland (1984), Bogdan and Biklen (1992), and, for a contrasting view, Douglas (1976).

12. Compare these questions with the more earthy ones that typically are raised by prospective participants (Bogdan & Biklen, 1992):

> What are you actually going to do?
> Will you be disruptive?
> What are you going to do with your findings?
> Why us?
> What will we get out of this?

4

Early Steps in Analysis

In this chapter we describe methods for qualitative data analysis that are useful during the early stages of a study, often while data collection is going on. They help organize data for later, deeper analyses, such as those using the displays described in Chapters 5 through 8.

Why is early data analysis important at all? Some qualitative researchers put primary energy into data collection for weeks, months, or even years and then retire from the field to "work over their notes." We believe this is a mistake. It rules out the possibility of collecting new data to fill in gaps, or to test new hypotheses that emerge during analysis. It discourages the formulation of "rival hypotheses" that question a field-worker's routine assumptions and biases. And it makes analysis into a giant, sometimes overwhelming, task that demotivates the researcher and reduces the quality of the work produced.

We strongly recommend early analysis. It helps the field-worker cycle back and forth between thinking about the existing data and generating strategies for collecting new, often better, data. It can be a healthy corrective for built-in blind spots. It makes analysis an ongoing, lively enterprise that contributes to the energizing process of fieldwork. Furthermore, early analysis permits the production of the interim reports that are required in most evaluation and policy studies. So we advise interweaving data collection and analysis from the start. And even in studies with one round of data collection, these early analysis methods can be very helpful.

In this chapter we describe eight main methods useful for early analysis, along with seven supplementary ones. Each of the *main methods* is presented in this format:

- *Name of method.*
- *Analysis problem.* The problem, need, or difficulty faced by a qualitative data analyst, for which the method may be a useful solution.
- *Brief description.* What the method is and how it works.
- *Illustration.* In more detail, a "minicase" showing how the method is developed and used. Usually this section has a variety of nonstandard subheadings, such as "Making the form," "Assembling the data," "Drawing conclusions," "Revising and editing," and "Using the results."
- *Variations.* Alternative approaches that use the same general principle. Work of other researchers is cited.
- *Advice.* Summarizing comments about the use of the method, and tips for using it well.
- *Time required.* Estimates to guide the researcher. These will vary according to subject matter, the researcher's skill, the research questions being asked, the number of cases, and so on.

The *supplementary methods* are illustrated in boxes, with brief explanations. The aim is to suggest simple methods that can be used in conjunction with the main method being discussed. The format varies, but usually has a brief statement of the sorts of problems the method can help with, plus a brief exhibit or illustration of the method and concluding advice.

Our Assumptions About "Data"

For the methods in this and following chapters, we assume that the field-worker has collected information in the form of handwritten or typed field notes,[1] notes dictated after field contact, or tape recordings of interviews or other events in the field setting. In all cases, we are focusing on words as the basic form in which the data are found.[2]

We further assume that the basic, raw data (the scribbled field notes, the dictated tapes, the direct tape recordings) must be processed before they are available for analysis. Field notes must be converted into "write-ups," either typed directly or transcribed from dictation. A write-up is an intelligible product for anyone, not just for the field-worker. It can be read, edited for accuracy, commented on, coded, and analyzed using any of the methods we are about to describe.

Raw field notes are usually fairly illegible and contain private abbreviations. They are also sketchy. Field notes taken during an interview usually contain half or less of the actual content. Notes made afterward are even worse. But a write-up usually will add back some of the missing content because the raw field notes, when reviewed, stimulate the field-worker to remember things said at that time that are not in the notes.[3] Such additions should, of course, be marked to guard against bias. Transcribed dictation is subject to the same sort of processing.

Direct tape recordings of field events also must be processed in some way. For example, the field-worker listens to or watches the tape, makes notes, selects excerpts, and makes judgments or ratings. More typically, the tape is transcribed into text. This process, however, is fraught with slippage; it is dependent on the knowledgeability and skill of the transcribing person. Note, too, that transcripts can be done at different levels of detail, from the "uhs," "ers," pauses, word emphases, mispronunciations, and incomplete sentences of an apparently incoherent speaker (whose facial expressions, explanatory gestures, and tone of voice cannot be typed up), to a smooth, apparently straightforward summary of the main ideas presented by the speaker.[4]

So we are focusing on *words* as the basic medium, and are assuming that the words involved have been *refined* from raw notes or tape recordings into a text that is clear to the reader or analyst. Note, however, that this text may be reduced and simplified considerably from the raw events.

Now, on to the methods. They are arranged roughly from earlier to later in data collection and from simple to complex. Beginning with the *contact summary sheet,* a simple way to summarize time-limited data, we proceed through first-level *coding,* second-level or *pattern codes,* and the process of deriving even more general themes called *memoing.* As more and more data pile up, the *case analysis meeting* and the *interim case summary* are crucial for understanding. The *vignette* presents focused, time-limited information on one or several episodes. The *prestructured case* is an economical way of integrating case data.

We conclude this chapter with a section on *sequential analysis,* illustrating how methods of this sort can follow each other in a flow over time.

A. Contact Summary Sheet

Analysis Problem

After a field contact (from one to several days) and the production of write-ups, there is often a need to pause and ponder: What were the main concepts, themes, issues, and questions that I saw during this contact? Without such reflection, it is easy to get lost in a welter of detail. And communicating important things about a contact to one's colleagues is essential for any project with more than one field-worker.

Brief Description

A *contact summary* is a single sheet with some focusing or summarizing questions about a particular field contact. The field-worker reviews the written-up field notes and answers each question briefly to develop an overall summary of the main points in the contact.

Illustration

Deciding on the questions. The main thing here is being clear about what you (or your colleagues) need to know now about a particular field contact, with a write-up that may run to dozens of pages. Which questions will go to the essence of the data set? Here are some possible questions:

- What people, events, or situations were involved?
- What were the main themes or issues in the contact?
- Which research questions and which variables in the initial framework did the contact bear on most centrally?
- What new hypotheses, speculations, or hunches about the field situations were suggested by the contact?

- Where should the field-worker place most energy during the next contact, and what kinds of information should be sought?

Making the form. The questions should be arranged on a single sheet of paper (using more than both sides of one sheet defeats the purpose), with space for the field-worker's answers. Identifying information on the case, the particular contact, the field-worker, and the date should be indicated as well.

Entering the data. A contact summary sheet is usually best filled out as soon as written-up field notes have been reviewed and corrected. At that point you have a perspective that combines immediacy with a reflective overview of what went on in the contact. You can include your own reflective remarks (see Box 4.2), as well as questions to be answered during the next contact.

Waiting until a contact has been coded thoroughly and fully is probably too late. In addition, the process of coding usually adds so many subsequent thoughts about the contact that summarizing what was originally in the notes may get buried or distorted.

The data on a contact summary sheet are essentially phrases or sentences that the field-worker considers to be an answer to the form's questions after the complete write-up of the contact has been reviewed. It is helpful to take notes while the write-up is being reviewed.

Figure 4.1 shows excerpts from an example. Note that the second and fourth questions of this form indicate that the field-worker started with a focused set of "target questions." Those are useful when your time is limited. Information on each question is summarized, and new target questions are posed for the next visit. Some of these questions come from the background research questions ("How do users really perceive the innovation?"), and some are provoked by data collected during the visit (e.g., English teacher Reilly's "fall from the chairmanship").

Using the results. The filled-out sheet can be used in several ways: (a) to guide planning for the next contact, (b) to suggest new or revised codes (see sections B, C, and D, following), (c) to help with coordination when more than one field-worker is involved in the study, (d) to reorient yourself to the contact when returning to the write-up, and (e) to help with further data analysis (the summary sheets for a number of contacts can themselves be coded and analyzed). All of these uses are easier if the forms have been entered into a computer database.

It is helpful to attach a copy of the summary form to the top page of the write-up, so it's close to the data it summarizes. If you're working alone, it is very useful to share copies of the filled-out form with your critical friend and to build a case file including all contact summary forms for that case. Here, too, a computerized approach makes life easier.

Variations

Contact summary sheets, as just noted, can be used in a more systematic way by applying codes to them. An excerpted illustration appears in Figure 4.2. Here the analyst had a list of codes (called "themes" or "aspects") that were applied to the "salient points" selected from the write-up. New codes were also generated.

We have experimented with doing a "first impressions" contact summary immediately after the contact and before the write-up, but we do not recommend it. There are risks of (a) being overimpressed with vivid incidents and (b) putting off the write-up, with the consequence of distortion of memory. Like Lofland and Lofland (1984), we encourage doing write-ups no later than the day after a field contact. It's a good rule, which we aren't always able to respect.

Advice

The contact summary form looks rather simpleminded. That is because it is. It's a rapid, practical way to do first-run data reduction without losing any of the basic information (the write-up) to which it refers. It captures thoughtful impressions and reflections. It pulls together the data in the "soft computer"—the field-worker's mind—and makes them available for further reflection and analysis not only by the field-worker but also by others.

Keep contact summary forms simple. Focus on the primary concepts, questions, and issues. You need an instrument that makes it easy to do a rapid retrieval and synthesis of what the contact was all about.

During the first uses of the form, it's a good idea to have someone else read the basic write-up and independently fill out a summary form. In that way you can surface bias or selectivity. You need to be able to rely on summaries, to be reasonably sure they are a good capsule of the write-up.

Time Required

Filling out a good contact summary involves reading and reviewing the write-up (typically, 3 or 4 minutes per single-spaced page), plus less than an hour to fill in the form. If more time is needed, the form is probably too complex or demanding.

Figure 4.1
Contact Summary Form:
Illustration (excerpts)

Contact type:		Site:	Tindale
Visit	X	Contact date:	11/28-29/79
Phone		Today's date:	12/28/79
	(with whom)	Written by:	BLT

1. <u>What were the main issues or themes that struck you in this contact?</u>

   Interplay between highly prescriptive, "teacher-proof" curriculum that is top-down imposed and the actual writing of the curriculum by the teachers themselves.

   Split between the "watchdogs" (administrators) and the "house masters" (dept. chairs & teachers) vis a vis job foci.

   District curric. coord'r as decision maker re school's acceptance of research relationship.

2. <u>Summarize the information you got (or failed to get) on each of the target questions you had for this contact.</u>

<u>Question</u>	<u>Information</u>
History of dev. of innov'n	Conceptualized by Curric. Coord'r, English Chairman & Assoc. Chairman; written by teachers in summer; revised by teachers following summer with field testing data
School's org'l structure	Principal & admin'rs responsible for discipline; dept chairs are educ'l leaders
Demographics	Racial conflicts in late 60's; 60% black stud. pop.; heavy emphasis on discipline & on keeping out non-district students slipping in from Chicago
Teacher response to innov'n	Rigid, structured, etc. at first; now, they say they like it/NEEDS EXPLORATION
Research access	Very good; only restriction: teachers not required to cooperate

3. <u>Anything else that struck you as salient, interesting, illuminating or important in this contact?</u>

   Thoroughness of the innov'n's development and training.

   Its embeddedness in the district's curriculum, as planned and executed by the district curriculum coordinator.

   The initial resistance to its high prescriptiveness (as reported by users) as contrasted with their current acceptance and approval of it (again, as reported by users).

4. <u>What new (or remaining) target questions do you have in considering the next contact with this site?</u>

   How do users really perceive the innov'n? If they do indeed embrace it, what accounts for the change from early resistance?

   Nature and amount of networking among users of innov'n.

   Information on "stubborn" math teachers whose ideas weren't heard initially -- who are they? Situation particulars? Resolution?

   Follow-up on English teacher Reilly's "fall from the chairmanship."

   Follow a team through a day of rotation, planning, etc.

   CONCERN: The consequences of eating school cafeteria food two days per week for the next four or five months...

<u>Stop.</u>

Figure 4.2
Contact Summary Form:
Illustration With Coded Themes (excerpt)

CONTACT SUMMARY

Type of contact:	Mtg. Principals (Who, what group)	Ken's office (place)	4/2/76 (date)	SITE Westgate
				Coder MM
	Phone ______ (With whom, by whom)	______ (place)	______ (date)	Date coded 4/18/76
	Inf. Int. ______ (With whom, by whom)	______ (place)	______ (date)	

1. Pick out the most salient points in the contact. Number in order on this sheet and note page number on which point appears. Number point in text of write-up. Attach theme or aspect to each point in CAPITALS. Invent themes where no existing ones apply and asterisk those. Comment may also be included in double parentheses.

PAGE	SALIENT POINTS	THEMES/ASPECTS
1	1. Staff decisions have to be made by April 30.	STAFF
1	2. Teachers will have to go out of their present grade-level assignment when they transfer.	STAFF/RESOURCE MGMT.
2	3. Teachers vary in their willingness to integrate special ed kids into their classrooms—some teachers are "a pain in the elbow."	*RESISTANCE
2	4. Ken points out that tentative teacher assignment lists got leaked from the previous meeting (implicitly deplores this).	INTERNAL COMMUNIC.
2	5. Ken says, "Teachers act as if they had the right to decide who should be transferred." (would make outcry)	POWER DISTRIB.
2	6. Tacit/explicit decision: "It's our decision to make." (voiced by Ken, agreed by Ed)	POWER DISTRIB/ CONFLICT MGMT.
2	7. Principals and Ken, John, and Walter agree that Ms. Epstein is a "bitch."	*STEREOTYPING
2	8. Ken decides not to tell teachers ahead of time (now) about transfers ("because then we'd have a fait accompli").	PLAN FOR PLANNING/ TIME MGMT.

Document Summary Form

Field-workers often pick up a wide range of documents from their sites: meeting agendas, evaluation reports, newspaper articles, budgets, brochures, lunch menus, minutes of meetings, rosters. The list is not endless, but very large. Documents are often lengthy and typically need clarifying and summarizing. You need to know the document's significance: what it tells you and others about the site that is important.

It helps to create and fill out a document summary form, which can be attached to the document it refers to. Box 4.1 is an illustration. This form puts the document in context, explains its significance, and gives a brief summary. Note the field-worker's reflective commentary, set off in double parentheses.

Document summary forms also can be coded, not only for later analysis but also to permit rapid retrieval when needed. A number of computer programs permit "off-line" indexing, so you can ask for certain codes and be directed

Box 4.1
Document Summary Form: Illustration

DOCUMENT FORM

Site: Carson
Document: 2
Date received or picked up: Feb. 13

Name or description of document:

The Buffalo (weekly sheet)

Event or contact, if any, with which document is associated:
Paul's explanation of the admin. team's functioning
Date: Feb. 13

Significance or importance of document:

Gives schedule for all events in the district for the week.
Enables coordination, knits two schools together.

Brief summary of contents:

Schedule of everything from freshman girls' basketball to "Secret Pals Week" in the elementary school.

Also includes "Did you know" items on the IPA program (apparently integrating the IPA News).

And a description of how admin. team works (who is on team, what regular meetings deal with, gives working philosophy (e.g., "we establish personal goals and monitor progress" . . . "We coordinate effort, K-12, and all programs" . . . "We agree on staff selection"). Concluding comment: "It is our system of personnel management."

Also alludes to the 26 OPERATIONAL GUIDELINES (Document 16)

((I'll guess that the admin. explanation does not appear every week—need to check this.))

IF DOCUMENT IS CENTRAL OR CRUCIAL TO A PARTICULAR CONTACT (e.g., a meeting agenda, newspaper clipping discussed in an interview), make a copy and include with write-up. Otherwise, put in document file.

to the file drawer or folder where the original document exists.

For good reviews of methods of document analysis (including content analysis), see Carley (1990), Weber (1990), Bailey (1982), and Krippendorff (1980b).

B. Codes and Coding

Analysis Problem

As soon as the field researcher begins to compile information, challenges appear. A big one comes from the multiplicity of data sources and forms. Some information comes from structured or casual observations. More, probably, comes from interviewing—open-ended, structured, ethnographic (Spradley, 1979) or "elite" (Marshall & Rossman, 1989).[5] There are also less obtrusive measures (Webb, Campbell, Schwartz, & Sechrest, 1965): everyday or special documents, archival records, diaries, and physical artifacts. In some studies there can be information from questionnaires and surveys, films and video, or test data.

All of this information piles up geometrically. Worse still, in the early stages of a study, most of it looks promising. If you don't know what matters more, everything matters. You may never have the time to condense and order, much less to analyze and write up, all of this material—unless you have a few years ahead of you.

That's why we think conceptual frameworks and research questions are the best defense against overload. They also reflect a point we made earlier: that data collection is inescapably a *selective* process, that you cannot and do not "get it all" even though you might think you can and are.

Some transcripts can be "thick," but many can, and probably must, be "thin." And transcriptions often erase the context along with some crucial nonverbal data.[6] What you "see" in a transcription is inescapably selective. A critical theorist sees different things than a deconstructivist or a symbolic interactionist does. The instrumentation, too, will selectively determine much of the data collected. Informants themselves are selective, too, sometimes deliberately, sometimes unwittingly. They may gloss over important parts of their behaviors and perceptions, and the researcher may not be aware of this.

Observation is also selective: The researcher is constantly making choices about what to register and what to leave out, without necessarily realizing that—or why—one exchange or incident is being noted but another is not.

So selectivity is endemic to data collection and does not, in itself, resolve the problem of overload. In fact, if you calculate, as we (Chapter 3) and others (Lofland & Lofland, 1984) have, that you need roughly two to five times as much time for processing and ordering the data as the time you needed to collect it, there are two more dilemmas: data overload and data retrieval.

Data overload. You might well find yourself, 6 months before the end of your study, with an alpine collection of information that might require a week just to read over carefully. A chronic problem of qualitative research is that it is done chiefly with words, not with numbers. Words are fatter than numbers and usually have multiple meanings. This makes them harder to move around and work with. Worse still, most words are meaningless unless you look backward or forward to other words; take, for example, the pronouns *this* and *them* in the preceding sentence. Or take the noun *board* in the phrase "The board is on the fence." Are we talking about a piece of wood or a decision-making body?

Numbers are usually less ambiguous and can be processed more economically. Small wonder, then, that many researchers prefer working with numbers alone or getting the words they collect translated into numbers as quickly as possible.

We argue that although words may be more unwieldy than numbers, they render more meaning than numbers alone and should be hung on to throughout data analysis. Converting words into numbers and then tossing away the words gets a researcher into all kinds of mischief. You thus are assuming that the chief property of the words is that there are more of some than of others. Focusing solely on numbers shifts attention from substance to arithmetic, throwing out the whole notion of "qualities" or "essential characteristics."

Also, when word-derived numbers do not make sense, there is usually no very satisfactory way of making them more intelligible with more numbers. A solution to this problem, as we will see in later sections, is to keep words and any associated numbers *together* throughout the analysis.

A week at a field site often can result in something like 200-300 pages of typed-up field notes and ancillary materials. Everything looks important, especially at the outset. What at first seemed simple gets rapidly more complex and has to be fleshed out. New leads surface and need checking out. All of these tasks add bulk; the danger is getting overloaded with more data than can be processed. Furthermore, the extended narrative nature of field notes makes them difficult to use during analysis. It is difficult to retrieve the most meaningful material, to assemble chunks of words that go together, and to condense the bulk into readily analyzable units.

Data retrieval. The second dilemma is embedded in the first; within that mass of data, you may not know—or may not have tagged—the pieces that matter the most for the purposes of your study.

The challenge is to be explicitly mindful of the purposes of your study and of the conceptual lenses you are training on it—while allowing yourself to be open to and reeducated by things you didn't know about or expect to find. At the same time, in collecting the data, you will have to resist overload—but not at the price of sketchiness. This is easier said than done.

To do it, you will need a variety of safeguards against tunnel vision, bias, and self-delusion (see Chapter 10, section B). Just as important, you will have to accompany each wave of data collection with a corresponding exercise in condensation and analysis. This step is where coding and other forms of ongoing, iterative reflection come in.

Brief Description

Coding is analysis. To review a set of field notes, transcribed or synthesized, and to dissect them meaningfully, while keeping the relations between the parts intact, is the stuff of analysis. This part of analysis involves how you differentiate and combine the data you have retrieved and the reflections you make about this information.

Codes are tags or labels for assigning units of meaning to the descriptive or inferential information compiled during a study. Codes usually are attached to "chunks" of varying size—words, phrases, sentences, or whole paragraphs, connected or unconnected to a specific setting. They can take the form of a straightforward category label or a more complex one (e.g., a metaphor).

For our purposes it is not the words themselves but their *meaning* that matters.[7] Bliss, Monk, and Ogborn (1983) tell us that a word or a phrase does not "contain" its meaning as a bucket "contains" water, but has the meaning it does by being a choice made about its significance in a given context. That choice excludes other choices that could have been made to "stand for" that word or phrase, and that choice is embedded in a particular logic or a conceptual lens, whether the researcher is aware of it or not. It is better, we think, to be aware.

Codes are used to retrieve and organize the chunks mentioned earlier. The organizing part will entail some system for categorizing the various chunks, so the researcher can quickly find, pull out, and cluster the segments relating to a particular research question, hypothesis, construct, or theme. Clustering, and, as we will see, display of condensed chunks, then sets the stage for drawing conclusions.

Until recently, qualitative researchers outside of linguistics and its subfields (e.g., discourse analysis) were not very explicit about how they went about assigning units of meaning to pieces of data. The conventional advice was to go through transcripts or field notes with a pencil, marking off units that cohered because they dealt with the same topic and then dividing them into topics and subtopics at different levels of analysis (cf. Agar, 1980, p. 104ff.). These identifiable topics (or themes or gestalts) presumably would recur with some regularity. They would be given a "name," and instances of them would be marked with a shorthand label (a code). With scissors, file cards, or computers, the topics could be more finely differentiated, clustered, and even relabeled.

The process can be that straightforward. Let's take a look.

Illustration

Types of codes. Let's assume you were interested, as we were in our school improvement study, in the reasons why a new educational practice is adopted. You might begin by asking informants why they or others decided to try the practice. A piece of the field notes might look like this:

> I asked him what the need for the new program was, and he responded that the students coming into the 9th grade were two years below grade level and that the old curriculum was ineffective. Through testing (the Nelson Reading Test) it was determined that students were growing academically only 5 or 6 months during the 10-month school year.

Assuming that you found it possible to apply a single summarizing notation to this chunk, it might be "MOT" to indicate "motivation" (other codes could be applicable). That code would appear in the left-hand margin beside the segment (the right-hand margin might be used for a comment (see Boxes 4.2 and 4.3).

If you wanted more differentiation, the code might separate teachers' motivations from administrators'; we then get "ADM-MOT." Or perhaps you might want to specify the time period or phase in which that motivation appeared (e.g., the "adoption" phase, by lengthening the code to read "AD/MOT"). Or to include all of these things, "AD/ADM-MOT."

These are *descriptive* codes; they entail little interpretation. Rather, you are attributing a class of phenomena to a segment of text. The same segment could, of course, be handled more *interpretively.* Let's assume that, as you become more knowledgeable about local dynamics, a more complex, more "backstage" web of motives turns up. Some people may have adopted the new practice chiefly to attract attention to themselves and to set up a desirable promotion. We then have the public, official motive, such as the one in the segment shown above, and the more private or backstage motive. The segment we just saw could then be coded "PUB-MOT" (for public motivation) and the other segments "PRIV-MOT."

A third class of codes, *pattern* codes, is even more inferential and explanatory. A coded segment of field notes illustrates an emergent leitmotiv or pattern that you have discerned in local events and relationships. These codes can be called, for example, LM (leitmotiv), PATT (pattern), TH (theme), CL (causal link)—and should include a word indicating the inferred theme or pattern. They typically are used later in the course of data collection, as the patterns come clearer.

Here is an example. In the innovation field study, this segment appeared:

> But he (Mr. Walt) says that he does not know that much about what is exactly involved in the SCORE-ON program. He thinks that "it is a combination of a lot of things." The resource lab appears to be used in the morning for the FACILE program, which Mr. Walt knows a great deal more about. . . . In the afternoon, Mrs. Hampshire uses the lab for SCORE-ON purposes. Mr. Walt says that this is a different program, and therefore it is a different use.

That chunk looks innocent enough to be taken descriptively, which is the way the field researcher saw it at first. The segment was coded IP-SALIENCE (how salient the innovation was for users). Several interviews and some casual observations later, however, it looked different. There was apparently an intense power struggle between different factions or "teams" in the district central office—which the researcher later likened to a "lot of rattlesnakes in a jug"—and people were lining up in one or the other camp. Mr. Walt was in the FACILE camp, not in the SCORE-ON camp; both projects were competing for funds and supervisory posts. So the segment got an added code: PATT-TEAMS.

Note that the inferential code is applied here to a segment collected and coded earlier in the analytic cycle. This means periodic rereading of some coded field notes when that code becomes salient. With computer retrieval (see Appendix), this is easy to do.

These illustrations tell us three important things about codes. First, codes can be at different levels of analysis, ranging from the descriptive to the inferential. Second, they can happen at different times during analysis; some get created and used at the start, and others follow—typically the descriptive ones first and the inferential ones later.

Third and most important, codes are astringent—they pull together a lot of material, thus permitting analysis. The PATT-TEAMS code, for example, signals a theme that accounts for a lot of other data—makes them intelligible, suggests thematic links, and functions like a statistical "factor," grouping disparate pieces into a more inclusive and meaningful whole.

We have gone rather quickly through these coding illustrations, working with fairly "molar" codes. Before moving on, we should point out that far more molecular coding schemes are possible and that they open useful analytic avenues. A good exemplar is work on the analysis of conversations (e.g., recursive frame analysis, as used by Chenail, 1991).

Creating codes. One method of creating codes—the one we prefer—is that of creating a provisional "start list" of codes prior to fieldwork. That list comes from the conceptual framework, list of research questions, hypotheses, problem areas, and/or key variables that the researcher brings to the study. In our school improvement study, for example, we conceptualized the innovation process, in part, as one of "reciprocal transformations." Teachers change the characteristics of new practices. Those practices, in turn, change the teachers and modify working arrangements in the classroom, which, in turn, influence how much of the innovation can be used, and so on.

We began, then, with a master code—TR—to indicate the transformational process we had hypothesized and some subcodes—TR-USER, TR-CLASS (classroom changes), TR-ORG (organizational changes), TR-INN (changes in the innovation)—to mark off segments of data in each class of variables. The list was held lightly, applied to the first sets of field notes, and then examined closely for fit and power. Quite a few codes were revised, but the conceptual orientation seemed to bear real fruit—to fit and account well for what we saw and heard.

A start list can have from a dozen or so up to 50-60 codes; that number can be kept surprisingly well in the analyst's short-term memory without constant reference to the full list—*if* the list has a clear structure and rationale. It is a good idea to get that list on a single sheet for easy reference.

Table 4.1 is an example. The first column has a short descriptive label for the general categories and the individual codes. The second column shows the codes, and the third keys the code to the research question or subquestion from which it derives. You also need a list of code definitions (see Figure 4.5) that will be clarified further as the study proceeds.

In this study, we were coding "by hand"—thus the need for short labels. Many computer programs let you assign meaningful phrases as codes, such as "welcomed outside ideas" instead of a more cryptic "AP-IN/OUT."

We know of at least two other useful methods of building codes. First, a more inductive researcher may not want to precode any datum until he or she has collected it, seen how it functions or nests in its context, and determined how many varieties of it there are. This is essentially the "grounded" approach originally advocated by Glaser and Strauss (1967), and it has a lot going for it. Data get well molded to the codes that represent them, and we get more of a code-in-use flavor than the generic-code-for-many-uses generated by a prefabricated start list.[8] The analyst is more open-minded and more context-sensitive, although, here, too, the ultimate objective is to match the observations to a theory or set of constructs. It is not at all the "completely unstructured" process that is so daunting to new researchers.

Of these inductive coding techniques, one of the most helpful is that of Strauss (1987), described best in Strauss and Corbin (1990). Initial data are collected, written up, and reviewed line by line, typically within a paragraph. Beside or below the paragraph, categories or labels are generated, and a list of them grows. The labels are reviewed and, typically, a slightly more abstract category is attributed to several incidents or observations. The incidents then can be put onto a qualitative data category card. Figure 4.3 is an illustration from Turner (1981) about a transport agent (Porcelli) attempting to have his lorries loaded first.

Data retrieval software lets us create such a card, along with the links, very quickly (hypertext programs do this best—see Appendix)—if we have the right key words for moving down from the inferential code (role/status) to gather in the more descriptive incidents.

To pick out provisional codes, Strauss (1987) suggests rereading field notes of contrasting groups, so you're sensitized to what is different about them. He also proposes cutting up a copy of field notes into segments, each containing a potentially important aspect. Having to sort the material into piles creates both categories and differentiations, and gives an idea of the frequencies of each category. Another idea: Read field notes for regularly occurring phrases, and with an eye to surprising or counterintuitive material that needs to be clarified elsewhere in the notes or in the field.

As a way of getting started, Strauss suggests coding for "conditions," "interactions among actors," "strategies and tactics," and "consequences." To find the conditions, for

Table 4.1
Illustration of a Start List of Codes

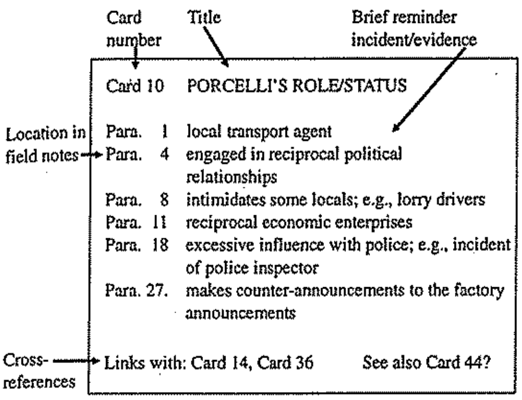
INNOVATION PROPERTIES	IP-OBJ	3.1
IP: OBJECTIVES	IP-OC	3.1.1
IP: ORGANIZATION	IP-ORG/DD, LS	3.1.1
IP: IMPLIED CHANGES–CLASSROOM	IP-CH/CL	3.1.4
IP: IMPLIED CHANGES–ORGANIZATION	IP-CH/ORG	3.1.5
IP: USER SALIENCE	IP-SALIENCE	3.1.2
IP: (INITIAL) USER ASSESSMENT	IP-SIZUP/PRE, DUR	3.1.3, 3.4, 3.5
IP: PROGRAM DEVELOPMENT (IV-C)	IP-DEV	3.1.1, 3.3.3, 3.3.4
EXTERNAL CONTEXT	EC (PRE) (DUR)	3.2, 3.3, 3.4
EC: DEMOGRAPHICS	EC-DEM	
In county, school personnel	ECCO-DEM	3.2.3, 3.3, 3.4
Out county, nonschool personnel	ECEXT-DEM	3.2.3, 3.3, 3.4
EC: ENDORSEMENT	EC-END	3.2.3, 3.3, 3.4
In county, school personnel	ECCO-END	3.2.3, 3.3, 3.4
Out county, nonschool personnel	ECEXT-END	3.2.3, 3.3, 3.4
EC: CLIMATE	EC-CLIM	3.2.3, 3.3, 3.4
In county, school personnel	ECCO-CLIM	3.2.3, 3.3, 3.4
Out county, nonschool personnel	ECEXT-CLIM	3.2.3, 3.3, 3.4
INTERNAL CONTEXT	IC (PRE) (DUR)	3.2, 3.3, 3.4
IC: CHARACTERISTICS	IC-CHAR	3.2.2, 3.4, 3.5
IC: NORMS AND AUTHORITY	IC-NORM	3.2.2, 3.4.3, 3.5
IC: INNOVATION HISTORY	IC-HIST	3.2.1
IC: ORGANIZATION PROCEDURES	IC-PROC	3.1.1, 3.24, 3.3, 3.4
IC: INNOVATION-ORGANIZATION CONGRUENCE	IC-FIT	3.2.2
ADOPTION PROCESS	AP	3.2, 3.3
AP: EVENT CHRONOLOGY–OFFICIAL VERSION	AP-CHRON/PUB	3.2.4, 3.3.1
AP: EVENT CHRONOLOGY–SUBTERRANEAN	AP-CHRON/PRIV	3.2.4, 3.3.1
AP: INSIDE/OUTSIDE	AP-IN/OUT	3.2.5
AP: CENTRALITY	AP-CENT	3.2.2
AP: MOTIVES	AP-MOT	3.2.6
AP: USER FIT	AP-FIT	3.2.7
AP: PLAN	AP-PLAN	3.3.3
AP: READINESS	AP-REDI	3.3.4, 3.2.1
AP: CRITICAL EVENTS	AP-CRIT	3.3.1
SITE DYNAMICS AND TRANSFORMATIONS	TR	3.4
TR: EVENT CHRONOLOGY–OFFICIAL VERSION	TR-CHRON/PUB	3.4.1, 3.4.2, 3.4.3
TR: EVENT CHRONOLOGY–SUBTERRANEAN	TR-CHRON/PRIV	3.4.1, 3.4.2, 3.4.3
TR: INITIAL USER EXPERIENCE	TR-START	3.4.1, 3.4.2, 3.4.3
TR: CHANGES IN INNOVATION	TR-INMOD	3.4.1
TR: EFFECTS ON ORGANIZATIONAL PRACTICES	TR-ORG/PRAC	3.4.3
TR: EFFECTS ON ORGANIZATIONAL CLIMATE	TR-ORG/CLIM	3.4.3
TR: EFFECTS ON CLASSROOM PRACTICE	TR-CLASS	3.4.2
TR: EFFECTS ON USER CONSTRUCTS	TR-HEAD	3.4.2, 3.4.3
TR: IMPLEMENTATION PROBLEMS	TR-PROBS	3.4.1
TR: CRITICAL EVENTS	TR-CRIT	3.4.1, 3.4.2, 3.4.3
TR: EXTERNAL INTERVENTIONS	TR-EXT	3.4.3
TR: EXPLANATIONS FOR TRANSFORMATIONS	TR-SIZUP	3.4.1, 3.4.2, 3.4.3
TR: PROGRAM PROBLEM SOLVING	TR-PLAN	3.4.1, 4.4.2, 3.4.3

(continued)

Table 4.1 (Continued)

NEW CONFIGURATION AND ULTIMATE OUTCOMES	NCO	3.5
NCO: STABILIZATION OF INNOVATION–CLASSROOM	NCO-INNOSTAB/CLASS	3.5.1
NCO: STABILIZATION OF USER BEHAVIOR	NCO-STAB/USER	3.5.2
NCO: USER FIRST-LEVEL OUTCOMES	NCO-USER 1OC	3.5.4
Positive and negative	NCO-USER 1OC/+, –	
Anticipated and unanticipated	NCO-USER 1OC/A, U	
Combinations (when appropriate)	NCO-USER 1OC/A+, A– U+, U–	
NCO: USER META OUTCOMES	NCO-USER META	
Positive and negative	NCO-USER META OC/+, –	
Anticipated and unanticipated	NCO-USER META OC/A, U	
Combinations (when appropriate)	NCO-USER META OC/A+, A– U+, U–	
NCO: USER SPINOFFS AND SIDE EFFECTS	NCO-USER SIDE	3.5.5 (3.5.2)
Positive and negative	NCO-USER SIDE OC/+, –	
Anticipated and unanticipated	NCO-USER SIDE OC/A, U	
Combinations (when appropriate)	NCO-USER SIDE OC/A+, A– U+, U–	
NCO: CLASSROOM INSTITUTIONALIZATION	NCO-INST/CLASS	3.5.5
NCO: STABILIZATION OF INNOVATION–ORGANIZATION	NCO-INNOSTAB/ORG	3.5.6
NCO: STABILIZATION OF ORGANIZATIONAL BEHAVIOR	NCO-STAB/ORG	3.5.7
NCO: ORGANIZATIONAL INSTITUTIONALIZATION	NCO-INST/ORG	3.5.8
NCO: ORGANIZATIONAL FIRST-LEVEL OUTCOMES	NCO-ORG 1OC	3.5.9
Positive and negative	NCO-ORG 1OC/+, –	
Anticipated and unanticipated	NCO-ORG 1OC/A, U	
Combinations (when appropriate)	NCO-ORG 1OC/A+ A– U+, U–	
NCO: ORGANIZATIONAL META OUTCOMES	NCO-ORG META	3.5.9
Positive and negative	NCO-ORG META OC/+, –	
Anticipated and unanticipated	NCO-ORG META OC/A, U	
Combinations (when appropriate)	NCO-ORG META OC/A+, A– U+, U–	
NCO: ORGANIZATIONAL SPINOFFS AND SIDE EFFECTS	NCO-ORG SIDE	3.5.9 (3.5.7)
Positive and negative	NCO-ORG SIDE OC/+, –	
Anticipated and unanticipated	NCO-ORG SIDE OC/A, U	
Combinations (when appropriate)	NCO-ORG SIDE OC/A+, A– U+, U–	
NCO: INSTITUTIONAL EXPANSION	NCO-INNOGRO/ORG	3.5.8
NCO: ORGANIZATIONAL REDUCTION	NCO-INNODWIN/ORG	3.5.8
EXTERNAL AND INTERNAL ASSISTANCE (SEPARATE CODES FOR EXTERNAL, PEER, ADMINISTRATIVE)		
ASS: LOCATION	ASS-LOC	3.6.1
ASS: RULES, NORMS	ASS-RULE	3.6.1
ASS: ORIENTATION	ASS-ORI	3.6.2
ASS: TYPE	ASS-TYPE	3.6.3
ASS: EFFECTS	ASS-EFF	3.6.4
ASS: ASSESSMENT BY RECIPIENTS	ASS-ASS	3.6.5
ASS: LINKAGE	ASS-LINK	3.6.6
EMERGING CAUSAL LINKS	CL	
CL: NETWORKS	CL-NET	N.A.
CL: RULES	CL-RULE	N.A.
CL: RECURRENT PATTERNS	CL-PATT	N.A.
Within site	CL-PATT/LS	N.A.
Intersite	CL-PATT/OS	N.A.
CL: EXPLANATORY CLUSTER (researcher)	CL-EXPL	N.A.
(respondent)	SITECL-EXPL	N.A.
QUERIES	QU	
QU: SURPRISES	QU-!	N.A.
QU: PUZZLES	QU-Q	N.A.

Figure 4.3
Example of Category Card (Turner, 1981)



example, the analyst looks for such words as *because* or *since.* To find the consequences, you follow up on such terms as *as a result of* and *because of.* Phrases that are used repeatedly by informants ("in vivo" codes) are also good leads; they often point to regularities in the setting.

A second main coding alternative, partway between the a priori and inductive approaches, is that of creating a general accounting scheme for codes that is not content specific, but points to the general domains in which codes can be developed inductively. Lofland's (1971) scheme suggests that codes in any study can deal with the following sorts of phenomena, graduated from micro to macro levels:

1. *Acts:* action in a situation that is temporally brief, consuming only a few seconds, minutes, or hours
2. *Activities:* actions in a setting of more major duration—days, weeks, months—constituting significant elements of people's involvements
3. *Meanings:* the verbal productions of participants that define and direct action
4. *Participation:* people's holistic involvement in or adaptation to a situation or setting under study
5. *Relationships:* interrelationships among several persons considered simultaneously
6. *Settings:* the entire setting under study conceived as the unit of analysis

Another mid-range accounting scheme comes from Bogdan and Biklen (1992), who divide codes in the following way:

1. *Setting/Context:* general information on surroundings that allows you to put the study in a larger context
2. *Definition of the situation:* how people understand, define, or perceive the setting or the topics on which the study bears
3. *Perspectives:* ways of thinking about their setting shared by informants ("how things are done here")
4. *Ways of thinking about people and objects:* understandings of each other, of outsiders, of objects in their world (more detailed than above)
5. *Process:* sequence of events, flow, transitions, and turning points, changes over time
6. *Activities:* regularly occurring kinds of behavior
7. *Events:* specific activities, especially ones occurring infrequently
8. *Strategies:* ways of accomplishing things; people's tactics, methods, techniques for meeting their needs
9. *Relationships and social structure:* unofficially defined patterns such as cliques, coalitions, romances, friendships, enemies
10. *Methods:* problems, joys, dilemmas of the research process—often in relation to comments by observers

Such schemes help the researcher think about categories in which codes will have to be developed. Any particular study, of course, may focus on only a few of the categories. And there is a slight danger in distributing coded segments within a nominal frame that is little more than a catalogue—unless the categories are meaningfully interrelated at some point.

Many researchers use a simple two-level scheme: a more general "etic" level, like those suggested above; and a more specific "emic" level, close to participants' categories but nested in the etic codes.

Revising codes. For all approaches to coding—predefined, accounting-scheme guided, or postdefined—codes will change and develop as field experience continues. Researchers with start lists know that codes will change; there is more going on out there than our initial frames have dreamed of, and few field researchers are foolish enough to avoid looking for these things.

Furthermore, some codes do not work; others decay. No field material fits them, or the way they slice up the phenomenon is not the way the phenomenon appears empirically. This issue calls for doing away with the code or changing its level.

Other codes flourish, sometimes too much so. Too many segments get the same code, thus creating the familiar problem of bulk. This problem calls for breaking down codes into subcodes.

With manual coding, revision is tedious: Every chunk you have coded before has to be relabeled. But the search-and-replace facility of your word processor can accomplish this easily. Most analysis software has routines for revising codes.

Still other codes emerge progressively during data collection. These are better grounded empirically, and are especially satisfying to the researcher who has uncovered an important local factor. They also satisfy other readers, who can see that the researcher is open to what the site has to say, rather than determined to force-fit the data into preexisting codes. Most field researchers, no matter how conceptually oriented, will recognize when an a priori coding system is ill molded to the data or when a rival perspective looks more promising.

Some analysts have names for coding procedures later in the study cycle. For example, Lincoln and Guba (1985) talk about these operations:

1. "filling in": adding codes, reconstructing a coherent scheme as new insights emerge and new ways of looking at the data set emerge
2. "extension": returning to materials coded earlier and interrogating them in a new way, with a new theme, construct, or relationship
3. "bridging": seeing new or previously not understood relationships within units of a given category (that relationship will itself have a name, and it may call for a new configuration of the categories)
4. "surfacing": identifying new categories

Like Strauss (1987), Lincoln and Guba suggest that coding and recoding are over when the analysis itself appears to have run its course—when all of the incidents can be readily classified, categories are "saturated," and sufficient numbers of "regularities" emerge. Be careful here. Fieldwork understanding comes in layers; the longer we are in the environment, the more layers appear to surface, and the choice of when to close down, when to go with a definitive coding scheme or definitive analysis, can be painful. That choice may be dictated as often by time and budget constraints as on scientific grounds. When those constraints are relaxed, saturation can become a vanishing horizon—just another field trip away, then another. . . .

Throughout this process of revision, three sources of knowledge are being weighed. First, researchers are refining or recasting parts of the conceptual structure they brought to the study—or, if they went with a more grounded approach, they are feeling their way to a set of relationships that account for important pieces of what they are seeing and hearing.

Second, the field site has a life of its own that becomes more meaningful and decipherable as you spend time there, sharing the daily routines of actors in the setting.

Third, the field site emits a continuous stream of leads, mysteries, themes, and contradictions that need to be pursued and that will never fit perfectly into a precoded conceptual frame or even into a more grounded, emerging coding system.

The tension between these three streams projects the study forward. The coding is a way of forcing you to understand what is still unclear, by putting names on incidents and events, trying to cluster them, communicating with others around some commonly held ideas, and trying out enveloping concepts against another wave of observations and conversations.

The importance of structure. Whether codes are created and revised early or late is basically less important than whether they have some conceptual and structural order. Codes should relate to one another in coherent, study-important ways; they should be part of a governing structure. Incrementally adding, removing, or reconfiguring codes produces a ragbag that usually induces a shapeless, purely opportunistic analysis. It also makes the codes harder to memorize and use, and makes the retrieval and organization of the material burdensome.

Figure 4.4 is an illustration of a poorly structured set of codes. It comes from a study of the creation of new schools (Miles et al., 1978). The study developed inductively, and over 2 years the number of codes mushroomed. The only structure is a sorting of 175 different codes into four general bins (formal and informal actors, processes, and aspects of the new school); the codes are arranged only alphabetically within each bin. The scheme proved more and more difficult to remember and use; eventually it was collapsed in sheer desperation to a series of 28 more general categories, including "adaptiveness," "time management," "sociopolitical support," "internal communication," and the like.

Figure 4.4 can be compared profitably with the conceptual structure underlying the coding scheme shown in Table 4.1; that scheme was focused around families of variables, was easily remembered and usable, and led directly to our analysis.

Table 4.1 and Figure 4.5 are examples of a well-structured code list: a start list of codes, keyed to research questions and (in this case) to "bins" of conceptual variables, and defined precisely enough so that researchers have a common language and can be clear about whether and how a segment of data actually fits into a category. The actual coded segments then provide instances of the category, and marginal or appended comments begin to connect different codes with larger wholes (see section C).

Gradually the structure, as revised, will include codes that are "larger" (more conceptually inclusive) and "smaller" (more differentiated instances), but it will need

Figure 4.4
Illustration of a Poorly Structured List of Codes (excerpt)

Actors		*Planning/Implementation Processes*		*Aspects of New School*	
Formal		201	commitment	301	boundary maintenance
101	administrator	202	complexity management	359	budget (district)
127	advisory group	203	conflict management	302	budget (school)
128	architect	204	constituency development	303	collective sentiments
102	Board (central)	205	cooptation	360	community control
129	Board (district)	253	decision-making	304	communication (formal)
130	builder	207	designing	305	communication (informal)
131	chairperson	208	energy depletion	306	conflict management
103	citizen	209	energy mobilization	307	curriculum
132	community liaison	210	energy overinvestment	308	data collection/feedback
...		211	future-envisioning	309	discipline
118	principal (focal)	212	goal clarification (stud. outcome)	310	departmentalization
119	principal (other)	213	goal clarification (sys. prop.)	311	equipment
130	researcher (other)	214	goal clarification (benefits)	312	evaluation (of school)
120	researcher (SA)	215	goal succession	313	extracurricular activities
131	salesman	254	group-building	314	food service
121	specialist (central off.)	...		315	friendship grouping
122	specialist (school)	227	planning	316	goals (student outcomes)
123	supt. (central)	231	planning horizon	317	goals (system properties)
135	supt. (district)	228	planning/implementation linkage	318	goals (benefits)
124	student	229	planning model	319	governance
125	teacher	232	policy-making	320	group definitions
136	teaching team	233	power base-building	321	influence (informal)
126	union representative	234	power struggle	361	inter-organizational linkage
112	voluntary organization	237	recruitment	...	
		238	redesign	341	role definition
Informal*		239	reflexivity	342	rules
151	buffer	240	rehearsal	343	salaries
106	core group	257	research relationship	365	school-community linkage
107	core member	242	resource acquisition	366	security
152	evaluator	241	resource allocation	344	space use
113	implementer	243	resource identification	345	staff assignments
153	leader (socio-emotional)	258	role accumulation	346	staff characteristics
154	leader (task)	235	role confusion	347	staff selection
155	linker	236	role strain	367	staff utilization
156	mass media	246	start-up	353	status/prestige
157	opinion leader	260	task behavior	368	student characteristics
117	planner	247	thoroughness	369	student grade levels
				...	
				357	transportation
				372	zoning

AFTER CODING, LOOK AT PROCESSES AND ASPECTS LISTS, AND PUT (*) BY THE MOST IMPORTANT KEY WORDS (MAXIMUM = 6).

*For actors actually present at the contact, circle the number. If an actor was not present, but is *discussed* in the contact, put parentheses around the key word.

to maintain a relational structure. An operative coding scheme is not a catalogue of disjointed descriptors or a set of logically related units and subunits, but rather a conceptual web, including larger meanings and their constitutive characteristics. Some software is especially helpful in displaying the structure of coding schemes, either in hierarchical form (NUDIST, Kwalitan) or in a network (ATLAS/ti, SemNet). See the Appendix.

Definitions of codes. Whether codes are prespecified or developed along the way, clear operational definitions are indispensable, so they can be applied consistently by a single researcher over time and multiple researchers will be thinking about the same phenomena as they code. A first-level code is usually a single term—*incentives, institutionalization, linkage*—that can easily suggest different meanings to different analysts. Because codes will drive the retrieval and organization of the data for analysis, they must be precise and their meaning shared among analysts. Defining them helps on both counts.

Figure 4.5 is excerpted from the list of definitions for the codes shown in Table 4.1. Such definitions will be improved/focused further as the study proceeds. However, a *conceptual structure,* whether prespecified or evolving, must underlie the definitions. In the new schools study, we developed clear, operational, and reliably usable definitions. But the codes as a whole had little intellectual shape and thus proved unfruitful.

Figure 4.5
Definitions of Selected Codes From Table 4.1 (excerpt)

Site Dynamics and Transformations–TR

Event chronology–official version: TR-CHRON/PUB	Event chronology during initial and ongoing implementation, as recounted by users, administrators or other respondents.
Event chronology–subterranean version: TR-CHRON/PRIV	Event chronology during initial or ongoing implementation, as recounted by users, administrators or other respondents, and suggesting (a) a consensual but different scenario than the public version or (b) varying accounts of the same events.
Changes in innovation: TR-INMOD	Reported modifications in components of the new practice or program, on the part of teachers and administrators, during initial and ongoing implementation.
Effects on organizational practices: TR-ORG/PRAC	Indices of impact of new practice or program on: (a) intraorganizational planning, monitoring, and daily working arrangements (e.g., staffing, scheduling, use of resources, communication among staff) and (b) interorganizational practices (e.g., relationships with district office, school board, community, and parent groups).
Effects on classroom practice: TR-CLASS	Indices of impact of new practice or program on regular or routine classroom practices (instructional planning and management).
Effects on user constructs: TR-HEAD	Indices of effects of new practice or program on teacher and administrator perceptions, attitudes, motives, assumptions or theories of instruction, learning, or management (e.g., professional self-image, revised notions of what determines achievement or efficiency, other attitudes toward pupils, colleagues, other staff members, stance toward other innovative practices).

Naming codes. Give a code a name that is closest to the concept it is describing. If you have the term *motivation,* the code should be MOT and not, for example, AIM or INC (for *incentive*). And do not use numbers—162, 29, or 29A. The rationale is that the analyst must be able to get back to the original *concept* as quickly as possible, without having to translate the code into the concept. It is also important that a second reader—a colleague, teammate, or secondary analyst—be able to do the same.

Check-coding. Definitions become sharper when two researchers code the same data set and discuss their initial difficulties. A disagreement shows that a definition has to be expanded or otherwise amended. Time spent on this task is not hair-splitting casuistry, but reaps real rewards by bringing you to an unequivocal, common vision of what the codes mean and which blocks of data best fit which code.[9]

Check-coding not only aids definitional clarity but also is a good reliability check. Do two coders working separately agree on how big a codable block of data is? And do they use roughly the same codes for the same blocks of data? If not, they are headed for different analyses. Or the cross-case analyst working from field notes coded idiosyncratically by different researchers will soon be going crazy.

The best advice here is for more than one person to code, separately, 5-10 pages of the first set of transcribed field notes and then to review each rendition together. At first, you usually don't get better than 70% intercoder reliability, using this formula:

$$\text{reliability} = \frac{\text{number of agreements}}{\text{total number of agreements + disagreements}}$$

Each coder will have preferences—sociologists see organization-level codes for the same blocks of data that are intrapersonal for psychologists and interpersonal for social psychologists—and each vision is usually legitimate, especially for inferential codes. Clarifying these differences is also useful; each analyst tends to be more ecumenical during later coding for having assimilated a colleague's rival vision of data that initially looked codable in only one way.

Similarly, each coder is well advised to check-code the first dozen pages of field notes, once right away and again (on an uncoded copy) a few days later. How good is the internal consistency? We should look here for higher initial code-recode reliability—something closer to 80%—than was the case for between-coder agreements. Eventually both intra- and intercoder agreement should be up in the 90% range, depending on the size and range of the coding scheme.

Finally check-coding about two-thirds of the way through the study is also beneficial. You may drift into an idiosyncratic sense of what the codes—revised or original—mean in light of new, often unshared insights. And the more interpretive codes, up one notch from the literal sense of the text, are more slippery when more than one researcher is scanning for them.

Levels of detail. How fine should coding be? That depends on the study. Some linguistic analyses require line-by-line or even word-by-word coding. More typically, codes get applied to larger units—sentences, monothematic "chunks" of sentences, or paragraphs in the written-up field notes.

The important thing is that the researcher be reasonably clear about what constitutes a unit of analysis. In our work B.C. (before computers), we usually defined the unit of analysis as a sentence or multisentence chunk and used the following rule of thumb: Assign the single most appropriate ("better," more encompassing) code among those related to a given research question.

Any block of data—a clause, sentence, or paragraph—is usually a candidate for more than one code. But if you are coding manually and the margin gets piled up with multiple codes for each block, you are in for heavy sledding when the notes are reviewed for site-level analysis.

This problem is not critical when computer retrieval is used (Becker, Gordon, & LeBailly, 1984). In fact, multiple coding is actually useful in exploratory studies. Some programs, such as The Ethnograph, use line-by-line coding to permit "nesting" and "overlapping" codes, rather than keeping chunk boundaries fixed. Others, such as ATLAS/ti, let you highlight and code material of varying sizes.

A good case can be made for multiple-coding segments with both a descriptive and inferential code; these are legitimately two necessary levels of analysis. But keep in mind that inferential codes need not be exhaustive. The analyst is looking for good explanatory exemplars, not for all instances.

The codes themselves shouldn't be overbuilt. In another early study, we experimented with multiple-facet coding. For example, if a segment showed administrators providing problem-solving assistance to teachers who found that help useful, the code was AD/PS-T/+. This method made the codes difficult to scan and absolutely maddening to cluster during final write-ups. Two or three facets in a single code seem to be as much as the brain can process efficiently. But here again, note that software such as NUD-IST, ATLAS/ti, Orbis, or Sonar Professional can easily sort such coded chunks. (They help clarify the hierarchical structure of a coding scheme, if that is the way you have built it.)

Finally not every piece of the notes must be coded. And, if done carefully, coding of later material can be more sparing. Field notes usually contain much dross—material unrelated to the research questions, either prespecified or emerging. There are such things as trivial, useless data. The aim is to keep the dross rate down, but it will never be zero. And it is true enough that some dross will appear, later on, to be of real value.

When to code. This is important. Some analysts reserve coding for the end of data collection. We think that is a serious mistake, because late coding enfeebles the analysis. Coding is not just something you do to "get the data ready" for analysis, but, as we have said several times, something that drives ongoing data collection. It is a form of early (and continuing) analysis. It typically leads to a reshaping of your perspective and of your instrumentation for the next pass. At the same time, ongoing coding uncovers real or potential sources of bias, and surfaces incomplete or equivocal data that can be clarified next time out.

Furthermore, coding is hard, obsessive work. It is not nearly as much fun as getting more good stuff in the field. Trying to do the coding all at one time tempts the researcher to get sloppy, resentful, tired, and partial. This attitude damages the robustness of the data and the quality of the analysis.

One simple rule of thumb here: Always code the previous set of field notes before the next trip to the site. Always—no matter how good the excuses for not doing it. This rule can be threatened by backlogs and slow turnaround in typing write-ups, but ongoing coding is the right plan.

Perhaps the more important point is this: The ultimate power of field research lies in the researcher's emerging map of what is happening and why. So any method that will force more differentiation and integration of that map, while remaining flexible, is a good idea. Coding, working through iterative cycles of induction and deduction to power the analysis, can accomplish these goals.

Advice

To summarize: Codes are efficient data-labeling and data-retrieval devices. They empower and speed up analysis.

Creating a start list of codes prior to fieldwork is helpful; it forces the analyst to tie research questions or conceptual interests directly to the data. But the analyst should be ready to redefine or discard codes when they look inapplicable, overbuilt, empirically ill-fitting, or overly abstract.

You also can work more inductively by waiting for the field notes to suggest more empirically driven labels. You should not, however, wait too long, or move capriciously from one coding scheme to another.

Make sure all of the codes fit into a structure, that they relate to or are distinct from others in meaningful, study-important ways. Do not casually add, remove, or reconfigure codes.

Keep the codes semantically close to the terms they represent. Do not use numbers as codes.

Have all of the codes on a single sheet (or computer display) for easy reference.

Define the codes operationally. Be sure all analysts understand the definitions and can identify, quickly and easily, a segment fitting the definition. Check-coding the same transcripts is very useful for the lone researcher (get code-recode consistencies over 90% before going on) and is essential for studies with more than one researcher. Do this once later in the study, too. Don't assume consensus.

For starters, use a single code for a segment. Multiple coding is warranted if a segment is both descriptively and inferentially meaningful, and it's helpful if you have the right software to manage it.

Coding should not be put off to the end of data gathering. Qualitative research depends heavily on ongoing analysis, and coding is a good device for supporting that analysis.

Remember: Codes are category labels, but they are not a filing system. Every project needs a systematic way to store coded field data and a way to retrieve them easily during analysis. Notebooks, file folders, and index cards are the time-honored means. But, as we note in Chapter 3, section B, good computer software is far ahead of them when it comes to data management. See the Appendix for ideas on choosing software.

Time Required

The time required for generating initial codes and definitions depends, of course, on how many you start with, and on the clarity of the conceptual framework and research questions. For the start list and definitions shown in Table 4.1 and Figure 4.5, the first cut took 2 full days—1 for the codes and 1 for the definitions. Revisions and completions added another 2 days.

Coding itself takes varying amounts of time, depending on the code's conceptual structure and complexity, the quality of the field notes, and the coder's skill. Here is some typical arithmetic from our experience.

A single-spaced page of transcribed field notes has about 50 lines. It might contain 5-10 codes. A 2-day field trip usually generates 40-80 such pages, even when the researcher is disciplined. Coding each page might run about 5-10 minutes, once the codes are familiar; in the beginning, you should count 10-15 minutes. So "inexperienced" coding of a 2-day data set might take up to 3 days; later it could be done in a day and a half or so. Taking more than 2 coding days per contact day is a signal of too-fine units of analysis, too much multiple coding, too many codes, or a weak conceptual structure.

Coding is tiring. It often feels longer than it really is. It helps to intersperse coding with written-in marginal remarks (see Box 4.3) of an active, musing sort, rather than just dully plowing ahead. Breaking from time to time to do other related work—such as arguing with other coders, writing memos (see section D), or jotting down notes about what to look for in the next field visit—also helps.

Reflective Remarks

Raw field notes (the scribbles and jottings that enter your notebook as you are watching a situation or talking with someone) must, as we have pointed out, be converted into a write-up, a transcription that is legible to any reader.

The temptation during a write-up is to slog along, converting raw notes into a coherent account. But that method misses an important resource: the field-worker's reflections and commentary on issues that emerge during the process.

As a write-up is being produced, reflections of several sorts typically swim into awareness. For example:

- what the relationship with informants feels like, now that you are off the site
- second thoughts on the meaning of what a key informant was "really" saying during an exchange that seemed somehow important
- doubts about the quality of some of the data; second thoughts about some of the interview questions and observation protocols
- a new hypothesis that might explain some puzzling observations
- a mental note to pursue an issue further in the next contact
- cross-allusions to material in another part of the data set
- personal reactions to some informants' remarks or actions
- elaboration or clarification of a prior incident or event that now seems of possible significance

When something like any of these examples arises, it's useful to enter it directly into the write-up. It may or may not be fodder for a deeper reflection (see Box 4.3) or a more extended memo (section D). A good convention is to mark off the remark with double parentheses to signal that it is of a different order from the data it comments on. The material in Box 4.2 gives some examples. Remarks such as these add substantial meaning to the write-up, not least for other readers. And they usually strengthen coding, in pointing to deeper or underlying issues that deserve analytic attention.

Reflective remarks can also be added while you are jotting down raw field notes; that technique improves the usefulness of field notes considerably. You are simultaneously aware of events in the site, and of your own feelings, reactions, insights, and interpretations (cf. Patton, 1990). For a suggestive typology, see Bogdan and Biklen (1992), who divide reflective remarks into those on analysis, method, ethical dilemmas, own frame of mind, and points of clarification. Some researchers use "O.C." (observer comment) instead of double parentheses to indicate reflective remarks.

Marginal Remarks

Coding, as we have noted, can become tedious if you treat yourself as a sort of machine scanning the page methodically, picking out small segments of data and assign-

Box 4.2
Reflective Remarks: Illustrations

Mike joked, "Maybe I could go and act like a senior." He made a dumb monkey face as he was speaking. ((This staff does not seem to put down students, really, but they cannot resist occasional jokes of this sort—more of this later.))

. . . .

Jim indicated that they had unofficially done their own analysis of attendance data and said, "I'm sure it's been effective," (that is, CARED in increasing attendance rates). ((This sounded pretty soft and vague to me.))

. . . .

John went on to explain that during the second semester he would be doing pretty much the same, that is, "nothing much at all." ((This denial of his activity was later picked up informally by Jim, I think, in a conversation with me. It was, in fact, a sort of minimization, and perhaps a deflection from the fact that he was away a great deal of the time, when presumably, he might be able to be helpful with issues in the program itself.))

ing categorical labels to them. The sensation of being bored is usually a signal that you have ceased to think. One way of retaining mindfulness in coding is the marginal remark. It's analogous to the reflective remark.

As coding proceeds, if you are being alert about what you are doing, ideas and reactions to the meaning of what you are seeing will well up steadily. These ideas are important; they suggest new interpretations, leads, connections with other parts of the data, and they usually point toward questions and issues to look into during the next wave of data collection, and to ways of elaborating some of these ideas (see next section).

If your codes appear in the left margin, it is useful to put preanalytic remarks of all sorts in the right margin. Box 4.3 presents a few examples. Marginal remarks, like reflective remarks, add meaning and clarity to field notes. They also point to important issues that a given code may be missing or blurring, suggesting revisions in the coding scheme.

Marginal notes also can be made on transcripts by critical friends or co-researchers. They query segments of field notes, often in order to suggest new leads, a deeper analysis, or alternative ways of looking at the data. In the excerpt shown in Figure 4.6, the reader asks a series of questions, most of them about local meanings and typologies.

Note that some of the marginal notes point the researcher toward new or more differentiated codes: definitions of responsibility, "dangers," typology of dead bodies.

Box 4.3
Marginal Remarks: Illustrations

IC-NORM-DVR CL-PATT	Jim looked past me for a minute and asked Dawn, the aide, to go out and check the kids in the hall. I asked what that was about and he explained that "we don't release them until the bell," and that some students had already stepped outside in the hall even though the bell for the next period had not yet rung.	Control Concerns.
CL-PATT	. . . Largely the group free-associated from topic to topic. Certain themes recurred from time to time, including Mary's tendency to reassure the other two that the community relations work would be easy for them to carry out, and her steady advice not to give "professional" tasks to the aides. There was a good deal of advice-seeking from Mary. Finally, there did not seem to be any very specific planning or decision making about particular procedures or how things would occur in detail.	Reassurance. Conflict over aide role "Organic Coping"
CL-PATT	. . . ((I think it was along through here that Mary said for the first of several times that "I'll still be available," implying that she wasn't really going to go away, this wasn't irrevocable, she would still be accessible.))	Reassurance theme
CL-PATT	. . . John mentioned that there would be more girls than boys and suggested "should we go through the first day?" as if to launch discussion of how the program would start up. But this did not get a very direct response.	Organic Coping theme

Figure 4.6

A Colleague's Marginal Notes: Excerpt (Bogdan & Easterday, n. d.)

An Important Contact 7

Lois M. Easterday
Field Notes #10
Observed 10/20/74
8:00 a.m. to 11:15 a.m.
Recorded 1:00 p.m.

As in the last observation, I decided to call early in the morning to see if there was going to be an autopsy or something. I decided to call about 8:00 a.m.

Why are autopsies important to you? Are they equally important to the setting, i.e. are days structured around them?

What is the schedule for a typical day

"Medical Examiner's Office."

What does this title and introduction tell you about how they define their work

"Hello. Is this Bill?"
"Yes it is."
"Hi, Bill, this is Lois."
"Oh, hi."
"Is anything going on down there today?"
"Yea, they're gonna do a post in a few minutes."

How exclusive is the use of that term to this setting? What exactly does it mean?

What does "going on" mean to the people in the setting?

"Oh. Is it okey if I come down?"
"Sure. Come on down."
"Ok, I'll see you in a few minutes."
"ok, goodbye."

When I arrived at the M.E.'s office, Bill was seated in the reception area.
"Hi, Bill. How are you? Are you feeling better now?"
"What? (He kind of laughed) Oh, yea."
"Have they started the post already?"
"Yea, they just went back. Go ahead."
"Can I leave my coat here?"
"Sure just put it on the coatrack in the next room."
"Does the doctor know I am coming?"
"Yea, I told him."
"Is it Ok/"
"Sure. He said fine, come on in."
"Well I guess I'll go back then. See you later."
"See you."

Do the other workers in the setting show as much deference to him as you do?

What things do they consult the doctor about and what don't they consult him about

Definitions of responsibility

When I walked into the autopsy room, I immediately noticed that they had already started, as they had the chest cut and the skin pulled back. Chuck, George, and Dr. Ostman were there. They all said, "Hi."

Why didn't you describe the corpse here?

Field Notes #10 9

Page 3

Dr.: "Let's get some blood. Pull the heart up a little." He took a syringe and extracted some blood. He put it into a bottle George was holding. He did it again. This time the syringe slipped and blood went on George's apron and glove.

Notice the division of labor.

Dr.: "Oh, I'm sorry. Did I get you?"
George: "Oh, it's ok."
Dr.: "No, if it went through your glove, you'd better take it off and put something on your hand."
Chuck: "You'd better put some formaldehyde on it."
(OC: This reminded me of the conversations we had had previously about how persons sometimes have diseases and you can still catch them. I also remembered that the doctor said he had had some sort of fever. I thought, "Boy, I hope we don't all catch something and die. That would be too much. That would be above and beyond for PO class.")

Is this job defined as dangerous? If so, how are the "dangers" controlled? What are the dangers?

The doctor probed around some more and tried to get more blood. They put some more in that bottle and some in another bottle. Then they could not get anymore.
Dr.: "Ok, let's tie off the vessels and look at the aorta."
They did this. When the doctor probed around in the aorta, he pulled out a large clot.
Chuck: "That's it. That's what killed him."
Dr.: "No, it didn't. That's just an after death chicken fat clot."
Chuck: "Oh, I thought we had it."

What was the tone of voice? Was the doctor pulling rank or what?

Does a nurse feel as "we" during an operation? Does the Dr. use "we" also?

The doctor continued to probe around and while he did this, Chuck moved over beside me and said in a low voice, "Boy, you shoulda been here last week. We had a beaut."
"Oh, yea?"
"Yea, maybe you heard about him. Our friend back here. The jolly green giant. They pulled him out of the river."

What is a "beaut"? One of the words in their typology of bodies.

Under what circumstances do they make up names?

"Oh, yea. I heard about that. That's the guy they couldn't identify, huh? Do you still have him here?"
"We sure do."

Important part of the work? "Identified" and "unidentified" are these important categories?

The doctor pulled out the lungs. He cut one and handed it to George to weigh. "Here is the right lung." He handed him the other lung. "Here is the left lung."

An external review of field notes by a knowledgeable critic is invariably useful.

Marginal notes can also be used later in the coding cycle. For example, after assembling chunks of field notes (original or condensed) that share a common code, new marginal notes and underlines may be added. Chesler (1987) used this simple method in an interview study of the dangers of self-help groups. Summarized data segments relating to possible "dangers" were assembled, and the key clauses were underlined. These key phrases were boiled down still further in the margin. Figure 4.15 (p. 88) presents how the method works.

Although Chesler did not use it in this way, the marginal phrases could be referred back to the coding scheme itself to develop or test subcategories. The phrases could also be tried out on some data sets not yet examined, to see whether they recur among the same or other informants. And the next wave of data collection could use the list as a series of probes during interviews and observations.

Some computer programs, such as Metamorph and ATLAS/ti, permit you to write "annotations" (really, marginal notes) and attach them to particular chunks.

C. Pattern Coding

Analysis Problem

Given a working set of codes that describe the phenomena in transcribed field notes, how can the researcher move to a second level—one that is more general, perhaps more explanatory? Just naming or classifying what is out there is usually not enough. We need to understand the patterns, the recurrences, the plausible whys. As Kaplan (1964) remarks, the bedrock of inquiry is the researcher's quest for "repeatable regularities."

Brief Description

Pattern codes are explanatory or inferential codes, ones that identify an emergent theme, configuration, or explanation. They pull together a lot of material into more meaningful and parsimonious units of analysis. They are a sort of meta-code.

First-level coding is a device for summarizing segments of data. *Pattern coding* is a way of grouping those summaries into a smaller number of sets, themes, or constructs. For qualitative researchers, it's an analogue to the cluster-analytic and factor-analytic devices used in statistical analysis. The quantitative researcher works with sets of variables that either put people into distinct families built around what they do or say (Q analysis) or, alternatively, cluster such actions and perceptions across informants (R analysis).[10]

For the qualitative analyst, pattern coding has four important functions:

1. It reduces large amounts of data into a smaller number of analytic units.
2. It gets the researcher into analysis during data collection, so that later fieldwork can be more focused.
3. It helps the researcher elaborate a cognitive map, an evolving, more integrated schema for understanding local incidents and interactions.
4. For multicase studies, it lays the groundwork for cross-case analysis by surfacing common themes and directional processes.

Illustrations

These four functions can be clarified as we discuss how pattern codes are generated, what they look like, and what the field researcher does with them in the course of data collection.

Generating pattern codes. This is easy—sometimes too easy. As in everyday life, the researcher needs to reduce and channel data into a small number of concepts that can be mentally encoded, stored, and readily retrieved.

During initial fieldwork the researcher is looking for threads that tie together bits of data. For example, if two or three informants say independently that they resent a decision made by their boss, we may be on to several different phenomena—a conflict, an organizational climate factor, or a disgruntled subgroup of employees. Any of these interpretations involves chunking and sorting data (function 1, above). For starters, is there anything else in common between these informants or in the grounds given for resenting the decision? Is there a different or opposing semantic content among informants who are *not* resentful?

These first bits of data and the review of the coded segments being pulled together are leads; they suggest important variables to check out—factors that may account for other local perceptions and behaviors (function 2, above). Seeing the "resentment" data in any of these alternative ways also helps the researcher make sense of puzzling or surprising observations. These several bits come together into an initial plot of the terrain (function 3). Finally, if a colleague in a multicase study comes across a similar profile of resentment or, alternatively, finds no resentment of decisions at all in a place otherwise similar to the more "resentful" case, we have the first threads of cross-case comparisons (function 4).

The danger is getting locked too quickly into naming a pattern, assuming you understand it, and then thrusting the name onto data that fit it only poorly. Premature analytic closure is hard to shake, in part, because the analyst often is not aware of what is happening, or the new data collected to verify the pattern are being sought out selectively. But a second analyst, reading over the field notes, usually picks this up. Patterning happens quickly because it is the way we habitually process information.[11]

The trick here is to work with loosely held chunks of meaning, to be ready to unfreeze and reconfigure them as the data shape up otherwise, to subject the most compelling themes to merciless cross-checking, and to lay aside the more tenuous ones until other informants and observations give them better empirical grounding.

Sometimes, however, the data just don't seem to toss up any overarching themes; each code looks almost distinctive. In these cases it helps to go back to the research questions just to remind yourself of what was important and then to review the chunks bearing those codes.

In a more inductive study, it helps to look for recurring phrases or common threads in informants' accounts or, alternatively, for internal differences that you or informants have noted. Typically those differences will bring forth a higher-level commonality. For example, Heritage (1988), in a study of responses to invitations, reviewed the reasons given for accepting and rejecting them, and looked for a code that could encompass both. Such a code, of course, is conceptual in nature (e.g., people accept or reject invitations because of status, person perception, social proximity) and, as such, will pack more explanatory power.

What pattern codes look like. Pattern codes usually turn around four, often interrelated, summarizers: themes, causes/explanations, relationships among people, and more theoretical constructs. Here are some concrete examples from a recent study, with codes we assigned in capital letters.

Themes:

PATT (pattern): All supervisors seem to be using benevolent, fatherly terms when talking about employees ("my" staff, "my" people, "my" junior guys), but employees use mostly bureaucratic, regulation-type terms ("the office . . . upstairs," "the management").

RULE: You don't talk earnestly about your problems or your successes in the staff lounge.

PATT/OS (theme appearing in other sites, as well as in this one): It seems easier to get new projects adopted among lower-class students or in vocational tracks.

Causes/Explanations:

EXPL: Multiple role of the "helping teacher" seems to be an important ingredient of success.

SITE-EXPL (informants' explanations): The best projects are ones that put together the best practitioners' recipes.

MET (metaphor): The idea of career "trajectories"—people are using these projects to get away from some jobs and places to other ones.

Relationships Among People:

NET (social network): The money-and-support club: A. Becker, P. Harrison, V. Wales.

Emerging Constructs:

BARG: Negotiating or bargaining, but implicitly, seems to be the way decisions get made; a conflict model is a more plausible account of how decisions are made than a rational-technological model.

Such themes or configurations are various: repeatedly observed behaviors, norms, relationships; local meanings and explanations; common sense explanations and more conceptual ones; inferential clusters and "metaphorical" ones; single-case and cross-case. Almost anything is grist for a pattern code.

Using pattern codes in analysis. Pattern codes may be used in at least three ways. First, they are added in tentative form to the list of codes and are tried out on the next set of transcribed field notes or documents to see whether they fit. (Lincoln and Guba, 1985, call this "discriminant sampling.")

It may be useful at that point to "map" the codes—that is, to lay out the component codes that got you the theme—along with segments from the field notes. It helps to do it visually, in a network display, seeing how the components interconnect (for some help here, see Bliss et al., 1983). The mapping is a new take on your conceptual framework.

Although it is not hard to do this by hand, mapping by computer has some powerful advantages. ATLAS/ti does this well. Here, too, is an example (Carley, 1993) using MECA, software that extracts concepts and the relationships between them. The program also allows for judgments of relationship strength, valence (positive or negative), and directionality. Figure 4.7 is an excerpted map for one case (Joe) on dimensions that appear in statements about a far larger number of people:

Joe is defined as a gnerd, we see, because he studies in the library, does not fit in (negative relationship), is not friendly, works too hard, is not accessible, and does not interact with other students.

Next, the most promising codes to emerge from this exercise are written up in the form of a memo (see next section) that expands on the significance of the code. This process helps the writer become less fuzzy about the theme

Figure 4.7
Mapping of Concepts:
Illustration (Carley, 1993)

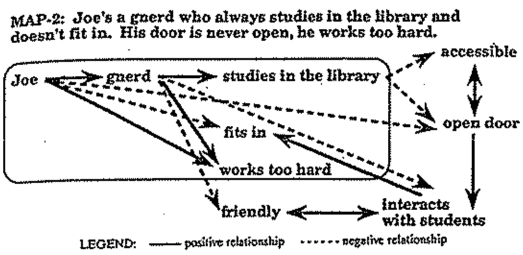
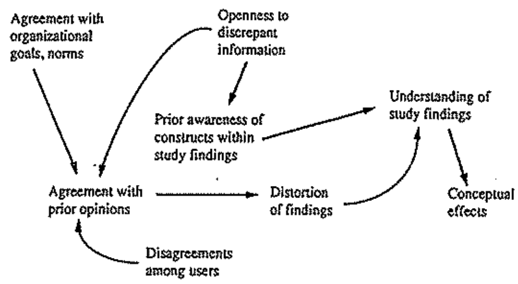


Figure 4.8
Interrelation of Pattern Codes
(Huberman & Gather-Thurler, 1991)



or construct, helps other researchers think summatively about their own data set, and gets cross-case analytic energy flowing.

Finally, pattern codes get checked out in the next wave of data collection. This is largely an inferential process. The analyst tries out the theme on a new informant or during an observation in a similar setting, engages in **if-then** tactics, as discussed in Chapter 10 (if the pattern holds, other things will happen or won't happen), or checks out a **rival explanation.**

(The **boldface** terms refer to specific tactics of drawing and verifying conclusions, which are discussed in detail in Chapter 10. We use this convention as a way of pointing to tactics as they occur in context.)

Usually a pattern code does not get discounted, but rather gets *qualified:* The conditions under which it holds are specified. For example, the rule "No earnest talk in the lounge" can be bent in cases of conflict, crisis, or socializing of new members. This clarification points to more precise ways of verifying the pattern, and strengthens its external validity. What we are doing is both generalizing and specifying: getting more explanatory power, while better defining the parameters within which that explanation will hold.

The idea of displaying core codes and subcodes on a single sheet is useful as a coding structure evolves. Here is an example drawn from Huberman and Gather-Thurler (1991). The authors were tracing the flow of research knowledge about vocational training and counseling into schools, small firms, and factories (22 cases). One issue was the *degree of understanding* people had of research findings. Which factors *influenced* that understanding? The researchers had a set of first-level codes and some emerging pattern codes, such as "distortion of research findings" and "disagreement among users."

For the analytic work, each of the four researchers reviewed one case and drew the pattern codes in network form, as in Figure 4.8.

The main thread here is the effects of institutional pressures (agreement with organizational goals and norms, disagreements among users) on individual responses to the research knowledge. These affect levels of understanding and, from there, the impact of the research findings on respondents' knowledge base and conceptualizations.

But this display was not meant to achieve closure. It was the prelude to further analysis:

- Each researcher went back to one case's field notes to assess whether the inferential chain held. If so, where were the supporting data?
- Then the researcher wrote a memo briefly discussing each variable and its influence in the case at hand and showing the connections between codes—both on and off the original figure.
- For each case, the researcher drew a figure analogous to Figure 4.8. Then the figures were overlaid to see where they were similar and different.
- The figures were taken to the field for testing and revision, followed by a second series of memos.

The researchers were near the end of their fieldwork and were working with relatively few pattern codes on a major research question. Such a funneling procedure is most rewarding during final within-case and cross-analysis report writing. In effect, the work is nearly done.

Variations

If a general pattern code (such as RULE) is being used a good deal, it is helpful to create subcodes that explain the content and enable easy retrieval:

RULE—INF Rules about informant behavior

RULE—PUB Rules about behavior in public settings

RULE—WORK Rules that specify how formal work tasks are to be carried out

Stay open to the idea of inventing new types of pattern codes. For example, we developed the pattern code QU!, meaning a query about something surprising that happened in the case. Being surprised is an important event in fieldwork, and we wanted to track it in our notes.

Finally, pattern coding can lead usefully to recoding or multiple coding of earlier data. When we come up with a construct such as the one we coded BARG ("bargaining in conflict mode") above, we need to retrieve and analyze its component parts—the segments from which we derived the pattern. We also need to preserve some conceptual coherence in the coding scheme as the study evolves. With a computer search-and-replace command, recoding or multiple-coding can be done easily. Some software (e.g., HyperRESEARCH, The Ethnograph, NUDIST) can be used (by working in a copy of the database) to allow two coding structures to be maintained until the analyst sees more clearly.

Advice

Pattern coding is crucial for getting to the next level above (or below) the immediate ebb and flow of events in the case. It should be done regularly as the initial sets of first-level codes are being applied.

Don't try to force the use of pattern codes; don't assume that, because they are a meta-level code, they can in principle be applied to *every* bit of data that already has a first-level code.

How many pattern codes, and when to generate them? This is largely a matter of analytic style. Some analysts are unregenerate pattern coders; others are more cautious. Some prefer to generate pattern codes very early and then check them out and qualify them; others are more resolutely inductive and wait until enough data accumulate to support a construct unequivocally. The important point is that pattern codes are hunches: Some pan out, but many do not.

Pattern coding is an intellectually pleasurable process. Those codes that survive the onslaught of several passes at the case and several attempts to disqualify them often turn out to be the conceptual hooks on which the analyst hangs the meatiest part of the analysis.

Time Required

Developing and applying pattern codes is an integral part of first-level coding; the activities are concurrent. Early on, doing pattern coding might occupy only 5-10% of total coding time; later, somewhat more, as the analyst becomes more and more preoccupied with making sense of the data.

D. Memoing

Analysis Problem

Fieldwork is so fascinating, and coding usually so energy absorbing, that you can get overwhelmed with the flood of particulars—the poignant remark, the appealing personality of a key informant, the telling picture on the hallway bulletin board, the gossip after a key meeting. You find it nearly impossible to step back, to make deeper and more conceptually coherent sense of what is happening. Reflective remarks, marginal remarks, and pattern coding are all a step away from the immediate toward the more general. But how do you take this step?

Brief Description

Glaser's (1978) definition is the classic one: "[A memo is] the theorizing write-up of ideas about codes and their relationships as they strike the analyst while coding. . . . it can be a sentence, a paragraph or a few pages . . . it exhausts the analyst's momentary ideation based on data with perhaps a little conceptual elaboration" (p. 83-84).

Memos are primarily conceptual in intent. They don't just report data; they tie together different pieces of data into a recognizable cluster, often to show that those data are instances of a general concept. Memos can also go well beyond codes and their relationships to any aspect of the study—personal, methodological, and substantive. They are one of the most useful and powerful sense-making tools at hand.

You are writing memos to *yourself*, secondarily to colleagues.

Illustrations

Here are some memos written during the school improvement study, showing different facets of memoing. We'll comment on them as we go.

The first memo, A, responded to a colleague's earlier memo suggesting the "welcoming structures" concept and adapting it from the field of cognitive psychology. Notice that the memo writer (a) aims at clarifying the idea, (b) ties it to information from a case, and (c) differentiates the idea from already existing codes.

A. On "welcoming structures" (March 6)

Your idea of a durable structure (typically combined with learned skills, procedures, etc.) at the organizational level which would facilitate adoption of innovations is a useful one, I think. We should be looking for it. In Perry-Parkdale there are so many Gov't programs that the concept clearly exists at the district level, at least for attracting money. At the building level there is prior experience with work experience programs, demonstration programs, etc.

The sheer ORG-FIT concept (good fit between the innovation and organizational properties) is not it; that only implies some sort of congruence. What we are talking about here is more of an active capability. I suggest a label that would recognize that what is at issue here is not merely a structure or mechanism, but working procedures, flows, and the associated skills and techniques. The cue you get is when you notice that people are in effect telling you that "we know how to handle these things."

Memo B, below, is an example of a "place-holding" memo—always a useful idea when an idea strikes.

B. Memo: Comparison processes (March 19)

Introducing a new program inevitably induces a comparison process, notably comparison-for-alternatives (see FACILE and SCORE-ON cases). Just wanted to hold a place for this idea—more to come.

Memo C, below, is a more integrative discussion, pulling together data from several cases and reformulating them around the issue of career patterns. It came partway through data collection, setting up a major addition to the study's outcome measures. In the last paragraph, the author deals with specific means of data collection on the topic.

C. Memo: Career patterns (22 February)

In a general sense, people are riding the innovations while in a state of transition; they are on their way from somewhere to somewhere via the project.... Where could people be going? They could be going:

- *up:* from a classroom to a supervisory or administrative role or to a higher administration slot. Innovations are a lot faster than waiting for someone else to move on or going back for a degree. They get you visibility and good positioning. If it gets institutionalized, you get institutionalized with it in the new role. Also, they're less brutal than getting promotions by doing in the person above you and more convenient than having to move away to move up.
- *away:* from teaching by easing into a part-time or more flexible job. These projects tend to be marginal, loosely administered (although Tindale is the contrary), transition-easing. They also can allow for permutations, as in the up-and-away pattern that C. may be following at Plummet.
- *in:* the remedial programs are less choosy about formal credentials. They provide access to civil services like education to people with weird backgrounds. Aides can get positioned to become certified teachers; people from business or the arts can come into a marginal or experimental universe and ease gradually into a more formal role incumbency....

It is especially worth keeping track, as we dictate and code, of where these people have come from and where they are, or think they are, on their way to. I suggest we ask each informant:

- a little more directly, why he/she is doing this, in terms of roles and role changes.
- what he/she expects to be doing in 2-3 years.
- if he/she has a sense of being in a transitional period.

Memos D and E, on "barometric events," illustrate how preexisting concepts can be useful in clarifying an idea. The memos tie the basic idea to case events and to the coding system. They also show the importance of using memos as a dialogue among staff.

D. "Barometric events" (19 March)

We can sometimes see a noticeable/identifiable change in the (weather) conditions of the system. It reminds me of Lewin's term "region of instability/uncertainty" or Redl's concept "dynamics of a focal event." The event has a future-shaping quality: thereafter, things won't be the same.... Or they lead to a new developmental stage. Events are preludes in prologues and fulfill a linking (transitional) function in time.

Key actors in the event provide a clue: What is the domain or subsystem affected by the event? For example, Ms. Spiess's seminar attendance (Banestown) was a sort of boundary activity, a crucial event linking the school system with external professional information sources, not to mention "inspiration."

E. Return memo on barometric events (April 4)

I think the idea is felicitous. It's very true that there is a sea change thereafter in several sectors of the subsystem. The codes AP-CRIT and TR-CRIT help to flag this. We can stay with a diachronic approach, even while doing some cross-sectional comparisons.

Memos should always be *dated,* entitled with *key concepts* being discussed, and *linked* to particular places in the field notes, to previous case analysis discussions, or to case summaries. Software that permits hypertext functions (e.g., HyperQual, MetaDesign, HyperRESEARCH, ATLAS/ti) can make such links easily. You can go to or from memos almost instantaneously.

Memos should be coded according to the concept(s) they are about, and kept separate from data files. As a study proceeds—especially a strongly inductive one—memos will accumulate and can themselves be sorted (or computer-searched) to build more comprehensive categories (see Glaser, 1978, p. 87).

Memoing helps the analyst move easily from empirical data to a conceptual level, refining and expanding codes further, developing key categories and showing their relationships, and building toward a more integrated understanding of events, processes, and interactions in the case.

Memoing is especially crucial when you are taking a strongly inductive approach, but it is equally important for other reasons when you begin with a preliminary framework. Without memoing, there is little opportunity to confront just how adequate the original framework is, and where it needs to be revised.

In the inductive approach, memos often serve a "clustering" function; they pull together incidents that appear to have commonalities. Here are some excerpts from Martin and Turner (1986):

> Based on incidents 3, 4, 5 and 7 . . . workers in adult services feel they are misperceived, not viewed with respect or as equals by service workers in other program areas. Mental health workers see them as "cab drivers" and people who are to be managed by them. This suggests that adult health workers *perceive* they are treated with less respect than they deserve. (p. 152)

Variations

Memos also can be written (a) on what is intensely puzzling or surprising about a case, (b) as alternative hypotheses in response to someone else's memo, (c) to propose a specific new pattern code, (d) to integrate a set of marginal or reflective remarks already made on field notes, (e) when the analyst does not have a clear concept in mind but is struggling to clarify one, and (f) around a general theme or **metaphor** that pulls together discrete observations (see Chapter 10, section A).

Different types of memos can be written for different phases of the study. For example, Lee, Kessling, and Melaragno (1981, p. B43) describe the use of "insight journals"—memos of a page or less, perhaps stored in a card file format—for the later stages of a study, when you are going specifically after cross-case comparability and synthesis. They can include cross-case patterns, "bright ideas," policy implications, and ideas for next-step syntheses. Insight journals can themselves be coded for later retrieval.

Advice

Here we draw on Glaser (1978) and Strauss and Corbin (1990). Our advice is an amalgam of their experience and ours.

1. Always give priority to memoing. When an idea strikes, STOP whatever else you are doing and write the memo. Your audience is *yourself*. Get it down; don't worry about prose elegance or even grammar. Include your musings of all sorts, even the fuzzy and foggy ones. Give yourself the freedom to think. Don't self-censor.

2. Memoing should begin as soon as the first field data start coming in, and usually should continue right up to production of the final report. Just as codes should stabilize reasonably well by one-half or two-thirds of the way through data collection, the ideas raised in memos usually will start settling down then or shortly afterward, as the analyst approaches what Glaser calls "saturation" (no significantly new explanations for data). Memoing contributes strongly to the development/revision of the coding system.

3. Keep memos "sortable." Caption them by basic concept, and mark or underline other concepts discussed during the text of the memo. Like coded data, memos can be stored and retrieved by using a wide variety of methods. Keep your codes by your side when doing the memo, especially the PATT-codes; some meaningful relationships may emerge that you had not seen at first.

4. Once again, memos are about *ideas*. Simply recounting data examples is not enough. By the same token, memos are not chiefly about people or events or interactions; these are all used as indicators for an analysis that is set in a conceptual frame.

5. Don't standardize memo formats or types, especially in a multiple-researcher study. Memoing styles are distinctive, and memo types are as various as the imagination can reach.

6. Memo writing is fun. And it often provides sharp, sunlit moments of clarity or insight—little conceptual epiphanies.

Time Required

Any given memo usually takes only a few minutes. Even one synthesizing a lot of data, such as Memo C in our examples, may not occupy more than a half hour. Memos are typically a rapid way of capturing thoughts that occur throughout data collection, data reduction, data display, conclusion drawing, conclusion testing, and final reporting.

Later in the study, however, memos can be more elaborate, especially when they piece together several strands of the data or look across multiple measures of a construct. Denzin (1978) has a good term for this: "inspection." The idea is to work with a finding that appears to have some range and power and to approach it from different angles, ask different questions of it, take it out of context and set it against a body of literature, put it back in its context and have another look, and see where it is incomplete and where it is well elaborated. As Denzin remarks, it is a bit like handling a strange physical object.

Developing Propositions

Memoing captures the thoughts of the analyst on the fly, so to speak, and is precious for that reason. As a study proceeds, there is a greater need to formalize and systematize the researcher's thinking into a coherent set of explanations. One way to do that is to generate propositions, or connected sets of statements, reflecting the findings and conclusions of the study.

A good illustration appears in Stearns, Greene, David, and associates (1980). They were studying 22 school sites implementing a new special education law. Five field-workers prepared case study reports for each site. But how to describe commonalities, note the differences, and develop explanations across all 22 sites?

Stearns et al. rightly assumed that much useful material of this sort resided in field-workers' heads, and took an ingenious approach to eliciting the material, clarifying it, synthesizing it, and verifying it. The process had seven steps, and was managed by three analysts who had themselves visited all of the sites.

1. Each field-worker made an unstructured list of statements he or she "would like to see in the final report." Example:

> Although teachers spend a lot of time doing individualized education plans for special education students, they don't find them all that useful on a daily basis.

Other statements were retrieved from documents and staff meeting notes; the total was 1,500, one statement to a card.

2. The 1,500 were reduced to 1,000 through sorting and reduction of duplication.

3. The more general, abstract statements were retained ($N = 250$), and specific instances were set aside for later use.

4. The 250 cards were sorted into "assumptions," "findings," and "conclusions" and were divided into 30 general categories. The analysts displayed the 30 sets of cards on the wall to see the interrelationships. Considering the information needs of their report audience as well, they developed a working outline for presenting their findings in the report.

5. Now for verification. The analysts developed from the 250 a draft list of propositions for field-worker review. Example:

> The greatest impact of the law at the school level has been to add new duties to old ones.

The propositions were listed in organized sequence under each of 21 headings.

6. Field-workers examined the proposition list, commenting on the validity of each and on needed qualifications or conditions, and noting "don't know" or "doesn't apply" when relevant.

7. The analysis staff wrote a "findings report" for each of the 21 topics, using only the sites where relevant and valid data were available. They checked back against a prior list of site characteristics for further explanations and added explanations that emerged during the final step.

The Stearns et al. illustration is a fine example of an inductive approach to propositions, with safeguards against premature and unwarranted closure. Although this was a large database (which could have been more easily managed with computer software), the approach can be used at any level, down to and within the individual case.

Another quick example: Kell (1990) did a multiple-case study of the effects of computers on classroom instruction. At the first analytic meeting, field researchers recorded their case-specific propositions on index cards, keyed to the research questions. The propositions then were clustered thematically, and evidence was sifted for each case.

In this study the propositions took the form of emerging hypotheses. Here are two illustrations from project data charts.

> Teachers' preferences for different software programs are greatly influenced by their theoretical orientations to reading, i.e., phonics or whole-language.

> Individualized learning and self-direction, as well as cooperation and peer teaching, are promoted through computer use, and some transfer of these learning styles to other class activities may occur.

The degree of support for the proposition in each case was then rated as "strong," "qualified," "neutral," or "contradictory."

After the next wave of data collection, which attended to missing or equivocal data, the propositions were revisited. For a matrix with rows showing each teacher at each site, column entries included data that supported the proposition and data that did not. As it turned out, the second proposition (above) was not supported.

At the end, the propositions were tested further with other data sources (notably surveys and observations), and cases that did not fit the modal patterns were reexamined carefully.

Although this illustration describes proposition generation in the later stages of a study, it can be used productively much earlier—even after the first round of site visits. Writing one proposition to a card, posting cards on the wall, and then clustering them helps a study staff see rather clearly what their preliminary understandings look like, as a guide for next-step analysis and further data collection.

E. Case Analysis Meeting

Analysis Problem

In any study that has multiple cases, the meaning of what is happening in each case tends increasingly to get lost in the welter of fieldwork, write-ups, coding, and other preliminary analysis. Reynolds and Wineburg (1990) looked at the methodological logs of graduate students in a yearlong qualitative methods course. Here is a this representative quote:

> I found to my cost when I began to analyze the data for the end-of-quarter progress report that my notes were almost useless because I hadn't gone back to them after the interviews. I am now faced with the job of listening to all the tapes (some 12 hours' worth) and doing what I should have done in the first place—preparing relevant summaries of information from each interview. (A. Reynolds, personal communication, Nov. 29, 1991)

How can a lone researcher or a research staff understand quickly and economically what is happening in a case, keep themselves current, and develop coherent constructs to guide later analysis?

Brief Description

At a case analysis meeting, the field-worker most conversant with a case meets with one or more people—a critical friend, a colleague, or co-researchers—to summarize the current status of the case. The meeting is guided by a series of questions, and notes are taken on answers to the questions as the meeting progresses.

Illustration

Our study of the creation of new schools included six sites. We wanted to keep as current as we could on events in the planning and implementation of each new school. We also were seeking explanations and hypotheses—and we were feeling strongly that our too-complex, unhelpful coding scheme needed to be revised.

Structuring the meeting. We settled on the idea of a case analysis meeting that would be held for each of the six sites in rotation. To help focus the meeting, a note-taking form was needed, which appears in compressed form in Figure 4.9. The actual form, of course, had its questions spread out over three or four pages to allow space for note taking. This particular example was focused on a complex case—a school—that we were following over time, but it can be adapted easily if the case is an individual or a small group.

Figure 4.9
Case Analysis Meeting Form

Case Analysis Meeting Form Date ______ Case ______

Recorder ______ Meeting Attendance ______

1. MAIN THEMES, IMPRESSIONS, SUMMARY STATEMENTS about what is going on in the case. Comments about the general state of the planning/implementation system.
2. EXPLANATIONS, SPECULATIONS, HYPOTHESES about what is going on in the case.
3. ALTERNATIVE INTERPRETATIONS, EXPLANATIONS, DISAGREEMENTS about what is going on in the case.
4. NEXT STEPS FOR DATA COLLECTION: follow-up questions, specific actions, general directions fieldwork should take.
5. Implications for REVISION, UPDATING OF CODING SCHEME.

Assembling the data. In using the form, the meeting can begin profitably with the most involved field-workers launching a discussion of item 1, "main themes." Others ask questions for clarification. The recorder follows the discussion, taking notes under that heading and asking for further clarification if needed.

Often the discussion will jump forward to later questions (e.g., a theme suggests an interpretation), and the recorder should enter those data under appropriate headings. Points or items under each heading should be numbered to mark them off and aid reference to them during discussion.

If the group does not move gradually to later questions, the recorder should ask them to do so.

The recorder should summarize the notes from time to time to be sure the discussion is being represented accurately.

Using the results. Photocopies of the notes are made for everyone; they can be reviewed at the end of the meeting. Specific plans can be made (to revise codes, how to collect new data of a certain sort), although such review and planning can be done afterward, too.

Figure 4.10 presents some excerpts from a filled-out case analysis form for our new schools study. The field-worker had been observing the start-up of a new, open-space elementary school. In this exhibit we can see that a main theme was the researcher's effort to describe (item 1) and then understand (item 2) why early implementation of the open-space teaching was going relatively smoothly even though the advance preparation had been poor. The hypotheses and hunches in item 2 (e.g., the "retreatability" concept, the principal-teacher relationship, or teacher professionalization) lead to additional data collection plans in item 4 (e.g., teacher interviews), as do the alternative, rival hypotheses suggested in item 3.

The meeting allows people to entertain opposing views (e.g., the idea of doing a retrieval interview on summer planning in item 4 stems from the view that there may have been more advance planning and preparation than the field-worker had thought).

Variations

Many other questions can be generated to guide case analysis meetings:

- What is puzzling, strange, or unexpected about recent case events?
- What is the state of our rapport with various people in key roles?
- What additional analyses do we need of existing data to understand the case better?
- What is definitely *not* true of the case at this point?
- What probably will happen over the next few days/weeks in the case?

These are content-free examples; the research questions for the study may also generate additional substantive issues that can go on the case analysis meeting form (e.g., What are the current outcomes of the innovation? How politically stable is the program? What is the level of parent involvement? What are the main channels of information transfer?).

The notes from case analysis meetings, as well as guiding specific next steps in data collection, can be revisited after the next round or two of data collection for confirmation/disconfirmation. In our illustration, it turned out that Ed's preoccupation with "technical" issues and apparent nonsupportiveness was seen by teachers as helpful; they believed he was granting them much autonomy as professionals, and they appreciated that.

Case analysis meetings can also be focused on a *single theme* in one case, such as "stabilization of the innovation," or can treat such a theme across *several cases.* See Stiegelbauer, Goldstein, and Huling (1982) for further suggestions.

Advice

Case analysis meetings are good devices for rapid retrieval of impressions and "headnotes," and for forming preliminary descriptive and inferential generalizations. The back and forth of interaction among colleagues helps keep the field-worker honest. Even so, care should be taken not to get locked into premature generalizations. The themes and suggestions from case analysis meetings should always be checked against events in the case, as noted in carefully coded write-ups of field notes.

Don't let a field-worker's generalization or impression go unquestioned or unillustrated. The tone should not be one of arguing, but of friendly skepticism and efforts at concreteness and shared clarity. A balance should be maintained between getting reasonable consensus and testing alternative, rival hypotheses. Summarize frequently to check understandings.

If the group is bigger than three or four, it will help to have someone chairing, as well as recording.

Like contact summaries, the summaries drawn from case analysis meetings can themselves be coded and drawn on for analysis. Figure 4.11 shows how this can work.

Time Required

A case analysis meeting longer than 1½ hours or so begins to lose focus and bite. The frequency of such meetings depends on such factors as the number of people and cases involved and the frequency of case contacts. In this illustration each of six sites was being visited once per week, and case analysis meetings for a given site were held every 3 weeks (thus we held two case analysis meetings per week, each covering two or three visits). The rule of thumb is this: Don't let a large amount of case data pile up before holding an analysis meeting. In our school improvement project, we found it useful to hold short case analysis meetings after each site visit (which usually had occupied 2 or 3 days and accumulated plenty of information).

F. Interim Case Summary

Analysis Problem

Researchers have four recurring nightmares about data analysis. In the first nightmare, the data are no good. They have not illuminated what they were supposed to. In the second nightmare, systematic error has occurred (commonly in the form of biased responses) in the most important data. In the third nightmare, conclusions come out of the wringer of successively more sophisticated analyses looking either trivial or trite ("You spent $75,000 to tell us *that?*"). And in the last nightmare, the data resist analysis, are opaque, even inscrutable.

Figure 4.10
Case Analysis Form: Exhibit With Data

---

1. MAIN THEMES, IMPRESSIONS, SUMMARY STATEMENTS about what is going on in the site.
   1. Ed (principal) is efficient "technical" manager, not dealing with social system; doesn't think about it. When Ken (asst. supt.) pointed out need for Ed to work with Janet, a complaining teacher ("treat her with kid gloves . . . good luck."), Ed said, "She'll be the one needing good luck."
      Not supportive especially: One teacher asked the field-worker for help, seemed reluctant when FW referred her back to Ed.
   2. Implementation of the open space approach is incredibly smooth in light of the minimal advance preparation and training. There is still a "walking on cracked eggs" feeling, though.
   3. Teachers seem cautiously willing to see how it will work out, not directly optimistic. Uncertainty, feeling unprepared. "If it doesn't work out I hope we can undo it" suggests weak commitment; is called "retreatability."
   4. Children relaxed.
   5. Teachers feel principal had no idea of what would be involved, really, in start-up.

   ....
2. EXPLANATIONS, SPECULATIONS, HYPOTHESES about what is going on in the site.
   1. Ed's "efficiency" emphasis helps smoothness.
   2. People know who to go to for support.
   3. Many teachers were students of asst. supt. and trust him.
   4. Things aren't being imposed by outsiders.
   5. Teacher attitudes may be related to the "retreatability" concept.
   6. Principal knew teachers well enough to compose workable teams to implement the open space concept. Also sent complaining teachers to another school.
   7. Principal respects the teachers—even though during the administrative planning they were treated like cattle.

   ....
3. ALTERNATIVE EXPLANATIONS, MINORITY REPORTS, DISAGREEMENTS about what is going on in the site.
   1. Perhaps the teachers' considerable past experience and training, their professionalization makes for the smooth implementation.
   2. The size of Ed's faculty has doubled; there are many strangers. That may be increasing the uncertainty as much as the lack of preparation.

   ....
4. NEXT STEPS FOR DATA COLLECTION: follow-up questions, specific actions, general directions field work should take.
   1. Ask Ed about Janet, how she's adjusting. Get to know her.
   2. Need time to talk with teachers, not just observe the start-up. Teachers are probably bothered more than their "professional" surface behavior shows.
   3. Will or can Ken give the teachers technical help?
   4. What happened in yesterday's faculty meeting?
   5. We should do a careful retrieval interview with Ken and Ed about the summer work, planning decisions, etc. that preceded the start-up.
   6. Ask key people: What are your hopes for the way the school will be by Christmas? by June? What indicators would they use for good teacher collaboration? humanization of teaching?

   ....
5. Implications for REVISION, UPDATING OF CODING SCHEME.
   1. Consider a code for *support.*
   2. Something on teacher *commitment or ownership* of the innovation.
   3. Use a pattern for the "retreatability" idea, which seems quite key.
   4. Our codes on "planning-implementation linkage" are too complicated; need to simplify them considerably.

   ....

Figure 4.11
Summary-Aided Approach to Analysis

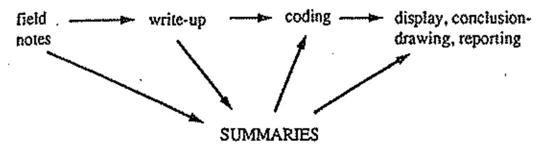


Figure 4.12
Interim Case Summary Outline: Illustration

Table of Contents

A. The Site
   1. Geography, setting
   2. Demographics of community and district
   3. Organization chart (showing key actors and their relationships)

B. Brief Chronology
   1. Adoption (includes brief description of the innovation)
   2. Planning (anything postadoption and pre-actual use with pupils)
   3. Implementation up to present

C. Current Status of Research Questions
   1. The innovation (deal with all subquestions; summarize what is currently known / if unknown, say so / if puzzles, describe them.)
   2. The school as social organization, preimplementation
   3. The adoption decision
   4. Site dynamics during implementation/transformation
   5. New configurations/outcomes
   6. Role of external and internal assistance

   (CONCLUDE THIS SECTION WITH A LIST OF UNCERTAINTIES/PUZZLES)

D. Causal Network
   1. Graphic network of variables, at this site, seen as affecting outcomes (draws on pattern codes)
   2. Discussion of network, including ties to other previous conceptual/empirical work on dissemination that seems especially salient or relevant

E. Brief Methodological Notes (how analysis was done, problems encountered, etc.; confidence in results, suggestions for next summary, etc.)

In conventional survey research, these nightmares may materialize *too late* (after the close of data collection). As a result, much preventive care is given earlier to proper sampling, validated and reliable instrumentation, and methodical data collection. In qualitative research the nightmares typically appear *early* in the game, and the analyst works on correcting them during further data collection. But these problems do not always appear spontaneously; they become clear only as the analyst examines the data as they come in.

These are methodological worries. Usually you have substantive concerns as well. What is really going on in the case so far? What's the big picture? Are patterns and themes emerging?

Typically, interim data examining is done on the run or is done for some subsets of data but not for others—as, for example, in generating *pattern codes* (section C) or writing *memos* (section D). You need an integrative exercise that obliges you to audit what is known and how well it is known—to collate the main findings to date, to estimate the confidence held in those findings, and to list gaps, puzzles, and data that still need to be collected. The interim case summary serves these purposes.

Brief Description

The *interim case summary* is a provisional product of varying length (10-25 pages) that provides a synthesis of what the researcher knows about the case and also indicates what may remain to be found out. It presents (a) a review of findings, (b) a careful look at the quality of data supporting them, and (c) the agenda for the next waves of data collection. The summary is the first attempt to derive a coherent, overall account of the case.

Illustration

We have used interim summaries in several field studies. Figure 4.12 presents the table of contents given to each researcher in our school improvement study as an outline for the interim case summary. Note that common formatting like this will enable cross-case comparability—which may suggest promising avenues for other analysts for their next site visits and will certainly dredge up themes and concepts that exist in more than one case.

The interim summary pulls together what you know about a case. In a multicase study, it is the first formalized shot at cross-case analysis, and has the big advantage of yielding emergent explanatory variables that can be checked out, rather than generated post hoc.

Organizing the summary. Whether the codes have been derived directly from the research questions or evolved during early exploratory work and are more distantly connected, it makes sense to scan the write-ups, looking for the primary codes of interest, jotting down notes as you go, and then writing the summary. (This process will go quite rapidly if your database is in computer files rather than only on hard copy; you can search for coded chunks, see them in context, and switch to another file to write up the themes you see emerging.)

Working by hand on hard copy is slower, but something like this process seems the simplest way of synthesizing the findings to date, and of becoming aware of the questions still unanswered or equivocally answered. Some analysts prefer to reread the write-ups carefully and then tackle the research questions en bloc. They then use the pattern codes to pull together the material for the summary.

Doing the summary can also be the occasion for setting up a data accounting sheet (see Box 4.4).

Using the results. The interim summary exercise, as we have noted, forces the researcher to digest the materials in hand, to formulate a clearer sense of the case, and to self-critique the adequacy of the data that have been collected. This process leads to next-step data collection, planning, and usually reformulation of codes and further analysis plans.

Whether the researcher is working alone, with occasional advisory-colleague contact, or has colleagues working on other parallel cases, the interim case summary is helpful. Exchanging interim summaries with colleagues brings them up to date. And the exchange is a good occasion to subject your emerging constructs or recurrent themes to a critical review. Blind spots are usually obvious to a second reader. Exchanging and discussing interim summaries helps cross-case analysis: People can better align their visions, argue on the basis of shared and documented instances, and resolve fuzzy or shadowy issues that need clarification for the study as a whole to move forward.

Variations

Interim summaries come in all shapes and sizes. The best ones are shapely and small—something on the order of 15-20 pages. (The outline shown earlier produced summaries of 25-35 pages.) Summaries also can be more specialized. For example, rather than collating material for both individual research questions and overarching themes, you might do two consecutive summaries, one reviewing the research questions and, a month or so later, another tackling the larger issues that, by then, should have become clearer.

Kell (1990) reports a study of classroom computer use with several interim summaries, each limited to "general trends since last visit" and two major research topics.

Very brief summaries also can be produced rapidly by the method outlined in Stiegelbauer et al. (1982): the *case study interview.* One field-worker interviews another for an hour, using a standard set of questions, such as, "How would you characterize the principal's interactions with people? What do you think of first when someone says ________ School?"

The interviewee prepares by reviewing all available data, but leaves them aside during the interview. The transcribed interview is then edited by the interviewee, referring back to available data as needed. This method helps the field-worker be integrative by pulling together impressions of the case and core themes that are beginning to appear. It is also a way to guard against the "holistic fallacy"—the assignment of a monolithic label to your case ("She's a 'traditional' teacher," "This is a 'freewheeling, innovative' school), ignoring counterexamples and exceptions.

Advice

There is never a "good time" to draft interim summaries, because they usually get done on time stolen from data collection. Strategically, the best time is about a third of the way into fieldwork, when there are initial data to report and enough time left to atone for the gaps or weaknesses the summary has revealed.

In a multicase study, be sure to allow time for individual researchers to study and discuss one another's summaries. These are usually focused and informed interactions springing from a common exercise, and they are typically more intellectual—and therefore more mind-expanding—than logistically oriented staff meetings. Discussion of interim summaries is a fertile, risk-free arena for individual researchers to try out their sense of how the data—theirs and others'—are coming together, and to get the analytic juices flowing.

Time Required

Even if the researcher has a leisurely schedule, the exercise should be brief. Two days should do the basic job, one for review and note taking, the second for drafting. Reading and discussion take another 2 or 3 hours. The most difficult part seems to be accepting the fact that interim summaries are *interim*—and likely to be incomplete, rapidly written, and fragmented. To do them "well" would require upwards of a week, which is too much time proportionate to the yield. Do them rapidly, and then think about them with your colleagues.

Data Accounting Sheet

Doing an interim case summary can be the opportunity for setting up a data accounting sheet. The sheet simply arrays each research question by informant or class of informants, as shown in Box 4.4. As the legend shows, the analyst checks the cell when a set of data is in hand, with the ultimate objective of filling all of the cells.

For a concrete example of how this method worked out in a study of statewide testing programs by Corbett and Wilson (1991), see Box 4.5. We can see quite clearly the coverage of particular study topics over different informant groups. This technique may look laborious, even overzealous, but it pays handsome rewards. In field research you lose sight all too quickly of how much—and which sort of—data have been collected from different informants. Because these data are often corroborative—verifying an explanation given by others, testing an emerging thesis—their absence is more serious than just having "missing data," as in a quantitative survey. They are the evidential bricks on which the analysis must be built.

Box 4.4
Data Accounting Sheet: Abstract Example

Research Questions \ Data Sources	Background Materials 1	2	3	Informant Group 1 1	2	3	4	5	Informant Group 2, etc. 1	2	3	4
Q 1.1	✓	✗	✓		N.A.		N.A.	N.A.	N.A.	N.A.	✗	
Q1.2	✗		✓	✓	✓	✗				✗		✓
Q 1.3		N.A.	✓		✗		✗	✓	✓		✗	✗
Q 2.1, etc.	N.A.	N.A.	N.A.	✗	✗							✓

Legend: blank = missing data; ✗ = incomplete data; ✓ = data complete; N.A. = not applicable

The accounting sheet accompanies subsequent coding; the analyst checks off each cell while coding each interview, observation, or document. At the end of coding a particular case contact, a photocopy of the data accounting sheet can be attached to the contact summary form (section A) and used in planning the next steps in data collection.

G. Vignettes

Analysis Problem

It's easy for any qualitative study to provide more data than can be understood easily in the short run, let alone the long run. But not all data are equal. During early data collection, as a researcher becomes more familiar with how things work in the case at hand, he or she often finds rich "pockets" of especially representative, meaningful data, far short of an interim case summary, that can be pulled together in a focused way for interim understanding. These are data *subsets:* Examples are "a day in the life" of an intensive care unit nurse, the events in a typical college faculty meeting, the story of how a key management decision was reached over a period of several weeks, and the way a student solves a particular math problem.

Sometimes these pockets are time limited, and data collection is straightforward; in other cases they are spread out over time and space and thus make it difficult to observe events directly. Vignettes offer a way to mine such pockets fairly easily.

They are also helpful in formulating core issues in a case—your theory of what is happening—for yourself, for your study colleagues, and for external consumers of interim reports that may be required. They can be embedded usefully in a longer case report as well.

Finally, vignettes offer an opportunity to engage study participants actively in producing, reflecting on, and learning from the data.

Brief Description

A *vignette* is a focused description of a series of events taken to be representative, typical, or emblematic in the case you are doing. It has a narrative, storylike structure that preserves chronological flow and that normally is limited to a brief time span, to one or a few key actors, to a bounded space, or to all three. As Erickson (1986) suggests, it is a "vivid portrayal of the conduct of an event of everyday life, in which the sights and sounds of what was being said and done are described in the natural sequence of their occurrence in real time." (pp. 149-150)[12]

Illustration

In several studies of professional "facilitators" working to help individual teachers improve their practice (Miles, Saxl, James, & Lieberman, 1986; Miles, Saxl, & Robinson, 1987; Saxl, Miles, & Lieberman, 1985), we wanted to look at "serial" interventions—how a facilitator worked repeatedly with a client teacher on a particular problem over the course of a few weeks—such as keeping your gym class from running wild when you are from Haiti and speak very limited English. In this case we could not watch either all of the contacts between facilitator and teacher, or every gym class. Thus we decided on a collaborative model to engage study participants actively.

Setting up the outline. It's important to identify what the structure of the vignette will be like. In this study we used an outline such as the following to examine a situation that turned out well, where the facilitator's work was successful:

The context
Your hopes
Who was involved
What you did
What happened as a result
What the impact was
Why this happened
Other comments, such as expectations for the future, predictions, what was learned, and so on

Producing the data. It helps to meet with several study participants to explain the vignette idea. Each person then chooses a situation to be described, makes some notes, and dictates or writes an account using the outline headings. The directions emphasize noncritical writing ("Do not worry about good language, just describe the situation as realistically as possible. If you have some second thoughts as you go along, just stick them in. Do not worry about

Box 4.5

Data Accounting Sheet: Illustration (Corbett & Wilson, 1991)

Data Summary Chart

Information Sources By Information Categories

Using the interview guide categories, indicate from whom information has been obtained. Place the number of people talked to within a position category about each topic in the space provided. Also, on the first line give the total number of people within each position category you talked to during the site visit.

Category	*Supt.*	*Asst. Supt.*	*Other Central Office*	*Bldg. Adm.*	*Teachers*	*Other Prof. (GCs)*	*Students*	*????*
Total interviewed	1	1	10	3	8	3	12	1
Local Testing Program				1				
– Planning								
– Implementation								
– Institutionalization								
State Testing Program							†††	
– Levels Standards	†		††	†	†††	†	†††††††	
– Competencies							†††	†
– Consequences	†		†††	†	†	††	†	
Internal District Context								
– Instructional – Prior	†				†			
– Effects		†	††††	†††	†††††	†	†	†
– Organizational – Prior								
– Effects		†	†††††	††	††††	††		
Cultural – Prior	†	†	†		†			†
Environmental Context								
– SEA	†	†	†††	†				
– Community		†		†††	††††	†		†
– Media				†				
Test Scores / Alternative Criteria								

having to stay too closely with the outline."). The typical vignette for an outline like this runs 6-15 pages.

Revising/editing. The researcher reads the typed or transcribed account, makes marginal notes and queries on it, and sends it back to the writer for review. The notes and queries are discussed over the phone; the researcher produces a revised and expanded version. This is sent back for further review and discussion; a final version, with pseudonyms replacing real names, then can be circulated to others in the work setting—an extra benefit in terms of recognition and potential learning for colleagues.

Variations

Vignettes can be (and more typically are) produced by the researcher alone, assuming a good database of processed field notes. The best discussion of this method we have found is Erickson (1986).

Seidman (1991) describes a more extended version called a "profile," which is a narrative summary using an informant's own words to describe experience over an extended time period.

Merryfield (1990) has experimented with a "constructed" form of vignette called the "narrative scene." It is a composite narrative, written not in the interim, but near the end of a study, after findings have largely been worked out, verified with participants, and "audited" by a third party. Box 4.6 shows an excerpt from her study of social studies education in Kenya.

Published along with explicit explanation of the ground rules used in construction, it is a potentially useful integrative device for reconstructing and communicating key phenomena—and interpretations of them—in a case.

Box 4.6
Narrative Scene:
Excerpt (Merryfield, 1990)

Returning from lunch, Washington Ombok wipes his face and looks down the hill to where the blue sky and blue lake shimmer on the horizon. He searches for clouds in hopes that the rains will come soon. Last year's drought caused many hardships for the farmers and fishermen of his lake district, one of the poorer agricultural areas in Kenya.

....

A messenger greets Mr. Ombok and sets two big cartons beside his desk.

"What's this?" asks Mr. Ombok.

"They are from the Institute," replies Mr. Charles Obonyo, the District Education Officer, as he enters the office. "I guess our PEP school has some new materials."

Mr. Ombok breaks open the top box and pulls out a draft pupil's book, mimeographed and stapled together with a light green cover. "Yes, our PEP school certainly can't complain about teaching materials. I wish all of our other schools had as many books. Is the Land-Rover back yet? I might as well get these books on out to Jalo."

Caution: Even with safeguards such as those mentioned, the risk of a vivid, convincing report that is false to the reality of the specific situations from which it was drawn is substantial. For example, the sentences attributed to "Washington Ombok" and to "Obonyo" were never uttered as such—though sentences like them obviously were heard by the field-worker. The sentences have a didactic intent, teaching the reader about the PEP program, but do real inspectors and education officers in Kenya talk like this? As Merryfield (1990) cautions, "In constructing scenes and dialogues the researcher literally puts words into people's mouths based not only on raw data but also on the study's major findings" (p. 23).

Advice

The collaborative vignette requires a trusting relationship between researcher and participant, one that is nonjudgmental but strong enough to tolerate occasional pressure to get the job done. Writing something on your own for an hour or so is quite different from being interviewed for the same time period.

It's also important to watch for potential bias: Collaboratively written "success" vignettes can be defensive or self-serving. And they may "co-opt" the researcher into unwarranted positive judgments—a form of "going native."

Any type of vignette, because it is a concrete, focused story, will be vivid, compelling, and persuasive. If it is not really representative, you and your readers run the risk of misunderstanding the case it refers to. Using multiple vignettes helps, but the question "Is this really typical?" must always be asked. P. Wasley (personal communication, 1991) reports that, after reviewing a week's field notes to find answers to research questions and key themes, she chooses a "day in the life" to write up—but then scans the entire corpus of notes to be sure she is not pulling out the most sensational incidents or unique ones.

Erickson's (1986) helpful discussion points out: "Even the most richly detailed vignette is a reduced account, clearer than life. . . . it does not represent the original *event itself,* for this is impossible. . . . [It] is an abstraction; an analytic caricature (of a friendly sort) . . . that highlights the author's interpretive perspective" (p. 150).

Thus, he notes, the vignette is a "potentially dangerous tool that can be used to mislead as well as to inform," whose interpretive validity "cannot be demonstrated within the vignette itself." Multiple vignettes and accompanying commentary are needed.

At the same time, Erickson acknowledges that writing a vignette after reading through field notes can be a powerful means for surfacing and clarifying your own perspective on what is happening.

Finally, vignettes can be a useful corrective when your data—coded, displayed, and pondered-on—somehow lack meaning and contextual richness. (For a good account of this, see Maxwell and Miller's [1992] account of their analysis of adolescents' friendships through the creation of "narrative summaries.")

Time Required

The collaboratively produced vignette described above (Miles, 1990) takes about 2 or 3 hours of the participant's time and 4 hours of researcher time (not including transcription; that adds another 3 or 4 hours).

The time for researcher-produced vignettes will vary considerably, depending on the size of the database being scanned and the degree of focus of the vignette.

H. Prestructured Case

Analysis Problem

We've alluded often to the problem of data overload in qualitative studies, which is exacerbated by the time required for processing field notes and for coding. In a study where time is limited and research questions are well specified, are there ways to focus and streamline data collection and analysis that are "quick and clean," and will provide trustworthy results?

These questions take on more importance in multiple-case studies, where comparability across cases is critical for warranted conclusions.

Brief Description

Assuming that the researcher has established an explicit conceptual framework, a rather precise set of research questions, and a clearly defined sampling plan, the prestructured case begins with a case outline, developed *before* any data are collected. The outline includes detailed data displays, as well as narrative sections accompanying them. The outline is, in effect, a shell for the data to come. Over several rounds of field visits, the researcher fills in successive drafts of the case, revising steadily; the final version of the case is ready shortly after the last field visit.

Illustration

In a study of reform processes in six urban high schools carried out by a staff of five researchers (Louis & Miles, 1990), we developed a detailed conceptual framework and a set of 14 detailed research questions, such as:

What *barriers, problems, and dilemmas* were encountered during planning/initiation and implementation of the school improvement project? These impediments may be "inherent" in the nature of high schools and the improvement process or may be conditioned by the local context.

What *management and coping strategies* were employed to deal with the barriers/problems/dilemmas? Which were technical, which were cultural, and which were political in nature?

Constructing the outline. With the research questions in hand, you can put together a clearly specified outline. Figure 4.13 shows an abbreviated form of the one used in the study, with an eye to a 40-page case. A more detailed version was used by all case researchers; the aim is to "drive the data collection by the anticipated product" (Miles, 1990). That is, the field-worker, knowing what the case has to look like, collects the data needed to fill the shell. Specific data displays (matrices, organization charts, etc.) are designed in advance.

Planning for data collection. Because time normally is limited when this method is used, it helps to do advance planning for within-case sampling. For example, in this study, for the two research questions above, we wanted a good sample of:

Figure 4.13
Prestructured Case Outline:
Abbreviated Version

A. Beginning note: case methods and data
- I. The context:
  - a. The school: an overview
  - b. The community context
  - c. The school district
  - d. The state context (SEA and legislature)
  - e. The school: a more detailed picture
  - f. Preconditions for change
- II. The improvement program as planned: an overview
- III. Why this program
- IV. The story of planning and implementation
  - a. Chronology
  - b. The process of planning and implementation
  - c. The problems
  - d. The assistance provided
    - 1. Sources, types, and adequacy
    - 2. Why this assistance?
  - e. How problems were dealt with
    - 1. Managing and coping strategies
    - 2. Why these strategies?
- V. The program implementation effort
  - a. Overall program extent and quality
  - b. Prospects for the future
  - c. Why implementation occurred as it did
- VI. The results
  - a. Interim results
  - b. Long-term results
- VII. Why these results?
- VIII. Lessons for improving the urban high school

key actors: people in coordinative positions

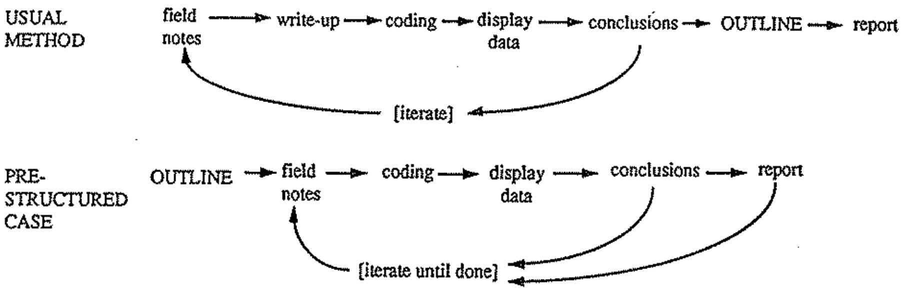
key events: meetings where problem solving and coordination occur

core processes: problems encountered during initiation, as well as early and later implementation; coping strategies as seen by a wide range of affected roles (teachers, building administrators, central office personnel)

Given limited time, it's also useful to plan a sequenced approach to data collection: Which research questions will be focused on during successive visits?

Data collection and analysis. With the outline clearly in mind, the researcher begins the first round of data collection. The raw field notes are coded without being transformed into write-ups. The researcher reviews the coded field notes and enters data directly into displays by using a word processor (for an illustration from the study, see Box 5.8) and writing accompanying analytic text—that is, the conclusions drawn from the displayed data. There will be instances of missing or unclear data, and of unanswered research questions; these provide the targets for the next round of data collection.

Figure 4.14
Traditional Analysis Sequence Compared With Prestructured Case



This procedure is iterated until data collection and the completed case are done. Thus the processes of data collection, analysis, and report writing are collapsed into one evolving procedure. With some experience the researcher has a constant sense of being on top of the data, and remains open to checking and extending findings. Feedback from study participants can be sought and incorporated along the way to verify and revise preliminary conclusions.

Figure 4.14 shows how this procedure differs from the more traditional one. Both are iterative, but the fieldwork in the prestructured case is driven by the outline, as well as the emerging conclusions—and the report draft itself.

Variations

Each case can be developed by a pair of researchers to ward off the potential difficulties discussed below. If time permits, the method can be used with fully written-up data.

Advice

Figure 4.14 points to a major weakness: Coding is done directly from scribbled field notes, not from a "public" write-up, and is much less subject to critique and refinement. So the method should *not* be used by inexperienced field-workers or by those unfamiliar with the type of setting being studied. Nor should it be used if the conceptual framework and research questions are not well spelled out. "In short, as with most things that matter, to do a pre-structured case *well,* you have to know what you are doing" (Miles, 1990, p. 48).

Furthermore, the method leans the researcher toward early—maybe too early—conclusion drawing. This tendency makes tunnel vision and undetected bias more likely. Even with experienced, knowledgeable researchers, we advise corrective tactics such as ensuring wide data sampling from a range of informants and settings; **"triangulating"** (see Chapter 10, section B) with different data collection methods, theories, or respondents; and—above all—using skeptical colleagues to see whether different conclusions can be derived from your displays.

The case outline itself also can help by limiting early versions to descriptive material, encouraging "reflective remarks," and including a subsection on "puzzles and unanswered questions" in each chapter of the case.

Time Required

In its "pure" form (coding from field notes), the prestructured case method reduces processing, analysis, and writing time from the range of four to seven times the contact time to perhaps two to four times. During early stages the ratio of office to field time is about 2 to 1, increasing to 3 or 4 to 1 later. Longer, more ambitious cases take proportionately more than those with limited field contact.

Once research questions are crisp, the front-end work of designing the case outline and instrumentation usually takes only 2 or 3 days.

So the method saves a lot of time. BUT the quality of the study may well suffer, as we have said, unless careful attention is paid to warding off biased and poorly founded conclusions. That carefulness will in itself add some time costs (e.g., for colleague critique and discussion).

I. Sequential Analyses

In this chapter we've reviewed a series of methods best suited for early analytic work, and often used as data collection continues. We should emphasize that a single interim analysis is rarely "enough." If we listen carefully, our cases will say to us, as Norman Mailer once said to his readers, "Do not understand me too quickly." It's easy to fall into the trap of premature closure, a feeling of "rightness" that is grounded falsely—especially because early data collection is usually partial, flawed, and simplistic in some important respects.

The first interim analysis should not only point to improved data collection but also lead to successively deeper, fuller waves of analysis by swinging back to pattern coding, memoing, and proposition developing as more detailed, better-quality data become available, and your cognitive map of the case gets richer and more powerful.

As Campbell (1979) suggests, we need to keep a "box score" along the way. As our emerging theory of the case generates multiple implications, predictions, and expectations *within the case*, how many are actually confirmed? He also reminds us of the need for an "adversary process": "Other experts are encouraged to look for other implications of the theory and other facts to contradict or support it" (p. 61).

Basically, a good interim analysis is one that helps us reorient our view of the case. And any interim analysis should be the first of several. Their strength is their exploratory, summarizing, sense-making character. Their potential weaknesses are superficiality, premature closure, and faulty data. These weaknesses may be avoided through intelligent critique from skeptical colleagues, feeding back into subsequent waves of data collection.

Illustrations

We have presented a series of techniques beginning with the collection of data and leading to interim summaries. By looking at the titles of each section, the reader can get an idea of how each technique could follow another in an iterative sequence: Each wave of data collection leads to progressively more molar clustering and analysis. Our illustrations, however, came from different studies.

Let's look at some studies showing how early analysis methods are used in sequence around a single set of research questions and proceed from single-case to multiple-case analyses.

In a Belgian study of primary health care (Schillemans et al., n. d.), the researchers focused on treatment of victims of incest. They began with a group of 10-15 medical practitioners, meeting in a weekly "exploration group" for 2 years. Here are the successive steps:

- Review of material from the first patient case (videotapes, documents, notes)
- Generation of a first batch of codes from the case review, such as "fear of the dark," "speech difficulties," "aversion to sexuality"
- Addition of more clinical codes from specialized literature, such as "boundary between man and woman," "violence"
- Examination of further cases ($N = 32$, with 32 controls); developing further codes via analytic induction
- Development of key hypothesis (e.g., that incest victims have weakly developed boundaries)
- Verification of hypotheses via interviews with several patients from original set
- Integration of findings with existing empirical and conceptual literature (e.g., on dissociated personalities)
- Extension of findings to new sample (screening new cases for instances of incest) and new analytic cycle

Each step was a funnel for gathering more data on fewer, but progressively more vital, dimensions in the data set.

Here is a second, very different example, in the tradition of Giorgi's (1975) work in phenomenological psychology. It is holistic and does not use coding directly, but relies on a series of deeply understood summaries.

Fischer and Wertz (in Giorgi, 1975) interviewed 50 victims of crime (e.g., assaults, robbery, burglary, vandalism), asking them to describe what was going on prior to the crime, what it was like to be victimized, and what happened afterward. The transcriptions then underwent a five-step analysis: (a) familiarization through rereadings, (b) demarcation into numbered units, (c) casting of the units into temporal order, (d) organizing clusters of units into scenes, and (e) condensing these organized units into narratives with nonessential facts dropped.

Along the way, there were four analytic transformations of the data. Here they are:

1. *Individual case synopses.* These aimed to disclose what was essential to each person's experience, while reducing the original transcript by about one-third. They used close approximations to the person's own words ("Mr. R felt the police investigation was a sham. They dusted for fingerprints, but they didn't seem to care. The R's felt ignored. They had expected the police to sit down and get a list of what was missing, but they said they were too busy.").

2. *Illustrated narrative.* These narratives work across cases by enumerating all of the possible sequences, then picking out the modal ones, and then moving up still another notch to find a generic sequence. Here it is: (a) living routinely, (b) being disrupted, (c) being violated, (d) reintegration, (e) going on. The narrative core for, say, "being disrupted," is this:

> Even as the person copes with the threatening or discovered crime, he or she scans imaginatively and perceptually for a still worse outcome.

Each part of the sequence is followed by a series of illustrative excerpts ("You imagine the worst when it's happening. . . . I just kept thinking my baby's upstairs and I might never see her again.").

Let's pause here for a moment. Notice that the authors have not resorted to explicit coding. They seem to be looking for key words, themes, and sequences to find the most characteristic accounts. This approach helps keep the connections between different segments—the strands that explain what came before the core theme you have identified and what came after.

By simply extracting and clustering the segments in which a theme appears, as we do with conventional coding and pattern coding, we are much better able to look across a data set within single or multiple cases. But we may lose the contextual information that tells us why and how the pattern appeared specifically. We also run the risk of combining codes for analysis, when the context of the code would have told us that they are not in the same thematic or conceptual family. Ideally, interim analyses will preserve both the contextual, over-time dimension and the more "paradigmatic" or variable-oriented view—a distinction we keep returning to.

3. *General condensation.* This is a compact description (in two pages) of the characteristics common to the transcriptions. The authors are not clear about how they managed to "encompass all individual cases." But the account does read as a general summary. Here is an excerpt:

> Being criminally victimized is a disruption of daily routine. It is a disruption that compels one, despite personal resistance, to face one's fellow as predator and oneself as prey, even though all the while anticipating consequences, planning, acting and looking to others for assistance. (Fischer & Wertz, 1975, p. 149)

4. *General psychological structure.* Here the analysis is nested in a more conceptual frame—is connected to a body of knowledge lying outside the data set. The authors worked iteratively in several ways: the analysis of several transcriptions, the extraction of a general psychological profile, then use of that profile on another set of transcriptions. At the same time, the synopses, condensation, and narratives were used as a cross-checking device. If these pieces do not cohere and do not involve the same constituents of victimization, conceptual analysis would be unhinged. Here is an excerpt from the general psychological structure:

> [There are] three ways in which victimization must be surpassed for complete recovery: the victim's active efforts, the world's repeated reassertion of social harmony, and the active assistance by others. All three aspects are mutually implicit. . . . For example, in discussion with a caring friend, the victim becomes actively involved (thus overcoming loss of agency), and makes sense of the crime (overcoming shock and confusion); while the friend allows reciprocity. (p. 154)

Now for a final illustration of sequential analysis. Using inductive coding and a grounded approach to the derivation of theory, Chesler (1987) studied the dangers of self-help groups as perceived by 63 professionals (physicians, social workers, psychologists). Their interview had three straightforward questions, such as "What do professionals mean when they talk about the dangers of self-help groups?" Let's follow the analytic trail.

Step 1. Underline key terms in the text.

Step 2. Restate key phrases. Figure 4.15 shows how this worked. The idea is to remain as descriptive and literal as possible. Note, however, that Chesler often resorts to paraphrase.

Step 3. Reduce the phrases and create clusters. Chesler reports that this step was done several times, as different clustering patterns were tried. As one coder completed the clustering process, another coder redid it independently. The two then were compared. Completion of this step resulted in 40 apparently distinct clusters—too many to analyze, writes Chesler. Here are two:

Control will be taken away
- proprietary control
- concerned with retaining control
- fear loss of control

Create misunderstanding/misinformation
- generate misinformation
- repeat misinformation
- misinformation circulating
- misinformation can be exchanged
- won't understand what's happening

Note, before we move on, the difference in approach from Fischer and Wertz (1975), who virtually never abandon the narrative form. Chesler has far more fragmentary, decontextualized data, but he can move it around readily. The cluster names are, in effect, codes.

Step 4. Reduction of clusters, and attaching labels. This is the process of pattern coding. As clusters are reduced in number and are combined to form "meta-clusters," comparisons are made "at the boundaries of each cluster." The decisions involve both implicit and explicit comparisons and thereby move to higher levels of inference. Here are a few of the meta-clusters showing "dangers" of self-help groups, with the incidence for each shown in parentheses:

Figure 4.15
Marginal Notes as a Coding Aid (Chesler, 1987)

Step 1: Underline key terms in the text	Step 2: Restate Key Phrases
Social worker, Gp. 3: The professionals are afraid people will be <u>repeating misinformation</u>, that people will <u>compare one diagnosis to another</u> and come back and say, "Why aren't we getting XXXX?" There is a fear that they will <u>get people who are obsessed with the disease</u>, and <u>not coping well</u>, and totally <u>fixated in getting the secondary gains</u> from the disease. Frankly, I've seen that happen in a few individual cases.	repeat misinformation compare diagnosis obsession with disease cope poorly fixation on secondary gains
Social worker, Gp. 7: Professionals are afraid that a group could <u>get out of hand</u>, <u>take power</u> or just <u>be harmful</u> in some way.	get out of hand take power be harmful

1. Challenge the power of professionals (17)
2. Create emotional problems for parents (15)
3. Parents learn/know too much (11)
4. Spread misinformation (8)
5. Take over professionals' job (social work) (6)
6. Transfer doctors or increase physician competition (6)
7. Question medical authority/judgment (6)

Step 5. Generalizations about the phrases in each cluster. These correspond to the "propositions" examined earlier (section D). Analytically, they are the plausible next step once a clustering procedure and "patterning" exercise have identified and qualified a set of core themes. A few examples from the meta-cluster "Parents learn/know too much":

> Professionals have fears from the sharing/comparing of information.
>
> Doctors are worried that parents will get too educated.
>
> Professionals are afraid that parents will compare notes, compare protocols, and learn of experiments.

Step 6. Generating minitheories: memo writing that poses explanations. On to the memos (as in section D). Chesler (1987) notes that as these pattern codes and propositions are created and refined, and then contrasted with one another, the researcher is "well into the process of generating theory that explains their meaning." Here is a fragment of a memo clearly bridging from one cluster to a more conceptual explanation for it:

> Why is too much information a danger? I was prepared to hear professionals state that parent misinformation or lack of information was a danger, but why too much information? What is involved here? I remember an article I read many years ago, in which Rieff wrote about knowledge being the basis of professional practice. . . . So perhaps the danger to professionals is that as parents get informed, the professionals no longer have that edge in expertise and thus status and control. (Chesler, 1987, p. 17)

Step 7. Integrating theories in an explanatory framework. Here Chesler turns toward the literature on professionals' ideologies and roles. He uses it as an orienting framework, applying it to the original clusters, the pattern codes, the propositions, and the memos. The central theme becomes the "image of control exercised by professionals and their perception that autonomous parent self-help activity (and groups) threatens professional control." Beyond that is the intention to exercise a monopoly over the health care knowledge base, delivery of service, and value claims.

Summary Comments

In this chapter we've focused on methods most useful during early stages of analysis. They serve to organize your data for later and deeper analysis, while also clarifying your ideas about the meaning of your data, and suggesting leads for added data collection.

We've reviewed several methods: those for *summarizing* data (contact summaries, the case analysis meeting, and interim case summaries); approaches to *coding* (attaching meaningful labels to data chunks) at descriptive and more inferential, general levels; methods for *thinking* about your data via annotations and memos; and methods for producing focused or extended *reports*, such as vignettes and prestructured cases. We concluded with examples of how such early analysis methods can be strung together over time during a project.

We now turn to the deeper descriptive analysis of cases that involves systematic display of data for drawing and verifying conclusions. That work, supported by early analysis, begins during data collection and continues afterward.

Notes

1. Some researchers are beginning to type field notes directly on a laptop or notebook computer during interviews or observations (Fetterman, 1989; P. Wasley, personal communication, 1991). Such notes, however, like handwritten ones, require later processing, depending on the systematization of entries, typing and spelling skill, perceived sketchiness, and use of abbreviations.

2. There is a long and well-worked-out tradition of photographs as a form of data, which we do not explore in this book. For good treatments of problems of photographic data collection and analysis, see Harper (1989), Ball and Smith (1992), Bogdan and Biklen (1992), Templin (1982), Wagner (1979), and Becker (1978). On film and videotape, see the compendium of methods, with annotated bibliography, by Erickson and Wilson (1982).

Similarly, data sometimes appear in the form of drawings made by the field-worker (e.g., the set-up of a room) or by field informants (e.g., an organization chart). See Chapter 4, section A, for further discussion.

Finally, data may come in the form of documents collected from a field site. See Box 4.1 for ideas on summarizing these.

We do not deal with the extensive range of techniques available for analyzing texts (in either document or transcript form) from the fields of discourse analysis (Potter & Wetherell, 1987; Van Dijk, 1985), semiotics (Pecheux, 1982), or the immense field of linguistics (Akmajian, 1990; Rosenberg, 1982; Yngve, 1986).

3. Sanjek's (1990) examples of field notes are rich and fascinating and include a wide range of types, from scratch notes to write-ups, texts written or dictated by respondents, field-worker journals, and many others.

4. See Kvale (1988) for a thoughtful discussion of the transcript and its problems. For illustrations of micro-level transcription, see Potter and Wetherell (1987)—they describe in their Appendix some standard conventions for showing pauses, emphasis, inaudibility, and so on.

5. See Mishler (1986) on issues of interviewing quality. The informant and interviewer, he says, co-construct meaning, producing a "story" around the "facts" as each person "reads" signals: phrases, pauses, digressions, initiation of a new question, insistence on a line of questioning, asking for more on one item but not another, cutting off the discussion, and so on. The informant "learns" what the interview is about and decides what can be said—what this story will be about—and how it will be represented.

Thus the looser the interview strategy, the less comparable your data. And Mishler cites studies showing that about one-third of structured questions are not asked as planned. Furthermore, "one-shot interview[s] without local knowledge of a respondent's life situation and following a standard schedule" are suspect; they are essentially a "meeting between two strangers" who cannot make out one another's socially organized frames of meaning.

6. See Kvale's (1988) admonition to "beware of transcripts" (p. 97). Transcriptions, he argues, are transformations of one mode—a conversation or oral discourse—into another mode: narrative discourse. There are dangers of superficial coding, decontextualization, missing what came before and after in the informant's account, and what the larger conversation was about.

7. Technically, coding has associations with semiotics, a subfield of linguistics that emphasizes the analysis of communicative "signs" (e.g., in art, music, literature, everyday language). To find out more about the relationship between linguistics and fieldwork, see Manning (1987).

8. The classic reference on approaches to building a systematic set of codes (categories) is Lazarsfeld and Barton (1972).

9. Mishler (1986) points out that coded data have no "meaning" in themselves; they are the result of gradually built-up understandings among the research team, which are themselves abstracted from the original meanings conferred by informants. The process is insidious: All parties, using common vernacular language, are convinced intuitively that they understand the same things by the same terms. Only gradually, through training and periodic, continuing coding reliability checks, is a coder's subculture built up.

10. The Q versus R distinction was first made by Stephenson (1953). For the reader new to the idea, an example may help. If you measured several attitudes in a population of, say, college students and correlated the attitude measures, you might find that conservative political attitudes were somewhat positively related to attitudes toward beer drinking. That would be an R analysis.

Using the same data set, you also could see whether there were clusters or families of students. It might turn out that the students fell into four main clusters: (a) conservative beer drinkers (the largest group), (b) progressive beer drinkers, (c) total abstainers, and (d) middle-of-the-roaders. That would be a Q analysis. Pattern codes for qualitative data can be used for either Q or R analysis.

11. For a good review of how people tend to cling to their beliefs, even when faced with countering evidence, see Ross and Lepper (1980); the best full-scale treatment is by Nisbett and Ross (1980). "People" decidedly include researchers, as McEwan and Bull (1991) note:

> The history of science is littered with examples in which our most revered scientists—Galileo, Newton, Priestley, Einstein—refused to abandon their theories in the face of apparently refuting evidence. . . . Even when scientists understand that a particular phenomenon is inconsistent with a theory, they still rarely abandon that theory. Most often they simply ignore the apparently refuting evidence, especially if no other available theory can explain it . . . [or] will make ad hoc adjustments in beliefs not central to the theory (such as beliefs about how their instruments function) in order to render the evidence consistent with the theory. (p. 322)

12. For a good discussion of the "story" or narrative mode of thought, as contrasted with abstract "propositional" thought, see Vitz (1990) and Howard (1991); see also Mishler (1990) and Polkinghorne (1988).

5

Within-Case Displays: Exploring and Describing

In this chapter we explore a series of displays for drawing and verifying *descriptive* conclusions about the phenomena in a bounded context that make up a single "case"—whether that case is an individual in a setting, a small group, or a larger unit such as a department, organization, or community.

These analysis methods can be used either during or after data collection. They depend on the sorts of early analysis we describe in Chapter 4—particularly, coding. At first they provide preliminary conclusions about what is happening in the case—and how—and suggest leads toward new data. Later, as fuller, more complete descriptions are in hand, these displays can supply the basic material for *explanations*—plausible reasons for why things are happening as they are. We explore that topic in Chapter 6.

First we talk about the analytic progression from what and how to why, and then about the basic idea of *data display*. Then we describe a series of within-case descriptive display types, illustrating how they can be designed and used to draw and verify conclusions.

Describing and Explaining

Looking at a situation, any researcher wants to know clearly *what* is going on and *how* things are proceeding—and usually wants as well to understand and explain coherently *why* things occur as they do. This innocent formulation can be embroidered endlessly, which we do not propose to do. We can, however, define some terms and make some useful distinctions.

Description, in Bernard's (1988) terms, means "making complicated things understandable by reducing them to their component parts." The issue is making a clear accounting of the phenomena at hand. *Explanation* is dicier: As Bernard suggests, it means "making complicated things understandable by showing how their component parts fit together according to some rules"—that is, *theory*.

Qualitative studies are often mounted to *explore* a new area and to build or "emerge" a theory about it. But they also can be designed to *confirm* or test an existing theory. In the confirmatory mode, as Gherardi and Turner (1987) point out, data are used to fill in gaps in a puzzle. In the exploratory mode, it's as if we are trying to solve an unstated or ambiguous problem, which has to be framed and reframed as we go. Wolcott (1992) talks about this as the "theory first" or "theory after" approach. Both are workable.

So what is theory? We can talk about *implicit* theory, the preconceptions, biases, values, frames, and rhetorical habits that (for example) lead us or the people we are studying to refer to a situation as a "single-parent family"—or to the same situation as a "broken home" (Rein & Schon, 1977). In this case the implicit theories are each biased. The first puts a certain kind of human group into a class called "family" and implicitly asserts its right to exist. The second classifies the situation as imperfect or damaged and implies that "a home" must have more than one parent.

We can speak also of *explicit* theory: a set of concepts that might be organized in list form, or in a hierarchy, or in a network of propositional statements. As Wolcott (1992) notes, these might be at the level of "grand theory" (e.g., symbolic interactionism) or, more typically, include modest middle-range concepts such as *culture* or *commitment* or *innovation adoption.* Conceptual frameworks, as we describe them in Chapter 2, are the researcher's first cut at making some explicit theoretical statements.

One more distinction, to which we alluded before, seems important. Efforts to describe and explain may be "paradigmatic" (Maxwell, 1992a; Maxwell & Miller, 1992)—they involve a *variable-oriented* approach (Ragin, 1987) that deals essentially with the relations among well-defined concepts. For example, you might study adolescents' decisions to attend college by examining the relationships among variables such as socioeconomic class, parental expectations, school grades, and peer support.

Or studies may be "syntagmatic," or *process-oriented,* following the events in a specified *case* context over time (Maxwell, 1992a; Mohr, 1982). For example, you might study a particular adolescent, Nynke van der Molen, over a period of several months to follow the events and conditions (e.g., a poignant discussion with her mother on why she had never worked outside the home; Jane's experience in dissecting a frog) that were related to her decision to go to veterinary school. (For further discussion of this distinction, see Chapter 7, section A, on cross-case analysis.)

We concur with these authors that *both* approaches need to be combined for careful description and explanation. This chapter includes methods of both sorts.

From describing to explaining: The analytic progression. Usually it is hard to explain something satisfactorily until you understand just what the something is. Thus a natural progression, as Rein and Schon (1977) suggest, is from telling a first "story" about a specified situation (what happened, and then what happened?), to constructing a "map" (formalizing the elements of the story, locating key variables), to building a theory or model (how the variables are connected, how they influence each other). We have constructed a deeper story, in effect, that is both variable-oriented and process-oriented.

The progression is a sort of "ladder of abstraction" (Carney, 1990; see Figure 5.1). You begin with a text, trying out coding categories on it, then moving to identify themes and trends, and then to testing hunches and findings, aiming first to delineate the "deep structure" and then to integrate the data into an explanatory framework. In this sense we can speak of "data transformation" as information is condensed, clustered, sorted, and linked over time (Gherardi & Turner, 1987).

Naturally there is no clear or clean boundary between describing and explaining; the researcher typically moves through a series of analysis episodes that condense more and more data into a more and more coherent understanding of what, how, and why. In this chapter we mainly emphasize description, but point forward to more explanatory methods in Chapter 6.

A. How Data Displays Work

The idea of a display is central to this book. By *display* we mean a visual format that presents information systematically, so the user can draw valid conclusions and take needed action.

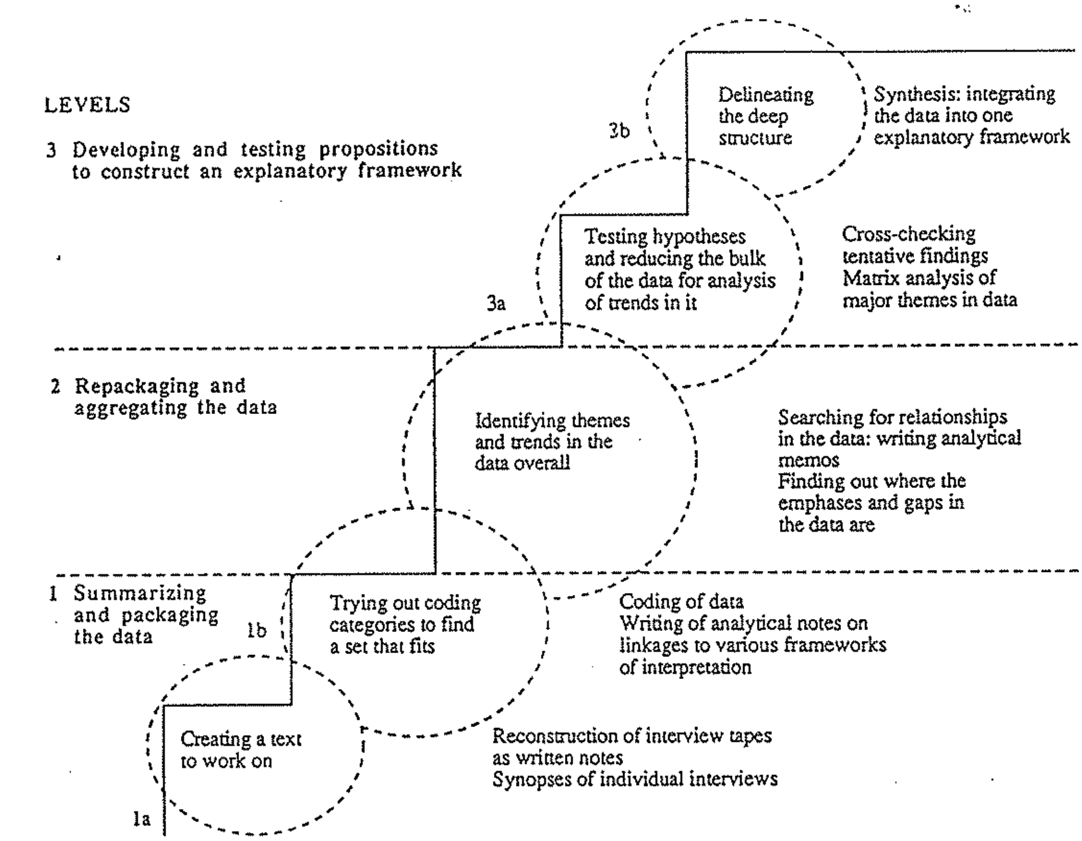
For qualitative researchers, the typical mode of display has been extended, unreduced text, usually in the form of written-up field notes, which the analyst scans through, attaching codes and then extracting coded segments and drawing conclusions. The analyst then writes a second form of extended text: a case study report.

Our experience tells us that extended, unreduced text alone is a weak and cumbersome form of display. It is hard on analysts because it is *dispersed* over many pages and is not easy to see as a whole. It is *sequential* rather than simultaneous, making it difficult to look at two or three variables at once. It is usually *poorly ordered,* and it can get very *bulky,* monotonously overloading. Comparing several extended texts carefully is very difficult.

The same objections apply with even stronger force for final readers. Indeed, it is been claimed (Mulhauser, 1975) that long case studies are almost useless for policymakers, who cannot afford the time required to comprehend a long account and to draw conclusions for their work.[1]

The argument of this book is, You know what you display. Valid analysis requires, and is driven by, displays that are focused enough to permit a viewing of a full data set

Figure 5.1
The Ladder of Analytical Abstraction (Carney, 1990)



in the same location, and are arranged systematically to answer the research questions at hand. A "full data set" does not, of course, mean the complete field notes. Rather the condensed, distilled data presented are drawn from the full range of persons, events, and processes under study. With extended text, there can easily be "selective stacking" of the data—even with good intentions. An organized display wards off this problem.

Although such displays may sometimes be busy, they will never be monotonous. Most important, the chances of drawing and verifying valid conclusions are much greater than for extended text, because the display is arranged coherently to permit careful comparisons, detection of differences, noting of patterns and themes, seeing trends, and so on.

We can see the importance of coherence analogically by looking at statistical packages such as SPSS and BMD. Their computational speed is amazing and convenient, but they have an equally important function: They display information in ways that (a) show the data and analysis in one place, (b) allow the analyst to see where further analyses are called for, (c) make it easier to compare different data sets, and (d) permit direct use of the results in a report, improving the credibility of conclusions drawn.

These virtues are commonplace to survey researchers, who simply expect to see computed data appear in well-labeled histograms, correlation matrices, scatterplots, factor plots, and box-and-whisker displays. Good displays permit the researcher to absorb large amounts of information quickly (Cleveland, 1985). But the qualitative analyst has to handcraft appropriate data displays. As yet there are few familiar, agreed-on data setups among qualitative researchers, so each analyst has to adapt those of others or invent new ones. A major purpose of this book is to encourage the creation and dissemination of innovative and reliable data displays for qualitative data.

We already have discussed some displays. For example, a *contact summary form* and a *report of a case analysis meeting* are data displays. They show reduced, focused, organized information on a single page rather than in extended text. The displays we discuss in this chapter present information in more sophisticated ways to aid conclusion drawing.

We begin with a discussion of how to use displays: building the format, entering data, drawing conclusions, writing analytic text, and cycling onward to revised or new displays.

Building the Display Format

Generating formats for displaying qualitative data turns out to be fairly easy and enjoyable. Formats can be as various as the imagination of the analyst. They fall into two major families: *matrices,* with defined rows and columns, and *networks,* with a series of "nodes" with links between them.

The data entries are also multiform: short blocks of text, quotes, phrases, ratings, abbreviations, symbolic figures, labeled lines and arrows, and so on.

The display format and shape of the entries will depend on what you are trying to understand: a general situation, detailed chronologies, the behavior of people in different roles, or the interplay of conceptual variables.

And it depends on how far along you are in the study and what has priority right now. The need might be for eyeballing data in an exploratory way. Or it could be for carrying out detailed analyses; for setting up data to use in another, more differentiated display; for combining parallel data for a single case; for combining data from several cases; or for reporting findings. A good format will allow all of these uses to some degree, but inevitably will do some well and others less well.

In this chapter we emphasize displays of single-case data—but displays that can be folded easily into multiple-case analysis, which is discussed in Chapters 7 and 8. If the formats of displays in a multicase study are comparable, the work of the cross-case analyst is much easier.

We can make several general points about the process of building appropriate displays. They are illustrated more fully as we proceed. First, as we have suggested, it is an interesting and not overly demanding task. Second, creating a good display format usually requires a number of iterations (sometimes with new data collection intervening). Third, form follows function: Formats must always be driven by the research questions involved and your developing concepts, often in the form of codes.

Illustrative display formats: Matrices. A matrix is essentially the "crossing" of two lists, set up as rows and columns. Let's take a look at some formats, explaining and labeling them as we go.

Table 5.1 focuses on understanding a chronology: It is *time*-ordered. Taken from our school improvement study, it lists the events at Banestown over a 4-year period related to the adoption and implementation of a reading program called SCORE-ON. The time periods were defined conceptually, but were adjusted according to the actual events at Banestown. The events are also sorted according to where they took place: at a specific school (Smithson), other local schools, the school district, and the state/macro level. Events that proved "barometric"—leading decisively to a changed situation—are marked with an asterisk. To find "events," the analyst consulted coded field notes (looking for the "AP-CHRON" codes described in Chapter 3) and then wrote a distilled summary phrase for entry in the matrix. Perhaps 10-20 pages of material have been condensed into the display.

This display is especially helpful for understanding the flow, location, and connection of events. It is good for exploratory eyeballing. It could lead to later, more causal explanation. It can be compared easily with similar matrices from other cases in the study. And, with analysis and commentary attached, it can provide a good thumbnail sketch of the change process for use in the final report.

Here the display is used for a complex case; it can be used also for an individual as the case (see, for example, Table 7.1).

Table 5.2 shows a matrix that's only partially ordered (it specifies a series of unordered "conditions" supporting preparedness for using an innovation, in a checklist format, sorting data by roles). The display calls for heavier data transformation by the analyst. Much condensation and standardization has gone on to get the data ready for entry into this format. A mixture of direct quotes and summary phrases is used. In addition, informants' responses have been pooled and crudely scaled (from *none* to *strong*). The matrix lets you sum up the components of "preparedness" vertically, and also compare different roles' preparedness on specific conditions by looking across rows.

For example, it is striking to note that central office staff who understood and were committed to the new program (a new reading laboratory) apparently did not supply adequate materials, training, or in-service to the people using the program; the users, naturally, were quite bewildered. (We show more of the analysis shortly.)

If there are multiple cases, you can assemble a full set of such tables, one for each case, and compare them on

Table 5.1
Event Listing, Banestown Case

LEVEL	TIME PERIODS CONTEXTUAL PRESS 1976-78	EMERGENCE OF THE PROBLEM Oct. 1978	AWARENESS AND PROPOSAL OF SOLUTION Nov. 1978	APPROVAL AND PREPARATIONS Jan.-Feb. 1979	TRAINING AND BEGINNING OPERATIONS March-April 1979	EXPANSION, NEW OPENINGS Sept. 1979	BUDGET REDUCTION, DISRUPTION May 1980
State/ Macro	minimal competency levels, testing introduced in state schools			proposal discussed, approval at state level	Smithson middle school teachers, 2 admins. trained at D/D site (4 days, early March)		Reduction in Title I allocations
District	supplemental skills program introduced in reading and math	- alarm at failure rate - internal solutions proposed, found unacceptable	* officials see SCORE-ON at 'awareness fair' - IV-C proposal rapidly drawn up, submitted	* Smithson pupil folders screened - appointments made of Smithson lab teacher and aide	- 30 4th grade pupils selected for Smithson lab - materials, technical assistance intènsified for Smithson lab	- staff active in extending Smithson, launching new labs * funding for all lab staff at Smithson taken over by Title I	* reduction in county, Title I budgets - proposed staff cuts, transfers in elementary schools
Local Schools	pressures begin to raise minimal levels			* continuation for following year planned for 5th grade in 2 middle schools; teachers named	- rooms, staffing completed for middle schools - 2 other elementary schools authorized to implement in the fall	lab opens at Carrington, Banestown Middle. modified version opens at Smith Camp, South End	- middle schools unaffected by cuts - threat of discontinuation at Banestown Middle (conflicts)
Smithson School	large numbers of low achievers placed in FACILE classes	* 4th grade teachers report 40 pupils 1-3 grade levels behind - teachers unfavorable to central office proposals	- teachers approve pull-out lab formula	- lab teacher and aide replaced; some disgruntlement - lab room created, minimally equipped	- Smithson lab opens (late March) - preparations inadequate, materials not arrived, scheduling difficulties	Smithson expands to 45 pupils in 3rd, 4th grades - new teacher added for morning sessions	* major shifts in lab staffing announced to teachers - program to be cut back, focused on grades 1-3, limited to 1.5 posts

* barometric event

each supporting condition and/or on a total "preparedness" index. We discuss such examples in Chapters 7 and 8.

Matrices can be more heavy-duty. Table 5.3 is aimed at understanding the effects of "assistance" supplied to a school site by various sources. Its format calls for the researcher to address five related variables, to distinguish two of them according to time, to pool responses, to align some responses along an evaluative scale, and to explain the response pattern for each type of assistance source. Here condensed information from 20-30 pages of field notes has been packed into a single page.

Note that the data are more abstract: there are no quotes, and generalizations and other inferential remarks appear in the last two columns.

The "consequences" and "explanations" are not direct condensations of informants' remarks or of researchers' observations. Rather, for any given consequence, such as the one in the top row, "users are helped administratively and substantively, feel obliged to do ECRI with minor adaptations," the researcher has looked at the data segments in the three preceding columns, checked to see whether they covary in some patterned way, and drawn a second-order generalization. In this case (see first row) the themes of "relieving pressure," "easing schedules," "feeling backed-up," "controlling fidelity," and users' positive assessment all suggested the reception of help and a sense of user obligation for reasonably faithful implementation.

A similar process of inductive inference occurs for "researcher explanations." Here, too, we see that the table collects data for easy viewing in one place, permits detailed analysis, and sets the stage for later cross-case analysis.

Illustrative display formats: Networks. A *network* is a collection of "nodes" or points connected by lines ("links"). They are very helpful when you want to focus on more than a few variables at a time. Networks can be set up easily and can hold a great deal of readily analyzable information. (For helpful discussion see Bliss et al., 1983).

Table 5.2
Checklist Matrix: Conditions Supporting Preparedness at Smithson School, Banestown Case

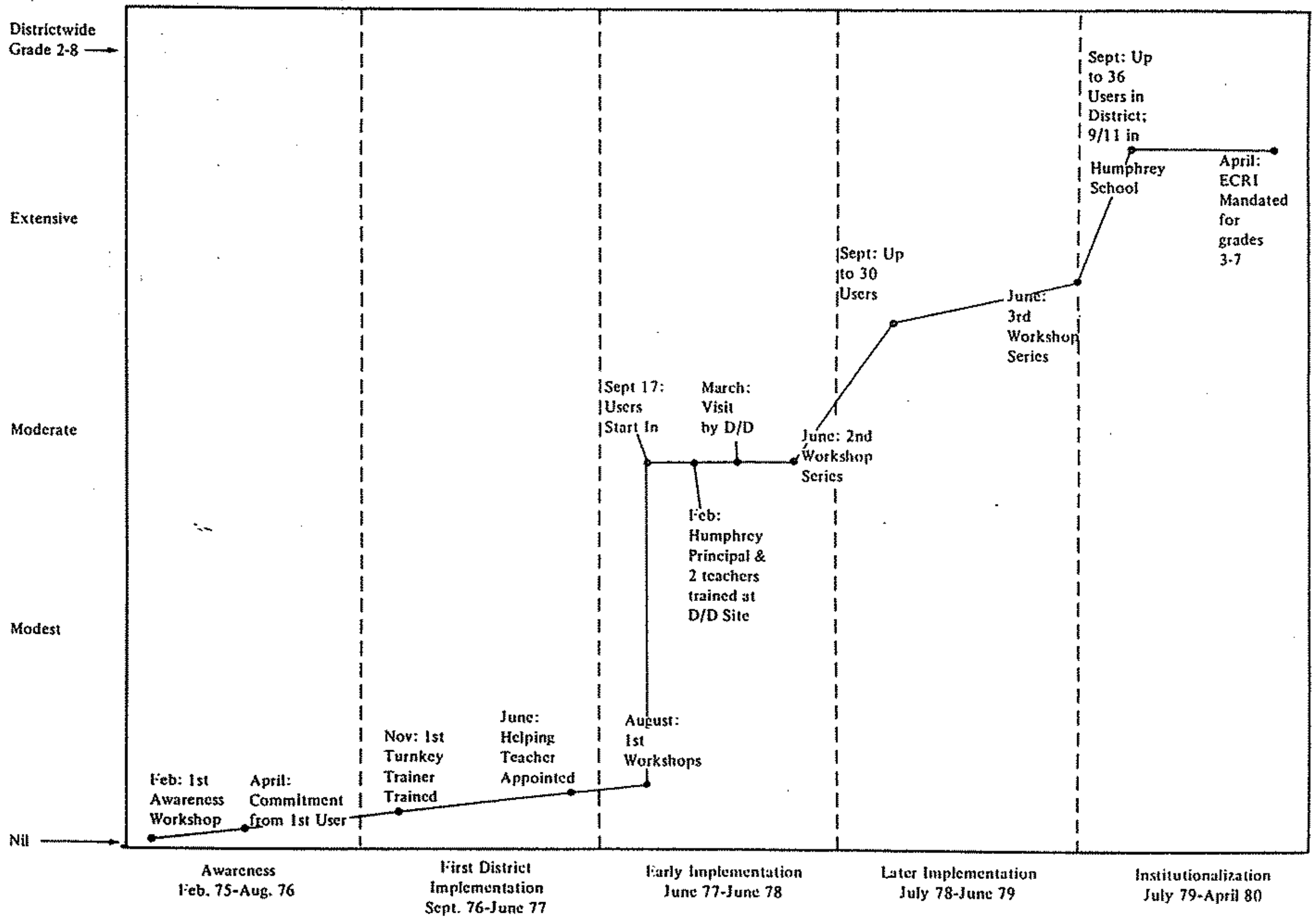
CONDITION	Presence of Supporting Conditions FOR USERS	FOR ADMINISTRATORS
Commitment	Strong - "wanted to make it work."	Weak at building level. Prime movers in central office committed; others not.
Understanding	"Basic" ("felt I could do it, but I just wasn't sure how.") for teacher. Absent for aide ("didn't understand how we were going to get all this.")	Absent at building level and among staff. Basic for 2 prime movers ("got all the help we needed from developer.") Absent for other central office staff.
Materials	Inadequate: ordered late, puzzling ("different from anything I ever used"), discarded.	N.A.
Front-end training	"Sketchy" for teacher ("it all happened so quickly"); no demo class. None for aide: ("totally unprepared.. I had to learn along with the children.")	Prime movers in central office had training at developer site; none for others.
Skills	Weak-adequate for teacher. "None" for aide.	One prime mover (Robeson) skilled in substance; others unskilled.
Ongoing inservice	None, except for monthly committee meeting; no substitute funds.	None
Planning, coordination time	None: both users on other tasks during day; lab tightly scheduled, no free time.	None
Provisions for debugging	None systematized; spontaneous work done by users during summer.	None
School admin. support	Adequate	N.A.
Central admin. support	Very strong on part of prime movers.	Building admin. only acting on basis of central office commitment.
Relevant prior experience	Strong and useful in both cases: had done individualized instruction, worked with low achievers. But aide no diagnostic experience.	Present and useful in central office, esp. Robeson (specialist).

Table 5.3
Effects Matrix: Assistance Location and Types (Masepa Case)

LOCATION	USER'S ASSESSMENT	TYPES PROVIDED	SHORT-RUN EFFECTS (USER'S 'STATE')	LONGER-RUN CONSEQUENCES	RESEARCHER EXPLANATIONS
Building Administration	++ ++ - +	1. authorizes changes 2. eases schedules 3. controls fidelity 4. consults, offers solutions	1. relieves pressure, encourages 2. helps early implementation 3. feeling policed 4. feeling backed-up, substantially helped	users are helped administratively and substantively, feel obliged to do ECRI with minor adaptations	administration, authority, servicing, availability and flexibility lead to sustained, faithful implementation model
Central Office Administration	+ ++	1. promotes ECRI 2. answers building administration, trainers' requests	1. pressures non-users 2. building administrators have material, administrative support	program is perceived as supported, assisted, 'protected' by central office	central office able to push program and answer requests, yet not perceived as main actor by users
Helping Teacher	++ + ++ ++ +-	1. provides materials 2. demonstrates, models 3. answers requests 4. encourages 5. circulates, controls	1. reduces effort, increases repertoire 2. trains, facilitates use 3. problems solved rapidly 4. maintains level of effort 5. ambivalent: helped yet coerced	new, experienced users receive systematic instruction, follow-up, materials; stay with program and are careful about making changes in it	personalized in-service mechanism, with both training and assistance allows for mastery and spread of ECRI in 'faithful' format
User-Helping Teacher Meetings	++ + + +	1. comparing practices with others 2. debugging, complaining 3. learning about new parts 4. encouragement	1. encourages, regulates 2. cathartic, solves short-run problems 3. expands repertoire 4. gets through rough moments	creates reference group, gives users a voice, solves ongoing problems and lowers anxiety	multi-purpose forum which consolidates use and users, defuses opposition
Teacher-Users in Other Schools: Target Schools	+ + +	1. sharing materials 2. exchanging tips, solutions 3. comparing, encouraging	1. increases stock 2. new ideas, practices; problems solved 3. motivates, stimulates	increases commitment, regulates use (decreases deviance)	additional source of assistance, which increases as number of users grows
Trainers in Target School, Other School	++ ++ + +	1. tips for presentations 2. solution to short-term problems 3. encourages 4. serves as successful model	1. facilitates practice 2. helps expand beyond core format 3. maintains effort 4. stimulates	reliable, unthreatening backup provided in school	elaborate and effective lateral network: trainers seen as peers

<u>Legend</u>:
++ = very effective
\+ = effective
+- = mixed effective
\- = ineffective

Figure 5.2
Growth Gradient for ECRI Innovation, Masepa Case



In Figure 5.2 we see a simple network display designed to show the "growth gradient" of an innovation's use in one case in our school improvement study. At first it doesn't look like a "network" at all. But it is: The nodes or points are events, and the lines or links have the implicit meaning "is followed by." The network is ordered (the vertical dimension shows numbers of users, and the horizontal one is time). Each point is labeled. You can get a clear idea of the tempo of expansion and can see which events were especially critical (e.g., the August 1st workshops in 1977). More generally, you can see that workshops appear to be the main mechanism for expanding use of the innovation.

In another part of our study, we were interested in how to display the sources of "assistance" that users of an innovation received (and requested). Because assistance occurs in an organizational context, we decided on showing a quasi-organizational chart with arrows showing where assistance of various types came from and who received it (Figure 5.3).

This network incorporates a large amount of information (see legend). It shows the organizational position of givers and receivers of help; depicts whether the assistance was solicited or not, mutual or not; labels the type of assistance received; and indicates how the receiver felt about it.

In this display we can see patterns readily: that virtually all assistance was received positively; that assistance received included all types except process facilitation; that much of it was solicited; that the helping teachers and the curriculum coordinator form a mutually helping team; and that the principal, too, gives assistance, but of a more ambiguous sort (more control, advocacy, resource giving, along with support and solution giving). Note, too, the within-building help among teachers in the form of support and solution giving.

This network was produced from about two dozen coded pages of transcribed field notes. The analyst retrieved the relevant chunks via codes and condensed them into categories, ratings, and directional influences.

Timing. When should formats be generated? If the format is at hand during data collection, it helps save energy by encouraging focus: You know what information you will be needing. But there are some cautions here. First, qualitative data evolve; later accounts round out, qualify, put in perspective, and disqualify earlier ones. Analysts do not march through data collection as in survey methodology; they scout around, sleuth, and take second and third looks. So there are risks in entering the data into a set format too soon.

Furthermore, for any given research question or issue, you can develop many different displays (like, dozens), using the same set of variables. Each display makes somewhat different assumptions; each has trade-offs among advantages and costs. We give an example of the typical pluralism among possible display types in section B below.

Another caution is that display formats nearly always evolve, too. The later ones are more data-sensitive than the earlier ones, as things come clearer. For instance, the headings in the display format of Table 5.2 came, in part, from relationships observed in the course of data collection; the display allows you to test those relationships more rigorously. You could have built such a format directly from a conceptual framework, but real clarity typically comes later, when the empirical data begin to take shape.

Formatting is a decisive operation; it determines which variables will be analyzed in which ways. If a variable isn't in the format, it won't get compared to another variable. Displays can help you summarize and compare findings within (and across) cases, but they also can be straitjackets. They may bully the data into shapes that are superficially comparable across cases, but you actually may be comparing intrinsically different things—on dimensions that turn out to be trivial.

So our general advice is to generate rough formats early during data collection, and revise them to a firmer state closer to the end of data collection, when they can be grounded more contextually and empirically. Expect to make several passes or iterations before the first format of the display is working right. Test a proposed format by entering data. Unworkable or confusing formats, or those that do not incorporate all of the relevant data, will show themselves rapidly.

Entering Data

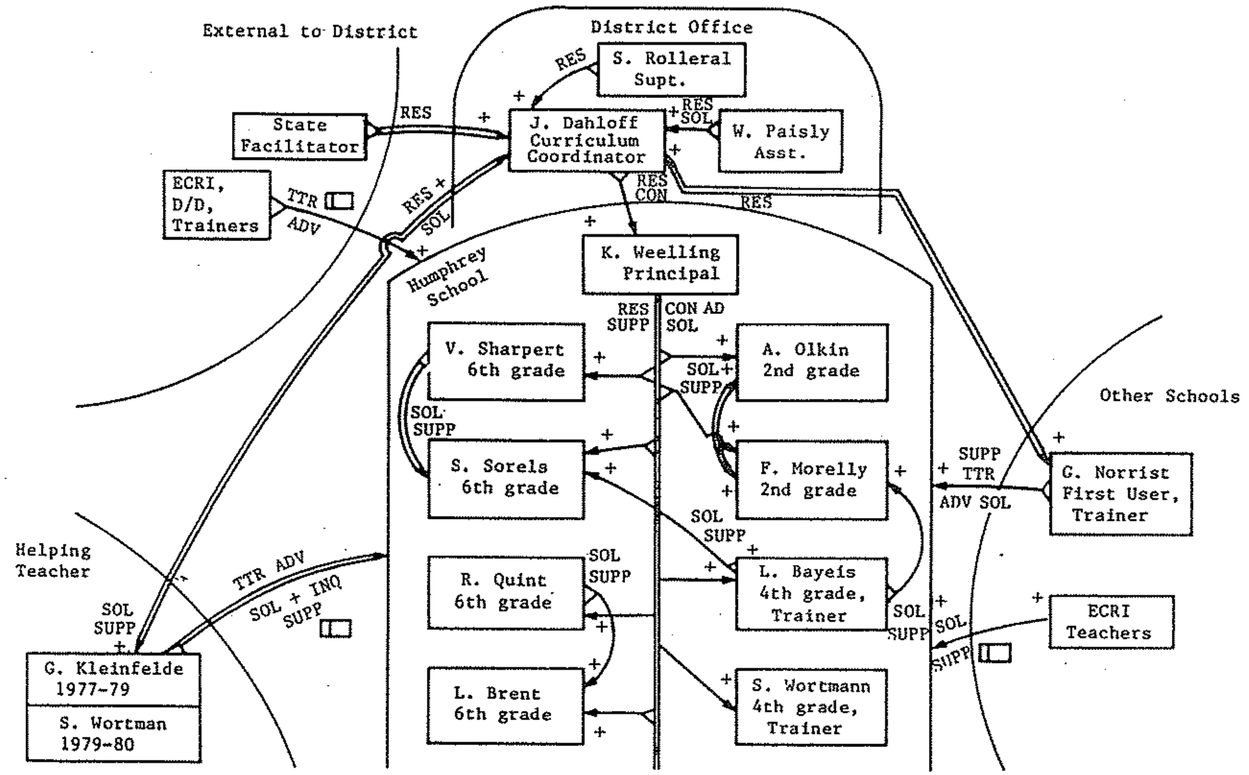
Creating formats is usually a matter of a few minutes. Revising them as the early items of data are entered is also quick work, especially if your word processor can easily add and switch rows and columns. Just as exploratory data analysis programs (Hartwig & Dearing, 1979) for quantitative data permit the "sculpting" of analysis and display frames to fit the shape of the data as they undergo progressive transformations, so good qualitative software programs can enable reformatting as data are entered and considered.

What takes time is data entry itself. Coded data segments have to be located in the transcribed field notes; they have to be extracted, condensed, and summarized. Some further data transformation and selection may also be involved: doing ratings, making judgments, picking representative quotes. The speed and quality with which data entry is done depend on the technology of storage and retrieval being used (see Chapter 3, section B). Generally speaking, computer-assisted methods both improve quality and reduce the time taken.

The time taken to enter data also depends on (a) the number of variables or dimensions in the display, (b) the

Figure 5.3

Illustration of Context Chart with Assistance Flows (Masepa Case)



Legend

Direction of assistance

- assistance received
- mutual assistance
- solicited and received
- heavy, frequent

Orientation

TTR = teaching, training
FAC = facilitating, process-helping
SOL = solution-giving
RES = resource-adding
CON = controlling
SUPP = supporting, encouraging

☐ = ECRI materials
ADV = advocating clients' interests
INQ = inquiring, feeding back formatively

number of respondents, and (c) the kind and number of transformations made. We'll give time estimates as we proceed with specific methods. As an illustration, a complex matrix like Table 5.3, assuming already-coded data, might take an experienced researcher a full day for data entry, given a database of a few dozen interviews. On the other hand, one like Table 5.1, given information from a few informants and some documents, could be managed in a few hours. These figures might be cut by 40% or so with the right software to pull out coded chunks quickly and to create condensed or transformed versions—and build rows and columns or a network structure for easy data entry.

We suspect that, for most people, a traditional extended-text approach to the same data set would not be accomplished more quickly. And the formats illustrated here, in our view, substantially increase the analytic power of the results—and their readability.

In this overview we do not discuss in detail just how you go about selecting/reducing data from field notes for entry into matrices or figures. A good rule of thumb has been offered by Tufte (1990), the dean of display makers: "to clarify, add detail." In other words, *over*reducing data can obscure understanding.

We come back to this issue as we discuss specific display methods. Generally, the choice of data for matrix entry must be driven by the particular row and column headings involved, or by your definition of network nodes and links. Keeping a precise record of the actual criteria and decision rules used (which will vary from display to display) is essential. Typical decision rules cover how ratings or judgments are arrived at, the basis for selecting particular quotes, and how you balance reports from different people.

Stay open on the question of whether the format you use for analysis should be the same one used for final reporting. The best answer is often yes (just as a quantitative research report normally includes data displays, as well as analytic text); readers can see for themselves how conclusions were drawn, rather than being handed study results to be taken on faith. To put it another way, as Wainer (1992, p. 18) does: "Tables are for communication, not data storage." Sometimes, of course, a display will be used for intermediate purposes and will not necessarily be shown to readers of the final report.

Drawing Conclusions

The displayed data need to be examined with care, and conclusions need to be drawn. Those conclusions normally appear in what we called analytic text. For example, the analyst looking at Table 5.1 wrote paragraphs such as these:

> Beginning in 1971 and intensifying in the years 1976-78, the State issued a series of increasingly precise and constraining documents. These texts called at first for adequate "standards of quality" in education. These standards were then outlined in the major subject matter areas. Competency testing was announced, first for grades K-3, then for 4-6. . . . After some delays, the first statewide tests keyed to these objectives for reading and math were scheduled for grade 4 in 1979 and for grades 5-6 in 1980-81.
>
> The response at the district level was to create a "supplemental skills program" in order to intensify remedial work in reading and mathematics so that minimal competency levels could be attained. . . .
>
> The implication was in the air that teachers might be held accountable for excessive rates of failure on statewide competency tests, but the pressures were still diffuse within individual schools. Teachers routinely attributed low performance to subnormal aptitude or to conditions at home over which they had no control.
>
> A special impetus was given in the fall of 1978, when the six fourth-grade teachers at Smithson noticed that they had an unusually large cohort of incoming pupils who were one or more grade levels behind in reading achievement. . . . It is not clear how so many of these pupils got to the fourth grade, but no one was surprised. The Vice-Principal says, "They got too tall, they got too bad."
>
> The teachers were worried that either promoting or retaining so many pupils would cause problems; they were leaning toward retention, but feared a massive protest by parents. Essentially, they were covering themselves, by announcing early in the year that they had inherited, not created, the problem.

These excerpts show us that the analytic text draws attention to the features of the displayed data and "makes sense" of them, knitting them together and permitting the analyst to draw conclusions and to add interpretations. It also encourages a return to the field notes to consult information not in the display and to add it to the text for further clarity. In short, the display does not "speak for itself"—and the analytic text does not stand alone without reference to the display.

How are conclusions actually drawn? In this example the analyst was using several tactics. One was **noting patterns, themes** (the idea that teachers were covering themselves in advance); another was **building a logical chain of evidence** (tracing the downward impact from state actions to the local school). These and other tactics are discussed in detail in Chapter 10. As we illustrate their use in specific situations in the next few chapters, tactics will be marked off in **boldface.**

The time required for examining data displays and writing the corresponding analytic text usually runs to 20% or less of the time required for data entry. There are many reports of the anguish of trying to write text from an un-displayed database of hundreds of pages. By contrast, well-formatted displays, with data entries in which you have confidence, make writing an easier, faster, more pleasurable and productive experience for the analyst, judging from our experience.

Moving to the Next Analytic Steps

The act of writing text as you ruminate over the meaning of a display is itself a focusing and forcing device that propels further analysis. We quote Lofland and Lofland (1984, pp. 142-143) here:

> It seems, in fact, that you do not truly begin to think until you attempt to lay out your ideas and information into successive sentences.... For better or for worse, when you actually start writing you begin to get new ideas, to see new connections, to remember material you had forgotten.... You are never truly inside a topic—or on top of it—until you face the hard task of explaining it to someone else. . . . (Thus "writing blocks" are perhaps better thought of as "thinking blocks.")

Or, as Mills (1959) points out in his essay "On Intellectual Craftsmanship," writing begins in the context of *discovery* and then must turn to the context of *presentation.* That effort often turns us back to the context of discovery once more. Writing, in short, does not come after analysis; it *is* analysis, happening as the writer thinks through the meaning of data in the display. Writing is thinking, not the report of thought.

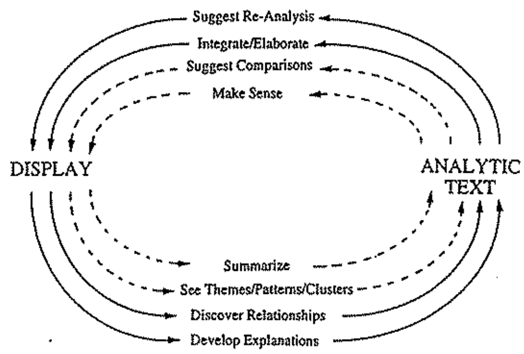
In Figure 5.4 we summarize the interaction of the display and the emerging analytic text.[2] Looking at the display helps you summarize and begin to see themes, patterns, and clusters. You write analytic text that clarifies and formalizes these first findings, helps make sense of the display, and may suggest additional comparisons to be made via the display.

Those new comparisons, in turn, help you discover new relationships and propose explanations. The resulting analytic text then may suggest ways of elaborating the display, integrating or rearranging it to permit what amounts to a reanalysis. That reanalysis helps deepen the initial explanations.

Choosing a Display Type

This discussion of displays and formatting has served as a general introduction to the substance of this chapter, which is concerned with displays for within-case descriptive analysis. We now move on to specific displays.

Figure 5.4
Interaction Between Display and Analytic Text



Eight major methods, along with ten others described briefly in boxed format, are involved. How do you choose an appropriate display when the aim is to describe a case well?

A project may be primarily in an exploratory, opening-up stage: If so, it usually makes sense to build a *partially ordered* display, one that does not demand much pre-specification of variables. In section B we discuss the *context chart* and the *checklist matrix.*

It may be that time is a crucial aspect of the study at hand. If you need to describe the flow of events and processes carefully, then a *time-oriented* display is essential. In section C we explain the *event listing, event-state network,* and other time-oriented matrices and networks.

It may also be that describing a setting carefully will require close attention to the interaction of people in their *roles.* In section D we deal with *role-ordered matrices.*

Finally, when a study is less exploratory and involves a fairly clearly defined set of key variables, it pays to develop a *conceptually oriented* display. In section E we review the *conceptually clustered matrix,* the *cognitive map,* and the *effects matrix.*[3]

These are four main types of descriptive display. Some displays combine features across these types (e.g., a role by time matrix). The names may sound like jargon, but don't worry about them; the issue is how any given display works and how it can further your analysis. All of the displays can be adapted to meet the needs of any particular study; if they seem too complex, simplify. Take what is useful.

You may want to review Chapter 9 as well, which includes summary ideas on matrix building and use. Chapter 13 will help you keep an overall, longer term perspective on the flow of work in your study.

B. Partially Ordered Displays

Some displays deliberately have some—but not too much—internal order. They aim to uncover and describe what is happening in a local setting, no matter how messy or surprising it may be. And they impose minimal conceptual structure on the data they display. Take, for example, the context chart.

Context Chart

Analysis Problem

Most qualitative researchers believe that a person's behavior has to be understood in a specific context, and that the context cannot be ignored or "held constant." The context can be seen as immediately relevant aspects of the situation (where the person is physically, who else is involved, what the recent history of the contact is, etc.), as well as the relevant aspects of the social system in which the person appears (a classroom, a school, a department, a company, a family, a hospital ward, a local community). Focusing solely on individual behavior without attending to contexts runs a serious risk of misunderstanding the meaning of events. Take the individual behavior noted below:

> Bill picks up telephone, listens for a while, says, "No, Charlie, I don't think we should do that. See you tomorrow. Bye."

Your interpretation of this event, without context, has to be limited to matters such as individual decisiveness, economy of speech, and the like. So, as you read this, you probably constructed a context to help you interpret it. Reflect on that context for a moment and decide what the event means to you.

Now consider some alternative contexts:

> Charlie is Bill's attorney, and they are going to court tomorrow. (The issue is something about a strategy to be followed, but Bill is certainly not a passive client.)
>
> Charlie is Bill's boss. (Is Bill giving requested advice or, in effect, vetoing something?)
>
> Charlie is Bill's subordinate, who is proposing that someone new should be invited to a meeting. (It feels like an abrupt dismissal.)
>
> Charlie is Bill's wife, calling from out of town to suggest meeting for a gallery visit before dinner tomorrow. (What kind of marriage is implied? Distant? Routine? Independence respecting?)

Understanding contexts is usually critical. Even that adjective is too mild. We should quote here the title of Mishler's (1979) thoughtful article: "Meaning in Context: Is There Any Other Kind?" Contexts drive the way we understand the meaning of events, or, as Mishler notes, meaning is always within context and contexts incorporate meaning.

The problem a qualitative researcher faces is how to map the social context of individual actions economically and reasonably accurately—without getting overwhelmed with detail. A context chart is one way to accomplish these goals.

Brief Description

A *context chart* is a network, mapping in graphic form the interrelationships among the roles and groups (and, if appropriate, organizations) that go to make up the context of individual behavior.

Illustration

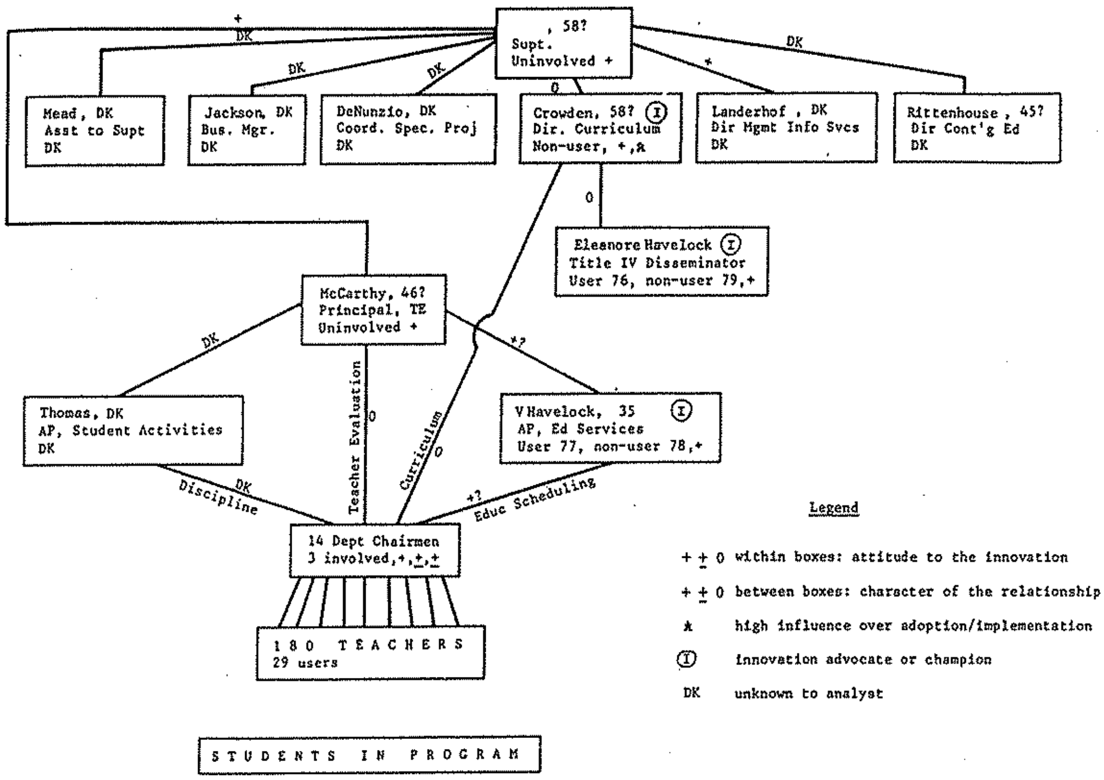
Most people do their daily work in organizations: They have superiors, peers, subordinates; their work is defined in a role-specialized way; and they have different relationships with different people in other roles in their social vicinity. (Context charts also can be drawn for people in families or informal groups or communities.)[4]

Suppose you were interested, as we were, in organizations called schools and school districts—and with the general problem of how innovations enter and are implemented in those organizations. What are some simple ways to display data to help our understanding of those contexts?

Building the display. Such a network ought to reflect the core characteristics of organizations: authority/hierarchy and division of labor. So it ought to show who has formal authority over whom and what the role names are. But those things don't tell us very much. We should also know about the quality of the working *relationships* between people in different roles.

And because we're talking about the introduction of an innovation, the display should show us who advocated it, who is actually using the innovation, and people's attitudes toward it (whether or not they are using it). The display should show us how the specific school we are studying is embedded in the larger district organization. Above all, we need a display that will not overload us with information,

Figure 5.5
Context Chart for Tindale East High School and District



but will give us a clear, relevantly simplified version of the immediate social environment.

Figure 5.5 shows how these requirements were met after a field-worker made a first visit to Tindale East, a high school involved in implementing a new reading program. The analyst selected out the roles and groups that are most critical for understanding the context. District office roles are above, school roles below. The network is thus partially ordered by roles and by authority level.

For each individual, we have a name, age (a feature the analyst thought was important in understanding working relationships and career aspirations), a job title, whether the individual was a user of the innovation or not, and his or her attitude to the innovation (positive, ambivalent, neutral). Special symbols are applied when the individual was an innovation advocate or influenced implementation strongly. The relationships between individuals are also characterized (positive, ambivalent, neutral). Once past the upper echelons, the display simply counts individuals without giving detail (a secondary context chart at the level of individual teachers was also developed, but is not shown here).

Entering the data. To get the data, the analyst consults field notes and available organization charts and documents. The decision rules look like this:

1. If there is ambiguous or unknown information, enter "DK."
2. A relationship rating (how X gets along with Y) should not be disconfirmed by the other party to the relationship, though it need not be directly confirmed.
3. The "innovation advocate" and "high influence" ratings should be given only if there is at least one confirmation and no disconfirmations.

4. For information such as job title, number of persons, and so on, assume accuracy for the moment and enter it.

Drawing conclusions. The analytic text for this display looked like this, in part:

> Looking at lines of authority, we can see that only one central office person (Crowden) has direct authority over department chairmen as they work on the innovation. Crowden is not only an advocate but also has high influence over implementation, and seems to have a license from the superintendent to do this.
>
> The department chairs, it appears, have three other masters, depending on the immediate issue involved (discipline, teacher evaluation, scheduling). Because, in this case, the innovation does involve scheduling problems, it's of interest that V. Havelock is not only an advocate, but has actually used the innovation and is positive toward it. We might draw the inference that Crowden serves as a general pusher, using central office authority, and V. Havelock aids directly with implementation issues; the field notes support this.
>
> Note, too, that principal McCarthy (a) is not accountable to the superintendent for curriculum issues and (b) has a good relationship with V. Havelock. Perhaps McCarthy gets his main information about the innovation from Havelock and thus judges it positively.
>
> The fact that the three chairs involved are ambivalent to positive about the innovation will require more data than displayed here to elucidate. A next-level context chart should display the three chairs, plus the 29 users and *their* attitudes.

So the chart shown in Figure 5.5 helps us place the behavior of individuals (e.g., Crowden, V. Havelock) in a clear context and understand its meaning. For example, when Crowden, discussing the innovation, says, "It is not to be violated; its implementation is not based on the whim of a teacher at any moment in class, and its success is not dependent upon charismatic teachers," the chart helps us understand that this prescriptive stance is backed up with direct authority over department chairs for curriculum issues—an authority that is accepted neutrally. In short, the analyst has been employing the tactic of **seeing patterns or themes,** as well as **subsuming particulars into the general** (see Chapter 10, section A, for more on these tactics).

So it is not surprising when a department chair from Tindale says she tells teachers, "If you want to depart from the guide, ask me and also tell me why you want to do it and how it will fulfill the guide's objectives." And she admits to the researcher, "The Tindale West department chair and I mostly decide 'no' before teachers even ask to make changes."

Variations

The specific types of *behavior* involved in the relationships between roles can be identified on the chart as well (see Figure 5.3, where the primary focus is on the different types of assistance [e.g., teaching/training, facilitation]: who gave what kind of help to whom).

Dotted lines can be used to show informal influence. Circles can be drawn enclosing informal groups. In one study of a rural high school, we asked people to draw pictures of the faculty as they saw it, and interviewed them about the meaning of their drawings. This process helped us find six major informal subgroups among the faculty. They included the "old new guard" (progressive teachers who had been around a good while); "boy coaches" ("the last great disciplinarians in the world"); "girl coaches"; "newcomers"; "structuralists" ("a 1950-1960 kind of teacher"); and "goners" ("people who might just as well be in a labor union . . . at 3:30 they're gone"). We also found one person, James Colley, who was a group unto himself, according to several respondents.

Linkages to other organizations in the environment can be shown.

In an organizational context that has changed a lot in a short time, successive context charts can be drawn (e.g., in our study the structure and relationships in an alternative school changed rather substantially for each of the 3 years of its existence, and we needed three charts).

In Figure 5.5 certain *social-system* aspects of the context are emphasized. It is sometimes important to map the *physical* aspect of an immediate context (e.g., a classroom, with indications of the teacher's desk, resource files, student tables and chairs, entrances) in order to help understand the ebb and flow of events in the setting.

Advice

If you're new to qualitative research, keep your first context charts simple. They can be embroidered as you go on. Context charts work well when your case is an individual—they show you the real richness of a person's life setting.

Use context charts *early* during fieldwork to summarize your first understandings and to locate questions for next-step data collection.

Keep the study's main research questions in mind, and design the context chart to display the information most relevant to them.

Remember that you are not simply drawing a standard organizational chart, but are mapping salient properties of the context. Remember, too, that your chart will not be exhaustive or complete: It is a collection of organizational fragments or excerpts. (In Figure 5.5, for example, custodians, secretaries, and the immediate subordinates of most of the district office personnel are excluded.)

Time Required

If the fieldwork has involved focused attention to the issues mapped in a context chart, one of the size shown in our illustration can normally be put together quite rapidly—in an hour or so. You may often find that an implicit chart has been materializing in your head as you talk to different people in the site. Drawing it and filling in the associated details from coded field notes helps you express and test your implicit ideas.

Checklist Matrix

Analysis Problem

Let's look at some contrasting stereotypes. The *survey researcher* is seen as a purposeful, efficient worker who designs instrumentation, marches into a site to administer it to everyone in the sample, marches out, codes and processes the data, and analyzes the output. By contrast, the *field researcher* is seen as a near-aimless loiterer who throws together some orienting instrumentation, hangs around a field site for days, collects all manner of data, fiddles with the instrumentation, talks to some people more than to others, observes without an observation schedule, accumulates pages and pages of words, and then spends days coding and entering them in a complicated homemade chart for analysis—and does all this for only one or two cases, to boot.

Of course, we could turn these stereotypes inside out, speaking of the methodological rigidity and blindness of the survey researcher and the "groundedness" of the fieldworker, alert to local reality. There are grains of truth in stereotypes, whichever way they're slanted.[5]

Sometimes field researchers behave like survey researchers by collecting comparable data from all key respondents and entering them in a prespecified format. This may happen when:

- The variable is conceptually important.
- The variable can be unbundled easily into distinct indicators or components.
- You need the variable badly (e.g., as a good outcome measure).
- The study has multiple cases requiring comparability of formatting and measurement.
- You want to relate field data to survey measures of the same variable.

Brief Description

A *checklist matrix* is a format for analyzing field data on a major variable or general domain of interest. The basic principle is that the matrix includes several components of a single, coherent variable, though it does not necessarily order the components.

Illustration

Here is a checklist for assessing "preparedness" prior to executing a new practice (Table 5.2). In our school improvement study, we met most of the conditions just described. Preparedness was tied to implementation success in our conceptual framework and had strong backing in the literature. It also made good common sense that people do better those things for which they are prepared. Empirical studies had broken down preparedness to distinct components that correlated well with good project execution. In short, the variable had a lot going for it. In addition, we were studying 12 cases and needed parallel measures. We also wanted to link our findings with a large survey that had a preparedness section in its questionnaire.

We did have some doubts about the variable and saw fieldwork as the ideal way of resolving them. We wondered whether the preparedness → good execution logic was too mechanistic. We also suspected that other variables could get mixed up with preparedness. A highly experienced user could look well prepared, although the technical training itself may have been dismal. And we believed that many unpredictable school-level factors could well affect project execution. So a largely individual and technical factor such as preparedness could hardly account by itself for varying degrees of successful use of the new practice. These were all things that on-site observation and interviews could capture, and we needed a display to help us sort out these matters.

The two major research questions we wanted to answer were:

What were the components of the original plan for implementation?

> These might have included front-end training, monitoring and debugging/troubleshooting unexpected problems, and ongoing support. How precise and elaborate was this plan? Were people satisfied with it at the time? Did it deal with all of the problems anticipated?

Were the requisite conditions for implementation assured before it began?

These might have included commitment, understanding, materials and equipment, skills, time allocation, and organizational backup. Were any important conditions seen as missing? Which were most missing?

Building the display. Table 5.2 is repeated here. The rows are drawn from the various components of the implementation plan and the supporting conditions mentioned in the research questions. During early fieldwork we realized that "relevant prior experience" was an important component of preparedness, and added it as a row.

There is no effort to sort or order the components at this point. We believed that preparedness might well differ for users of the new program and administrators, so the columns reflect that distinction. The matrix is thus partially ordered by role.

In this case the innovation was a remedial lab program in reading, being used for the first time by a teacher and an aide. The cell entries include summary phrases, direct quotes, and an overall adequacy judgment by the analyst, ranging from "absent/none" to "sketchy," "basic," "adequate," "present," and "strong." The quotes are useful devices; they both help justify and illuminate the rating.

Entering the data. The analyst reviews relevant sections of the write-ups, retrieves material by codes, forms a general judgment or rating of the level of adequacy of the component at hand, and locates relevant quotes.

The quotes and the ratings are kept together in the cell. The idea here is that ratings only tell you how much there is of something, not what that something *means*. Brief quotes are enough to communicate—and to help another analyst judge, by going back to the write-ups, whether the rating is justified.

Some decision rules used for data entry:

- Because there are only two users and they occupy different roles (teacher and aide), don't force a general judgment. The same applies to administrators.
- Use judgments of users and administrators if they offer them. Put them in quotes.
- Accept direct reports and do not require verification by another person.

Drawing conclusions. Now we can look down the columns of the matrix and form a general idea of the level of preparedness for different components. For example, the analyst wrote, in part:

> The image is one of solid commitment and administrative support (except for the building principal), but minimal adequacy in requisite skills and understanding of the practice, and absence of forward-looking mechanisms for training, debugging or planning.

The analysis tactic here is one of **noting patterns or themes** (see Chapter 10). Note that simply adding up judgments in the column is not the way to proceed.

But the display suggests new questions. What's going on here? The central office administrators were committed and had more basic understanding of the program than the users (tactic: **making comparisons**). Why didn't they help? The analyst went back to field notes and noted:

> For Mrs. Baeurs [the administratively oriented, less skilled prime mover], all the lab teachers required were the provision of materials and initial training. . . . The teachers, she reasoned, had already done individualized work with children "so they didn't really have to do any kind of mind change." . . . At one point, however, she says, "I don't know. Maybe we got into it too fast."
>
> Even Mrs. Robeson, the reading specialist, acknowledged that the teachers "didn't think they were ready," but saw few ways in which initial preparation could have been better. She also played down the size of the needed changes in teachers.

Thus the analyst found more information: The weak, hasty training and support came because distant administrators minimized the actual needs of new users.

The analyst also consulted the survey data, finding that all respondents judged the list of conditions to have been "only partly in place." Administrators emphasized funds and materials; teachers had more doubts about whether there was motivation to achieve project purposes.

That finding, in turn, helped the analyst realize that a major missing row in the matrix was *support of other classroom teachers,* whose lack of commitment to and understanding of the lab program led to poor coordination, skepticism, and disgruntlement that was offset only partly by their personal liking for the lab teacher and aide.

Variations

Any given research question and set of variables can have many different displays. Each has advantages and limitations. Let's illustrate. The research questions here asked whether conditions supporting preparedness through an implementation plan were assured before implementation began. The variables included a set of "conditions" with data for users and administrators. Table 5.2 is only one possible display. Let's look at three others.

First, Table 5.4 recasts Table 5.2 by transposing rows and columns so we can easily follow the responses of a specific individual across all conditions, not having to lump them in one cell as in Table 5.2. Table 5.4 also breaks out the building principal separately. But note: We have much more detailed, diverse data, rather than a general judgment such as "adequate" for any given condition, as we had in Table 5.2.

Table 5.2 (Repeated)
Checklist Matrix: Conditions Supporting Preparedness at Smithson School, Banestown Case

CONDITION	Presence of Supporting Conditions: FOR USERS	FOR ADMINISTRATORS
Commitment	Strong - "wanted to make it work."	Weak at building level. Prime movers in central office committed; others not.
Understanding	"Basic" ("felt I could do it, but I just wasn't sure how.") for teacher. Absent for aide ("didn't understand how we were going to get all this.")	Absent at building level and among staff. Basic for 2 prime movers ("got all the help we needed from developer.") Absent for other central office staff.
Materials	Inadequate: ordered late, puzzling ("different from anything I ever used"), discarded.	N.A.
Front-end training	"Sketchy" for teacher ("it all happened so quickly"); no demo class. None for aide: ("totally unprepared.. I had to learn along with the children.")	Prime movers in central office had training at developer site; none for others.
Skills	Weak-adequate for teacher. "None" for aide.	One prime mover (Robeson) skilled in substance; others unskilled.
Ongoing inservice	None, except for monthly committee meeting; no substitute funds.	None
Planning, coordination time	None: both users on other tasks during day; lab tightly scheduled, no free time.	None
Provisions for debugging	None systematized; spontaneous work done by users during summer.	None
School admin. support	Adequate	N.A.
Central admin. support	Very strong on part of prime movers.	Building admin. only acting on basis of central office commitment.
Relevant prior experience	Strong and useful in both cases: had done individualized instruction, worked with low achievers. But aide no diagnostic experience.	Present and useful in central office, esp. Robeson (specialist).

Table 5.4
Checklist Matrix on Preparedness (alternative format 1)

		PSYCHO-SOCIAL CONDITIONS						IMPLEMENTATION PLAN				
		Relevant Prior Experience	Commitment	Understanding	Skills	School Admin. Support	Central Admin. Support	Materials	Front-End Training	Ongoing In-Service	Planning, Coordination Time	Etc.
Users	1											
	2											
Building Administrator (principal)	1											
	2											
Central Office Administrator	1											
	2											
Other Central Office Staff												

Table 5.5
Checklist Matrix on Preparedness
(alternative format 2)

CONDITIONS	USERS	ADMINISTRATORS
Missing		
Weak		
Adequate		
Strong		

Table 5.6
Checklist Matrix on Preparedness
(alternative format 3)

CONDITIONS	EXAMPLES[a]	HOW IMPORTANT[b]	WHY IMPORTANT[c]
Commitment			
Understanding			
Materials			
Training			
Etc.			

a. Specific illustrations, marked with A or U for administrator or user, respectively.
b. Rating -- very, quite, somewhat, or not important.
c. Explanations of reasons for importance of condition, given by respondent (A or U) or researcher (R).

Second, the analyst has clustered the conditions into two general domains: psychosocial and the implementation plan. That clustering permits comparing these more easily.

The cell entries for Table 5.4 are a rating of "strong," "moderate," "weak," and "absent," accompanied by a brief explanatory phrase showing what the condition consisted of, as the respondent saw it. (Under "materials," for example, a user might have said, "basic notebook, plus the file cards and prep sheets.")

Table 5.5 takes a different cut. The cell entries here are the names of *conditions,* together with the N of users or administrators reporting. An asterisk can appear beside condition names in the first row seen as "most" missing, and a number symbol (#) beside those seen by respondents as critical for success.

Table 5.5 thus sorts the conditions by *role,* according to how "present" they were, and lets us see which were important. But note that we now have lost the data for individuals that appeared in Table 5.4. It is also much harder to track the status of any particular condition. Matrix building is full of this sort of trade-off.

Table 5.6, a third possibility, de-emphasizes the differentiation among administrators and users and takes a decisive turn toward the *dynamics of conditions*—how and why they are important. It also brings in the researcher's view of the situation. We have retained the idea of specific examples, but we have lost identifiable individuals and the capacity to make easy comparisons between users and administrators.

Any data set can be displayed in many different ways. Each way gains you some things and loses you others. The moral: Do not close up too rapidly on a display format. Try several iterations; get colleagues' ideas and reactions.

Components in a checklist matrix, although they begin as unordered, may sometimes have a meaningful structure. They may fall into several clusters, as in Table 5.5, or even can be ordered from peripheral to central, or weak to strong.

Columns also can refer to levels of a site (e.g., classroom, school, district).

Especially for exploratory work, it pays to have a column headed simply "Remarks," where relevant commentary of any sort can be included to aid understanding.

Advice

Checklist matrices are good when you are exploring a new domain. If you have some rough ideas about some key variable and some first components of it, you can begin and then amplify the display as you learn more.

In fact, a checklist format itself does a good deal to make data collection more systematic, enable verification, encourage comparability, and permit simple quantification where you think it is appropriate.

Always consider several format possibilities: Ask a colleague for ideas. Look at what different formats give you and what they withhold from you.

Be explicit about the decision rules you follow in (a) selecting quotes and (b) making judgments or ratings. Otherwise colleagues or readers will be dubious about your conclusions—and you probably will come to doubt them yourself. Keep quotes and ratings together in the matrix so that their mutual pertinence can be assessed.

Time Required

If the written-up field notes have been coded to permit easy retrieval of chunks related to each component of a checklist matrix, and if fewer than a dozen or so informants are involved, a display like this can be assembled in 1-3 hours and analyzed rapidly. Most of the hard work—refining the components of the matrix—will have been done during instrumentation and early data collection.

Again, the time estimates are a reasonable range, in our experience. Just how much time it takes depends on the computer software in use, the researcher's skill, the theory underlying the study, the particular research question at hand, the completeness of the data, the number of informants, the number of cases, and so on.

The Transcript as Poem

As a break from the sorts of complexly crafted displays we have been looking at, consider Box 5.1. Richardson (1992) interviewed Louisa May, an unmarried mother, and converted 36 pages of interview notes into a focused, partially ordered display: a 3-page, 5-stanza poem, of which the first appears here. Richardson notes that she used *only* Louisa May's "words, her tone and her diction, but rel[ied] on poetic devices such as repetition, off-rhyme, meter, and pauses to convey her narrative" (p. 126).

The display is striking. It brings the reader very close to a condensed set of data, has a compelling flow, and forbids superficial attention by the analyst. You have to treat the data set—and the person it came from—seriously because a "poem" is something you engage with at a deep level. It is not just a figurative transposition, but an emotional statement as well.

As Richardson notes, such a display "breaches sociological norms" by ignoring defined variables, by emphasizing the "lived experience," illuminating the "core" of the case involved, engaging the reader (and the researcher) emotionally, and shifting the concept of authorship.

The time taken was substantial: 4 weeks, with nine drafts and four critiques by others. Two points: (a) the selection, organization, and presentation of data in a display are decisive analytic actions, and (as in this case) they need to be done in a thoughtful, lucid way; (b) displays owe as much to art and craft as they do to "science."

Richardson wrote very little separate "analytic text," in our terms, but there is nothing in the method to discourage that.

Box 5.1
Excerpt From a Transcript as Poem
(Richardson, 1992)

LOUISA MAY'S STORY OF HER LIFE

i

The most important thing
to say   is that
I grew up in the South.
Being Southern shapes
aspirations   shapes
what you think you are
and what you think you're going to be.

(When I hear myself, my Ladybird
kind of accent on tape, I think, OH Lord.
You're from Tennessee.)

No one ever suggested to me
that anything
might happen *with* my life.

I grew up poor in a rented house
in a very normal sort of way
on a very normal sort of street
with some nice middle-class friends

(Some still to this day)

and so I thought I'd have a lot of children.

I lived outside.

Unhappy home. Stable family, till it fell apart.
The first divorce in Milfrount County.

So, that's how that was worked out.

C. Time-Ordered Displays

A second major family of descriptive displays orders data by time and sequence, preserving the historical chronological flow and permitting a good look at what led to what, and when.

Event Listing

Analysis Problem

Life is chronology. We live in a flow of events. Some of these events occur before other events, some after. We usually believe that connections exist between events (e.g., deciding to renovate a bathroom leads to taking out a loan, the arrival of workers, the knocking down of a ceiling, and the intrusion of plaster dust into other rooms).

But although we can think of ourselves as being in the midst of a river of events, that metaphor breaks down, because the river's flow is not one-dimensional. Some events occur in one domain of life, others elsewhere (the plaster dust intrusion occurs while one is revising the last chapters of a new book). Some events are close to us, some distant (while the loan is being taken out, the 1984 U.S. federal budget deficit is projected as $191 billion). Some events are related coherently to other events (the ceiling removal and the plaster dust), and others are adventitious (a loved old cat dies).

Life goes on. Years later, the plaster dust is long gone, the book is being rewritten, and the federal budget deficit for 1992 is projected at $334 billion. Events long ago in time have consequences for the present (there is a new cat closely resembling the old one). Distant events can have consequences in close events (one writes a letter to the *Times* denouncing the newspaper's typically misleading allusion to "the" deficit, rather than to "*this year's* deficit").

Qualitative researchers are always interested in events: what they are, when they happened, and what their connections to other events are (or were)—in order to preserve chronology and illuminate the processes occurring. A *process,* after all, is essentially a string of coherently related events. Typically these interests lead to the production of an extended narrative, arranged in proper time sequence (usually without flashbacks or flashforwards). But extended, unreduced text, as we have noted, has severe disadvantages (its size, its dispersion, and its lack of structure). Extended text only feebly resolves the problems of multidimensionality, inter-event influence, and differential salience of events.

However, narratives are probably indispensable if we are to understand a complex chronology in its full richness. The problem is that going straight to an extended narrative from written-up field notes runs an acute risk: You can tell a story that is partial, biased, or dead wrong—even though it may look vivid, coherent, and plausible to a reader. The event listing is a good way of guarding against false chronologies.

So the problems we face in understanding event flows are those of sorting out the different domains of events, preserving the sequence, showing the salience or significance of preceding events for following events—and doing all of this in an easily visible display that lets us construct a valid chronology.

Brief Description

An *event listing* is a matrix that arranges a series of concrete events by chronological time periods, sorting them into several categories.

Illustration

In the school improvement study, we wanted to display events during the adoption and implementation of an innovation at the school level, showing them by different phases or time periods of the process. How might this be done?

Building the display. Keeping the classic left-to-right convention for the passage of time, we might make columns of the matrix be successive time periods. These could be defined arbitrarily (e.g., Year 1, Year 2) or more organically, by empirically derived phases or stages of the adoption-implementation process.

Which categories might we define for different types of events? A simple set is one dealing with the *locale* of events: school level, school district level, state, regional, or federal level. They could be rows of the matrix.

Perhaps some events are more critical than others, serving to cause new events, or move the process forward into a new phase. They could be noted or marked in some way.

Table 5.1 is repeated on the following page. It shows how this technique worked out for an innovation called SCORE-ON, a laboratory for teaching remedial math and reading skills to children "pulled out" of their regular classes. The time periods ("contextual press," "emergence of the problem," etc.) were initially defined conceptually from a general adoption-implementation model, but labels for each period came from the actual core activities during that period. A new time period was defined when a significant shift in activities occurred.

The analyst focused mainly on Smithson School, and wanted to have that as the most "local" of the locales (bottom row). However, the innovation was also being implemented in other schools (next row up). And events could be sorted into "district" and "state/macro" levels, which, in turn, influenced the lower levels.

Finally, the analyst marked "barometric events" (those that moved the process on into the next time period or phase) with an asterisk.

Entering the data. An exploratory interview question had asked people to describe the history of the innovation ("Can you tell me how SCORE-ON got started in this school?"). Follow-up probes fleshed out the sequence from innovation awareness to adoption, how and by whom key decisions were made, and the reasons involved. Other questions dealt with outside agencies and events and "anything else going on at the time that was important." Similar questions were asked about events during the implementation process.

The analyst looked at coded field notes (here, the codes are any that include the subcode CHRON, for chronology) and extracted accounts of specific events, such as "4th grade teachers report 40 pupils 1-3 grade levels behind," or "officials see SCORE-ON at 'awareness fair.' "

Decision rules for data entry must be clear. In this case the analyst defined an event as a specific action or occurrence mentioned by any respondent and not denied or disconfirmed by anyone else. A brief phrase describing the event went in the appropriate cell. If at least two people said the event was important, crucial, or "made a big dif-

Table 5.1 (Repeated)
Event Listing, Banestown Case

T I M E P E R I O D S

LEVEL	CONTEXTUAL PRESS 1976-78	EMERGENCE OF THE PROBLEM Oct. 1978	AWARENESS AND PROPOSAL OF SOLUTION Nov. 1978	APPROVAL AND PREPARATIONS Jan.-Feb. 1979	TRAINING AND BEGINNING OPERATIONS March-April 1979	EXPANSION, NEW OPENINGS Sept. 1979	BUDGET REDUCTION, DISRUPTION May 1980
State/ Macro	minimal competency levels, testing introduced in state schools			proposal discussed, approval at state level	Smithson middle school teachers, 2 admins. trained at D/D site (4 days, early March)		Reduction in Title I allocations
District	supplemental skills program introduced in reading and math	- alarm at failure rate - internal solutions proposed, found unacceptable	* officials see SCORE-ON at 'awareness fair' - IV-C proposal rapidly drawn up, submitted	* Smithson pupil folders screened - appointments made of Smithson lab teacher and aide	- 30 4th grade pupils selected for Smithson lab - materials, technical assistance intensified for Smithson lab	- staff active in extending Smithson, launching new labs * funding for all lab staff at Smithson taken over by Title I	* reduction in county, Title I budgets - proposed staff cuts, transfers in elementary schools
Local Schools	pressures begin to raise minimal levels			* continuation for following year planned for 5th grade in 2 middle schools; teachers named	- rooms, staffing completed for middle schools - 2 other elementary schools authorized to implement in the fall	lab opens at Carrington, Banestown Middle. modified version opens at Smith Camp, South End	- middle schools unaffected by cuts - threat of discontinuation at Banestown Middle (conflicts)
Smithson School	large numbers of low achievers placed in FACILE classes	* 4th grade teachers report 40 pupils 1-3 grade levels behind - teachers unfavorable to central office proposals	- teachers approve pull-out lab formula	- lab teacher and aide replaced; some disgruntlement - lab room created, minimally equipped	- Smithson lab opens (late March) - preparations inadequate, materials not arrived, scheduling difficulties	Smithson expands to 45 pupils in 3rd, 4th grades - new teacher added for morning sessions	* major shifts in lab staffing announced to teachers - program to be cut back, focused on grades 1-3, limited to 1.5 posts

* barometric event

ference" for what happened subsequently, an asterisk was assigned.

Drawing conclusions. A quick scan across the display shows us that the process of change is strikingly rapid. A problem seen in one elementary school in the fall of 1978 by the fourth-grade teachers apparently leads to the discovery and introduction of an innovation (SCORE-ON) that was in place in five district schools by the following fall.

A look at the asterisks helps explain some of the speed: the active involvement of central office officials after they saw the innovation at an awareness fair, leading to justificatory events such as the pupil folder screening, and to specific school-level planning and the appointment of specific teachers to manage the remedial laboratory. We also can see that state-level competency requirements were the backdrop, and that the teachers' report of problems was probably an alarm or trigger that set off actions already fueled by concern at the district level to meet state requirements.

When we note the repercussions of an externally driven budget crisis during the following school year, we can draw the inference that the original availability of Title I funds might have played a strong part in the original changes.

These are plausible hunches about the meaning of the data in Table 5.1. To check them out, the analyst can now begin to piece together a focused narrative that ties together the different streams into a meaningful account, a narrative that could only with difficulty have been assembled—or understood—from the diverse accounts spread through the field notes.

Here are some excerpts from the focused narrative the analyst produced. They should be read in conjunction with Table 5.1.

Beginning in 1971 and intensifying in the years 1976-78, the State issued a series of increasingly precise and constraining documents. These texts called at first for adequate "standards of quality" in education [which] were then outlined in the major subject matter areas. Competency testing was announced, first for grades K-3, then for 4-6. . . . After some delays, the first statewide tests keyed to these objectives for reading and math were scheduled for grade 4 in 1979 and for grades 5-6 in 1980-81.

The response at the district level was to create a "supplemental skills program" in order to intensify remedial work in reading and mathematics so that minimal competency levels could be attained. . . .

Schools throughout the county became more sensitized to the minimal competency obligation. . . . The implication was in the air that teachers might be held accountable for excessive rates of failure on statewide competency tests, but the pressures were still diffuse within individual schools. Teachers routinely attributed low performance to subnormal aptitude or to conditions at home over which they had no control.

A special impetus was given in the fall of 1978, when the six fourth-grade teachers at Smithson noticed that they had an unusually large cohort (40) of incoming pupils who were one or more grade levels behind in reading achievement. . . . Thirty-eight of these forty had come out of the [existing] FACILE program in the first to third grades. It is not clear how so many of these pupils got to the fourth grade, but no one was surprised. The Vice-Principal says, "They got too tall, they got too bad."

The teachers were worried that either promoting or retaining so many pupils would cause problems; they were leaning toward retention, but feared a massive protest by parents. Essentially, they were covering themselves, by announcing early in the year that they had inherited, not created, the problem. . . .

The teachers contacted Mrs. Robeson, through the reading committee which she chaired. She met with them, then brought in Mrs. Bauers, the elementary supervisor. . . .

During this phase, a circular announcing Federal funding . . . came to the central office from the State superintendent. An awareness conference, presenting a series of projects, many of them keyed to remedial skill development, was to take place nearby. At Mrs. Bauers' initiative—and with an eye to a solution for the problem at Smithson School—a contingent from the central office (Mrs. Bauers, Mrs. Robeson, Mr. Rivers) attended the presentations and was attracted to SCORE-ON. It seemed a relatively flexible program, easy to integrate into the school in pull-out form. It was directed specifically to the bottom quartile in reading and math. . . .

A proposal was written by Mr. Walt . . . with a request of $5,600 for training and materials. The request was cleared through the State Department, following a visit to the capital by Mrs. Bauers and Mrs. Robeson, accompanied by Alan Simens, principal of Smithson School.

Note several things here: (a) the narrative, which is both straightforwardly descriptive and analytic, helps knit together and flesh out events at different levels of the chart, (b) the analyst can add explanatory conditions or states that show how one event led to another, (c) the return to the field notes often turns up other critical events or supporting information not originally in the display, and (d) the narrative is more understandable when read in conjunction with the display, and vice versa.

The event listing helps ground the analyst's understanding of a complex flow of events, and increases confidence in the associated chronological account. It also lays the basis, as we shall see, for the beginnings of a causal analysis: what events led to what further events, and what mechanisms underlay those associations. Full-fledged causal networks are discussed in Chapter 6.

Variations

Event listings can be limited much more sharply to "critical incidents" (see Box 5.3), defined as important or crucial, and/or limited to an immediate setting.

For a careful reconstruction, the events in a listing can be dated within cells. Or the time periods can be specified much more narrowly. Or an event listing can cover a very brief time span (e.g., 1 hour in a classroom). Microlevel events can be well shown in a networklike flow in the "activity record" (Box 5.5).

Events from a listing also can be shown as a networklike flow that includes more general "states" or "conditions," such as "lack of enthusiasm": see the description of "event-state network" below, and Box 5.4.

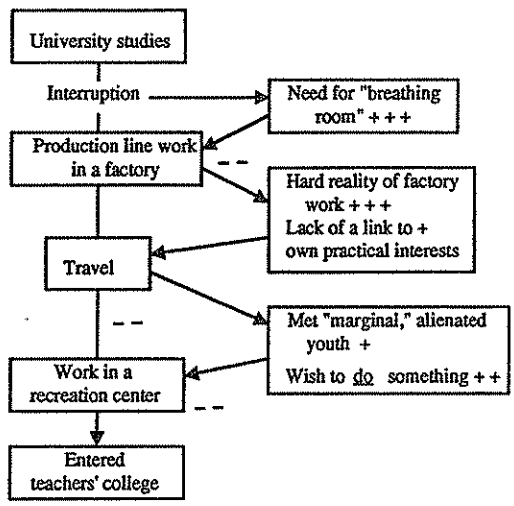
Even more selectively, events can be shown as a flow limited to major events, each with a shorthand explanation of the connections or causal pushes that moved the process from one event to another. Box 5.2 shows a network adapted from work by M. Pillet (personal communication, 1983), who was interviewing students who had interrupted their university studies, tracking them through subsequent experiences.

The case here is one student. The succeeding "experiences" are indicated in the left column; at the right we see the researcher's summary of the major forces moving the student to the next experience. The plus marks are an assessment of the strength of the various forces; the minus signs show the student's strength of dissatisfaction with the succeeding experiences.

Advice

Event listings are easy to do, and have good payoff in understanding any extended process. They can be done at

Box 5.2
Event Flow Network: A Student's Learning and Work Experience



many different levels of detail. Think first of how "micro" you want to be. Sorting out a series of candidate events on file cards (or in a computer subfile) can help stabilize the fineness of the approach.

In building matrices, keep the rows to a maximum of four or five levels and be sure they represent a meaningfully differentiated set of categories. Do not go finer than the study questions dictate. (In Table 5.1, we didn't need to use events in the classroom locale.)

If you know in advance that you want to construct a chronology of events, then something like a CHRON code is essential to keep you from being overloaded with specific events. Doing an event listing for ongoing (not retrospectively recalled) events can be even more time-consuming unless clear codes (e.g., CRIT) for various types of critical events are assigned along the way.

As always, keep decision rules for data entry explicit and clean. When you enter a phrase or a piece of data, it helps to have a "source tag" attached so that you can get back to its location in the written-up field notes (e.g., you might use 2-31 for interview 2, page 31) as you work on the production of the narrative. (Many software programs can assign source tags automatically.) Be sure you have exhausted the coded data chunks before you start taking a summary look.

When the first version of the display is filled in, start a draft of a focused narrative. That step will require you to return to the field notes as you go. Stay open to the idea of adding new events to the listing or subtracting events that seem trivial or irrelevant.

Consider the idea of some quantitative analysis of events and their sequence.[6]

Time Required

As always, much depends on how well coded the data are, the scope of time covered by the event listing, the fineness of detail, and the scale and complexity of the setting. An event listing like the one in Table 5.1 (one site, perhaps eight key informants, well-coded notes) can be assembled in half a day. The Pillet variant, on the other hand, was done in 15 minutes by a researcher closely familiar with the data in a single interview focusing on one person.

Writing the focused narrative usually takes about as much time as creating the event listing itself—more than for the typical analytical text accompanying other types of display. The illustration in Table 5.1 led to a 10-page narrative, which required about 5 hours of work. This is time well invested; an event listing and associated focused narrative represent your first basic grasp of what happened and suggest the subsequent lines of deeper description, analysis, and interpretation that will follow.

Box 5.3
Critical Incident Chart and Time Line (excerpt from Hawthorn School) (Stiegelbauer, Goldstein, & Huling, 1982)

Aug 1980		CBAM workshop for principals		District workshop for resource teachers re: composition
Sept. 1980		9/4; 9/24-29 Principal meets with teachers about teachers' goals for year	9/9;9/16 Principal meets with resource teacher to plan for year	9/3 District workshop for teachers on composition/ sourcebook
Oct. 1980	10/12-15 CBAM * interviews teachers	10/14 Principal has staff meeting re composition Teachers confused about program plan committee to review*	10/23 Principal meets with resource teacher re composition	10/22 Sixth grade teacher tells principal of other teachers' concerns about innovation
Nov. 1980		11/10* Principal & resource teacher meet to plan composition & committee Resource teacher tells principal of district ideas to clarify innovation	11/10 Principal & resource teacher meet with committee (includes 6th grade teacher) 11/25 Resource teacher meets with committee in 2 sections--upper and lower grades	Principal meets with various teachers to encourage work on composition-related activities 11/25 Committee teachers meet with others at grade level
Dec. 1980		12/1; 12/15 Principal & resource teacher meet to plan composition committee meetings 12/1 Composition discussed at staff meeting	12/10 Resource teacher meets with committee sections to review scope and sequence 12/18 Resource teacher meets with committee sections to plan *	12/16 Resource teacher distributes scope and sequence to all teachers

LEGEND * refers to events that were critical in initiating writing composition as a program separate from the language arts curriculum as a whole. These events led to the development of a HAWTHORN writing program, based on teachers needs and Sourcebook guidelines.

Events boxed $\boxed{}$ had a strong catalytic effect on focusing the program or on developing the principal's game plan for implementation.

Critical Incident Chart

Sometimes a researcher wants to limit an event listing to those events seen as critical, influential, or decisive in the course of some process. A useful approach was developed by Stiegelbauer et al. (1982), who extracted "critical incidents" occurring during implementation of a new language arts program at a school site. The process was aided by specialists in the CBAM (Concerns-Based Adoption Model) approach to school improvement.

In Box 5.3, time goes vertically. That orientation makes it a bit easier to examine and compare events that occurred during a given time period. The columns are not ordered, but easily could be. The dates for events make temporal ordering easy.

The chart selects only two to four critical events per month and marks (with boxes) those seen as having a "strong catalytic effect" on determining the need for the program or on the principal's "game plan" (strategy) for implementation. Certain events are also marked that affected a subprogram for writing skills. The authors' analytic comment is: "Relationships between events can be seen as both temporal and due to an exchange of information. In this school, teacher confusion about the program, discussion of teacher concerns, and new information about the program from the resource teacher all had a significant effect on program development" (p. 17).

Event-State Network

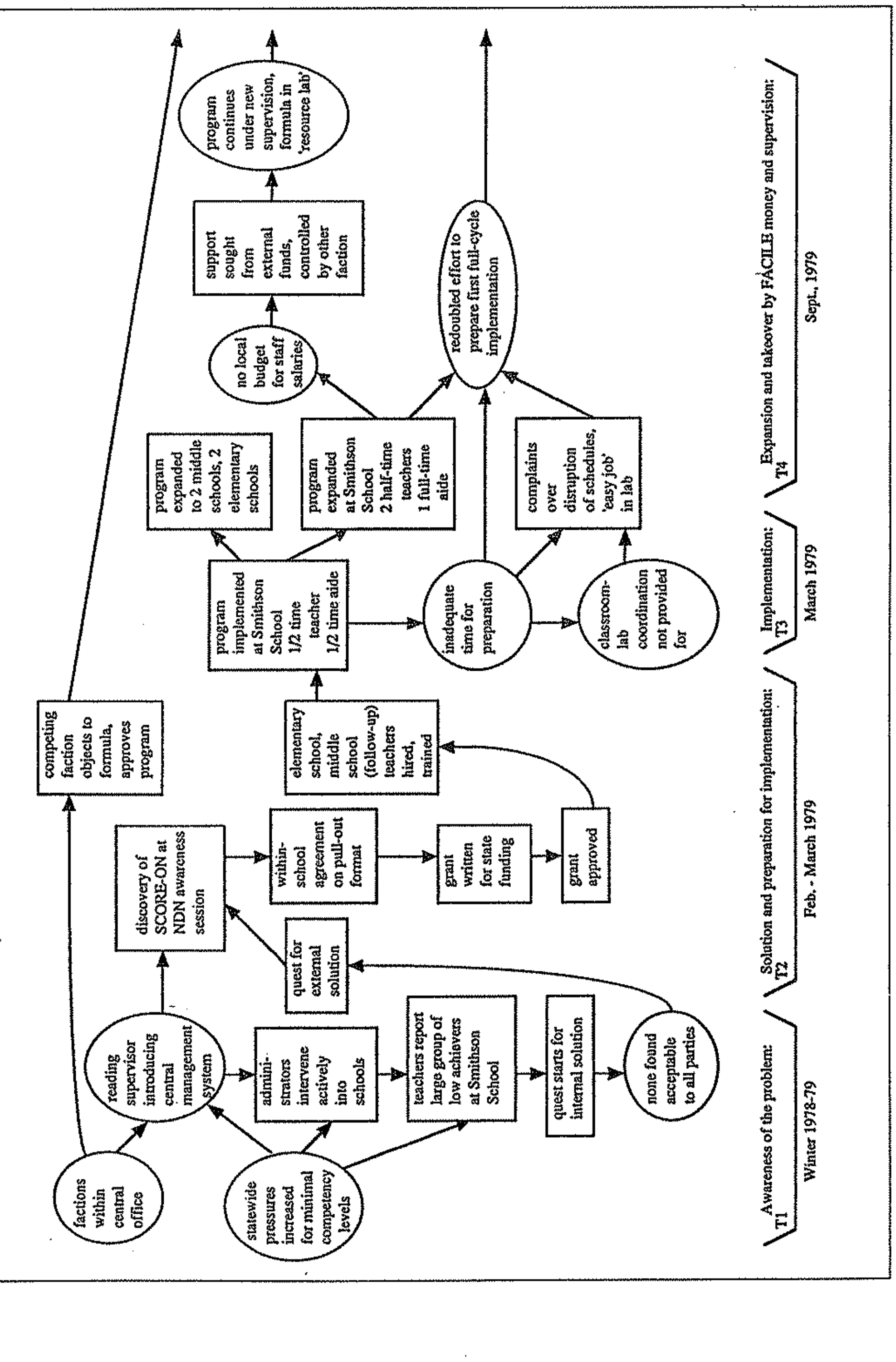
The alert reader will have noticed that some entries in Table 5.1 are not really specific events, but more general "states," such as "alarm at failure rate." Such states or conditions are not as time-limited as events, and often serve as the mediators or links between specific events.

You can display this difference by showing events as boxes (the sharp edges imply specificity and a narrow time span) and states as bubbles (round edges imply more diffuseness, less concreteness, existence over a longer time). You connect the boxes and bubbles by lines to show what led to what. Box 5.4 shows an excerpt from an event-state network drawn from the data in the event listing in Table 5.1.

Event-state networks are a simple, easy way to make sense of an event listing. Once an event listing is done, a network can be generated rather quickly. Write one event per small card, generate "state" cards as you go, and arrange the emerging network on a big sheet of paper, drawing arrows as you go. As we explain in Chapter 6, section D, this method is an especially helpful step toward assessing causal dynamics in a particular case. To create state cards: (a) recheck the field notes for evidence of the consequences and/or antecedents of specific events and (b) surface your generalized understandings of what was moving events along in the process you are studying. By the time an event listing is generated, you usually have a set of implicit generalizations about why things happened as they did. Event-state networks are a way of making those ideas explicit and verifiable. Programs such as Inspiration and MetaDesign (see Appendix) permit you to create such networks easily, shifting and changing them as your understanding grows.

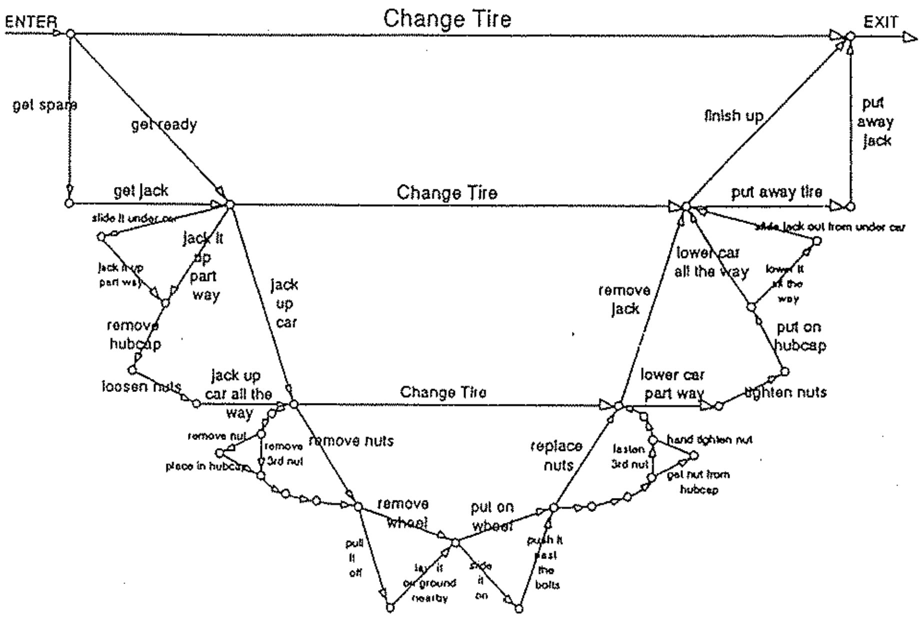
Event-state networks are enjoyable and rapid—an hour or so for the size shown, after the event listing is complete. It does not usually help to add a long narrative text (the

Box 5.4
Event-State Network, Banestown Case (excerpt)



Box 5.5

Activity Record: Example (Werner, 1992; Werner & Schoepfle, 1987b)



display itself is more economical and illuminating), but a summary set of analytic comments is useful.

Activity Record

It often happens that you want to describe a specific, recurring activity, limited narrowly in time and space: a surgical operation, the sale of a used car, a classroom lesson, an arrest booking, cashing a check, planting a garden. Such activities—really, predictably recurring sequential patterns—are the stuff of daily life.

The activity record, described by Werner (1992) and Werner and Schoepfle (1987b) displays specific actions, each assigned to a link, shown as an arrow (Box 5.5). Each "node" in the network is understood as signifying "and then. . . ." Thus you "get spare" and then "get jack" and then "jack it up part way."

Note the hierarchical levels: "get ready" is composed of "get spare" and "get jack," and so on. The same applies to "change tire": Each level of "change tire" subsumes the later activities beneath it. As Werner points out, the activity of *literally* changing the tire only appears at the very bottom; the levels above, however, are important contextual preconditions for the subactions below. (You cannot actually change the tire without having jacked up the car.)

Werner notes that the method makes activities very explicit, showing their time boundaries and sequential/hierarchical structure, and can be used easily with informants for checking accuracy.

Decision Modeling

The activity record is useful in mapping patterned actions, but it does not tell us much about the thoughts and plans, the alternative decisions made during a flow of actions embedded in a range of conditions. Decision modeling, also described by Werner and Schoepfle (1987b), uses displays that combine these internal events with public ones.

In this example, the authors begin with text from a cookbook (Jaffrey, 1975):

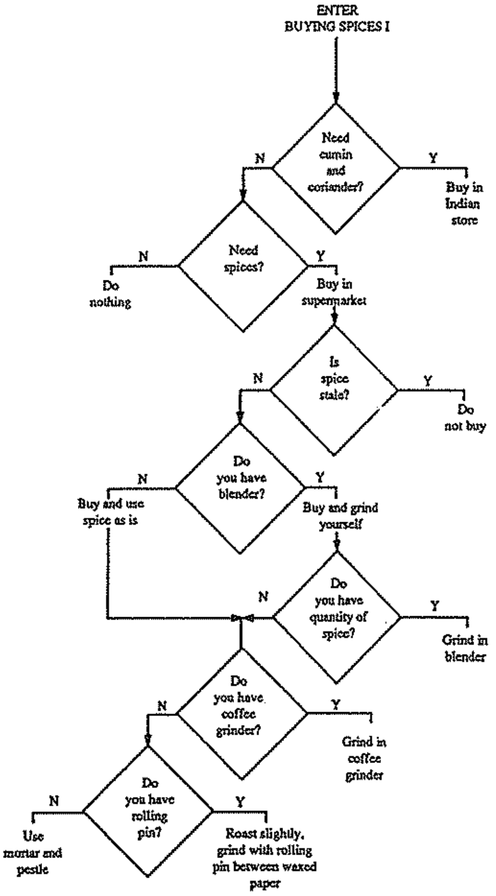
> There is a slight problem with supermarket spices which you might as well be aware of. Those spices which do not "move"—i.e., sell fast—tend to stay on the shelves and get stale. A few lose their aroma, others fade in the light, some get oily and rancid. Therefore try to buy only whole spices and grind them yourself in small quantities. The grinding can be done in a coffee grinder, or, if the spices are slightly roasted first, some can be crushed between waxed paper with a rolling pin. The electric blender will grind spices, if you do them in sufficiently large quantities. If all else fails, you could use mortar and pestle, though that tends to crush spices rather than grind them. Whole spices retain their flavor for very long periods. Make sure you store them in jars with tightly screwed lids, well away from dampness and sunlight. Ground cumin and coriander are fine if bought from Indian spice dealers in small quantities. (Werner & Schoepfle, 1987b, pp. 130-131)

Box 5.6 shows a flow chart indicating how Jaffrey's directions can be represented systematically. It strips away rich but nonessential information (e.g., different styles of becoming stale, methods of storage) but retains the key conditions and specifies contingently the decisions to be made.

The authors explain how to create such flow charts by analyzing the text sentence by sentence to locate key conditions, associated actions, and assumptions, usually checking all of these with key informants. For precision, the data then are displayed in "decision tables"; systematic elimination of redundancies leads to reduced tables, which then are used to draw the flow chart. (Such charts can be drawn by using many different software packages; one of the best is MetaDesign, which also can "hide" extended explanatory text at each of the nodes.)

The method is very useful for making the inexplicit visible and clear and can also be used for recurring decisions involving many people and complex situations (e.g., farming methods, patient care in an intensive care unit, serving a meal to a party in a restaurant, dealing with customer complaints, producing a contract research proposal). Repeated interviewing of multiple informants is usually needed, along with attempts to "disconfirm" the model through feedback and verification, and checking with a new, uninterviewed sample of people.

Box 5.6
Decision Modeling: Example (Werner & Schoepfle, 1987b)



Growth Gradient

Events can sometimes be conceptualized as associated with some underlying variable that changes over time. That is a rather abstract characterization, which will come immediately clear with a look back at Figure 5.1.

Here the analyst is interested in a main variable, *internal diffusion of an innovation*, defined as growth in the number of teachers using it. That variable is represented in network form as a single line between points. But the analyst has attached various critical events to the line—appointments of key personnel, training, and the like—that help expand our understanding of the movements of the main variable (e.g., the sudden jump following the August 1977 workshops).

Time-Ordered Matrix

Analysis Problem

With qualitative data you can track sequences, processes, and flows, and are not restricted to "snapshots." But how to display time-linked data referring to phenomena that are bigger than specific "events," so as to understand (and perhaps later explain) what was happening?

Brief Description

A *time-ordered matrix* (the more general form of the event listing we reviewed above) has its columns arranged by time period, in sequence, so that you can see when particular phenomena occurred. The basic principle is chronology. The rows depend on what else you're studying.

Illustration

In our school improvement study, we were concerned with how an innovation was changed and transformed over time during several years of implementation. We predicted that most innovations would show such changes as they were adapted to the needs of users and the pressures of the local situation.

Building the display. We broke down the innovation into specific components or aspects, using these as rows of the matrix. The columns of the matrix are time periods, from early through later use. If a change in a component occurred during the time period, we could enter a short description of the change. A blank cell would mean no change—a nice feature that permits seeing *stability,* as well as change.

Table 5.7 shows how this matrix looked. The innovation, CARED, is a work experience program for high school students. The official components were those specified by the original program developer. These components do not necessarily exhaust important aspects of the innovation, so there are rows for "other aspects," such as time and credits or student selection. Such aspects usually appear after initial fieldwork and direct experience with the use and meaning of the innovation.

The columns are time periods, starting with the initial planning period (because we expected changes while this relatively demanding and complex innovation was being readied for use). The three succeeding school years follow.

Entering the data. In this case the analyst was looking for changes in the innovation, component by component. Those changes could be found in the coded field notes, where innovation users had been asked whether they had made any changes in the innovation's standard format. Follow-up probes asked for parts that had been added, dropped, revised, combined, or selected out for use. In some cases documentary evidence, such as evaluation reports, could be consulted.

In this case six or seven people (users) were reporting on clearly defined changes in the innovation, which itself required rather close teamwork, with mutual monitoring and daily discussion. So we used the decision rule that if a reported change were confirmed by at least one other staff member and not disconfirmed by anyone else, it should be entered in the matrix.

Drawing conclusions. Rather than showing the final analytic text, let's sit at the analyst's elbow as he begins drawing conclusions. Looking across the rows of Table 5.7, you can begin to see "drifts," or gradual shifts expressing an accumulated tendency underlying specific changes. For example, the row "program requirements/curriculum" shows an increasing tendency to make *stiffer achievement demands* on students. (The tactic here is **noting patterns, themes**—see Chapter 10.)

The component "student responsibility and time use" suggests that a process of exerting more and more *control* over student behavior is occurring (e.g., the "accountability" scheme, the system of passes, etc.).

At this point the analyst can deepen understanding by referring back to other aspects of the field notes, notably what else people said about the changes, or reasons for them. In this example a staff member said: "Your neck is on the block . . . the success and failures of the students redound directly on you. . . . There's a high possibility of accidents."

The field notes also showed that some parents had complained that "students can't handle the freedom," and that staff were worried about students "fouling up" on their work assignments on the job. So the increased emphasis on control might come from the staff's feeling of vulnerability and mistrust of students (tactic: **noting relations between variables**).

A third drift can be noted: *increased routinization and reduced individualization,* occurring mainly in the second year of operation (student "batching," reduction of unique projects, addition of various standard forms). The field notes showed that the staff was preoccupied with the program's demandingness ("You keep going all the time . . . you can't get sick . . . it's more than a job."); after a full year's experience, they were ready to routinize and streamline the demands that a strongly individualized program made on them. The *reduction in parent involvement* may also be a connected issue.

Table 5.7
Time-Ordered Matrix: Changes in the CARED Innovation (a work-experience program)

		PLANNING PERIOD Jan.-June 77	FIRST YEAR 77-78	SECOND YEAR 78-79	THIRD YEAR 79-80
DEVELOPER COMPONENTS	Individualization			Some tendency to "batch" students, encourage standard projects, etc.	
	Student responsibility and time use	Added logs/time sheets	Added "accountability" scheme for discipline	Added system of passes for students to leave room	Refusal to accept late work; students were failed
	Direct exposure to work experience				Some tendency to reduce individualization/exploration aspects; more long-term placements; convenience location of jobs
	Program requirements/ curriculum		Added direct instruction on basic skills, separate from job	More competencies required	More competencies and project required
	Learning strategies	Did not include employer seminars; intensified "competencies" experience	Wrote new "competencies" Discarded learning style instrument New basic skills materials	Student learning plan only 1 semester; more basic skills emphasis	Journal more routine communication; less counseling emphasis
	Parental involvement		Parent advisory committee added	Advisory committee mtgs dwindled and stopped	Reduced level of detail in reporting to parents
OTHER ASPECTS	Time and credits	Reduced to 4-hr time block; students must take 1 other course			
	Student selection			Moved away from full-random selection to self-selection plus a few random	Lower SES
	Program size		32 students	64 students	70 students, cut to 50 in 2nd semester; 25 for next year (juniors only)

Table 5.8
Summary Table for Verifying and Interpreting Time-Ordered Matrix: Changes in the CARED Innovation

COMPONENT	TENDENCY	SIGNIFICANCE, MEANING FOR SITE
Individualization	Standardizing	Undergoes individualization, shows streamlining as response to demandingness
Student responsibility	Controlling, tightening up	Reflects staff feelings of vulnerability, mistrust towards students
Exposure to work	Simplifying, standardizing	Response to demandingness: cutting back on innovativeness of the project
Program requirements	Going academic	Shows vulnerability, marginality of project in the two high schools -- again cutting back on innovativeness of CARED
Learning strategies	Going academic, de-individualizing, simplifying, standardizing	More "routinization" overall
Parental involvement	Cutting back, simplifying	Response to demandingness
Time and credits	Cutting back, going academic	Done early: reflects principals' opposition to project, etc.
Student selection	Self-selecting	Project becomes "dumping ground" but also gets stronger endorsement from counselors, principals
Program size	Growing, then dwindling	Reflection of funding problems and possibly of under-endorsement

What else is happening? The portion of the matrix labeled "other aspects" provides some additional clues. The change in time and credits during the planning period came about, the field notes showed, as a result of negotiations with principals, who were very resistant to having students away from school for a full day. Here the analyst speculated that the underlying issue was the principals' opposition to the program, and suggested that the increasing achievement emphasis was a way to buy more endorsement.

We can also note an important structural shift in the second year: moving away from random student selection to self-selection. The field notes showed that this decision was precipitated by principals and counselors, who opposed entry by college-bound students, and wanted the program to be a sort of safety valve for poorly performing, alienated students. Thus the program became a sort of dumping ground or oasis (tactic: **making metaphors**) for such students, and principals' endorsement was somewhat higher. The addition of the summer program for the first 2 years was also an endorsement-buying strategy.

But look at the "program size" row. Though the program doubles in the second year, it cuts back substantially in the third year. Endorsement is not enough, apparently. In this case severe funding problems were beginning to develop (tactic: **finding intervening variables**).

We can see that this display helps us understand general patterns of changes over time and points us back to the original field notes to seek for explanations. The explanations can, in turn, be tested by looking at other changes in the matrix to see whether they fit the proposed pattern (tactic: **testing if-then relations**).

So the analysis could end here, and the report could either contain the analytic text pulling together the strands we just wove, or present the table along with it. But there may be a need to do a bit more. For example, the drifts look plausible, but we may have been too impressionistic in identifying them (tactic: **seeing plausibility**). And Table 5.7 may be too "busy" and unfocused for the reader—it is more an interim analysis exhibit than a report of findings.

One way of resolving these problems is to boil down the matrix, to (a) verify the tendencies observed in the initial analysis and (b) summarize the core information for the researcher and the reader.

Table 5.7 could be condensed in myriad ways. One approach is to "standardize" the several drifts by naming them—that is, finding a gerund such as *controlling, tightening up* that indicated what was going on when a change was made in the innovation (tactic: **clustering**) and then tying that drift to its local context, inferring what the changes mean for the case. The result appears in Table 5.8, a "summary table."

Reading the "tendency" column and using the tactic of **counting** the number of mentions for each theme confirms the accuracy of the initial analysis. The core themes are, indeed, stiffer achievement demands ("going academic"), more control, increased routinization ("simplifying"), and reduced individualization ("standardization"). You might even try for an overarching label to typify the set—something like "self-protective erosion," thus **subsuming particulars into the general**—and this labeling can kick off the final analytic text, usable in the report of findings.

A scan of the "significance" column helps in summarizing the underlying issues: exhaustion, lack of local endorsement, and the discrepancy between the demands of the innovation and the organizational procedures and norms in the immediate environment.

Variations

It might be helpful, especially for presentation purposes, to add a column to Table 5.7, after each time period, for the researcher's speculations about the issues underlying or possibly causing the changes (endorsement seeking, routinization, etc.).

Here the time periods were relatively long (a full school year), but depending on the phenomena under study, they could be shorter (semesters, months, weeks, days, hours).

The cell entries here were specific changes. But you can also enter specific events, such as decisions, actions, key meetings, or crises.

The rows of this matrix were aspects or components of an innovation. Many other types of rows can be used in a time-ordered matrix. They might be *roles* (principal, teacher, student, parent); *event types* (planned/unplanned, critical actions); *settings* (Classroom 1, Classroom 2, playground, principal's office, central office); or *activity types* (teaching, counseling, formal meetings).

Advice

Use a descriptive time-oriented matrix like this when your data are fairly complete, to begin developing possible explanations that can be tested by moving back to the coded field notes.

Be sure the time periods chosen are a good fit for the phenomena under study. They should be fine enough to separate events that you want to keep in sequence, rather than blending them into one big pot.

During analysis, remain alert to what this sort of matrix can uniquely give you: *timing and sequence.* What occurs early, afterward, later? Look for sequences. Look for stability (in this type of matrix, blank spaces).

After some preliminary analysis, consider whether the rows are organized in a sensible way. You may need to regroup them into "streams" or domains. Consider whether a boiled-down, next-step matrix is needed to advance your understanding—or that of the reader.

Time Required

If field notes have been coded to permit easy retrieval of the phenomena under consideration, a matrix of the general size and sort illustrated here usually can be assembled in 2 or 3 hours; analysis and writing may take another hour or so. But a vacuum-cleaner-like matrix trying to display a very large number of data points with a less clear conceptual focus will take longer.

D. Role-Ordered Displays

Our third main family of displays orders information according to people's roles in a formal or informal setting.

Role-Ordered Matrix

Analysis Problem

People who live in groups and organizations, like most of us, and people who study groups and organizations, like sociologists, health care researchers, anthropologists, and business theorists, know that how you see life depends, in part, on your role. A *role* is a complex of expectations and behaviors that make up what you do, and should do, as a certain type of actor in a setting—a family, a classroom, a committee, a hospital, a police department, or a multinational corporation.

Wives tend to see the world differently from husbands, and mothers from fathers and daughters. Doctors tend to see an illness, not a person, and patients often feel unattended to. Bosses tend not to see the frustrations faced by workers, partly because they are distant from them, but partly because subordinates censor the bad news when reporting upward. A teacher's high-speed interactions with several hundred children over the course of a day have a very different cast to them than the principal's diverse transactions with parents, vendors, secretaries, central office administrators, and teachers (cf. Huberman, 1983).

How can you display data systematically to permit comparisons across roles on issues of interest to a study? Or to test whether people in the same role do, in fact, see life in similar ways? One answer is the role-ordered matrix.

Brief Description

A *role-ordered matrix* sorts data in its rows and columns that have been gathered from or about a certain set of "role occupants"—data reflecting their views.

Illustration

Once again we draw on our school improvement study. The question of interest is, How do people react to an innovation when they first encounter it? This general question can be unbundled into several subquestions, such as:

- Which aspects of the innovation are *salient,* stand out in people's minds?
- How do people *size up* the innovation in relation to its eventual implementation?
- What *changes*—at the classroom or organizational level—do people think the innovation will require?
- How good a *fit* is the innovation to people's previous classroom styles, or to previous organizational working arrangements?

Building the display. Keeping in mind that we want to see answers to these questions broken out by different roles, we can consider which roles—for example, teachers, principals, central office personnel, department chairs—could be expected to attend to the innovation, and could provide meaningful reactions to it. The matrix rows could be roles, but if we want to make within-role comparisons, the rows should probably be *persons,* but clustered into role domains. It might be good, too, to order the roles according to how far they are from the actual locus of the innovation—from teachers to central office administrators. The columns can be devoted to the research subquestions.

Table 5.9 shows how this approach looks. The innovation involved is an intensive remedial program emphasizing reading, implemented in a high school, in the subjects of English, science, and math.

Entering the data. The researcher searches through coded write-ups for relevant data. But which people should be included: Only those who were directly involved with the innovation, or others with a stake but less direct experience? In Table 5.9 the researcher focuses on all available direct users of the innovation, plus department chairs who were not actually using the innovation but had a real stake in its use, plus key administrators, even though they did not have close contact with users.

The data entered in each cell are a brief summary of what the analyst found for each respondent in the coded field notes. The main decision rule was: If it's in the notes, and not internally contradicted, summarize it and enter a phrase reflecting the summary. There are also "DK" entries where data are missing because the relevant question was never asked of that person, was asked but not answered, or was answered ambiguously.

Drawing conclusions. Now we can begin looking down columns of the matrix, both within and across roles, to see what is happening. Scanning the first two columns for salient characteristics and how teachers size up the innovation shows us that many teachers—notably in English—see the new remedial program as *prescriptive,* with little latitude given for adaptation (tactics: **counting** and **making comparisons**). And the teachers who see the innovation as prescriptive are also those who have used it the longest, suggesting that prescriptiveness was highest when the program was first introduced (tactic: **noting relations between variables**). A number of teachers also mention *complexity* (though note that first-year users are more likely to see the program as simple and easy to use, suggesting program stabilization).

When we drop down to department chairs and central office administrators, the picture is somewhat different. They are more likely to take the "big picture" view, emphasizing the "curriculum," "strands," and the like. Although they, too, emphasize prescriptiveness ("works if followed . . . depends on being used as set up"), they either do not give clear answers on the issue of complexity or (as in the case of the curriculum director, a major advocate of the program) say that "any teacher can use it successfully." But teachers, faced with an initially demanding, rigid program, are not so sure, it seems (tactic: **making comparisons**).

Moving to the third column of Table 5.9, we can see role-perspective differences. Two teachers mention teaming as an anticipated change, one that curtailed their freedom and made them accountable to peers' schedules and working styles. Administrators, the field notes showed, considered the teaming to be necessary to implement the program's several strands, and as a way of helping weaker teachers do better through learning from stronger ones. Even so, they do not consider it a salient change, saying either that no organizational changes are required ("the

Table 5.9

Role-Ordered Matrix: First Reactions to the Innovation

		SALIENT CHARACTERISTICS	SIZUP	ANTICIPATED CLASSROOM OR ORGANIZATIONAL CHANGES @	FIT WITH PREVIOUS STYLE OR ORGANIZATIONAL SETTING @@
TEACHERS	+ REILLY 4th year	Highly structured	Little latitude	DK-1	DK-3
	+ KENNEDY 4th year	Frightening format Overload of objectives Reams of handouts	Difficult, complicated No latitude	Teaming (loss of independence) No Scope Magazine use	Poor; user felt she would be locked into structure and others' schedules
	+ FARR 1st year	Skill-oriented, organized Activities well planned Pre- & post testing good	Simple, clear Easy to use & understand	Less freedom to change direction	Fairly good; user views self as structured, organized
	+ ROGERS 2nd year	Prescriptive Rigid	Confusing Simplistic content	Working with basic students for 1st time	Composition assignments fit well; grammar, etc. simple-minded
	+ FLEMING 4th year	Prescriptive Use of media Teaming Heavy monitoring	Many materials Very complex, not quite clear	Working w/ other T's Mastering all materials	DK-2
	* BENNING 1st year	Objectives too broad Good content	Similar to prev. school's program, easy to use	Break down objectives Add games, activities Give objectives to kids	Close fit, when anticipated changes made
	* THATCHER 2nd year	(wrote science curriculum) Skill-oriented Reading emphasis	Too detailed	DK-1	Fair; reading was new; fewer labs
	# WOJICK 1st year	Variety of modes (workbooks, worksheets, computer terminals)	Easy to use; level & format right	DK-1	DK-2
	# MUSKIE 2nd year	Computer terminals Short worksheets	1st 1/2 flawed; 2nd 1/2 on target Good variety	DK-1	DK-1
DEPT. CHAIRS	VAN RUNKEL Science Chrmn	Science content revision Reading reinforcement Flexibility in activities	Questioned content reorg'n--would it fit together?	None -- same teachers, using new curriculum	Good fit: Program replaced old curriculum
	MANNHOELLER English Chrmn	Orderly curriculum w/ horizontal, vertical org'n Reinforcement of 3 strands	Concept is right Depends on its being used as set up	DK-1	Good fit: Program designed in part to fill basic English course slot
PRINCIPAL	MCCARTHY Tindale East	DK-2	DK-2	DK-2	Good fit: Order is maintained; no special requirements
CENTRAL OFFICE	CROWDEN Dir of Curriculum	3 strands of level I Sequential, comprehensive Reinforcement	Works if followed Any teacher can use successfully	None -- program designed to fit structure	Close fit: same staff, teachers wrote curriculum, same auth. structure
	MANN Superintendent	DK-2	DK-2	DK-1	Good fit: Program is a successful curriculum revision effort in district

+ = English teacher
* = science teacher
# = math teacher

DK = Don't know
DK-1 = question not asked of informant
DK-2 = asked, but not answered (strayed from question, didn't know answer)
DK-3 = ambiguous answer

@ Classroom changes question asked of teachers; organizational one asked of others.

@@ Personal style fit question asked of teachers; organizational fit question asked of others.

program is designed to fit the structure") or that they do not know whether organizational changes were anticipated.

Finally, if we continue the **making comparisons** tactic, the fourth column shows a range of "personal fit" for different teachers, depending on their views of the content, their own styles, and the organizational issues involved. The administrators, however, uniformly emphasize good fit at the organizational level, stressing the appropriateness of the curriculum and its fit into the existing structure; the director also invokes the fact that teachers wrote it.

In short, a matrix of this sort lets us see how perspectives differ according to role, as well as within role. In this case users from the English department who came in at the onset of the program had an initially tougher time than later users or than math and science users. Table 5.9 also hints at problems that arise from these differing perspectives (e.g., users' worries about teaming seem to be ignored by administrators—though they regard the teaming as critical to success).

A within-role analysis, moving across rows, shows that the superintendent, as might be expected, knows very little about the innovation. More surprisingly, the principal does not either. In this case a recheck with the field notes (tactic: **following up surprises**) told the field-worker that the formal role description for high school principals in this district actually forbids them to make curriculum decisions, which are the province of the curriculum director and department chairs.

We also can apply the tactic of **making if-then tests.** If the director and the chairs have a shared province of work (curriculum decisions), then their views of the innovation should resemble each other more closely than either resembles teachers' views. Looking vertically once again, we can see that department chairs' views are much more like those of central office administrators than those of teachers.

Variations

This display emphasizes different roles as *sources* of data and perceptions. It is also possible to develop a role-ordered matrix that treats roles as *targets* of others' actions or perceptions. For example, how are teachers treated by department chairs, principals, and central office personnel?

You also may want to have another, closer look at the data after identifying the main trends in the initial analysis. For example, it looks as if the findings are a function of *time of first implementation* (differences between first- and fourth-year users) and of *between-role differences* (users vs. administrators). So it may be useful to boil down Table 5.9, arraying the information according to these two contrasts (see Table 5.10).

The further condensation in Table 5.10 corresponds to the information in the analytic text, confirming and focusing it. The contrasts appear more sharply, both between earlier and later users (reflecting debugging and loosening of the innovation over time) and between users and administrators (reflecting their differences of perspective and priority).

The technique used to generate Table 5.10 had the following decision rules: (a) "Standardize" the cell entries to a more generic descriptor and (b) take the modal response. As the table legend shows, these rules were not always so easy to carry out. There are also some nagging problems: small and unequal respondent N's, collapsing of different species of administrator into one family, and missing data. All of these caveats should be acknowledged in text or on the legend for the table.

In retrospect, it's admirable that any useful conclusions could have been wrung out of data as diverse and spotty as those in Table 5.9. If data are thin or of doubtful quality, we have no business generating next-step matrices that give the misleading appearance of analytic order and adequacy. We see here the risk noted by Dawson (personal communication, 1983): that systematic methods like these can cause researchers to present findings more strongly than they would have otherwise. All the more reason for qualitative researchers to document their procedures, leaving a clear "audit trail" for others to follow (see Chapter 10, section D for more on this).

Advice

If your case is an individual, role-ordered matrices may well be helpful in showing how role partners view or interact with the person at the center of your case. They will be much simpler than this one.

Clarify the list of roles you consider to be most relevant to the issue at hand; avoid overloading the matrix with roles that are clearly peripheral. Differentiate the matrix by subroles (e.g., teachers of math or science) if relevant. If you wish to show the degree of within-role consensus or diversity, show person-level data within role domains.

Keep the questions in columns as parallel as possible for different roles.

Indicate clearly when data are missing, unclear, or not asked for in the first place.

Return to field notes to test emerging conclusions, particularly if the decision rules for data entry involve, as in this case, a good deal of reduction.

Ask for an audit of your analysis and conclusions from a colleague. Role-ordered matrices, because of our prior experience with role differences, can lend themselves to too-quick conclusion drawing. For example, in Table 5.9 the analyst originally emphasized teachers' reluctance to team-teach as a main conclusion, thinking of past work on teacher isolation (Lortie, 1975) and the vivid comments of several teachers. But a colleague pointed out that the team-

Table 5.10
Role-Ordered Matrix:
Example of a Next-Step Matrix to Analyze Initial Findings
(first reactions to the innovation)

ROLES	SALIENT CHARACTERISTICS	SIZUP	ANTICIPATED CHANGES	FIT
Early users now in 4th year (N = 3)	Structure, prescriptiveness	Complex, low latitude	Teaming	Poor*
Later users now in 1st-2nd year (N = 5)	Content focus, variety of modes	Easy to use	----#	Good
Administrators (N = 5) (all combined)	Reinforcement, order, integration	Fidelity of use is key	Few	Good

* Given "don't know" responses, this is inferred from analysis of other responses of 4th-year users to this set of questions

User-unique responses

ing issue appeared only for two early English users and that the "freedom" issue for another English user seemed more related to general prescriptiveness of the innovation than to peer constraint. Preliminary analyses, like cognac, need to be distilled twice.

Time Required

If, as in this case, there is a fairly clear semistructured interview with questions directly relevant to the columns of the matrix, and if the coded data are available for relevant roles, the matrix can be constructed and data entered relatively rapidly—2 hours or so for a matrix of this size. Analysis and writing are also fairly rapid—an hour or less.

Role-by-Time Matrix

As we saw in the preceding illustration, role-ordered matrices often turn out to involve time as well. *When* something was done by or to people in a certain role is often an important issue. Box 5.7 shows how assistance to users of an innovation was provided by different roles over time. The analyst developed a list of types of assistance and wanted to know when which types were supplied by certain roles (superintendent or whomever).

We can see immediately that the superintendent is out of the picture as far as assistance giving is concerned, while initial assistance is moderate to heavy early in the game from the principal ("building administration") and from peers. Both offer support (SUP), but the principal also emphasizes advocacy (ADV), while peers provide direct facilitation (FAC).

The innovation's developers provide much training (TTR), solution giving (SOL), resource adding (RES), and support (SUP) during the training, with some light follow-up, but nothing in the third and fourth years. By then the innovation is running with only light facilitation and support from peers.

Box 5.7 also shows us what is *not* being done in the way of assistance: no one is directly controlling (CON) or inquiring/feeding back data (INQ).

The eight assistance types, drawn from prior work on assistance functions (Nash & Culbertson, 1977), are arranged in decreasing order of directiveness; we can conclude that the assistance provided is largely client centered and less directive.

Box 5.7
Role-by-Time Matrix: Illustration Role Assistance, by Phase

	TRAINING Spring 1976	FIRST YEAR 1976-77	SECOND YEAR 1977-78	THIRD-FOURTH YEARS 1978-80
SUPERINTENDENT	None	None	None	None
BUILDING ADMINISTRATION	Heavy ADV Moderate SUP	Heavy ADV Heavy SUP	Moderate ADV Moderate SUP	None
DEVELOPERS	Moderate TTR, RES, SOL, SUP	Light SOL, SUP	Light SUP	None
PEER USERS	Moderate FAC, SUP	Moderate FAC Heavy SUP	Light FAC Moderate SUP	Light FAC, SUP

CON = Controlling
TTR = Teaching/training
SOL = Solution giving
RES = Resource adding
ADV = Advocacy, representing client interest
FAC = Facilitating
INQ = Inquiring, feeding back data
SUP = Supporting

Using cell entries of such a general, "boiled-down" sort means that (a) a good conceptual scheme should underlie the entries and (b) you must be very clear on decision rules for data entry. The analytic text should refer back to written-up field notes for illustration and clarification.

E. Conceptually Ordered Displays

Our last family of descriptive displays, rather than relying on time or role as the organizing principle, orders the display by concepts or variables. In a sense, *any* display is necessarily conceptually ordered, but some are centrally organized this way.

Conceptually Clustered Matrix

Analysis Problem

Many studies are designed to answer a string of research questions. Sometimes that string becomes a lengthy laundry list. As a result, doing a separate analysis and case report section for each research question is likely to tire out and confuse both analyst and reader. The obvious solution is to cluster several research questions so that meaning can be generated more easily.

Brief Description

A conceptually clustered matrix has its rows and columns arranged to bring together items that "belong together." This outcome can happen in two ways: *conceptual*—the analyst may have some a priori ideas about items that derive from the same theory or relate to the same overarching theme; or *empirical*—during early analysis you may find that informants answering different questions are tying them together or are giving similar responses. The basic principle, however, is conceptual coherence.

Illustration

For example, in our school improvement study, we had a general question about users' and administrators' *motives* for adopting a new educational practice, and a more specific question about whether these motives were *career centered* (e.g., whether informants thought they could get a promotion or a transfer out of the project). So here we had an a priori idea of a possible relationship. Then, during data collection, we saw some inkling of a relationship between the motives questions and two others: a *centrality* question (whether the innovation loomed larger than other tasks in the daily life of a user) and an *attitude* question (whether the informant liked the new practice when first

Table 5.11
Conceptually Clustered Matrix: Motives and Attitudes (format)

RESEARCH QUESTIONS

INFORMANTS	Motives (types)	Career relevance (none/some)	Centrality (low/mod/high)	Initial attitude (fav., neutr., unfav.)
Users U_1				
U_2, etc.				
Administrators A_1				
A_2, etc.				

introduced to it). We wondered whether a relationship existed between people's motives and their initial attitudes toward the practice.

The best way to find out would be to cluster the responses to these questions. Not only is there a relationship to probe, but there is also a general theme (initial attitudes) and a possibility of handling three research questions at the same time.

Building the display. We need a format that (a) displays all of the relevant responses of all key informants on one sheet, (b) allows an initial comparison between responses and between informants, (c) lets you see how the data can be analyzed further (e.g., repartitioned or clustered), (d) for multicase studies, lends itself easily to cross-case analysis and will not have to be redone, and (e) for multicase studies, provides some preliminary standardization—a set of content-analytic themes that all case analysts will be using.

When you are handling several conceptually/empirically related research questions together, a likely start-up format is a simple informant-by-variable matrix, as shown in Table 5.11. Thus we have on one sheet a format that includes all respondents and all responses to the four research questions. Note that we have set up comparisons between different kinds of informants (users and administrators), so it is role-ordered, as well as conceptually ordered. The format also calls for some preliminary sorting or scaling of the responses: types of motive, career-relevant or not, degree of centrality, valence of initial attitudes.

Entering the data. Next we go back to the coded segments keyed to the research questions. The analyst notes down the motives given by or attributed to an informant, and then tries to put a label on the motive. One informant, for example, gave several motives: She heard how good the new practice was (social influence); her principal was "really sold on it" and "wanted it in" (pressure); most other teachers were using it or planned to—"It's what's coming" (conformity); using the new practice was an occasion to "keep growing" (self-improvement). At this stage, it is best to leave the start-up labels as they are, without trying to regroup them into fewer headings that cover all informants; this practice gives you more degrees of freedom while still providing a preliminary shaping of the data.

Turning to career relevance, the analyst summarizes in a phrase or sentence the relevance of adopting the practice for each informant. The next task is to look for evidence on how central the new practice is for people and what their initial attitudes seemed to be. For these two questions, the analyst makes a general rating, backing it with specific quotes. When these data are entered in the matrix, we get something like Table 5.12.

For cell entries, the analyst reduced the coded chunks to four kinds of entries: *labels* (e.g., self-improvement), *quotations, short summary phrases,* and *ratings* (none/some, low/high, favorable/unfavorable). The labels and ratings set up comparisons between informants and, if needed, between cases. The quotations supply some grounded meaning for the material; they put some flesh on the rating or label and can be extracted easily for use in the analytic text.

The summary phrases explain or qualify a rating, usually where there are no quotations (as in the "career relevance" column). In general it's a good idea to add a short quote or explanatory phrase beside a label or scale; otherwise the analyst is tempted to work with general categories that lump together responses that really mean different things (as clearly seen in the "High" responses in the "centrality" column). If lumping does happen and you are puzzled about something, the qualifying words are easily at hand for quick reference.

Drawing conclusions. Reading across the rows gives the analyst a thumbnail profile of each informant and provides an initial test of the relationship between responses to the different questions (tactic: **noting relations between variables**). For example, L. Bayeis does have career-relevant motives, sees the practice as very important, and is initially favorable. But R. Quint's entries do not follow that pattern or a contrasting one. We have to look at more rows.

Reading down the columns uses the tactic of **making comparisons** between the motives of different users and administrators, as well as comparisons between these groups. It also enables similar comparisons between responses to the career-relevance, centrality, and initial-attitude questions.

Table 5.12 contains about as many data as a qualitative analyst can handle and a reader can follow. The analyst has ordered informants according to their time of implementation (early users, second generation, recent users) and their roles (users and administrators) and, within the group of users, has included a nonuser to set up an illustrative contrast between motives for adopting and motives for refusing the new practice. As an exercise, see what conclusions you might draw from Table 5.12. Then we'll suggest some that occur to us.

A scan down the columns of Table 5.12 provides both information and leads for follow-up analyses. In the first column, "social influence" motives are prominent. Also, a noticeable amount of pressure is being exerted on users (tactic: **counting**).

The tactic of **making contrasts/comparisons** leads to more conclusions. For example, there is some career relevance in adoption for users, but practically none for administrators. Centrality is high—almost overwhelming—for users, but less so for administrators. Users are less favorable initially than administrators. Looking across rows, we can use the tactic of **noting relations between variables** and see that, for two of three career-motivated users, a relationship exists among the variables: High centrality and favorable attitudes are also present. But the opposite pattern (low career relevance, low centrality, and neutral/unfavorable attitudes) does not apply. In fact, it looks as if some people who are neutral would have been favorable were they not so apprehensive about doing well (tactic: **finding intervening variables**). This is a good example of the merits of including a quotation or explanatory phrase in the cell so as not to pool what may be different responses behind the same label ("neutral," in this case).

Variations

In using conceptually clustered matrices for multicase studies, there is often a tension between the desire to run with the case and the obligation to set up the case report to facilitate later cross-case analyses. For instance, the data for the Masepa case in Table 5.12 seem to cohere in ways that do not follow the four-column structure religiously. Note that, for users, regardless of their date of initial use, the modal pattern is social pressure, high centrality, and neutral, often mixed initial attitudes; people are backing into what most call "the most important thing" or "biggest step" they have undertaken. Clearly, the conceptual core of the table lies here, and the single-case analyst should fasten onto it, looking, perhaps, for an appropriate conceptual framework such as "approach-avoidance conflict" (tactic: **making conceptual/theoretical coherence**) or a descriptive gerund such as "ambivalent innovating" to pull the material together (tactic: **subsuming particulars into the general**).

For a single case, the next step would be to reconfigure the table around this theme, dropping the "career relevance" column, looking more carefully at the mix of motives, and checking the number of informants who correspond to this profile. Table 5.12 is just a way station in the case-specific analysis, where the prime objective is—and should be—to follow the data where they beckon.

But suppose you have several cases. The display must be set up to permit a coherent comparison, looking across all cases: the motives, career relevance, centrality, attitude relationships. This may be a rogue case. It's important to hold on to the common set of categories, scales, and ratings for each case—even if the empirical fit is poor in one or another of these columns—until the full set of cases can be studied.

Table 5.12
Conceptually Clustered Matrix: Motives and Attitudes of Users, Nonusers, and Administrators at Masepa

		MOTIVES	CAREER RELEVANCE	CENTRALITY	INITIAL ATTITUDE TOWARD PROGRAM
EARLY USERS: 1977-78	R. Quint	Self-improvement: "To get better, I had to change"..."Maybe I wasn't teaching the best ways." Pressure: "They wanted us to do it." Social influence: "Everybody was saying what Gail's doing is great."	None -improvement of practice	High: "Biggest thing I've ever done that somebody else told me to do."	Neutral: "There wasn't any appeal. They said it worked so I was going to try it."
	L. Bayeis	Observation: Saw G. Norrist do it and "was impressed." Fit to personal style: "I like structure." Practice improvement: "looking around for a different way to teach reading." Novelty: "you get tired of always doing the same old thing."	Vehicle to turnkey trainer role: also became Title I Coordinator	High: "most important thing I've been involved with."	Favorable
SECOND GENERA-TION: 1978-79	F. Morelly	Social influence: heard from several friends about program Opportunity, effort justification: "I took the training for recertification credit. After all that, I had to follow through." Pressure: "He (Weelling) is the reason we do it here. He's so enthusiastic about it."	None - possibly stabilizing her job at the school	High: "This is the only new thing I've done since I've been out of school... I had to invest so much."	Neutral, apprehensive
	L. Brent	Social opinion, influence: "I heard how good it was." Pressure: "(Weelling) was really sold on it. They really want it in." Conformity: Most doing it or planning to in the school: "it's what's coming." Self-improvement: occasion to "keep growing."	None, possibly fear	High: "It's been a nightmare."	Unfavorable, once training began
RECENT USERS: 1979-80	V. Sharpert	Obligation: requirement to obtain teaching post: "I didn't have a choice." Practice-improvement: complementing pre-service training	Ticket to teaching job in the district	High: "My first job"	Neutral, apprehensive
	A. Olkin	Social influence: "heard it was good;"..."a good friend liked it." Pressure: "strongly encouraged" by Weelling and Dahloff Observation, modeling: saw G. Norrist. "She really impressed me."	None: felt obligated by administration	High: "This was really the big one for me."	Neutral, mixed feelings
	S. Sorels	Observation: "It was so good for my own kids...tremendous change in reading, spelling, work habits."	Ticket to full-time teaching position	High: "This was really a big step for me--a big move..[nothing else] as high as this in my career."	Favorable: "I was excited about it."
NON-USER	C. Shinder	Relative disadvantage: "My program was better." Poor fit with personal style: "too scholastic...too programmed."	None	N/A	Unfavorable
ADMINIST-RATORS	K. Weelling Principal	Met need: "I was looking for a highly structured, skill-oriented reading program." Novelty, promise of practical improvement: intrigued by reading about mastery learning; wanted to see it in operation.	None at first: later, appreciates the visibility	High: "Largest investment I've ever made."	Neutral, then favorable
	J. Dahloff Curriculum Coordinator	Relative advantage, face validity of program: "well organized;" could be used for other subject matters. Social influence: "impressed" that outstanding teachers favored the program. Practice improvement: beginning teachers ill-prepared in reading. "We didn't know what to do with them...They just had to learn on the job."	Another in a series of implementa-tions	Moderate: "It was one thing among a lot of things I was working on."	Favorable
	W. Paisly Asst. Sup't	Social influence: "talked into it" by J. Dahloff	None	Low: "It was no big deal.."	Neutral

Advice

As with preceding displays, conceptually clustered matrices can be used with less complex cases, such as individuals or small groups. They are most helpful when you have specified, or discovered, some clear conceptual themes.

Avoid using more than three or four questions when doing a conceptually clustered matrix. Otherwise the mind will boggle. There will be too many data to see inclusively at one time, and too much time spent manipulating blocks of data to find clusters and covariations.

Work with all of the data on a single page, even if that page covers a wall or a big table. You can progressively reduce a giant chart to a more economical one by using summarizing phrases, short quotes, category headings, ratings, and so on, as in Table 5.12.

Having all of the data in one readily surveyable place helps you move quickly and legitimately to a boiled-down matrix by making sure that *all* the data fit into a reasonable scheme, and that any ratings or judgments you make are well founded.

Time Required

If you have sorted all of the coded segments—either on cards, cut-up photocopies, or, most simply, via software such as Text Collector, Kwalitan, or Metamorph (see Appendix)—all of the "motives" chunks together, all of the "career relevance" chunks together, and so on—a display like Table 5.12 can be built in 4-6 hours. A good deal depends, of course, on the number of informants: More informants will fatten the piles of coded segments, enlarge the initial matrix, and make for a longer scrutiny until the patterns begin to come clear.

Thematic Conceptual Matrix

Conceptually ordered matrices need not be organized by persons or roles, as in Table 5.12. More general conceptual themes can be the ordering principle. Box 5.8 is a display from a study of urban high school reform (Louis & Miles, 1990; Miles & Rosenblum, 1987). The researchers were aiming to answer two research questions:

1. What *barriers, problems, and dilemmas* were encountered during planning/initiation and implementation of the school improvement project? These impediments may be "inherent" in the nature of high schools and the improvement process, or may be conditioned by the local context.
2. What *management and coping strategies* were employed to deal with the barriers/problems/dilemmas? Which were technical, which were cultural, and which were political in nature?

After visiting problem-solving meetings, interviewing coordinators, and asking people about typical problems and what was done about them, the researchers coded their notes and began entering data into the matrix in Box 5.8. A first step was to employ the tactic of **clustering** to define the specific problems used in row headings. For example, here are several quotes:

> The program is really a nothing. Nobody is excited about it.
>
> Here it comes again. What is MES/IPI? Just a new flag the central office is waving so they can spend the grant money.
>
> Every year we get a new one. MES/IPI is just the latest thing coming down the pike.
>
> I have a lot of doubt whether it will do any good.
>
> The coordinators are meeting every other week, but we don't see anything down in the trenches.

Here the analyst identified two problems underlying these comments: *teacher ownership* of MES/IPI (a "more effective schools" program) and *skepticism* from the staff. Similar clustering was carried out with the results of questions like, "What's been done about [this problem]?" and entered as strategies. The analyst had to make a conceptual sorting: Was the strategy basically technical, political, or cultural? The entries are essentially summary phrases, using the decision rule of two confirmations and no contradiction by anyone in the site.

Here is an excerpt from the text the analyst wrote:

> First, many of the problems encountered have minimal (or no) coping associated with them. In effect, the problems continue.
>
> Second, it is rare to see any cultural-type coping. The problems are largely treated as deserving technical or political effort, even though cultural norms are doing much to make implementation difficult (e.g., "here it comes again" . . . "don't rock the boat" . . . "we are not good finishers"). This may in part be the tendency we already have noted to downplay organizational issues as deserving of direct attention; issues of accountability and responsibility remain ambiguous.
>
> Third, back of many of the problems noted are two main issues: the fact that MES and IPI were essentially imposed on the school, and that very little assistance was provided. Administrators said things like "we have been directed to do it . . . every year we get a new one . . . it doesn't matter whether we are pro or con. What we're here for is making it work." Delay in one sense is the simplest coping tactic available to people who don't like something they have been handed to implement.

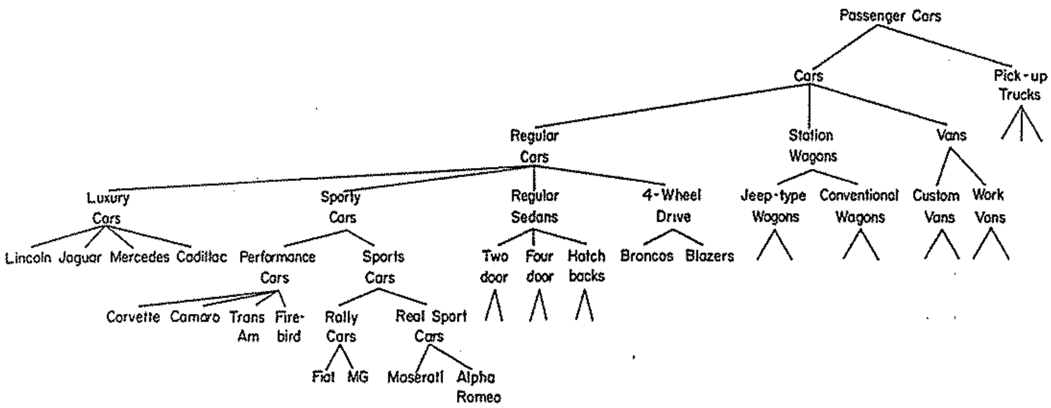
Here the analyst is blending inferences drawn directly from the displayed data (tactics: **seeing patterns, themes;** and **factoring**—that is, seeing a few general variables un-

Box 5.8

Thematic Conceptual Matrix: Problems and Coping Strategies (Example From Chester High School) (Miles & Rosenblum, 1987)

PROBLEMS	COPING STRATEGIES: Technical	Political	Cultural
Contextual problems			
1. Teacher ownership of MES/IPI 2. Skepticism from staff	Orientation mtg. for all faculty.	DK	Putting goals on the banner. [not effective]
3. Lack of coordination	Frequent mtgs. (princ., VPs, supv's). Princ. develops shorthand list of MES objectives for reference use.	Spec. ed. & voc. ed. negotiation about space for MES efforts. Power not much shared with admin. team.	
4. Thin school-level planning	State Dept. of Ed. MES director drafts HS plans for 3 objectives; princ. checks w. VPs and supv's.		
5. Uneven implementation	Discussion in admin. mtgs. Idea of using curric. funds for summer work to locate gaps in implementation (lesson plans).	Pressure from VP for curriculum; supv's required to use format she advocated.	
Problems generic to high schools			
1. Resistance to scanning		Idea that supv's should get together, do adaptation, agree on common start date. [not done]. Exhorting, pressuring by principal	
2. Poor fit of scanning to HS	Idea of making a new videotape. Idea that supv's should adapt, select from scanning procedures. [neither done] Supervisor holds info/training sessions with pairs of depts.	Allowing delay in implementation of scanning	
Problems stemming from program design			
1. Planning inadequacy			
2. Insufficient prep. at HS			
3. Slow implementation of MES	Funding delay partially buffered by business office ("they are understanding.")	Supt pushed SDE asst. comm. on funding delay, got action. Pressure from district coordinator to produce 86-87 plans (threat of SDE withdrawing funds).	
4. Work time required		Princ. negotiated with c.o. to get "program days" devoted to IPI. When c.o. reneged, princ. gave up fac. mtg. days for depts for IPI work.	

Box 5.9
Portion of Jack's Folk Taxonomy of Cars and Trucks (Bernard, 1994)



derlying many specifics), with illustrative comments drawn from elsewhere in the case.

Folk Taxonomy

We'd like to avoid the tacit implication that concepts are always properly sorted into rows and columns. Concepts also can be displayed in network form. Let's start with a hierarchical network display. Box 5.9, drawn from Bernard (1994), shows how one informant, Jack, classified cars and trucks when he was asked "frame elicitation" questions like "What kinds of cars are there?" As Jack continued, the researcher asked more such questions, like "What cars do you call luxury cars?" "Is an MG a real sports car?" "Is a performance car a type of sporty car?" to refine the categories and to understand links between them in the person's mind.

The display, a hierarchical tree diagram, lets us see how the person classifies important phenomena. Jack clearly has a relatively differentiated set of ideas, and has seven levels in his scheme (though he does not give us many specific exemplars for the cars and trucks on the right side). We might infer that he falls into the "car nut" category.

Bernard makes several useful points: The display is typically idiosyncratic (e.g., Jack's term "regular cars"); not all categories necessarily have labels, or exemplars; the categories often may be overlapping and indeterminate (Jack says in the interview that Jeeps are both station wagons and 4-wheel drive vehicles).

Bernard notes that folk taxonomies also can be created by starting with a batch of cards, which the person sorts into piles before "frame elicitation" questions begin.

Bernard also suggests "componential analysis," as used in linguistic and kinship studies, as a way to clarify categories, making it clear why a particular label is assigned to a particular object. You ask for distinctive features of the sorts of objects being "taxonomized" and aim to find the smallest group of features that discriminates them clearly, thus producing a matrix like that shown in Table 5.13.

Bernard comments on the risk of analytic shallowness with the method and says the issue is focusing on the person's real meanings, how they actually think about the phenomena involved. See also Spradley (1979) for more suggestions.

Table 5.13
Componential Analysis: Example for Seven Cars According to Jack (Bernard, 1994)

	1 *Expensive*	*2* *Practical*	*3* *Foreign*
Corvette	+	—	—
Firebird	—	—	—
MG	—	—	+
Maserati	+	—	+
Mercedes	+	+	+
Jeep	—	+	—
Dodge Van	+	+	—

Note: Researchers may use the method to sort out their own theoretical ideas—being careful, of course, to call it a "real," rather than a folk, taxonomy.

Several computer programs are especially good at generating taxonomies of either sort: NUDIST, ATLAS/ti, SemNet, MetaDesign; see Appendix.

Cognitive Maps

Analysis Problem

Many of our examples so far have been complex, multilevel cases. But cases are often focused at the individual level. We need good displays that show us the complexity of the person. People's minds—and our theories about them—are not always organized hierarchically, as in folk taxonomies. They can be represented fruitfully in nonhierarchical network form: a collection of nodes attached by links. Remember the conceptual network "Joe's a gnerd," in Figure 4.7. In studies of thinking, the nodes can be represented as concepts (e.g., Jack's idea of "luxury car"), and the links as relationships (a Mercedes is "a kind of" luxury car).

$$\text{Mercedes} \xrightarrow{\text{is a kind of}} \text{luxury car}$$

But there can be many sorts of networks. Should we use the person's own concepts, or a standardized set? Will we restrict the network to single words at the nodes, or use complex phrases? Will concepts be at several levels, or just one? Will the links be given explicit names, rather than being left tacit, as in the Jack example? If named, will the links be standard and uniform throughout, or diverse for each link? And will the links be directional (either one- or two-way), or not?

Even with these technical choices made, some other problems remain. People's concepts are not that easy to evoke or to display systematically for drawing conclusions. We need a concrete method to evoke the complexity of the individual's thinking.

Brief Description

A cognitive map displays the person's representation of concepts about a particular domain, showing the relationships among them. Descriptive text is associated with it.

Illustration

In a study of school restructuring efforts (Khattri & Miles, 1993), the researchers wanted to understand how individual teachers and administrators viewed such matters as what a "school" is, what a "restructured school" would look like, and how such a school could be designed, implemented, and stabilized.

Adapting methods used by Champagne, Klopfer, Desena, and Squires (1981) and Novak and Gowan (1984), the researchers began data collection with a "premapping" interview, asking a person such questions as, "What is restructuring? What cannot be called restructuring? What are examples of 'structures'? What is the right way to go about it? What is the wrong way to go about it? What decisions about restructuring have been made in the school?"

Transcripts of these interviews were culled for key concepts, phrased in the person's own words, such as *accountability, staff development, small group does new things, get conservative staff support* and *institutionalizing*. These concepts were divided into two lists, one for the *content* or substance of restructuring and one for the *process*—how it proceeds—and written on Post-Its.

Building the display and entering the data. At the mapping interview, the person is given the first bundle of Post-Its (typically under three dozen, alphabetized) and is asked to arrange them on a large sheet of cardboard "in a way that shows how you think about the words." When this task is complete, the researcher asks, "Why are they arranged this way?" The researcher draws lines around concepts that the person says belong together, and evokes a name for the group, which also is written on the display. The question "What relationship is there between . . . and . . . ?" leads to the person's naming of links between concepts and/or concept groups, and those, too, are written on the display. The process is repeated for the second bundle.

Drawing conclusions. During preliminary analysis, the researcher listens to the tape of the mapping discussion, clarifies any errors, and writes a descriptive text that walks through the complete map. The map and text are fed back to the respondent to be sure it is an accurate representation.

Note that the procedure engages the respondent and the researcher in joint work, simultaneously building the display and entering data.

Figure 5.6 shows the results of this process when Peter, a teacher in a small alternative high school, arranged his concepts of *how* restructuring proceeds. (It was redrawn for this book from the original production done with Peter, using MacDraw software.) Figure 5.7 shows the descriptive text the analyst wrote.

The text in Figure 5.7 is descriptive, staying close to what Peter said about his ideas. The analyst wanted to take some next analytic steps and began looking for core or underlying themes, writing these comments:

Figure 5.6
Cognitive Map: Peter's View of How Restructuring Proceeds (Khattri & Miles, 1993)

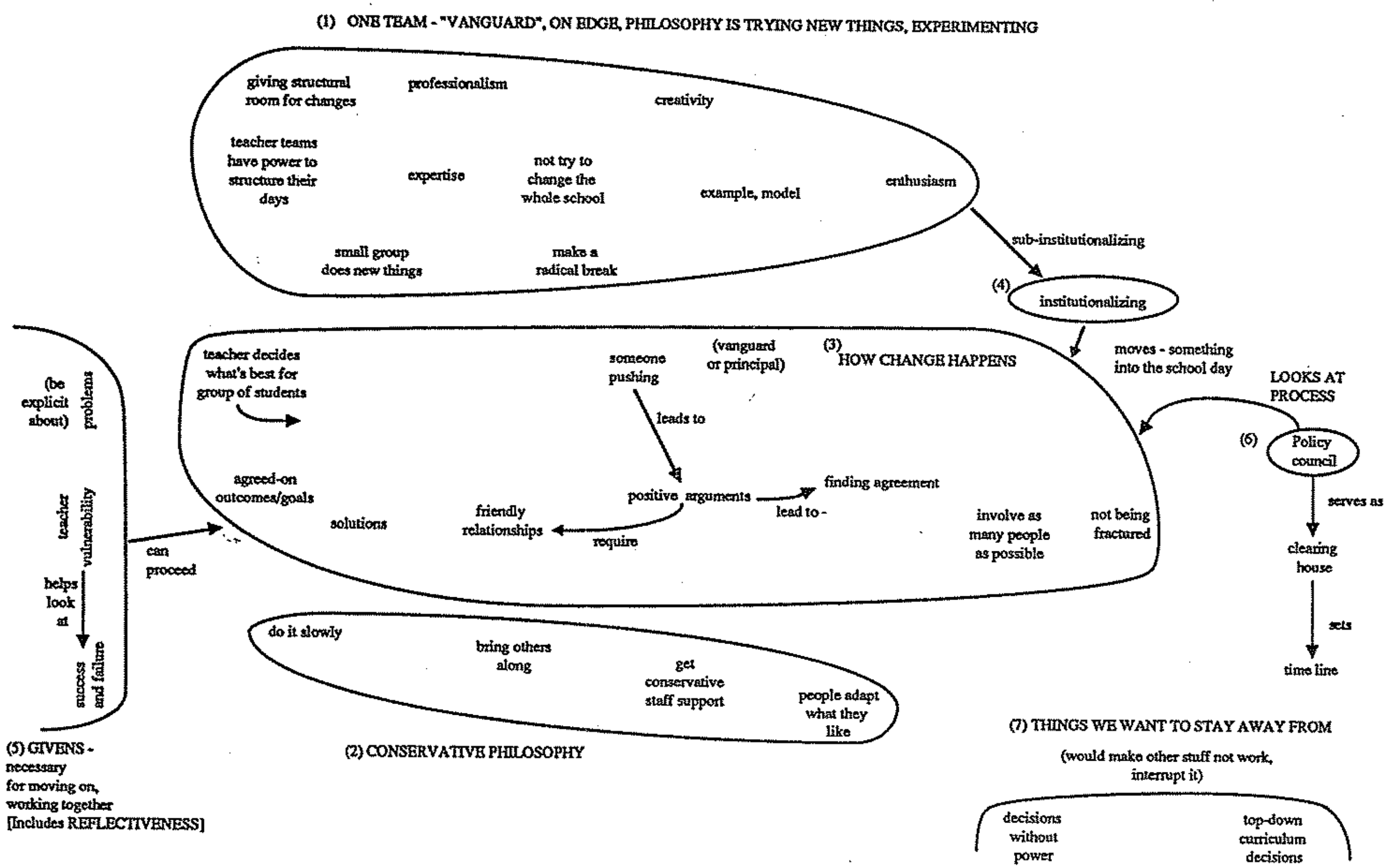


Figure 5.7
Descriptive Account for Peter's Map of How Restructuring Proceeds
(Khattri & Miles, 1993)

How restructuring proceeds: Peter, map 1

In one major cluster, Peter describes process features that characterize a "vanguard" team that's on the edge (1) with a philosophy of experimenting and trying new things: giving structural room for changes, professionalism, creativity. Teacher teams have power to structure their days, and the image is of a small group doing new things, making a radical break. The aim is not to change the whole school, but to use expertise, develop an example or model (of the new) and generate enthusiasm.

At the bottom of the map is a differently-oriented cluster, labeled conservative philosophy (2), aiming to "preserve things so they don't fall apart": it includes the concepts of do it slowly, bring others along, get conservative staff support, and people adapting what they like.

In the middle is a cluster called how change happens (3)—"the way when it's working, it works": it describes, in effect, what the schoolwide processes are for change. It includes the ideas that teachers decide what's best for a group of students and that there are agreed-on outcomes/goals. Solutions appear, and in the context of friendly relationships, there are positive arguments, which are the "heart" of the change process (positive arguments require friendly relationships). The arguments are also influenced by someone pushing (either the principal or a vanguard team), and they lead to finding agreement, though perspective differences exist. The idea is to involve as many people as possible, and not be fractured.

The experimenting, vanguard team also "subinstitutionalizes" changes; the institutionalizing work (4) is a sort of "end product" that moves the change into the regular school day.

At the left of the map, Peter identifies certain givens (5) necessary for "moving on, working together." These include being explicit about problems, and the presence of teacher vulnerability, which helps people look clearly at success and failure. The concept of reflectiveness about changes on the "What can be restructured" map is included here as well.

At the right, the Policy Council (6) looks at the overall process, serves as a clearing house, and sets a time line.

Finally Peter identifies things we want to stay away from (7) that would make the other processes not work, or interrupt them. These include decisions without power to implement, and top-down curriculum decisions.

The HOW map is essentially organized around a sort of *dialectic of the change process.* The sense is that an active, progressive team takes initiative, constrained somewhat by other conservative team(s), with the result that a sort of de facto school-wide *consensus re change processes* exists. Peter also notes, in effect, the need for certain *norms as pre-conditions* for effective change (problem orientation, openness/vulnerability, reflectiveness, etc.).

. . . .

A conceptual theme underlying both maps [the WHAT map and this HOW map] is *peer-controlled power,* the idea that teams of teachers are responsible only to each other for the control they exert over the shared learning environment. A second underlying theme is *experimentation,* which Peter sees as central not only to the work of his team, but to teaching. "The philosophy is that experimentation doesn't always bring a good thing, but the act of experimenting is good for everybody."

The WHAT and HOW maps also seem to share the concepts of *reflectiveness* and *initiative* (by the progressive team).

Here the analyst is using the tactics of **factoring** and **noting patterns, themes.** (The comments above were derived from a matrix that displayed possible core themes for the analyst.)

Variations

This version of cognitive mapping makes for maximum idiosyncrasy—and complexity—in the results. A simpler version (Morine-Dershimer, 1991) asks the respondent to generate a list of concepts related to a major topic (e.g., restructuring). The major topic is placed in the center, and then other concepts are placed around it, with unnamed links radiating out to them, and from them to other concepts in turn.

Leinhardt (1987) describes a method in which the respondent works with a batch of prewritten cards (standard across all respondents), sorting them into piles, naming the piles, and commenting on the relationships.

You can use cognitive mapping techniques, such as folk taxonomies, to clarify your own ideas about the meaning of a particular set of data. Such a map is really an evolved conceptual framework. Computer programs with hypertext features that permit linking any bit of data with another and that form higher-level categories are particularly helpful here; we think of HyperRESEARCH, ATLAS/ti, and Metamorph (see Appendix).

Cognitive maps also can be drawn from a particular piece of text, such as an interview transcription, a folk tale, or a curriculum (McKeown & Beck, 1990). Here the analyst is interrogating the text, rather than the person. Relevant software includes MECA and SemNet.

Advice

Think carefully about the idiosyncrasy versus comparability issue: How important is it that your displays can be used across several persons or situations?

Use supplementary displays such as matrices to further condense and examine the cognitive maps—especially if you contemplate cross-case analysis.

Cognitive maps have a way of looking more organized, socially desirable, and systematic than they probably are in the person's mind. Allow for those biases when making analyses and interpretations.

Make a software decision early. The learning curves and user friendliness of the ones we have cited show considerable variation; leave enough time for fluency to develop on your part. The time will be repaid many times over.

Time Required

The mapping procedure Khattri and Miles (1993) used is labor-intensive, even after transcription of the two interviews. Extracting concepts from the premapping interview takes 2 hours or so. Correcting the "raw" map and writing the descriptive text typically takes 3 or 4 hours. The analytic work following may occupy another 3 or 4 hours (if supplementary matrices are used; if not, 2 or 3 hours).

Effects Matrix

Analysis Problem

In many studies the researcher is interested in outcomes—a topic with strong conceptual overtones. An evaluator may want to know what changes a particular program or treatment brought about in its target population. A more descriptive researcher may simply want to know where things stood at the end of some process (e.g., how well stabilized a newly created school is after 2 years). A researcher aiming at explanations (the topic we discuss further in the next chapter) will usually want to have a good look at some main dependent variables of interest (e.g., reading scores, dropout rates, school social climate) before turning to predictor variables.

These are examples of "ultimate" outcomes. That label may sound pretentious, but it only means that the variable is the last one being looked at in a temporal or causal chain. A researcher will be interested, too, in preceding outcomes, which usually are labeled "intervening" or "intermediate."

The problem for the qualitative researcher is how to select and display data that will faithfully represent the changed state of persons, relationships, groups, or organizations, seen as one or more outcomes of interest. Words are much harder to manage in this respect than numbers—where the primitive decision rules often turn out to be something like "Subtract 'pre' from 'post' scores and proceed to the statistic of your choice." For qualitative data, clarifying just which outcomes have occurred is not always an easy process. Conceptual clarity is at a premium.

Brief Description

An effects matrix displays data on one or more outcomes, in as differentiated a form as the study requires. The label "effect" is used to remind the reader that outcomes are always outcomes of *something:* a global program, an independent variable, an intervening variable. There is always at least an implicit predecessor. The basic principle in an effects matrix is focus on *dependent* variables.

Table 5.14
Effects Matrix:
Organizational Changes After Implementation of the ECRI Program

	Early Use 1st and 2nd yrs.		Later Use 3rd yr.	
Effect Type	**Primary Changes**	**Spin-Offs**	**Primary Changes**	**Spin-Offs**
Structural	Scheduling: ECRI all morning, rescheduling music, phys. ed. Helping teacher named: has dual status (teach/admin)	Cutting back on math, optional activities Two separate regimens in school Ambiguity of status and role	Integrated scheduling, cross-age grouping in grades 2-6	Less individual latitude: classroom problems become organizational problems
Procedural	No letter grades, no norms	Parents uneasy 2 regimens in class Teachers insecure Loosens age-grading system	ECRI evaluation sheets, tightening supervision	Teachers more visible, inspectable
	Institutionalizing assistance via helping teacher	In-house assistance mechanism implanted	More uniformity in work in all classes	Problems, solutions more common, public
Relations/ Climate	Users are minority, band together	Cliques, friction between users, non-users	Tighter academic press	Reduction in "fun activities," projects (e.g., Xmas)
			Perception by teachers of collective venture	More lateral help More 'public' distress

Illustration

When an organization, such as a school, implements an innovation, the possibility exists that the organization may change in some way as a consequence. Although some innovations can be "dropped into" an existing structure, serving as a replaceable part, most innovations turn out to make demands on the system and to have ripple effects. The organizational response often shows up in the form of something new—new attitudes, new rules, new procedures, new structures. In the school improvement study, we wanted to study the changes in local schools that were traceable to their having implemented an innovation.

Building the display. First we need to get clear what makes up the bundle called an "outcome" (in this case, organizational change). What parts or aspects does it have?

Table 5.14 shows how this worked out. The analyst decided that the outcome "organizational change" has three basic parts: structural changes, procedural or operating changes, and more general relational or social climate changes. The conceptual sequence is from "hard" to "soft" aspects of change.

Second, the analyst, reflecting on the particular case at hand, believes that such aspects of organizational change should be displayed separately for the "early use" period (first and second years) and for "later use," when 9 of the school's 11 teachers were users of the innovation.

Third, the analyst wants to distinguish between "primary changes"—those directly following from the requirements of the innovation's implementation—and those termed "spin-offs"—secondary effects, some of which may not have been fully anticipated. Because the usual convention (at least in a culture where the printed word is read from left to right) is that later events should be displayed to the right of earlier events, the analyst puts the time dimension in columns and the outcome types in rows.

This matrix crosses not two, but three lists: effects, time of use, and primary versus spin-off outcomes. Note that the primary change/spin-off division could just as easily have appeared as a subdivision of the three change types. That would permit easier comparison of, say, structural spin-offs at Times 1 and 2. As we've seen, any given set of data usually can be configured in a dozen or more ways. Each way has its advantages and limitations.

Entering the data. The cell entries in Table 5.14 are brief phrases describing specific organizational changes found in the coded write-ups of field notes. The original question to teachers, their principal, and several central office personnel was, "Did any changes occur in the organization of the school or the district during this period?" Follow-up probes asked about specific structural or "set-up" changes, procedural changes, and climate or feeling shifts.

The decision rule was that any change reported and verified by a document or at least one other respondent would be reduced to a summary phrase for entry into the matrix.

Drawing conclusions. Again we encourage you to take a first cut at writing some analysis of Table 5.14 before reading ours. (For context, the ECRI innovation is a relatively demanding, rather structured, language arts program being implemented in an elementary school. It takes a behaviorist approach to word recognition, phonetics, composition, and vocabulary.)

A look at the left side of the table shows us that the major structural change was rescheduling previously separate language arts activities into one integrated morning period; this change required rescheduling of the subjects taught by circulating teachers (music, physical education). This change, in turn, (see "Spin-offs" column) created two separate daily regimens, because the scheduling was done for two ECRI users in the first year and five in the second year. Other teachers were on the old schedule.

Other early changes included use of mastery levels, rather than letter grades, and the creation of a new helping teacher role to assist new users of ECRI. We can see by moving vertically down the columns that (as has been found in much social psychological research) structural changes tend to lead to procedural changes and, in turn, to climate/attitudinal changes (tactic: **making conceptual/theoretical coherence**). In this case the pressure, insecurity, and special attention led to a "we unhappy few" climate change, where users banded together. Moving across the climate row, we can see the spin-off effect: cliques and friction.

Moving to the right half of the matrix, we can note that, in the third year, because 9 of 11 teachers were now involved, scheduling became more integrated, and close supervision and uniformity were more typical. And it is clear that the fuller-scale, more regulated implementation now is seen by teachers as a collective venture. Looking across from primary changes to spin-offs, we see that teachers are both more likely and more able to express need for help and to give and receive it (tactic: **building a logical chain of evidence**).

This display lets us see how structural changes induce procedural and attitude changes. We can see how first-level changes lead to later consequences ("spin-offs"), and we can see how organizational changes flowed and developed over the 3-year period. Naturally these conclusions all need to be confirmed, tested, verified—for example, by **looking for negative evidence** or **checking out rival explanations.** But we have a good running start on what the effects added up to. If we simply had bundled all of the reported organizational changes into one content-analytic summary, all of this information would have been lost.

Illustration

Let's turn to another example, presented briefly. Suppose you were interested in sorting out, rather inductively, all of the "ultimate" outcomes reported as the result of some substantial intervention—in our study, the implementation of an innovation.

Building the display. Outcomes can be sorted according to their directness. Some outcomes are "direct" effects, like the "primary changes" just noted. Others are more general; they might be termed "meta-effects," outcomes that go considerably beyond the immediate direct effects the innovation is supposed to have. Finally, we can think of "side effects," outcomes quite far away from the original intention.

Each type of effect, we realize, can be either positive or negative. An innovation's proponents usually want others to accentuate the positive. But unwelcome outcomes are typical in any area of human endeavor (Sieber, 1981), and they should be included here.

Now, where can we look for outcomes? They can be *reported by* different roles (teachers or administrators), and they can be *attached to* different roles (outcomes for students, outcomes for teachers, etc.).

We should look also at *actual* outcomes in relation to the *intended* program outcomes—those the innovation's developer believes it will ordinarily accomplish.

Table 5.15 shows how the display worked out. We actually have four crossed lists here: directness of effects, positiveness and negativeness of effects, program objectives, and effects on different populations.

The innovation involved is a high school environment studies program. The people interviewed included three users of the innovation and two administrators.

Entering the data. The researcher has entered summarizing phrases, including all outcomes mentioned. Those that receive strong emphasis by the respondent, or that are confirmed by another respondent, get underlined. Entries that represent an inference made by the researcher, rather than

Table 5.15
Effects Matrix: Direct, Meta, and Side Effects of Program
(Environmental Studies, Lido High School)

		DIRECT EFFECTS +	DIRECT EFFECTS —	META EFFECTS +	META EFFECTS —	SIDE EFFECTS +	SIDE EFFECTS —
PROGRAM OBJECTIVES	*Effects on Pupils*	Plan & conduct env'l tasks Awareness of env'l problems					
	Effects on Staff	Hands-on tasks Work with community After-class work		Interdisciplinary skills			
	Effects on Community	Investment in env'l ed. activities					
SEEN BY USERS	*Effects on Pupils*	Env'l awareness Hands-on work		As adults, will make good env'l decisions in community		Improved values Kept dropouts in + self images	
	Effects on Staff	Hands-on approach Looser style		Completes env'l program here Outside workshops	Off-campus program is dumping ground	Acquired more off-campus sites "Expert" in community. Kept user in ed.	Crystallized hard work & low pay because of its demands*
	Effects on Community	Env'l awareness Knowledge of school programs					
SEEN BY ADMIN'RS	*Effects on Pupils*	Hands-on work off-campus				Success experiences	
	Effects on Staff			Utilizes their creativity			
	Effects on Org'n			No more discipline problems Rounds out program	Not cost-effective to transport 14 kids to site	Orientation to community	
	Effects on Community						

__________ = claim made strongly by one person, or by more than one respondent

* = inference made by researcher

a direct paraphrase of respondent comments, are marked with an asterisk.

Drawing conclusions. Using the tactic of **counting,** we can see immediately in Table 5.15 that positive effects at all three levels outnumber negative ones. And we can see also, by **comparing/contrasting** intended objectives with actually noted outcomes, that there is a good deal of congruence.

But some negative meta- and side effects are closely intertwined with sought-for outcomes. For example, a side effect of the program is that it keeps dropouts in, but its off-campus, practical aspect also means that the program becomes a "dumping ground" (tactic: **making metaphors**) for low-achieving students with discipline problems. These are the kinds of real-life contradictions that qualitative data analysis is good at surfacing and keeping visible throughout analysis and interpretation.

Variations

Effects matrices also can be organized by specific *persons;* you can show, for example, the changes in concepts, attitudes, and behavior experienced by several teachers as a result of their use of an innovation. Under these circumstances the cell entries will typically be much finer and "thicker," including items such as these:

> I'm more sensitive to kids, more aware of them individually and how I affect them.
>
> I learned not to fight too hard when other teachers weren't doing it right.
>
> You should be looking at what's good, not worry so much about where the kids are failing.
>
> No real change in the classroom, no.

Effects data can be retrieved from observation or available documents, as well as from interviews. And the typical "program"—a treatment aimed at achieving outcomes—has some type of evaluation component attached to it, with some already-collected data, which can be invoked as well.

Advice

Effects matrices are used most easily in complex cases and pay off well when multiple cases are included.

Unbundle outcomes carefully and then work to "substruct" them (see Chapter 7, section C) into a differentiated matrix. Consider two or three alternative formats before firming up. Revise if needed after a few data have been entered.

Be explicit and crisp about the decision rules for data entry. Outcomes usually carry more intellectual and practical weight than other types of variables: Much rests on the clarity and validity of outcome assessments. If your effects matrix does not include negative (undesired) outcomes, be prepared to give a good justification for the reader.

Time Required

Much depends on the degree of differentiation in the matrix, the size of the database, and the practical accessibility of coded data. A matrix such as the one in our first illustration (data from 15 informants answering a specific interview question) can usually be assembled in a few hours, with rapid analysis following. The second illustration, which involved scanning coded interviews from four people, took slightly less time.

Summary Comment

Analysis of qualitative data rests very centrally on displays that compress and order data to permit drawing coherent conclusions, while guarding against the overload and potential for bias that appears when we try to analyze extended, unreduced text. Two major families of displays are matrices (rows by columns) and networks (nodes connected by links).

We've described several types of partially ordered displays, which are useful in exploratory work. Time-ordered displays are essential in understanding the flow and sequence of events and processes. Role-ordered displays sort out people according to their position-related experiences. Conceptually ordered displays emphasize well-defined variables and their interaction.

Building displays is an interesting, creative task involving the management of trade-offs; entering condensed data from coded field notes needs to be done according to clear decision rules. Conclusions can be drawn and verified by using a wide range of tactics explored in more detail in Chapter 10. Conclusions embodied in successive versions of written text interact closely with the display (which may change during the process).

This concludes our discussion of exploratory, descriptive, single-case displays. We've repeatedly foreshadowed issues of *explaining* phenomena, how the analyst comes to be able to answer "why" questions. We now turn to this topic.

Notes

1. Narrative text, as a context-preserving, syntagmatic presentation, is a powerful way to help a reader reexperience events meaningfully linked over time—as in a story, novel, play, or biography (cf. Josselson & Lieblich, 1993). But such texts are already analyzed—reduced, ordered, focused. It is extended, *un*focused material that is hard to draw conclusions from. Analysts aiming for a phenomenological understanding of a text (cf. Melnick & Beaudry, 1990; Polkinghorne, 1988) may well need to read and reread an extended text to experience its deeper messages and meanings. But they, too, move toward focused displays, such as the "synopses" described in Chapter 4, section I. See also Smith and Robbins (1982) for useful suggestions on presenting results to busy audiences, including matrices, vignettes, summaries, and findings-oriented headings.

2. This way of looking at display-text interaction was outlined originally by Eleanor Farrar and Mary Ann Millsap.

3. These families of descriptive display are similar to the levels of questions that Wainer (1992) says can be answered through graphic displays: (a) "data extraction" (what's the state of affairs?), (b) trends over time, and (c) "understanding the deep structure," seeing both trends and groupings.

4. Contexts come in still other types, of course. For useful ideas on mapping the physical context and the kinship context, see Crane and Angrosino (1974) and Schusky (1983).

5. For penetrating discussion of the differences (and similarities) between survey and fieldwork interviewing, see Mishler (1986) and Schuman (1982).

6. Statistical techniques for estimating trends and significance levels of event histories are described by Tuma (1982) and Allison (1984). The clearest discussions of how to analyze "sequences" we have seen are by Abbott (1990, 1992). For a simple way to display and analyze how several variables co-occur during a time period, see Frick (1990).

6

Within-Case Displays: Explaining and Predicting

The question of *why* things happen is at the forefront of the research experience, once you're past the basic problem of understanding just *what* is happening. That question is also a salient one for people in their day-to-day lives; questions of "why," "how come," and answers beginning with "because" appear constantly in day-to-day interaction. If you doubt this, ask the next person you meet any "why" question ("Why did the meeting end so abruptly?" "Why did the president veto the bill?" "Why does it get cooler when it rains?" "Why is the checking account overdrawn?" "Why did you say that?") and you will usually be presented with instant explanations.

In the group problem-solving technique called "why questions," a chain of questions, answers, and following questions gradually clarifies a problem—or at least the assumptions that people are making about the problem:[1]

> Why do we have graffiti on subway cars?
>   Because kids want to express their identity.
> Why do they want to express their identity?
>   Because they are alienated.
> Why are they alienated?
>   Because they have no jobs.
> Why do they have no jobs?
>   Because they are unskilled.

Note that quite another chain could start like this:

> Why do we have graffiti on subway cars?
>   Because the cars are not protected in their yards at night.
> Why are they not protected?
>   Because the transit budget doesn't permit it.
> Why . . .

We can see that *assumptions* have a driving, controlling quality. The first chain assumes that explanations are to be found at the individual level; the second assumes that causes are economic and probably political. And there is a terrifying multitude of possible answers to any "why" question:

> Why do they have no jobs?
>   Because they are unskilled.
>   Because they are lazy.
>   Because there is structural unemployment.
>   Because employers are prejudiced.

But these are, at best, explanatory *claims,* assertions. If, as researchers, we're interested in predicting graffiti occurrence, or in helping someone control it, or in offering alternative choices for action, we have to go far beyond assertion, showing an empirical basis for the claim that Y is explained or caused by X.

A. Explanations and Causality

We'd like to explore ideas about explanation, and to outline our views of causality as a frame for the specific displays we describe.

What Does It Mean to Explain Something?

As Draper (1988) points out, "explaining" can include a range of activities: providing requested information or descriptions, justifying an action or belief, giving reasons, supporting a claim, or making a causal statement.[2] "Scientific" explanation occupies a fairly narrow band, usually focusing on the latter two. And scientific explanation of human behavior is an even narrower sub-band.

We like Kaplan's (1964) view that explanation is a "concatenated description . . . putting one fact or law into relation with others," making the description *intelligible.* Even so, as we saw with the "why" questions, every explanation is intermediate, containing elements that are themselves, in turn, to be explained.

We should not delude ourselves that scientific explanation is somehow lofty, wholly determinate, precise. Here, too, Kaplan is illuminating: Explanations, he says, are always open; they depend on certain conditions and are partial, approximate, indeterminate in application to specific cases, inconclusive, uncertain, and typically limited to specific contexts (pp. 351-355).[3]

What sorts of scientific explanation are there? Kaplan (1964) distinguishes "purposive" explanations, which either depend on individual goals or motives or serve some function, from "historical" explanation, which makes sense of a series of events. Similarly, Kruglanski (1989) says that voluntary actions or other events can be explained purposively or "teleologically," or causally, by reference to internal or external conditions. Good explanations must take into account both personal meanings and public actions.

Furthermore, good explanations will need to link the explanations given by the people we are studying with explanations we develop as researchers. But this link presents problems.

Sense-making and fallibility. Explanations offered in daily life have many flaws—and so do researchers' explanations. Goleman (1992) reports that people in simulated jury studies assembled evidence into a plausible "story"—but a story that depended heavily (45% of explanations given for a verdict) on the juror's personal assumptions or inferences about the motives of people in the case. The same public evidence, with extra assumptions/inferences added, thus could lead to varying verdicts (e.g., first-degree murder, not guilty).

Similarly Read, Druian, and Miller (1989) found that people construct scenarios that make sense of a sequence or order of events. In other words, they often impose a causal structure on events as a way of making them meaningful. In Box 6.1 the same events are ordered differently in two sequences. How do we explain what is going on? What are Louie's motives?

In Sequence 1 we construct a scenario that says Louie wants to be friendly and supportive to Bob. In Sequence 2 we tend to conclude that he is threatening Bob for failing to pay protection money. Note that we must use special knowledge, beyond the events, about (for example) what "insurance" means.

Furthermore, as Gilovich (1991) remarks in his fascinating book *How We Know What Isn't So,* people often assign causal explanations to random events, "believe [that] something is systematic, ordered and 'real,' when it is really random, chaotic and illusory" (p. 21). He notes that "people are extraordinarily good at ad hoc explanation. . . . To live, it seems, is to explain, justify, and to find coherence among diverse outcomes, characteristics and causes" (p. 22). Yet Gilovich shows us abundant evidence that such explanations often miss the mark: People misperceive and misinterpret data, make too much of ambiguous data, and often end up with biased interpretations ("seeing what we want to see").

These fallibilities are crucial for us as qualitative researchers, who must traffic in meanings, as well as actions; who must rely, in part, on the explanations that people give us; and who must live with the knowledge that because we, too, are "people," our own explanations are equally vulnerable.

Theory and data. The claims we make for our conclusions are usually buttressed by three features: We say (a) we have evolved, or tested, a theory; (b) we have stuck to all of the available, relevant data; and (c) there has been, in Ragin's (1987) terms, a steady "dialogue" between our ideas (theory) and the evidence (data). Thus a good theory, as Glaser (1978) suggests, is one whose categories fit (or have come to fit) the data; that is relevant to the core of what is going on; that can be used to explain, predict, and interpret what is going on; AND that is modifiable.

Here we are again with "theory." If our natural tendency is to tell an explanatory story about events, we can say, with Rein and Schon (1977), that theory is a sort of map that seeks to *generalize* the story at hand. A more worked-

Box 6.1
Two Event Sequences (Adapted from Read, Druian, & Miller, 1989)

Sequence 1

That morning Bob received a phone call telling him that fire had destroyed the stockroom of his store. Bob became very upset. Later, Louie threw open the front door of the store and walked over to Bob. Louie said, "I heard that you didn't make the latest payment on the insurance for your store. I'm really sorry for you. You never know when something bad will happen."

Sequence 2

Louie threw open the front door of the store and walked over to Bob. Louie said, "I heard that you didn't make the latest payment on the insurance for your store. I'm really sorry for you. You never know when something bad will happen." Bob became very upset. The next morning Bob received a phone call telling him that fire had destroyed the stockroom of his store.

out theory, they suggest, might be called a "model," one with a series of connected propositions that specify a number of components and the relations among them. To put it yet another way, a theory can be seen as a predicted pattern of events (Yin, 1991), which we place alongside what happened to see whether the pattern matches.

Van Maanen (1979), however, reminds us that there aren't really two compartments—"theory" and "data." Rather there are first-order *concepts*—the so-called "facts" of a qualitative study, which never "speak for themselves"—and second-order concepts—the "notions used by the researcher to explain the patterning of the first-order concepts." So we have to acknowledge, with Van Maanen (1988), that the "facts" we discover are already the product of many levels of interpretation; that is, facts are events to which we have given meaning.

There is thus a profound influence of our theories—implicit or explicit—on what we notice and how we explain it. One of the best recent exhibits of this influence is the "debate" in *Anthropology and Education Quarterly* between Everhart (1985a, 1985b) and Cusick (1985a, 1985b) on the meaning of schooling in their respective studies, thoughtfully described and critiqued by Noblit (1988). Both studied high schools, but their interpretations varied vastly. Everhart's critical-theory assumptions led him to focus on students and on schooling as a way to "proletarianize" students and reproduce the structures of capitalism through hierarchy and the passing out of reified information to students as "empty vessels." Cusick's structural-functional assumptions led him to focus on the school as created by staff, and on schooling as driven by an "egalitarian ideal," compromised by racial animosity, with "unappealing knowledge" taking second place to "friendly relations" with students. Each researcher saw the other's formulations as "ideological." In this case neither Everhart nor Cusick was trying to develop, or test, theory; their allegiances to pre-existing "grand theories" led each to drastically different conclusions about very similar phenomena.

Our point here goes well beyond the ideas of "subjectivity," "bias," and "multiple interpretations," as in the film *Rashomon* or Durrell's *Alexandria Quartet*. We need not settle for an endless proliferation of interpretations. Everhart and Cusick, with Noblit's assistance, made their assumptions *explicit*, so the connections between theory and conclusions came clear. In that way the strength of explanations can be assessed.

A Way of Thinking About Causality

In what follows we lay out our working assumptions about causality and how it might be assessed in qualitative studies.[4]

Just as there's no clean boundary between description and explanation, there's no precise demarcation between general "explanation" and "causality." It's not fair to say that determining causality is necessarily a "stronger" form of explanation than others. At the minimum, we can note, as philosophers from Locke and Hume onward have, that causality decisively brings in the question of *time:* We are concerned with finding out what led to what. Prior events are assumed to have some connection with following events (Kim, 1981), even though that connection may not be neat and clear.

The case can be thoughtfully made (e.g., Lincoln & Guba, 1985) that causality is not a workable concept when it comes to human behavior: People are not billiard balls, but have complex intentions operating in a complex web of others' intentions and actions.

Pessimism about understanding social causality has other sources. Idealist epistemologists have lined up behind Wittgenstein (in the *Tractatus*) in claiming that causes are no more than "superstitions." Others have evoked the "stickiness" of social life, the idea that its parts are so close together in space and time that it is hard to distinguish causal relationships (Harrod, 1956). Another line of reasoning emphasizes the "disorderly" nature of social facts; they are more chaotic and less coherent than our minds make them. As Hanson (1958) says, "Causes are certainly connected with events, but this is because our theories connect them, not because the world is held together by cosmic glue" (p. 64).

Should we give up? Hardly. There ought to be ways for us to understand how human events and meanings, actions and intentions, are chained over time, as sticky and disorderly as they may be.

Determining causality. How do we decide that one thing causes another? We might begin with the classical rules, originally proposed by Hume:

Temporal precedence: A precedes B

Constant conjunction: When A, always B

Contiguity of influence: A plausible mechanism links A and B[5]

We are looking here at regular, contingent relations among events. This is a bare-bones view. As House says (1990b), thinking about the evaluation of educational and social programs, life is not that simple. A program (e.g., Follow Through for young children) is not an invariant "A" that inevitably will lead to result B (e.g., higher self-esteem). Rather, we are faced with a locally unique complex of interacting factors.

Let's turn to a less stern, more pragmatic approach. Here are ideas suggested by the medical statistician Sir Austin Bradford Hill (1965), thinking about environmental causes of disease, such as scrotal cancer in chimney sweeps. What would lead us to conclude that an observed *association* represented a causal link?:

Strength of association (much more B with A than with other possible causes)

Consistency (A is found with B by many studies in different places)

Specificity (a particular link is shown between A and B)

Temporality (A before B, not the reverse)

Biological gradient (if more A, then more B)

Plausibility (a known mechanism exists to link A and B)

Coherence (A-B relationship fits with what else we know about A and B)

Experiment (change A, observe what happens to B)

Analogy (A and B resemble the well-established pattern noted in C and D)

So there are many routes beyond Hume's. We cannot unthinkingly apply his three rules, but have to work specifically and concretely. What features do we have to consider?

Local emphasis. We argue that causality is ultimately local, linked with specific nearby events in time. For example: As you read this, pick up a pencil and drop it. What caused the pencil to drop? We might evoke distant and abstract forces such as gravity and the relative mass of Earth and the pencil, or the authors' suggestion that you drop the pencil, made months or years ago as they were writing this chapter. But, in fact, we have to invoke immediate, local causes:

your having read the suggestion just now

your inclination to take it seriously

the presence of a pencil nearby

your picking up the pencil and holding it above, not touching, some surface

the coefficient of friction between your fingers and the pencil

your decision to open your fingers

the outward movement of the fingers

It can be argued that causal factors are arranged from local/inside to distant/outside in a sort of concentric circles pattern (Strauss & Corbin, 1990). Or take Bartlett's (1990) question of what causes students to line up by gender outside their classrooms: Is it a question of teacher management, of a historical residuum in primary education, of students' prior socialization, of the meaning of "order" in schooling, or of historical differences in gender arrangements? We can speculate about such distant factors, but if they have an effect, it is here and now. The immediate causal nexus is in front of us: A teacher felt, said, or did something, and then several students felt, said, or did something over a period of time. As Campbell (1986) says, we must be concerned with the validity of our findings in a *particular* setting by achieving "thorough local acquaintance."

Causal complexity. The causes of any particular event are always multiple (Abbott, 1992a). That statement masks even more complexity (Ragin, 1987): The causes are not only multiple but also "conjunctural"—they combine and affect each other as well as the "effects." Furthermore, effects of multiple causes are not the same in all contexts, and different combinations of causes can turn out to have similar effects. That statement means we have to think of causes and effects as arranged in a *network* (Salomon, 1991) that we approach as a system changing over time, rather than as a study of the billiard-ball impact of A (as vs. C) on B.

Even the most elegant quantitative procedures, at bottom, deal with associations, not really causes. They can only develop plausible possibilities "smoothed" across many persons and situations.[6]

Temporality. Time is crucial in such a view of causality. As Faulconer and Williams (1985) note: "Human events depend on time. Without the presence of possibility and its manifestation in a movement from one set of possibilities to another, there can be no human events" (p. 1183).

Assessing causality requires us to understand what Abbott (1992a) calls the "plot"—the events arranged in a loose causal order. (Remember E. M. Forster's famous distinction: "The king died and then the queen died" is only a story. "The king died and then the queen died of grief" is a plot, bringing in the "why" question—that is, causality.) Abbott notes that there are always multiple, intersecting plots: "Every event lies in many narratives at once. Every event has multiple narrative antecedents as well as multiple narrative consequences" (p. 438).

Although a variable-oriented view will inevitably show a looping back of assorted effects onto causes that produce new effects (Eden, Jones, & Sims, 1983; Weick, 1979), the "plots" still unfold over time and must be understood in that way (as we saw with the alternate stories of Louie and Bob's encounter in Box 6.1).

Retrospection. It also becomes clear that assessing causality is essentially a *retrospective* matter. We are always concerned with showing "how some event has occurred in a particular case," as House (1991) notes. Even if we do a classic experiment, we cannot tell what the effects were until afterward. As Mohr (1982) points out, we are concerned with looking at "an end point whose existence connotes the occurrence of certain prior events." It is more "pull-type" than "push-type" causality: The later event implies the preceding ones. So we must be historians and/or learn to "write history while it happens" (Lindee, Speaker, & Thackray, 1992). We use the historian's method of "followability": Our explanations "follow the causal action" (Abbott, 1992b). Or, to quote Polkinghorne (1988): "The report is retrodictive rather than predictive . . . a retrospective gathering of events into an account that makes the ending reasonable and believable . . . more than a mere chronicling. . . . It configures the events in such a way that their part in the whole story become clear" (p. 171).

Scriven (1974) formulated retrospective analysis as a "modus operandi" approach, where you look to see whether certain antecedent events took place and had clear connections to later outcomes. (In police work this approach also involves the location of recurring patterns across cases.)

Variables AND processes. Understanding causality can proceed by our identifying abstract *concepts* and seeing their interaction—doing what Mohr (1982) calls "variable" analysis and what Maxwell and Miller (1992) call "categorizing" or "paradigmatic" analysis. And it can proceed, as we have just seen, by the creation and parsing of *stories,* the flow of connected events in context—what Mohr calls "process" analysis and Maxwell and Miller call "contextualizing" or "syntagmatic" analysis.

In the "process" mode, we'll be likely to assemble chronologies, pay attention to time, and look for connections within the big picture. In the "variable" mode, we'll be likely to code small chunks of data, retrieve them, and look for similarities and conceptual patterns, with less regard to setting, sequence, and the passage of time. Clearly, *both* stances will be needed at different points in a study. The issue is being aware of your analytic assumptions and acting on them appropriately, moving back and forth between story and concept modes to deepen each.[7]

The Power of Qualitative Causal Analysis

The conventional view is that qualitative studies are only good for exploratory forays, for developing hypotheses—and that strong explanations, including causal attributions, can be derived only through quantitative studies, particularly the classical experimental-control design. With Maxwell (n. d., 1984), we consider this view mistaken. Seeing that an experimental group had effect X and that controls did not tells us nothing about what went on in the "black box." We don't understand how or why it happened, and can only guess at the mechanisms involved.

We consider qualitative analysis to be a very powerful method for assessing causality. Let's look at the features of causality we have identified. Qualitative analysis, with its close-up look, can identify *mechanisms,* going beyond sheer association. It is unrelentingly *local,* and deals well with the *complex* network of events and processes in a situation. It can sort out the *temporal* dimension, showing clearly what preceded what, either through direct observation or *retrospection.* It is well equipped to cycle back and forth between *variables* and *processes*—showing that "stories" are not capricious, but include underlying variables, and that variables are not disembodied, but have connections over time.

Of course, a useful theory should apply to more than one case. The assessment of local causality in this case needs to be tested and deepened through application of the causal explanations to other cases. The process of analytic induction (Manning, 1982; Miller, 1982) starts with a tentative hypothesis to explain something (in this instance, a cause-effect linkage). After looking at data from a single case, the hypothesis may be supported or revised, or the case is excluded as unexplainable by the hypothesis. This procedure is repeated through a series of cases, lending greater and greater confidence to the (evolving) hypothesis.[8] We have more to say about multicase analysis in Chapters 7 and 8.

In the rest of this chapter, we describe displays that can support good explanations by using the features we have described. We discuss the *explanatory effects matrix*, which looks at outcomes or results of a process, and then turn to the *case dynamics matrix*, a way of seeing "what leads to what."

Then we turn to *causal networks*, which pull together independent and dependent variables and their relationships into a coherent picture. Finally we consider the ultimate test of explanatory power: *making and testing predictions*.

B. Explanatory Effects Matrix

Any effects matrix, like the ones discussed in Chapter 5, makes us ask: Why were these outcomes achieved? What caused them—either generally or specifically? An explanatory effects matrix is a first step in the direction of answering such questions.

Here is a quick example. The researcher talked to users of an innovation (the same environmental studies program that appears in Table 5.15), asking them from whom they had received day-to-day, ongoing assistance, what it actually consisted of, and what the effects (both short- and long-run) seemed to be. The results are displayed in Box 6.2. The researcher entered quotes and paraphrases, aiming to get at the essence of the interview material appropriate for each cell. In a last column, the researcher adds a general explanation.

Such a matrix helps clarify a domain in *conceptual* terms; it is a useful first-cut exploration, beginning to trace back—and forward—the emerging threads of causality. In this case, that tracing involves the issue of which types of assistance, from which roles, lead to which types of effect in users—and why. For example, the technical help provided by developers and the materials seems to induce clarity and legitimacy, while it may take peer and administrative support to increase actual user mastery of the new program (the tactic here is **making contrasts, comparisons**: See Chapter 10).

The matrix helps us understand things temporally, and we get a first feel for the causal mechanisms that may be involved. So there is some heuristic value here. But a look at the "researcher explanation" column shows us mostly straightforward summaries; we have not advanced the causal quest very much. It is hard to see the links between assistance types, roles, and effects more than impressionistically. Furthermore each row is looked at one at a time as a causal sequence. It is hard to grasp the complexity—the interactions among rows. More detailed and thorough causal analysis usually requires other methods, such as case dynamics matrices, to which we now turn, or causal networks (section D).

C. Case Dynamics Matrix

Analysis Problem

The analyst, during and after data collection, is constantly trying to link data with explanations, trying to understand why specific things happen as they do—and how people in the cases explain why things happen as they do. How can you display, in a preliminary way, the explanations that seem relevant to a particular question?

Brief Description

A *case dynamics matrix* displays a set of forces for change and traces the consequential processes and outcomes. The basic principle is one of preliminary explanation.

Illustration

Suppose you were interested, as we were in the school improvement study, in the question of how and why an innovation induces change in the organization implementing it. That is a nontrivial "why" question, because the history of many innovations is that they are either shrugged off or absorbed into routine operating procedure, disappearing without a trace. Under what conditions does this *not* happen?

Building the display. If we are looking for causes, we might start with "dynamic" issues—things with a pushing or demanding quality. In this instance they might be "demands" or "requirements" or "strains" that the innovation carries with it. They can be in rows of the matrix. For columns we could include underlying issues or assumptions about the meaning of these demands, the organizational coping responses, and how things finally settled down, in terms of resulting organizational changes. Table 6.1 shows how this works out. The innovation is a program for doing individualized educational planning for all students in a school district, not just those involved in special education. As such, it had been deliberately designed locally to be an organizational, districtwide change, with its procedures to be used routinely by all staff.

Entering the data. Looking through coded field notes, the analyst can pick out chunks of material marked by relevant codes. In this case the codes were as follows:

Box 6.2
Explanatory Effects Matrix: Ongoing Assistance

LOCATION	USER'S ASSESS-MENT	TYPES PROVIDED	SHORT-RUN EFFECTS ("STATE" OF USERS)	LONGER-RUN CONSEQUENCES (ABLE/UNABLE TO DO)	RESEARCHER EXPLANATION
Superintendent	0	None	"He's going along just for the money"	Building/expansion of program on their own	Supt was a money manager not involved with programs
Bldg Admin'n *Years 1-2*	+	Met w/ community Gave encouragement Buffered supt	"We have some help here" "We're not alone" "We'll fight our battles"	Left users more time to address program issues, students, site, buses	Admin'r supported innov. programs by providing needed help to teachers
Years 3-4	0	None	"We lost our friend"	Had to fight own battles and thus learned how	Admin'r's relative indifference probably solidified users
Developers	++	Gave ideas, help Provided framework Activity planning Positive outlook Resources	"They know their stuff" "It all makes sense" How it could work "Hey, I can do this!" "We can get equipment"	Expansion of science program to off-campus sites w/ community involvement	Dept needed money to do the program: one user needed direction and support: they got both and off they went
Peer Users	+	Helped w/ planning Gave ideas, sugg'ns Gave encouragement	How it could work Filled in gaps Not alone; there's help	Strong users of the off-campus program; they know how to do it	One user was experienced in this approach and brought others up to speed
Materials (Guides)	+	Overview, approach, and suggested activities in detail	This is really good stuff, well worked out, tried, useful	Good basis for development of own program	Mat'ls fulfilled their role: to stimulate & aid in development of local program

TR-PROBS: implementation problems

TR-ORG/PRAC: effects on organizational practice

TR-ORG/CLIM: effects on organizational climate

TR-SIZUP: explanations for effects on the organization

The analyst has also decided to sort organizational effects into three exhaustive general classes: structural, procedural, and climate. This sorting requires "unbundling" the TR-ORG/PRAC chunks into pieces that are really procedural and those that have a structural aspect.

The decision rules for data entry in Table 6.1 were (a) locate discriminably different "demands" and summarize in a phrase or two each the essence of the demand, the coping method, and the resolution; (b) use only data that are not contradicted by other informants; (c) for "underlying issue," move up a level of abstraction, drawing on your ideas about organizational theory and the explanations site people offer, and enter concepts that seem most relevant. The question is, What is the demand a specific example of?

As always, *any* display forecloses some issues. Here the analyst is looking at only those demand-caused effects that actually constituted an organizational change, and is tacitly excluding demands that cause no organizational effects, or are absorbed without a trace.

Drawing conclusions. In this sort of matrix, much analysis occurs during the actual data entry, which is done by moving across each row. By the time the row is filled in, the analyst has a first sense of what the dynamics have been (tactic: **noting patterns, themes**). Analytic, summarizing text can be written by cycling back as needed to the written-up field notes for amplification and clarification. Here is an example of such text for the first row of Table 6.1:

> The enthusiasm of the "early settlers" in IPA was strong; teachers would often spend a weekend with one of their five children, doing special activities. But as the program expanded, there was tension and role conflict: Parents somehow expected teachers to carry out all the activities. Official expla-

Table 6.1
Case Dynamics Matrix:
The IPA Innovation as a Force for Organizational Change in the District and Its Schools

STRAINS, DIFFICULTIES CREATED	UNDERLYING ISSUES (AS SEEN BY RESEARCHER)	HOW COPED WITH	HOW RESOLVED: TYPE OF RESULTING CHANGE
Conflicting expectations: should parents or teachers do activities?	Work load. Parent-teacher role conflict.	"Explaining" that teachers could not take primary responsibility for out-of-school activities.	Increased use of "batched" activities, many set up by coordinator (P).
View that forms and procedures were "extra," overloading.	Work load. Autonomy, resistance to control.	In-service assistance.	Repeated revision and simplification of forms and procedures; production of an operating manual. Reduction of expectations (no home visits, fewer conferences) (P).
User uncertainty and resistance to use.	Autonomy.	In-service assistance. Management Team interviews of all staff.	See above. Also creation of in-service committee (S), with coordination through Management Team (P).
Extra time requirements.	Work load.	Initially, via volunteerism, high commitment.	Use of substitutes (P). Dismissal of school during conference days (P). Reduction of expectations (above).
Program is complex, demanding, externally funded.	Authority, coordination, accountability.	Early creation of Management Team, addition of elementary teachers.	Institutionalization of Management Team (S). Heightened expectations for teacher upward influence (C). Lowered morale when expectations violated (C).
Enthusiasm of "advocate" teachers led to peer criticism.	Autonomy.	Quieting conflicts through informal discussion. Informal coordination and referral.	Norms supporting flexibility and colleague influence within schools, and cross-school interaction (C). Increased linkage and closer interaction between schools (C). Hobby Day (P).

(S) Structural change
(P) Procedural change
(C) Climate change

nations that parents and students would have to "take responsibility" were only partially effective; the most favorable estimate was that only 25-30 percent of parents actually helped out, and a few were actually resistant to "nonschool" activities. The gap had to be filled by the coordinator, who was increasingly expected to set things up and carry them out (a program on dinosaurs, visits from football heroes). Batching in classrooms began: units on trucking, or oceanography, were used for groups of interested children.

Similar analysis can proceed row by row. We can also look at the types of resulting change in the last column, noting that genuine structural changes are fewer than procedural or climate changes (tactic: **counting**). And we can look down the "issues" column to see that issues of work load and autonomy are occurring repeatedly (tactic: **clustering**). Cycling back over to the right column, we can see that work load issues were generally resolved through routinization and reduction of effort. But, using the tactic of **making contrasts, comparisons,** we can see that autonomy issues were resolved in an interesting, unexpected way: increased interdependence and colleague influence, *closer* linkage. When the analyst runs into such a conclusion, it poses a puzzle: How to shed light on an unpredicted finding? This calls for the tactic of **following up surprises.**

Perhaps we can go to other rows to find possible explanations. For example, in the second row from the bottom, we note that the Management Team was created to solve problems of coordination and also heightened expectations for upward teacher influence. Perhaps the team is a way of trading increased influence for decreased autonomy? In the field notes (tactic: **making if-then tests**), there is partial support for that idea: The team is seen by different parties as increasing ownership, even co-optation ("Teachers are less quick to criticize if there's a colleague on the Team"). That fits, roughly. But others mention power mobilization and institutionalization as the main *raisons d'etre* for the team. So we must look elsewhere.

Reviewing field notes further, the analyst comments:

> It also appeared that the program induced more elementary-secondary contact and collaboration—partly because many teachers were themselves parents, and experienced the IPA program in that role (and could notice poor or partial implementation as well as benefits).
>
> Dorry Hoffman was even able to get the revered Coach Covington to work with some of her third grade students by telling him she had some promising running backs.
>
> Several high school teachers came to the elementary school to help with Hobby Day, an IPA offshoot (substitutes were provided). And there was a well-worked-out program for sixth graders to orient them to what life in the high school would be like the following year.

So people's willingness to be more interdependent and less autonomous seems to have been driven by direct experience in working with others, with benefits accruing (tactic: **factoring**).

With this type of display, we aren't wholly stuck in row-by-row analyses, and can see how rows influence each other—even though it is not as rich as a network display would be.

The data in this sort of explanatory matrix are at several removes from the actual field notes, so the display will work best as a summarizing, clarifying, puzzle-posing (but not always puzzle-solving) device. Cycling back to the field notes is often a helpful move.

Variations

Case dynamics matrices, rather than starting with "strains," as this one did, can use "dilemmas"; rows also can be clustered conceptually. Table 6.2 is an example on the same topic as Table 6.1. Here the analyst sorted the dilemmas into the familiar "structure," "procedure," "climate" triad and also tightened the "coping" category to specify who actually did what. This matrix, too, is limited to changes that caused friction and disequilibria. A check back to Table 5.14 shows the full range of changes, some of which were well accepted.

This illustration excludes cases of nonchange. A column could be added easily that emphasizes stability and nonchange data—particularly on outcomes that might have been expected a priori to change as a result of the initial push or strain. That addition, in turn, suggests the need for an "explanations" column.

Columns also can be added that directly invoke explanations offered by site personnel.

Advice

Review (or have a colleague review) the first versions of your matrix to see what you are excluding and what analyses you will and will not be able to make.

Because this analytic work is quite inferential, it pays to have a colleague review your entries against field notes or even to try a replication. You also can ask someone else to look at the completed matrix and tell you what your analytic assumptions seem to be. Can alternative assumptions be made that are more fruitful?

Move back and forth regularly between the matrix and the coded field notes to (a) confirm and deepen conclusions and (b) unravel puzzles as they appear.

In writing analytic text with your conclusions, *link* them with the data excerpts that led you to them. That inductively based illustration is very different from what has been called the "sprinkling" function of anecdotes in text, where the researcher hunts for items that will entertain or convince the reader and then "sprinkles" them here and there as needed.

Time Required

Case dynamics matrices of this size and complexity usually can be assembled in 2 or 3 hours, if the coding scheme is well coordinated with the column headings. If not, the task of assembling a confident set of second-level generalizations/explanations becomes much more time consuming. And you have to rely more heavily on informal syntheses, intuitions, impressions—which, in turn, need to be checked back against the write-ups, making at least one complete pass through them. Expect this to take a day, or more.

D. Causal Network

Analysis Problem

Field research is a process of progressive focusing and funneling. As data are collected, we can see, more and more clearly, factors that bring the apparent flotsam and jetsam of local events into a meaningful pattern.

These local factors are not little lizards hiding under stones turned over by the researcher in the course of investigating the local terrain. Rather, they are a mental map that

Table 6.2
Case Dynamics Matrix: Organizational Dilemmas Posed by ECRI Program

	DILEMMA	ISSUE	WHO DID WHAT	HOW RESOLVED ORGANIZATIONALLY
STRUCTURE	- voluntary use vs. scheduling, curricular complications	2 regimens in same school	principal seeks continuity	non-users pressured to do ECRI or leave new slots filled with ECRI users
PROCEDURE	- voluntary, flexible use vs. perceived orthodoxy	classroom 'policed,' teachers unable to make adaptations	principal enforces evaluation sheet; users take liberties on the sly	principal: no evaluation for 1 year, then tightening prescriptive use; some latitude allowed
	- complexity, pace, redundancy of program	exhaustion, calls for aid, latitude, time off task	principal creates planning period, bends rules on Title I aides, brings helping teacher office into building	
	- "losing one's pupils" in morning scheduling	1/2 - 2/3 scheduled in homeroom classes	principal attempts to keep number down	satisfying 1-on-1 with own pupils; contact with new children
	- assistance vs. control	'helping' teacher also controls fidelity of implementation	shifting aid to helping teacher, control to building administrator	
CLIMATE	- users vs. non-users	users form cult, resented by others	coffee room conversation dominated by ECRI	most teachers become users non-users 'vindicated' (ECRI unsuited for lower grades)

you gradually elaborate in the course of piecing together discrete bits of data. This abstracted web of meaning is an achievement, not a given.

It is also an achievement for local informants, as we noted in Chapter 5, section E; they, too, walk around with a map in their heads that provides a frame for action and perception and explains what causes what in their world. This mental schema works for them; they can compare it with others' maps and use it to get what they are after (see Axelrod, 1976).

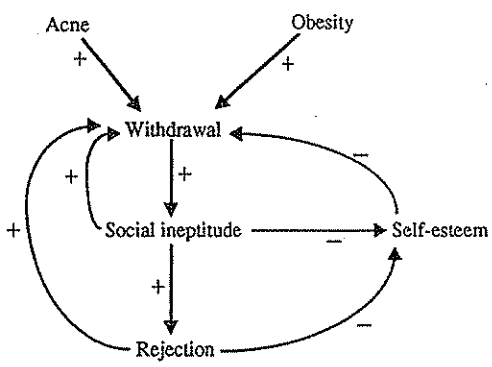
So, much of field research has to do with schema absorption and re-elaboration; you go around recording individuals' mental cause maps, putting them together, and making connections with your own evolving map of the setting.

The rub here is that you have to get up to speed, with a map that is initially far more primitive than informants'. So your first understanding of local cause maps is usually jumbled and vague and sometimes plain wrong. It takes several tries before things look brighter—and clearer. Trying to shortcut map iterations—most notably with an a priori theory—is a perilous, often self-defeating exercise, like jumping off a 40-meter ski jump after having practiced only on moguls.

Weick (1979) outlines another problem. When you have two related events, it is often hard to call one a "cause" and the other an "effect." The variables control each other in a loop. You can start your analytic sequence with a cause that triggers subsequent events, but they come back and modify the original "cause" you had identified. That "cause" is now an "effect," as Box 6.3 illustrates. An adolescent's acne and obesity might have led to withdrawal and failure to acquire social skills. Social ineptitude leads to further withdrawal, as well as to rejection; both reduce self-esteem.

Causation is not unilateral. The problem is how to get a reasonable handhold on the "plot"—the flow of events and

Box 6.3
Cause-Effect Loops (adapted from Weick, 1979)



states in a particular case—and not get lost in a blur of everything affecting everything else. Weick (1979) suggests "cause maps," which summarize a set of variables and connections that are inferred by an analyst after several exposures to a stream of experiences—much like the causal networks we are about to describe.

Developing good causal maps that respect complexity and are plausibly connected with those of respondents is the challenge. Meeting it helps the analyst avoid intellectual overload, and produces results that are systematic and verifiable. As we'll see, causal networks can be checked out thoroughly with site personnel long before the final report is written.

So the basic questions are as follows:

1. How can you build a progressively integrated map of case phenomena that has local causal significance?
2. If there are multiple cases, how can you align their maps to make cross-case maps containing more-general causal explanations?

For both questions the answer lies in the creation of a "causal network" that is an abstracted, inferential picture organizing field study data in a coherent way. In this section we review how to do this for a single case; in sections E and F of Chapter 8 we outline the multiple-case approach.

Brief Description

A *causal network* is a display of the most important independent and dependent variables in a field study (shown in boxes) and of the relationships among them (shown by arrows). The plot of these relationships is *directional,* rather than solely correlational. It is assumed that some factors exert an influence on others: X brings Y into being or makes Y larger or smaller. A causal network, to be useful, must have associated analytic text describing the meaning of the connections among factors.

Illustration

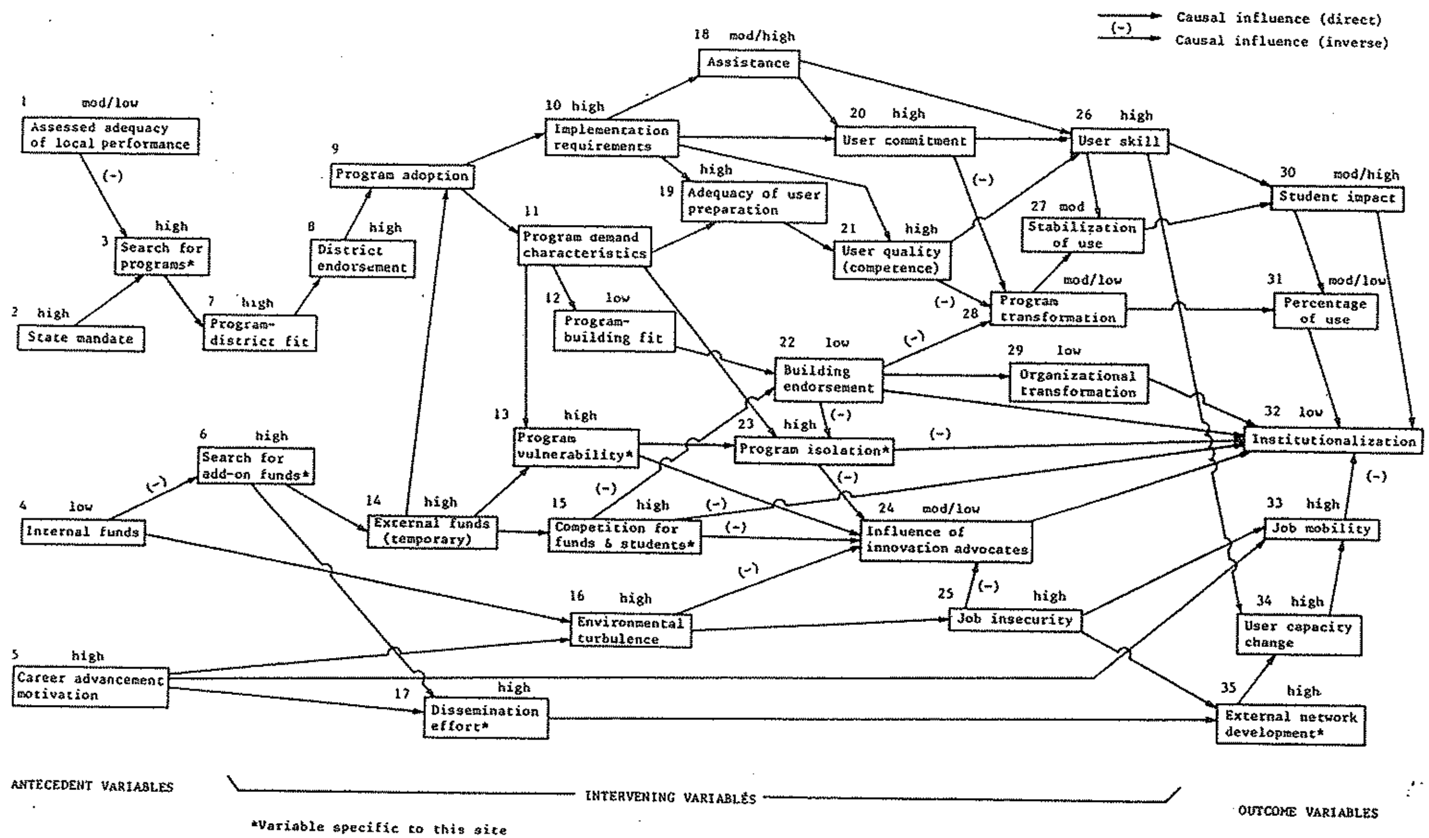
Figure 6.1 shows a causal network for one case. At first it may look complex and forbidding. But look at one section of it (say, starting at the upper left) in conjunction with some text, like this:

> The first three antecedent variables (1, 2, and 4) worked out this way. The state mandate (2) for well-planned career education programs, together with assessment of local performance as less than adequate (1), led to a search for new programs (3), which proved to be a good fit (7) with district characteristics, and hence to district endorsement (8), and to adoption of the program (9).
>
> But these were not sufficient causes of adoption. Inadequate local funds (4) to cover existing programs led to a ceaseless search for add-on funds (6) as almost a "way of life" for the district. That search led to getting temporary external funds (14) for a three-year period; they were the other basic cause of adoption (9).
>
> The program, when adopted, proved to have substantial implementation requirements (10), which dictated the need for assistance (18), and also exerted a good deal of pressure for careful selection of high-quality staff (21), and for careful preparation (19) of them to carry out the program. The heavy implementation requirements (10), and to some degree the assistance provided (18), induced a good deal of user commitment (20), and in turn user skill (26), which was also high because of staff quality (21). High user skill, in conjunction with the fact that the program was quite well stabilized (27) by late 1980, brought about a reasonable degree of student impact (30).

Text and network together communicate more than either could alone. We are moving toward an explanation—not just a description—of what happened at Perry-Parkdale. The network has taken 300-400 pages of field notes down to a few dozen boxed words and arrows on one page; it and the text tell economically how and why things turned out as they did. Both a sequential (case-oriented) and a variable-oriented optic have been used.

A causal network like this one can be corrected/verified not only by researcher colleagues but also by people you are studying. Of course, these "verifications" will come down to a matching of interpretations based on diverse experience (Denzin, 1989b), and interpretations will not always fully overlap. Still there will be grounds for assessing and adjudicating the plausibility of different accounts.

Figure 6.1
Causal Network for Perry-Parkdale CARED Program



We can testify that developing and refining causal networks is fascinating and enjoyable, with plenty of "aha" experiences. How do we get there? Let's walk through the process.

Two general approaches to building causal networks. Basically, you have to intend from the beginning of data collection to produce a causal map, so that successive rounds of data collection, interim analysis of all the data, and iterations of the map itself all build toward that end. But there are two different ways of approaching the task, which can be loosely labeled "inductive" and "deductive."

In the *inductive* approach, also called constructive or generative (Becker, 1958; Goetz & LeCompte, 1984; Zelditch, 1962), the researcher discovers recurrent phenomena in the stream of local experience and finds recurrent relations among them. These working hypotheses are modified and refined progressively in the next fieldwork pass. The local cause map emerges piecemeal and inductively. It has regularity and pattern; some things happen only when others do or don't. These things and the links between them acquire their own names or labels, cluster into probable causes and the effects they appear to engender—and the analyst has a causal network.

In the *deductive* strategy, also called "enumerative" or "conceptual" (Kaplan, 1964; Popper, 1968), the researcher has some orienting constructs and propositions to test or observe in the field. These analytic units are operationalized and then matched with a body of field data.

The deductive researcher *starts* with a preliminary causal network, and the inductive researcher *ends up* with one. Wolcott (1992) calls the approaches "theory-first" and "theory-later." In either approach, the initial version is amended and refined as it is tested against empirical events and characteristics. By the end of data gathering, both species of researchers are about at the same place. Both are substantiating a cause-and-effect map, but the "conceptualist" has a shaken-down model and the "constructivist" has a built-up one.

So induction and deduction are dialectical, rather than mutually exclusive research procedures. The constructivist's inductions are informed by a personal conceptual universe; the conceptualist's a priori frameworks contain more empirical data than at first meets the eye.

In earlier sections we tilted toward more front-end conceptualization. That tilt came from personal experience (Huberman, 1981b; Miles, 1979b) in studies where, without clear initial conceptualizing, we were drowned in tidal waves of shapeless data. We learned—inductively—that a somewhat more deductive approach would have reduced and focused our data set without losing juice or meaning, and helped us find causal relationships faster. In what follows, we approach causal network building from the more a priori side, while highlighting the values of doing it more inductively along the way. Most of the techniques we suggest can be used by analysts of either predilection, when they seem most useful. The main advice we offer is to avoid an either-or approach to the timing of conceptualization.

We'll also show how several methods for within-case analysis we've covered earlier can be building blocks for the causal network.

Getting started. What you find in any quest depends mostly on what you set out to find, and where you choose to look for it. In one form or another, you begin with a finite number of ideas and questions. You may take an initial cut at how some of these things might fit together, just as we did in the conceptual framework (Figure 2.3) that we generated for the school improvement study. Such a framework already carries some causal freight. It makes guesses about which factors logically *influence others,* which ones are likely to *appear together* and which not, and which ones *have to happen first* for others to happen later (directional influence).

The field researcher who begins with such a conceptual framework is not boldly going to ignore data that suggest otherwise. He or she is, however, going to generate research questions, codes, and samples of people, events, and processes that give the framework a chance of working out.

A more inductively oriented researcher, by contrast, would focus initially on a more general domain or macroconcept. For example, a field-worker might start with an interest in "servicing" and hang around a department store to see how it goes on—or, as Glaser (1978) describes, start with an interest in whorehouses and end up with "servicing" as the key to their generic processes.

In effect, no researcher can *really* begin with a tabula rasa mind. The tacit question is, What led me to start this study—what am I here for? The issue is, How soon and how systematically do I bring my assumptions and framing ideas to awareness? The following issue—a sharper one for "theory first" researchers—is, How do I permit data to elaborate, correct, and extend my ideas?

In our approach, stressing early conceptualization, the first steps toward causal analysis are the—deliberately general—conceptual framework (Chapter 2, section A) and research questions (Chapter 2, section B), along with the start list of codes (Chapter 4, section B). Next come the reflective remarks and added marginal remarks in the transcriptions of field notes (see Boxes 4.2, 4.3, and Figure 4.6). These are typically modest, data-shaping exercises. They alert the analyst to variables that go together and that contrast with other variables, and they invite a closer look at something that might be an underlying theme or pattern.

All these steps mean that discrete variables are getting clumped into tentative families. If they turn out to stay in those families through successive data-gathering and analyzing cycles, they are good candidates for a "box" in the causal network, with an arrow leading to or coming from another "box."

For example, in our study of Perry-Parkdale (Figure 6.1), we noticed that teachers using innovations described their experiences in such terms as, "Finally I felt really on top of it," "It got easier when I knew how," "It's a snap now," and "Basically, I learned how to do it right." These comments typically were more frequent *later* during the process of implementation. It didn't take too much insight to label the underlying variable "user skill" and to make a box for it.

Assembling fragments: Pattern codes, interim memos, and summaries. As we saw earlier, marginal and reflective remarks often translate into pattern codes (Chapter 4, section C). They are still bigger bites of data, and often become candidates for inclusion in a causal network. In essence, they signal a **theme** or **pattern** that makes a difference locally. For example, the TEAMS metaphor describing a conflict between two administrative factions in a school district (Chapter 4, section B) was later translated into a stream of variables such as "career advancement motivation" and "climate in district office" on the causal network that eventually affected key outcome variables such as "institutionalization" of the educational innovation.

Moving up another notch, pattern codes get extended into *memos* (Chapter 4, section D) and then into initial attempts to pull coded data and memos together in the *interim case summary* (Chapter 4, section F).

Of course, you don't work this hierarchically or deliberately. But on a cognitive level, this is roughly how discrete pieces of data are combining into an evidential chain that has a beginning causal logic. You think, roughly, "These variables are present or absent together (they covary), while others look random or unconnected." The tactic is **building a logical chain of evidence**, a sort of "abstracting induction" (Duncker, 1945). But with more knowledge, you begin to think, more elaborately, "Some of these variables are coming into play *before* others, varying *with* others, or having an *effect* on others, and that effect seems to change when *other* variables are taken into account." You have the rudiments of a cause map that contains guesses about *directions* of influence among *sets* of variables.

Operationally, however, these variables are little more than informed hunches about covariation that you need to check out during the next field visit—or play back to your critical colleagues for reactions. The *case analysis meeting* (Chapter 4, section E) is useful here. In a single-case study, it can surface alternative rough maps; in a multiple-case study, it can permit a trial of one analyst's emerging causal map on another's data (tactic: **replicating a finding**).

All of these devices support, even force, analytical activity. They are occasions for focusing and tidying up your thinking. So this is a good time to lay out some of the fragments of your emerging causal map in one place by doing the following:

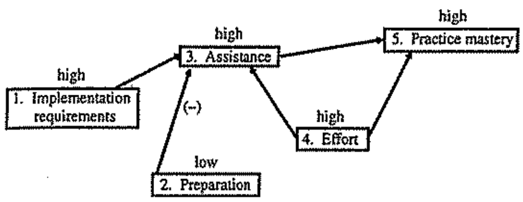
- Consider a *specific case* for which you've reviewed a lot of data.
- Translate the pattern codes into *variables*, that is, something that can be scaled (high to low, big to little, more to less).
- *Rate* the variable (e.g., high, moderate, low). How much of it is in the case?
- Draw a *line* between pairs of variables that *covary*, that is, that appear together consistently in the case, that have some kind of relationship—for example, more of one variable goes with less of another.
- Draw a *directional arrow* between each variable that comes first (temporally) and those later ones it appears to influence. *Influence* here means that more or less of one variable determines to some extent the rating of another. The rating of the second variable might have been different had the first one not been there—a reasonable "mechanism" is involved.
- If two variables covary but seem to have only a tepid or oblique influence on each other, probably another, latent variable needs to be invented to join the two (tactic: **finding intervening variables**). Review the full list of codes to see whether one fits here.

Figure 6.2 shows a causal fragment. This one puts together themes concerning the mastery of a new educational practice. The story can be told quickly. A demanding project with high implementation requirements (1) began with inadequate preparation (2) and was bailed out by high levels of assistance (3), which increased local efforts (4) and facilitated practice mastery (5). The minus sign indicates inverse causal influence: Low preparation led to high assistance in this case.

Once again, note: This fragment is *data*-based, not an a priori conceptual framework. The intent is to stimulate thinking, not to obtain closure. Try to assemble a few such fragments without necessarily connecting one to another. Play around; do some causal noodling, fiddle with other fragments. Don't try to connect variables that don't go together empirically, even if on logical grounds they "should." Some variables, like some blood types, don't commingle well, and you can get led easily into Byzantine reasoning to decide that they do go together.[9]

A useful next step is to take these fragments out to the field to see how and whether they work. This step can include some "member checks" (Guba & Lincoln, 1981)—that is, showing the rough diagram to a case participant (tactic: **getting feedback from informants**) to get confirmation and qualifications. We might find, for in-

Figure 6.2
Causal Fragment:
Mastery of a New Educational Practice



stance, that some preparation was decent, some assistance had no effect on levels of effort, and some users were competent and experienced enough to do without preparation or assistance and still get high practice mastery. That result might suggest that the next iteration of the network should have an arrow from a new box labeled "user competence" to the "effort" box (4) and to the "practice mastery" box (5).

Generating a causal network variable list. Late during data collection, you may be ready to assemble the remaining pieces of the causal network. A useful first step is to generate the full set of network variables. This, too, is an exercise in playful brainstorming.

The idea is to list all of the "events," "factors," "outcomes," "processes," and so on that seem to be important and then to turn them into variables. For instance, the several fights between employees will become "organizational conflict." The first pass should be exhaustive, the next one more selective. The first list should be combed for redundancies and overdifferentiation (e.g., three types of fighting between employees). As an illustration, Figure 6.3 presents a list of core variables generated in our school improvement study. (This was a complex, multicase study; the list for single, simpler cases would be much shorter.)

Such a list typically combines "constructivist" and "conceptualist" approaches. Some variables come directly from the field (e.g., external funds, program development); others are there because the initial constructs or research questions oriented the researcher toward them (e.g., program-district fit, assistance, program transformation).

In a single-case study, making and revising the variable list is fairly straightforward. In a multiple-case study, we are at a decisive moment. For cross-case comparisons, the same variables will be used to analyze 5, 10, or 20 cases. For this to happen, ideally each of the variables should be empirically meaningful in all cases. Of course, you should leave slack for two likely situations: (a) there will be some case-unique variables, and (b) some variables will be influential in most but not all cases. In the former instance,

Figure 6.3
List of Antecedent, Mediating, and Outcome Variables: School Improvement Study

Antecedent or Start Variables	*Mediating Variables*	*Outcomes*
Internal funds	External funds	Stabilization of use
Career advancement motivation	Program adoption (E)	Percentage of use
Assessed adequacy of local performance	Program concept initiative (L)	Student impact
Environmental turbulence	District endorsement	User capacity change
	Building endorsement	Institutionalization
	Influence of innovation advocate	Job mobility
	Implementation requirements	
	Adequacy of initial user preparation	
	Program-district fit	
	Program-building fit	
	Program-user fit	
	Assistance	
	User commitment	
	User skill	
	Program transformation	
	Teacher-administrator harmony	
	Validation effort (L)	
	Stability of program leadership	
	Stability of program staff	
	Organizational transformation	

Legend:
(E) externally developed innovation
(L) locally developed innovation

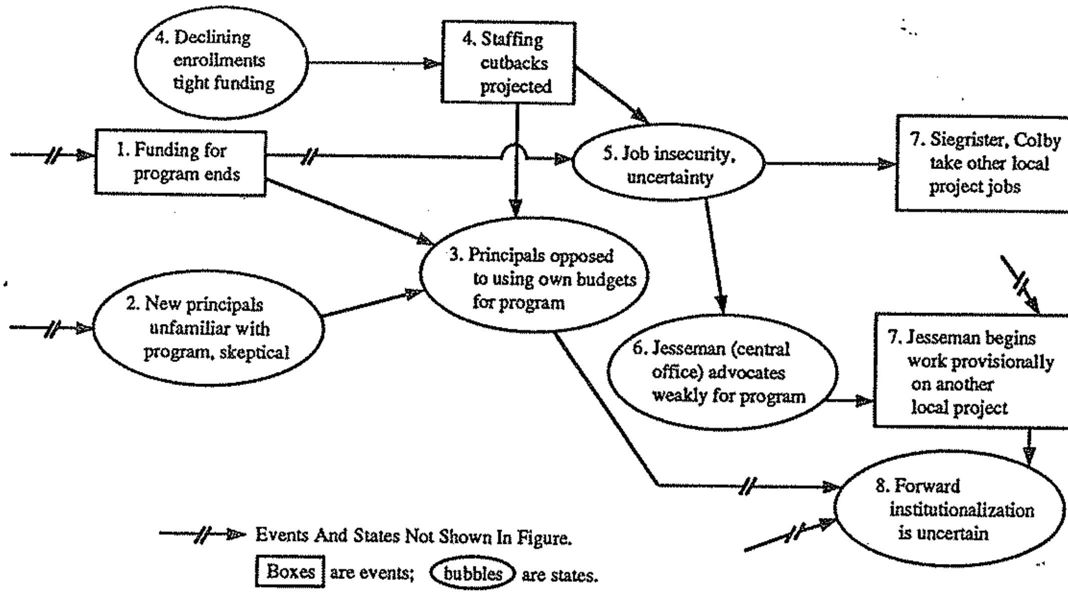
the final causal network will contain some case-specific variables, labeled as such. In the latter instance, some network variables will be dropped (with an explanation) from the cases where they contribute little or nothing to the analysis.

If you are a lone researcher doing a multicase study, the list-building process will go more easily if you've had plenty of prior contact with a critical colleague around field notes, pattern codes, memos, interim case summaries, and causal network fragments. If a research team is involved, expect an extended period of getting people's heads together through such discussion, so that variables have a common meaning.

Drawing the causal network. In principle, you could draw a full causal network for the case at this point, and then go

Figure 6.4

Excerpt From an Event-State Network: Perry-Parkdale School



back to the field for final testing and refining. We do not advise this. The risk is that network vision sets in too early. The analyst constructs a cause map and begins to use it to interpret all of the phenomena of interest. A final "coherent" picture emerges before the individual parts of the picture have been carefully studied individually and in combination. It's as if a survey researcher were to begin estimating a path model before looking at the distributions, cross-tabulations, and beta weights. The answers go looking for the questions.

The better alternative is to save full causal network drawing and analysis for late in the project, making it perhaps the last analytic exercise. This step does not rule out getting reactions to the network from the site later on, if necessary by mail, as we describe below.

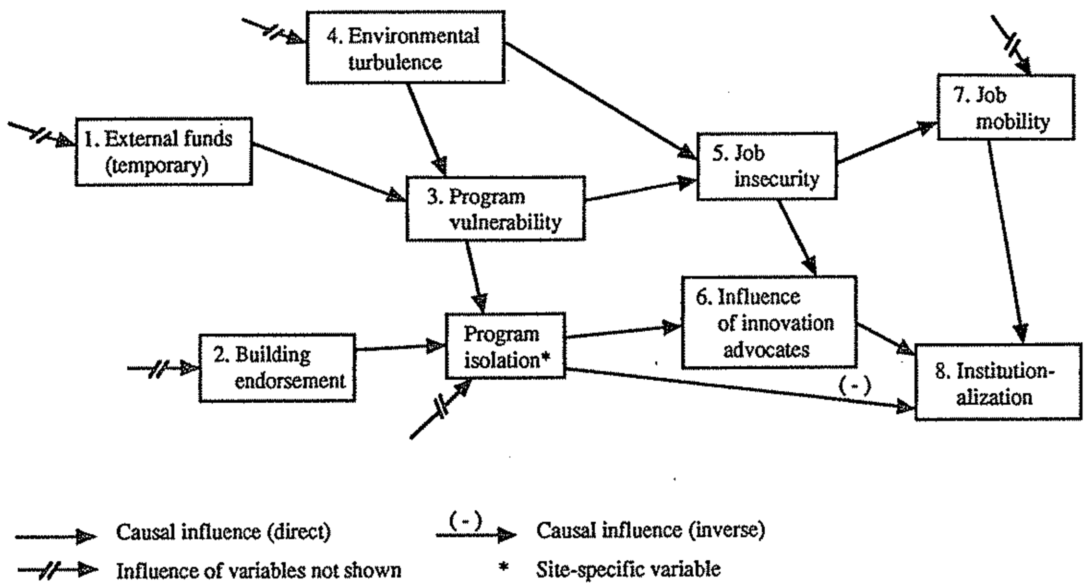
The within-case displays we have discussed up to now can feed causal network analysis. Analyzing *conceptually clustered matrices* (Chapter 5, section E), for example, lets you tease out the relationships within one of the variable families. *Effects matrices* (Chapter 5, section E; Chapter 6, section B) are exercises in identifying cause-effect relations. Both clustering and cause-effect inferencing are at work in *case dynamics matrices* (Chapter 6, section C).

In other words, causal analysis goes incrementally; you are testing individual paths more rigorously and, at the same time, building a cognitively meaningful, integrated cause map.

Finally the *event-state network* (Chapter 5, section C) gets us almost there. Let's illustrate how. Figure 6.4 shows an excerpt from the original event-state network for the case (Perry-Parkdale) in our illustration. Figure 6.5 shows an excerpt from the final causal network, with the variable numbers keyed to the corresponding variables in the event-state network. Usually there will be markedly fewer causal net variables than event-state boxes and bubbles. (This would be the case here, had we shown a fuller excerpt from the event-state network, and had we taken the three case-specific variables out of the causal network.)

In most instances case-specific events and states can be turned into more generic network variables. They give you more powerful explanations, usable in other cases. For instance, almost everybody has money problems ("external funds"), school faculties support or oppose new ventures ("building endorsement"), and the innovation's champions have more or less clout ("influence of innovation advocates").

Figure 6.5
Excerpt From a Causal Network: Perry-Parkdale School



There's much analytic power in working from a specifically temporal, process-oriented display (event-state network) toward a variable-oriented display that still retains temporality (causal network).

At the same time, we have a qualitative procedure that meets some traditional requirements (Judd & Kenney, 1981). We'll be showing the significant effects of the independent variables on the outcome variables, and seeing the clear effect of our predictors on the most important mediating variables, as well as the relationship between outcomes and mediators. We can see how our predictors work, both inferentially and temporally. But the components of a causal map are of a different nature than a disembodied beta coefficient or partial correlation; they are not probabilistic, but specific and determinate, grounded in understanding of events over time in the concrete local context—*and* tied to a good conceptualization of each variable.

Let's take a second look at the fully drawn causal network in Figure 6.1. Note (a) the over-time flow denoted along the bottom and (b) the direction of all the arrows. Causal modeling assumes temporal relationships; that is, early causes make for later effects.

Drawing the network is done most easily in the same way you analyze it: stream by stream. Some streams—unbroken chains of variables—are long: Look at the one that runs directly from box 1 through boxes 8, 9, 10, to 30, even to 32. Others are, conveniently, shorter, like the bottom stream from 5 to 16, 35, and 33. Within an unbroken stream, multiple channels usually lead in different directions or end up at the same place via a different route. Those also should be drawn in.

Streams can be drawn most easily—especially if the event-state network is used as the preceding try—from antecedents forward in time. It is also possible to take a dependent variable and work backward, but then it is important to run it forward again to be sure the links are coherent and empirically justified. You'll inevitably find cross-stream connections, and they can be drawn in as the network evolves.

Often these streams can be given a name—a scenario or theme—and this labeling makes them easier both to analyze and to report. For example, along the bottom stream, variables 5, 17, 35, and 34 have a "cosmopolitanizing" theme: People go outside their environments in connection with a new project, and the experience both stretches them and sets up career shifts. Or again, staying along the bottom of the network, boxes 4, 16, 25, and 33 add up to a "casualty" stream: low money, high turbulence (usually in the form of budget cuts), staff instability, and ultimate staff reassignments.

Streams also can be labeled according to the level of dependent variables, such as high student impact scenarios, low institutionalization scenarios, or high job mobility scenarios. These similarly named scenarios can be very different, even within the same case. For instance, as we just saw, one high job mobility scenario involves a promotion or desirable transfer, whereas another involves an unwanted reassignment or even a layoff.

These themes and scenarios should come as no surprise to the analyst. Most will have been foreshadowed in the pattern codes, memos, and interim summaries, and some will have been teased out of effects matrices and case dynamics tables. In the causal network, the analyst traces the emergence and consequences of a particular theme and orchestrates it with others.

Finally, causal networks can have a predictive function. In drawing a variable stream with two or three substreams leading out to the dependent variables, you are in a position to say something like "Effect Z is likely to obtain when predictors J, B, and L occur in sequence," or "To get high levels of Z, you need high antecedent levels of J, B, and L but not necessarily of V and Q." These statements have modest weight in a single-case analysis, but they perk up noticeably when several within-case nets show the same or a similar pattern. More on this later (see Chapter 8, sections E and F).

The causal network narrative. Drawing conclusions from the network goes along with its production. Like all time-related representations, this one tells a story—really, several stories. You can write out a chronological narrative including each of the major streams, starting at the beginning. Figure 6.6 presents the complete narrative for the causal network shown in Figure 6.1.

Writing the narrative usually does several things. First, it forces you to be less mechanistic and more coherent; turning a network into clear text requires you to be honest and explicit about what you think is causing what. Corrections and revisions are very typical, as we saw in our earlier discussion of display-text interaction (Figure 5.4).

Second, it provides an opportunity for expansion: You can explain why variables are related, why they are rated differently, why some precede others, which ones matter more, and so on.

Third, both products—the network display and the narrative text—are the basic material to be handed to a colleague for reaction and revision. It's nearly impossible to map a causal network "right" the first time. Someone else who knows the database needs to review it and suggest improvements. (Box 6.6 shows how this can be done with informants as well.)

When a causal network and its narrative are revised to coherent form, they can be used to generate more general explanations on a multicase level. Approaches to that task are outlined in Chapter 8, sections E and F.

Variations

We've encountered several interesting experiments with causal networking. For example, Owen and Hartley (1987) built causal nets subdivided by time periods (policy dissemination, program development, program implementation, etc.) and by level of system being analyzed (milieu, institution, program). A look at Box 6.4 shows that this subdivision does not overload the net, and shows clearly which levels of system were in play during each phase.

In another approach, Leithwood, Jantzi, and Dart (1991) described one case in detail, related the findings to survey data, and then derived a smoothed causal network for 12 of the schools under study. The idea was to understand the distinctions between "transformational" and "transactional" leadership as they played out to encourage teacher development activities (Box 6.5).

Note that this network has fewer variables than ours or Owen and Hartley's, because an intermediate analysis weeded out the cases with few similarities. For example, all but two of the variables here were influential in the single case first examined. The frequencies (legend, bottom right) show the number of cases out of 12 in which each variable appeared. Most are high-frequency variables.

Software useful for producing *conceptually coherent,* data-based maps (not just drawing pretty diagrams) includes MetaDesign, NUDIST, MECA, and ATLAS/ti (see Appendix).

Advice

Doing a causal network forces a more inferential level of analysis that pulls together the data into a single summarizing form. You have to look at all of the data and the preceding conclusions, and map them in a coherent way. If you've done it right, you will have respected the complexity of local causality as it has played out over time, and successfully combined "process" and "variable" analyses.

If you're new to qualitative research, start with causal fragments. They are enjoyable, and can be profitably done early and repeatedly, as a basis for discussion with your critical friends. Causal fragments are useful if your case is defined as an individual—and are *very* useful if you are studying several individuals.

The procedures we suggest for drawing networks are still evolving, and (thankfully) are not tightly defined. Two

Figure 6.6
Narrative for Causal Network: Perry-Parkdale CARED Program

The first three antecedent variables (1, 2, and 4) worked out this way. The state mandate (2) for well-planned career education programs, together with assessment of local performance as less than adequate (1), led to a search for new programs (3), which proved to be a good fit (7) with district characteristics, and hence to district endorsement (8), and to adoption of the program (9).

But these were not sufficient causes of adoption. Inadequate local funds (4) to cover existing programs led to a ceaseless search for add-on funds (6) as almost a "way of life" for the district. That search led to getting temporary external funds (14) for a three-year period; they were the other basic cause of adoption (9).

The program, when adopted, proved to have substantial implementation requirements (10), which dictated the need for assistance (18), and also exerted a good deal of pressure for careful selection of high-quality staff (21), and for careful preparation (19) of them to carry out the program. The heavy implementation requirements (10), and to some degree the assistance provided (18), induced a good deal of user commitment (20), and in turn user skill (26), which was also high because of staff quality (21). High user skill, in conjunction with the fact that the program was quite well stabilized (27) by late 1980, brought about a reasonable degree of student impact (30).

The stream of causality refers essentially to internal program dynamics. What was happening at the *district and building* level?

Moving back to program demand characteristics (11), we note that certain aspects of the program (such as its removing students from high school control, and from high school courses and activities) caused poor fit between the program and the sending buildings (12). That poor fit led to lack of endorsement (22) from building principals, counselors, and teachers. Poor endorsement was further weakened by the presence of competition for funds and for students (15), induced by the fact that the external funds (14) were temporary in nature.

Temporary funding, together with the program's demand characteristics (11) (for example, students were visible to employers, had to be responsible, and could easily behave like "assholes") also made for a good deal of program vulnerability (13). As a consequence, the staff tended to operate the program in a rather isolated fashion (23), to buffer it against the consequences of vulnerability when the immediate environmental endorsement (22) was weak. Certain programs demand characteristics (11), such as the intensive time block, reinforced isolation (23) as well.

An added set of causal variables was also in play. The career advancement motivation (5) of key central office staff and principals operated to induce a good deal of turbulence (16) in the district. This turbulence effectively reduced the influence of those who were advocating the innovation (24); for some advocates, influence was weakened by job insecurity (25).

So although the program was transformed and altered (28) to some degree to meet the objections stemming from low building endorsement (22), achieved a modest increase in percentage of use (31) for a while, and was, as we have seen, proving reasonably effective with students (30), these factors were not enough to ensure the program's institutionalization (being built into the system) (32).

Rather, the weak building endorsement (22), the program's isolation (23), the competition for funds and students (15), and the weak exercise of influence by innovation advocates (24) resulted in weak institutionalization (32). So, it seems likely, did the departure of program staff, whose mobility (33) was driven by both career advancement motivation (5) and job insecurity (25). It also seems likely that the very user capacity development (34) induced by experience with skillful use of the program (26), enhanced by the external network of contacts (35) generated through the dissemination effort (17), also contributed to the decision of staff members (and it seems, possibly the director) to move on.

Taken as a whole, these explanations seem baroque and complex. But there is fairly clear evidence that each causal link worked as described. The chart will look less complicated if one notes that the chart contains four basic streams: the *program development* stream across the top, the *building/district* stream in the middle, the *career* stream near the bottom, and the external *dissemination/networking* stream last. In many respects the final outcomes can be seen as stemming from conflicting pressures across the streams.

---

analysts working independently with the same data set, background matrices, and variable list might not draw quite the same network. (The same is true for quantitative researchers creating path-analytic models.)

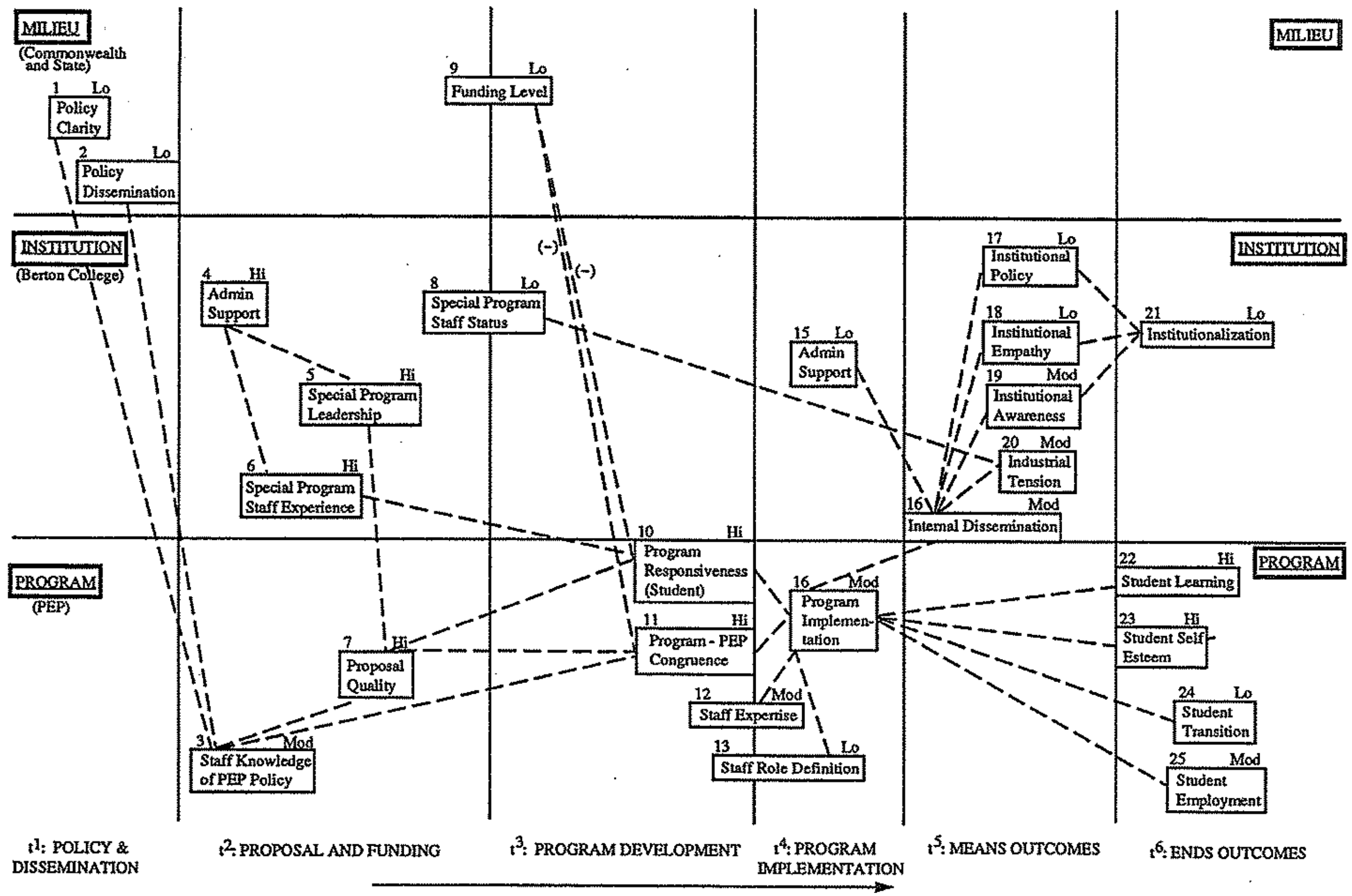
We advise (a) specifying carefully, as you go, which decision rules were employed in the construction of the network, (b) always writing accompanying text to clarify and summarize the flows, and (c) increasing the validity of any given network and narrative by subjecting them to colleague and informant critique.

Keep in mind that you may be building too much order and purpose into loose-ended, inconclusive events. Get a colleague to entertain a more chaotic view.

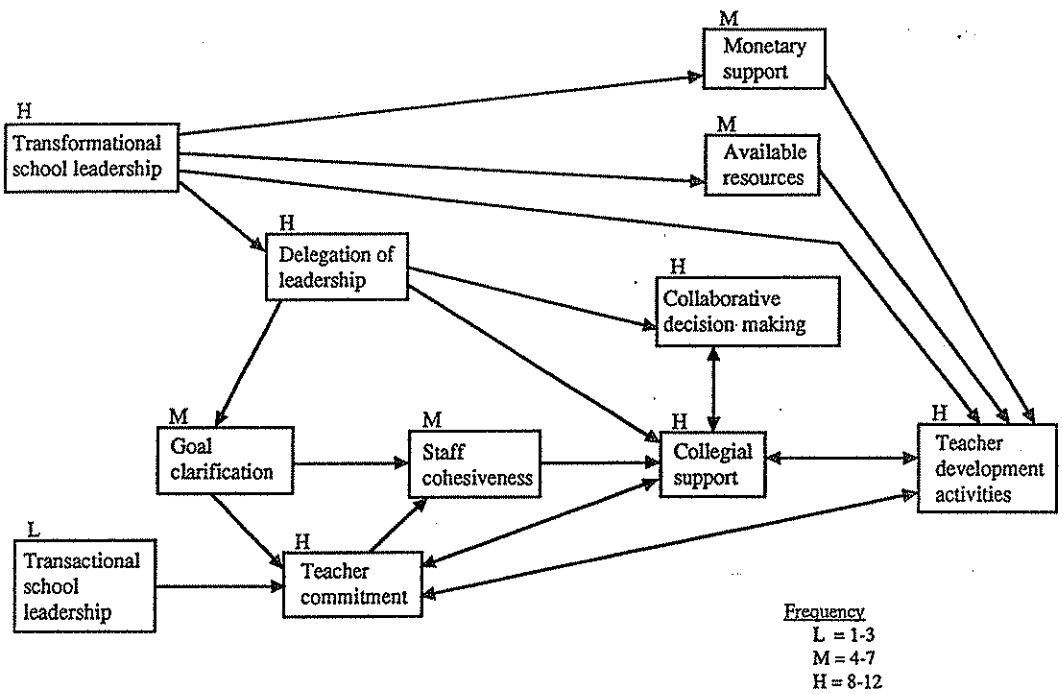
Causal networks aim to connect particular instances to more-general principles. If, when you and a colleague look at a network and its narrative, you cannot supply instances embodying the principle, you'll know that revision is needed.

Remember that a causal network is only as good as the analyses whose shoulders it stands on: the original codes,

Box 6.4
Segmented Causal Network:
The Influence of PEP Program at Berton College (Owen & Hartley, 1986)



Box 6.5
Smoothed Causal Network:
The Variables and Relationships Linking Administrators' Actions and Teacher Development Activities
(Leithwood, Jantzi, & Dart, 1991)



conceptual clusters, and matrices. That caveat sends us back to the quality and coverage of the data, and the trustworthiness of data collection methods. As in statistical analysis, good models are only as good as the measures and simpler analyses they derive from. Dross in, dross out.

Time Required

Causal network analysis is labor-intensive. Drafting a network of the size illustrated takes the good part of a day. Writing the first cut of the narrative may run 2 or 3 hours. Iterating and reconfiguring the network and the text takes another half-day or so. Remember, too, that this work is occurring late in the project: It presumes that (a) preceding analyses have been carried out reasonably cleanly and (b) the results of those analyses are easily at hand.

Verifying Causal Networks

The causal network is the analyst's most ambitious attempt at an integrated understanding of a case. The most important factors, their interactions, and their links to the key outcomes are all plotted on the network. As careful, plausible, and exhaustive as the network may appear to be, parts or even most of it could be quite wrong. One way of avoiding that outcome is to give back the network to people in the case, and ask for their corrective responses (tactic: **getting feedback from informants**).

The quality of the feedback depends on how it is solicited. The critic has to understand exactly what is being asked, and should work directly with the materials, rather than simply comment on them globally.

Box 6.6 shows a feedback form we've used for informants; it assesses the accuracy, from their perspective, of the

Box 6.6
Response Form for Causal Network Verification

School District ______________________ Your name ______________________

Your reactions to the materials you've read can be of much help in validating them. You can provide us with more confidence in our conclusions, and can show us where our explanations are partial or mistaken, and need to be revised for more accuracy.

Remember: we are asking you to think back to the way things were last June, and what led up to that state of affairs.

1. Looking at the project overview (Attachment A):

   a. What errors of fact do you see, as of last June?

   b. What differences in interpretation do you have?

   (Feel free to jot notes directly on the overview itself, if that will be easy.)

2. Now, looking at the causal network (Attachment B):

   a. Generally speaking, how accurate do you consider the network to be, as of last June? Please say a little about why you think so.

   b. Are there any important elements missing? List them here, and/or draw them in on the chart. Give each one a new number, larger than our list of numbers.

   c. Looking at the specific boxes we have used, are any of them unimportant, trivial, of little effect? Cross them off the network, list their numbers here, and explain briefly why they are of little value, from your point of view.

   d. Still looking at specific boxes, have we made mistakes in the ratings we have given (high, low)? If so, please write your corrections in on the network, note numbers here and give reasons for your revised ratings.

   e. Looking at the arrows between boxes, do you think they are accurate? Are they faithful to the events at your school and district? If not, please cross off arrows, or draw in new arrows. Write the numbers of the boxes involved here.
   For example:
   3 $\not\rightarrow$ 8 shows that you crossed off the arrow between boxes 3 and 8.
   3 $\longrightarrow$ 11 shows that you wrote in a new arrow between boxes 3 and 11.

   Please explain your revisions in arrows briefly.

   f. Is our discussion of outcomes at the end of the causal network description accurate from your point of view? What revisions would you suggest to make it more accurate?

   g. If you would like to draw a revised causal network that would show your explanations of why things happened as they did, please feel free to do so here. (Or use the back of the page if you need more space.) Yours might have fewer, more, or different boxes than ours. Events in any change process are complicated, and different people see them differently. So if you would like to play the network game, enjoy!

3. Any other concluding comments or suggestions:

MANY THANKS FOR YOUR HELP. Please return this form, and all the materials to us in the stamped envelope. (Feel free to make a copy for your files if you wish.)

causal network for the case. We have usually chosen informants (more than one per case, in different roles) who have proven reliable in the past, and who are in a position to reflect on the case's "big picture."

The informant gets a one-page case overview, the network, and the network narrative and is asked to work directly with these materials in giving responses. For economy of space, we have collapsed the 4-page response form into a single page (Box 6.6).

We've found that people enjoy doing this kind of work. It lets them see rapidly how the researcher put the pieces together, and they can run that view against their own. When their perspective is different, they can quite easily redraw parts of the network or redraft parts of the text.

However, causal networks are somewhat closed-ended. They have a peculiar logic—both in the narrative and in the diagram—that people sometimes have trouble breaking out of. Looking at a network, they cannot "make it their own" enough to seriously revise it. And for some people, diagrams and narratives of this sort can be overpowering or intimidating. Face-to-face debriefings, when possible, improve the quality of information exchanged from both sides, and can be more rewarding.

We've found that feeding back networks/narratives is rarely threatening or hurtful, because they are expressed more abstractly than usual case material—which can present ethical dilemmas for researchers and informants (see Chapter 11).

This is professional work, and we believe that people should be offered some recompense. We provided a modest reader's fee ($25-$50). The work time is not over an hour or two.

E. Making and Testing Predictions

Analysis Problem

Verifying a causal network and its accompanying text by feeding them back to informants is a good way to strengthen the validity of your findings. But it's possible that all actors, including the researcher, can succumb to a common perception—one that another researcher, less immersed in the case, might not share. Social life, as the ethnomethodologists from Garfinkel (1967) onward have taught us, is an issue of taken-for-granted and agreed-upon conventions that allow people to make sense of what they are doing individually and collectively. In many ways, social life is only possible through "groupthink." History, Napoleon is supposed to have said, is the version of past events that people have decided to agree upon. And field researchers, who depend heavily on local informants' perceptions in building a descriptive and explanatory map of the site, are likely to get sucked into collective perceptions that lend coherence to such a map.

There are many ways to avoid obvious biases in the data being collected (see the discussion in Chapter 10, section B). But there are fewer ways of getting around collective, consensual ones. A competent field researcher should be able to stay partially detached from site dynamics by virtue of outsider status and an analytical focus. But "going native" is not unknown.

The opposite problem with causal network feedback is that informants may not agree—either with parts of the network or with one another. People witnessing the same events may make sense of them in quite different ways and give discrepant accounts. In three studies where we fed back causal networks, we found only modest discrepancies among informants, or between them and us. But there was some, and we were left sitting on it, trying to think of a way to get resolution.

We did invent one good approach: testing the validity of your findings by predicting what will happen in your case after 6 months or a year have elapsed. Presumably, if the analysis is expressed clearly, it has some implications or consequences (using the **if-then** tactic). These can be spelled out and operationalized. The real-world results can be checked out later in a new field contact. (The basic tactic is **getting feedback from informants.**) The result is the qualitative researcher's rich, full-fledged version of a predictive validity coefficient that statisticians might use—for example, to see whether school grades predict later job success.[10]

Brief Description

Predictions are inferences that a researcher makes about the probable evolution of case events or outcomes over the following months or years. The predictions are drafted at the time of analysis, held, and then submitted to case informants later (we used 6-12 months, but timing obviously depends on the nature of the cases and predictions involved).

The informants respond to (a) the accuracy of the predictions and (b) the accuracy of the reasons given to justify them.

Illustration

A prediction exercise must be built into your study's budget and time line. In our school improvement study, we had been looking retrospectively at 3 or 4 years of change effort in schools. It made sense to plan for a follow-up a year after field work had ended, to see whether our analyses had predictive power.

Generating the prediction. To see how this process works, take a look back at Figure 6.1 and find variable 32 on the network, "institutionalization." It has a low rating, indicating that the innovation at this site has not become a durable, built-in, routine part of ongoing operations. The rating on this variable is a function of the several paths leading to it, both from other outcome variables (e.g., 30, 31) and from variables farther back on the four causal streams. Schematically, the causal network says that "institutionalization" is as low as the combined function of the variables leading up to it; they are the likely "causes" of the institutionalization "effect."

In testing the plausibility of the causal analysis, you then ask predictive validity questions: (a) What is likely to happen to the innovation between now and next year at this time? (b) On which factors are we basing that estimate?

Looking over the causal network and accompanying text, and reviewing an institutionalization checklist matrix (see Huberman & Miles, 1984, pp. 207-208) for the site and its analytic text, the analyst decided the program was in trouble. The checklist showed that the innovation was a peripheral one, that its benefits were less than its demands on staff, that its support was soft, that competing practices had not been eliminated, and that its hard money funding was uncertain.

There were some promising signs—the program was operating on a regular, daily basis, and it was getting positive results with turned-off students—but they were buried in a pile of pessimistic indicators. The analyst predicted that the project would be phased out gradually in the course of the following year.

Making this kind of prediction is not easy. Somehow researchers seem to escape having to make decisions about their data; findings "suggest," "indicate," "imply," or, to take the classic, cowering pose, they "make it not unreasonable to assume" something. By contrast, generating a single, precise prediction is a bottom-line operation. The prediction is either right or wrong, usually with only a small intermediate, right/wrong gray area.[11] And if the prediction turns out wrong, the chance is good that some of the data and analyses supporting it are less than solid.

Justifying the prediction. How do you puzzle out the "whys" of a prediction? Back to our illustration. In part, as we just noted, the institutionalization checklist gave the site a poor report card on the ingredients for locking in the practice organizationally. The streams on the causal network (Figure 6.1) leading to the institutionalization variable also look foreboding.

Specifically the project makes for a poor building-level fit (21) and is not supported by teachers or administrators (22). It has not worked through to a better fit by changing the organization (29). The project is vulnerable because of its demands, its isolation, and its dependence on soft money (11, 13, 14, 23). It competes for funds and students (15); its support in the central administrative office is uncertain (24) for a variety of reasons. Key project staff are about to leave (33). The number of students projected for the following year is lower than for the previous year (31). These are the reasons for the dire prediction.

Figure 6.7
Factors Supporting "Institutionalization" Prediction

A.	Poor district record of institutionalizing soft-money programs.
B.	Unlikelihood of strong support from new principals: some counselors indifferent: some elective-subjects teachers are hostile.
C.	Lukewarm support from central office.
D.	Financial squeeze in district: declining enrollments, staff cutbacks, lowered external funds. Program vulnerable because of high unit costs.
E.	"Tightening up" mandate to principals: Program may be seen as a "frill," with lax control over students.

Finally, you consider contextual factors and look over trends in the case that, if continued, could have an effect on institutionalization. For example, this district had a poor track record on institutionalizing soft-money programs, and the district was currently in a "tightening-up" phase while the project was a "loosening-up" type of operation.

You weight all of these indicators by their degree of influence; this is also a way of getting the number down to four or five, which is all that an informant can be asked to respond to. Figure 6.7 shows the five that were settled on.

Looking at contrary data. If you look only for the factors supporting a prediction, you'll probably make a poor one. Making a good guess is a function of weighting the odds for one outcome as against another—computing subjective probabilities on the fly. It's important to consider the factors that could work against the prediction (tactic: **looking for negative evidence**).

In the same example were indications that the project was well entrenched in routine operations, that it appeared to pay off for a normally hard-to-reach pupil public (in Figure 6.1, see the arrow from 30 to 32), and that it had some central office support (24). Pooling these indicators with a contextual and trend analysis of the case as a whole,

Figure 6.8
Factors Working Against
"Institutionalization" Prediction

A.	Possibility of dissemination funding, which would call for continued implementation.
B.	Program meets needs (oasis, dumping ground, safety valve) for alienated, undecided students who are not college bound.
C.	Program well installed (recruitment, job sites, monitoring and instruction), well stabilized.
D.	Program might be folded in to new vocational education center and be funded through it.

the analyst came up with the four factors shown in Figure 6.8 that could "spoil" the pessimistic prediction.

Selecting informants. The prediction and its justification are given to key informants at the site, with a request for data on the current, actual situation about which the prediction was made. Normally you choose informants who (a) have proven reliable, (b) are in a position to know the current situation, and (c) occupy differing roles and have somewhat differing perspectives.

Before the predictions are sent out, an ethical review is important: Can this information be damaging to these individuals or to their colleagues in any way? General predictions usually have little harmful potential, in contrast with person-specific information. But the question needs careful attention.

Analyzing feedback. How good is the prediction? How good are the supporting and spoiling factors? Figure 6.9 shows a response form from one of the three case informants. Factors A through E on the top part of the chart are the "supporting factors" we listed earlier; Factors A through D on the bottom part are the "spoilers" we just looked at. This informant agrees that most of the supporting factors were pertinent in determining the way things turned out 1 year later. The other informants made similar, but not identical, judgments on all of the predictors (e.g., two thought central office support was stronger and the "tightening up" was a more pertinent predictor).

As for the factors working against the prediction, the two high "pertinence" estimates suggest that the analyst had underrated the possibility of more external funding and the importance of the "dumping-ground" function of the project. (The two other informants agreed, but also felt that C and D were influential factors that bettered the chances of institutionalization. So we are still caught in the dilemma of differing perceptions among informants—a dilemma that sometimes can be resolved by follow-up phone interviewing.)

How about the prediction itself? This had been assessed on another sheet. You can see already, however, that the prediction is probably in trouble because some of the "spoiler" factors are pertinent and important. As it turned out, our prediction was too pessimistic. The project was reduced in scale and its status was transitional—but it was still around, though barely. Central office support apparently was stronger than estimated, owing to the successful and still necessary dumping-ground function, and the possibility of receiving new external funds to disseminate the project to other districts. The district office had decided to fund the project for one more year, but had cut project staff and students by half. We counted our accuracy as only moderate.

As the causal network in Figure 6.1 shows, this was a difficult prediction to make, in light of the number of streams and the fact that not all led to the same conclusion. The predictions for the other outcome variables shown at the outcome end of the network were easier to make and were on the nose. We also did better at predicting institutionalization at the other field sites; one other was moderately accurate, and the rest were high/moderate to highly accurate. Furthermore, the explanations given by us and the site people overlapped reasonably well (58% of assigned causes were in full agreement).

Note, too, that although the prediction was overstated, the causal network and the institutionalization checklist were still valid. Both predicted low institutionalization and were right; virtually all of the reasons given for a phasing out were pertinent and influential. Also the causal net had included countervailing evidence, such as the high student impact and the moderate to low central office support. But the influence weightings of these indicators were off; student impact was more important, support was higher and counted more, project isolation mattered less, and so on, than the analyst had thought.

Mechanics of feeding back predictions. Let's go through the process of submitting the predictions to case informants.

1. You make the predictions at the time of the analysis, by taking the main outcome variables and looking at the data bearing on their causes. These data come primarily from the causal network streams, from matrices or other displays in the section of the case report dealing with the outcome variable, and from contextual and historical information showing how the case evolved.

Figure 6.9
Filled-Out Response Form From Case Informant for "Institutionalization" Prediction

Pertinence Ratings:
1=not pertinent
2=pertinent, but not important in causing current situation
3=important factor in causing current situation

Our Prediction

The program will undergo a transitional year in 1980-81, during which time it will be phased out.

	FACTORS WE THOUGHT WOULD SUPPORT OUR PREDICTION	PERTINENCE (WRITE IN APPROPRIATE NUMBER BELOW)	BRIEF EXPLANATION (WHY YOU SAY THIS)
A.	Poor district record of institutionalizing soft-money programs.	2	It could have been different with a more aggressive leader sold on the idea.
B.	Unlikelihood of strong support from new principals; some counselors indifferent; some elective-subjects teachers hostile.	2	One new principal was supportive; the other not. Counselors reflected the principals' attitudes.
C.	Lukewarm support from central office.	3	This true. In four years neither the Supt. nor the Deputy have ever visited the room. The new Assistant spent 30 min. there last year.
D.	Financial squeeze in district; declining enrollments, staff cutbacks, lowered external funds. Program vulnerable because of high unit costs.	3	
E.	"Tightening up" mandate to principals: Program may be seen as a "frill," with lax control over students.	1	I wasn't aware of the "tightening up" toward the program. Student behavior - yes.

	FACTORS WE THOUGHT COULD WORK AGAINST OUR PREDICTION	PERTINENCE (AS THINGS TURNED OUT)	BRIEF EXPLANATION
A.	Possibility of dissemination funding, which would call for continued implementation.	3	As long as State funds are available, a demonstration program will be available.
B.	Program meets needs (oasis, dumping ground, safety valve) for alienated, undecided students who are not college bound.	3	This year more students were less motivated than ever before.
C.	Program well installed (recruitment, job sites, monitoring and instruction), well stabilized.	1	I don't think this matters at all to the decision-makers.
D.	Program might be folded in to new vocational education center, and funded through it.	1	I have not heard this said regarding the voc. ed center, though I've thought it has possibilities under State guidelines which allow for student community placement under contracts with employers.

Figure 6.10
Prediction Feedback Form, Part I

School District Dun Hollow Name ________

Our Predictions (Made with June 1980 data) Date ________

I. Institutionalization: that is, the degree to which the Eskimo units have become a durable, built-in or routine part of the ongoing operations in the district.

Our Prediction

The curriculum will not be incorporated into the district after the field test at Carr and Tortoise Area Schools.

Your Description of the Actual Situation Since Then

Actual Situation

There are probably many factors leading to the actual situation. The most important ones are:

A. ________

B. ________

C. ________

D. ________

E. ________

F. ________

Please rank these factors you have listed in their order of importance by putting next to the letter A - F nos. 1 for the most important, 2 for the next most important, etc.

Figure 6.11
Prediction Feedback Form, Part II

I. Institutionalization: that is, the degree to which the Eskimo units have become a durable, built-in or routine part of the ongoing operations in the district.

Pertinence Ratings:
1=not pertinent
2=pertinent, but not important in causing current situation
3=important factor in causing current situation

Our Prediction

The curriculum will not be incorporated into the district after the field test at Carr and Tortoise Area Schools.

FACTORS WE THOUGHT WOULD SUPPORT OUR PREDICTION	PERTINENCE (WRITE IN APPROPRIATE NUMBER BELOW)	BRIEF EXPLANATION (WHY YOU SAY THIS)
A. Consensus at the site is that field test was unsuccessful.		
B. Eskimo units were not written into social studies K-3 curriculum.		
C. The units have little support from building principals.		
D. No innovation "champion" at the site; the program developer is not pushing forcefully for adoption.		
FACTORS WE THOUGHT COULD WORK AGAINST OUR PREDICTION	**PERTINENCE (AS THINGS TURNED OUT)**	**BRIEF EXPLANATION**
A. The superintendent has come out vocally in favor of adoption.		
B. Eskimo center remains within the district; the need for a program is likely to persist.		
C. Superintendent wants to strengthen working relations with Regional Unit developing Eskimo program.		

2. For each prediction, you work out a set of factors supporting the estimate, and another set of factors working against it.

3. This material, with the predictors, is held. It will go out for feedback after a predefined period of time (e.g., 6 months, a year) to at least two case informants. After the informants are chosen, it's important to review the predictions carefully to be sure they bear no potential harm to those informants, or to others in the case (see also Chapter 11 on ethical issues).

4. An informant gets two envelopes. In the first is a form that defines the variable, gives the prediction, and asks the informant to judge the accuracy of the prediction by specifying the actual situation. The informant then explains his or her depiction by listing the most important contributing factors. Figure 6.10 shows a typical feedback form for the first envelope.

5. The informant then opens a second envelope, and finds the original prediction plus a pertinence table listing the factors that would support the prediction and factors that might invalidate it. The informant is asked to rate the pertinence of each factor in accounting for the actual situation, and to explain the reasoning behind that rating. The response form looks like Figure 6.11.

The logic behind the two-envelope procedure is to widen the database (seeing whether more contributing factors are at work than you had assumed) and to subject your predictors to an empirical test by comparing them with the informant's. Also, the informant is not "contaminated" by an externally generated framework when thinking about the prediction in the first envelope.

6. The informant sends the two sheets, plus any additional commentary, back to you for assessment of predictive accuracy, as outlined above.

Variations

Generating the predictions can be done in two ways—by the primary analyst, or by another analyst who reviews the case report. We encourage the latter procedure if possible.

If a second reader cannot generate precise and well-grounded predictions from the matrices and causal network of a case report, something is wrong with the report. Probably it is not explicit enough in delimiting causal streams, or in gauging the relative importance of the streams—or both.

We've outlined fairly detailed procedures for making reasons-based predictions, and for getting careful feedback from respondents. In some cases life can be much simpler. Stringfield and Teddlie (1991), for example, looked at quantitative data from 16 schools and classified them as low or high outliers on achievement measures. Their predictive accuracy was tested by subsequent qualitatively oriented observers who were "blind" to the predictions (as were the school personnel involved). The observers not only identified all highs and lows but also correctly predicted later improved achievement in four of the low schools.

Advice

If you want to try testing predictions, time and money must be set aside in advance.

Generating predictions is a good experience: You have to take a hard look at the data and analyses, and address their portent explicitly. It's a way, as Campbell (1979) points out, of clarifying a general theory about your case by generating multiple predictions about different aspects of the case. It's also a powerful validation device, even when informants' responses are equivocal. Predictions also bring in additional, well-targeted data that can be appended to a case report. Finally, as with the causal network feedback exercise, informants feel rewarded. Most enjoy doing the critique and believe they have learned from it.

The exercise, however, has some limits. Unless you go back to the field, you are relying solely on self-report data. And predictions can be wrong for reasons that have little to do with their internal validity. A few unanticipated events, such as the sudden death of a key actor or a budget windfall, can wreak havoc on the best-laid projections of an analyst. Though you may try to imagine the intrusion of unanticipated events as part of the predictive work, surprises, by definition, can never be fully anticipated.

Limit the number of dependent variables or outcomes for which you make predictions (say, to four or five) and keep the explanatory factors for each to the same number or less. Define each of the outcome variables specifically, or else informants will give confusing responses. For multiple variables, all can be defined in a single list.

As with causal network verification, you should not ethically expect site personnel to carry out what amounts to professional work without some form of recompense. We offered a reader's fee of $25-$50, given five predictions in the two-envelope mode.

Remember that a key informant's reflecting this intensely on the current situation and its causes is likely to have reactive effects: He or she may come to have new, perhaps painful, views of the situation, or take some new action as a consequence. You can never wholly anticipate such effects, but the try should be made. If you have a continuing relationship with the field site, such effects can be assessed through regular fieldwork. You also can ask about such effects on the feedback form itself (though that request, in turn, is likely to encourage even more reactivity) or through a follow-up phone call.

Time Required

Predictions take a good amount of time. The appropriate sections (e.g., the causal network) of the case report must be read closely, and you need to review sections with contextual and historical material. If several outcomes are being predicted (e.g., we were predicting not only institutionalization but also extent of use of the innovation, transformation of the innovation, student impact, and user capacity change), this reading can take 3 or 4 hours, depending on the detailed data being examined. Then the rough-cut predictions have to be made, the supporting and spoiling factors listed and weighted, and the final predictions made. This process usually adds another 4 or 5 hours. Setting up the charts, assuming the analyst does this alone, is a 2-hour job.

So for one case, assuming it is a complex one like ours, a prediction exercise usually will take 1 or 2 days, depending on the number of outcomes involved. An external analyst to whom the material is foreign will take 50-100% longer. Time must also be taken to draft a coherent cover letter explaining the exercise, to decide which informants will receive the mailing, and to preassess the potential impact on informants.

Is prediction making worth the time? Yes, certainly, if you want to know how good your analysis really is, or if much is riding on it. No, if your main thrust in a study is descriptive rather than predictive or explanatory.

Summary Comments

We began this chapter by reflecting on explanations and causality, proposing that answering good "why" and "how" questions requires us to go beyond sheer association to seeing the actual mechanisms of influence in a bounded local setting, which are always multifold, operating over time. We emphasized the importance of taking both a "variable-oriented," conceptual approach and a "process-oriented," storylike approach.

Displays such as the *explanatory effects* matrix and *case dynamics* matrix are helpful in beginning to sort out what leads to what over time, but *causal networks* are better at dealing with the complexity of causal processes. They can be verified via informant feedback, but the strongest test of their accuracy is whether *predictions* based on them are confirmed or not.

So far we have focused on displays for the analysis of single cases. In the next chapter we turn to multiple-case analysis and their use for descriptive purposes.

Notes

1. For a "causal tree" approach to diagramming "why" questions, see Werner and Schoepfle (1987b, pp. 188-192).

2. In this discussion we don't pretend to cover all the issues involved in explanation of social phenomena. For more, we encourage the reader to explore the references suggested, as well as Antaki (1988), Garfinkel (1981), Van Parijs (1981), and Turner (1980).

3. Indeterminacy, in fact, has powerful attractions. As Holland (1992) pointed out when the tenor Pavarotti was discovered to have been lip-synching, "People don't want to be two-timed. Everything we do in life is geared to cause and effect, and when Mr. Pavarotti opens his mouth, we insist on *not* knowing what will come out" (p. C13).

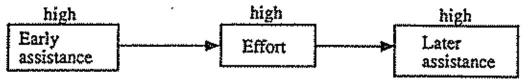
4. Here, too, we cannot be encyclopedic about causation. Useful general discussions, with helpful additional references, appear in House (1991), Abbott (1992a), Davis (1985), Putnam (1987), Lincoln and Guba (1985), Harré and Madden (1975), and Scriven (1966).

5. The existence of a plausible "mechanism" is critical if we are to go beyond sheer association. Casti (1992) gives a nice example: the high correlation between the number of storks nesting in English villages and the local birth rate. Hmmm. . . . It turns out that the mechanism is not what you would guess. Villages with high birth rates have more newly married couples and more new homes. Storks like to roost by chimneys not previously used by other storks.

6. One problem with quantitative "smoothing" is that it occurs near the start of analysis, bypassing and obscuring the understanding of dynamics in each case. If we understand local causality in each case first, it is possible to develop smoothed models late in the game that are more powerful, essentially generic. We give examples in Box 6.5 and Figure 7.3.

We concur with Geertz's (1983, p. 233) emphasis on "local knowledge" as the first order of business, *and* with his view that we need, in the end, something more than local knowledge:

Figure 6.12
Nonrecursive Relationship



> We need a way of turning its varieties into commentaries one upon another, the one lighting what the other darkens . . . bringing incommensurable perspectives . . . into conceptual proximity such that . . . they seem somehow less enigmatical than they do when they are looked at apart.

7. This broad distinction between variable-oriented and process-oriented analysis is not wholly new. Kaplan (1964) spoke of the "pattern" model for explanation (it fits into a coherent or known pattern) and the "deductive" model (it can be deduced logically from other known truths).

8. Miller (1982) points out that familiar probabilistic techniques such as confidence intervals can be employed in deciding when enough cases have been reviewed.

9. It often seems reasonable, as Weick (1979) notes, to add reciprocal or back-effect arrows. For example, in Figure 6.2, increased effort may, in turn, lead to more assistance because it encourages the assister. In the style of causal mapping we're describing, we do *not* advise such causal flows, which technically are labeled "nonrecursive"—a confusing term if there ever was one. They can be modeled by computer (see Forrester [1973, 1990] for rationale; Gaynor & Clauset [1985] show a detailed example built around the findings from our school improvement study).

But they rapidly bewilder the mind ("after all, everything affects everything else"). For graphic and intellectual clarity, we have found it better to stick with straight left-to-right flows that keep a temporal order. If you are sure you have a really clear looping relationship, it can be handled as shown in Figure 6.12.

10. Here again, Kaplan (1964) is illuminating. He points out that it's possible to make predictions without full understanding (explanation) and that it's possible to have rather clear explanations that do not necessarily generate a prediction. But, as he notes, "if we can predict successfully on the basis of a certain explanation we have good reason, and perhaps the best sort of reason, for accepting the explanation" (p. 350).

11. Note that the weather forecaster's ploy (There is a 70% chance of rain this afternoon) does not let the predictor off the hook either. Either it rains or it doesn't.

7

Cross-Case Displays: Exploring and Describing

Up to now our discussion has focused mainly on ways to describe, understand, and explain what has happened in a single, bounded context—the "case" or site. That is the task of the traditional ethnographic researcher, whose effort is to emerge with a well-grounded sense of local reality, whether the focus is on an individual, a family, a tribe, a formal organization, a community, or even a culture as a whole.

Yet for years many researchers have leaned toward multiple individual cases (e.g., teachers, alcoholics, middle managers, battered women, taxi drivers). And during the last few decades, there's been a substantial growth in studies of complex settings (schools, special programs, businesses) using multicase designs, often with multiple methods (Herriott & Firestone, 1983; Louis, 1982; Schofield, 1990). One aim of studying multiple cases is to increase generalizability, reassuring yourself that the events and processes in one well-described setting are not wholly idiosyncratic. At a deeper level, the aim is to see processes and outcomes across many cases, to understand how they are qualified by local conditions, and thus to develop more sophisticated descriptions and more powerful explanations.

But developing a good cross-case analysis/synthesis is not a simple matter. Alcoholic A turns out to be quite different in personality dynamics from Alcoholic B, and can't be facilely compared, as Denzin (1989b) eloquently shows us. Or suppose, for example, you have developed a good causal network explaining processes in a particular case. If you have a dozen such cases, just adding up separate variables, as in a quantitative survey approach, will destroy the local web of causality, and result only in a "smoothed-down" set of generalizations that may not apply to any specific case in the set—let alone others. Each case must be understood in its own terms, yet we hunger for the understanding that comparative analysis can bring.

In this chapter we focus on descriptive approaches to cross-case analysis, with explanatory methods described in Chapter 8. The questions of *what* and *how* usually must be asked and answered somewhat differently from the question of *why*.

A. Cross-Case Analysis: Overview

First, let's go deeper into why cross-case analysis is important, and what it actually involves.

Why Do Cross-Case Analysis?

One reason is to enhance *generalizability*. Although it's argued that this goal is inappropriate for qualitative studies (Denzin, 1983; Guba & Lincoln, 1981), the question does not go away. We would like to know something about the relevance or applicability of our findings to other similar settings, to transcend "radical particularism" (Firestone & Herriott, 1983). Just adding cases is a brute-force approach that will not help. But multiple cases, adequately sampled (are they typical? are they diverse? are they unusually effective or ineffective?) and analyzed carefully, as we'll see, can help us answer the reasonable question, Do these findings make sense beyond this specific case? (We have more to say on generalizability in Chapter 10, section C.)[1]

A second, more fundamental reason for cross-case analysis is to deepen *understanding and explanation*. Glaser and Strauss (1967, 1970), writing when few field researchers were studying more than one setting at a time, framed the issue well. They argued for using "multiple comparison groups" to find out "under what sets of structural conditions [the] hypotheses are minimized and maximized." The researcher can "calculate where a given order of events or incidents is most likely to occur or not occur." Multiple cases also help the researcher find negative cases to strengthen a theory, built through examination of similarities and differences across cases. That process, they suggested, is much quicker and easier with multiple cases than with a single case. Multiple cases not only pin down the specific conditions under which a finding will occur but also help us form the more general categories of how those conditions may be related.

As Silverstein (1988) puts it, we are faced with the tension between the particular and the universal: reconciling an individual case's uniqueness with the need for more general understanding of generic processes that occur across cases. That uniqueness, he suggests, resides in the individual's developmental history over time—but "encapsulated within the general principles that influence its development." Noblit and Hare (1983), in considering cross-case work, suggest that it "must have a theory of social explanation that both preserves uniqueness and entails comparison." In their more recent work (1988), they also emphasize that aggregating or averaging results across cases is bound to lead to misinterpretation and superficiality. So the intellectual needs here are clear, but specific approaches are not so obvious.

A Key Distinction: Variables Versus Cases

Our search for helpful cross-case analysis methods will be aided if we clarify two basically different approaches to inquiry (Ragin, 1987), to which we alluded earlier.

Figure 7.1
Hypothetical Quantitative Data Set: Factors Influencing University Attendance

FACTORS INFLUENCING UNIVERSITY ATTENDANCE

ID #	Gender	SES	Parent expectations	School performance	Peer support	Decision to attend university
001	1	2	4	6.3	12	1
002	1	4	6	5.2	8	0
003	2	4	5	8.9	7	0
004	1	3	8	6.5	9	1
005	2	3	7	7.0	10	1
...	:	:	:	:	:	:
282	1	1	5	5.6	8	0
283	1	2	3	8.2	4	0

For simplicity, let's consider Figure 7.1, a hypothetical quantitative data set. For each of 283 "cases," we have measures of six variables: gender, SES, and so on, that may be related to university attendance.

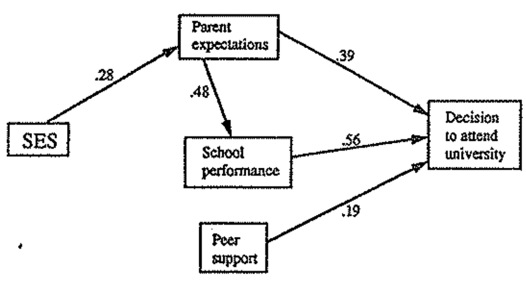
In a *variable-oriented* analysis, we'd read the table vertically, focusing on, say, parent expectations and computing the correlation between that variable and the decision to attend university. Or we might do regressions or a path analysis pulling all of the variables together and assessing their relative weight, ending with something like Figure 7.2. We'd see that deciding to attend university is mainly influenced by school performance, which is boosted by parent expectations (pushed up by SES); they also have a direct effect on the university decision. Peer support has a small effect as well.

But note that we have a smoothed picture. We've focused on *variables* and their relationship, and have no way of knowing whether the picture fits any particular case.

In a *case-oriented* analysis, we'd read across a row, looking at the measures for a particular case, #005, who is female, middle class, has parents with high expectations, and so on. These are "thin" quantitative measures, though. To do a real case analysis, we need to look at a full story of case #005: Nynke van der Molen, whose mother trained as a social worker but is bitter about the fact that she never worked outside the home, and whose father wants Nynke to work in the family florist shop. And we need to invoke chronology: Nynke's closest friend, Leonie, decided in the first semester of 1989-1990 to go to university, before Nynke started work in a stable. That occurred about when her mother showed her a scrapbook from social work school—a moving eye-opener for Nynke—and preceded Nynke's decision in the second term to enroll in veterinary studies.

These and other data in the story can be displayed in matrix form, of course, as in Table 7.1. They would let us begin to trace the flow and configuration of events to see

Figure 7.2
Hypothetical Path Analysis:
Factors Influencing University Attendance



how Nynke came to make her decision. And by looking at several of these cases, we might begin to see recurrent patterns, or perhaps families of patterns that occur as a young person is making an important life decision.

As Ragin (1987) emphasizes, a case-oriented approach considers the case as a whole entity, looking at configurations, associations, causes, and effects *within* the case—and only then turns to comparative analysis of a (usually limited) number of cases. We would look for underlying similarities and constant associations (e.g., in the cases of Nynke, Jaap, Leonie, and Arjen, would we see that there are critical parental interventions?), compare cases with different outcomes (those who do and don't attend university), and begin to form more general explanations.

The variable-oriented approach is conceptual and theory-centered from the start, casting a wide net (cf. Runkel, 1990)[2] over a (usually large) number of cases. The "building blocks" are variables and their intercorrelations, rather than cases. So the details of any specific case recede behind the broad patterns found across a wide variety of cases, and little explicit case-to-case comparison is done.

As Ragin notes, each approach has pluses and minuses. Variable-oriented analysis is good for finding probabilistic relationships among variables in a large population, but it is poor at handling the real complexities of causation or dealing with multiple subsamples; its findings are often very general, even "vacuous." Case-oriented analysis is good at finding specific, concrete, historically-grounded patterns common to small sets of cases, but its findings often remain particularistic, "while pretending to great generality."

The implication is not that one or the other approach is "better" for qualitative data analysis. Rather, the issue is one of making deliberate choices, alternating and/or combining/integrating methods as a study proceeds.

Strategies for Cross-Case Analysis

How do qualitative researchers proceed when it comes to analyzing data from multiple cases? Here we outline several approaches; consult the fuller descriptions later in the chapter or the cited references for more information. The aim here is to show what choices can be made as you approach the question of cross-case analysis.

Case-oriented strategies. Yin (1984) advocates a *replication* strategy: A theoretical framework is used to study one case in depth, and then successive cases are examined to see whether the pattern found matches that in previous cases. It's also useful to examine cases where the pattern is expected on a theoretical basis to be weaker or absent. For example, Gilgun (1992), studying adults who had been maltreated as children, began with couples in which the male was now a perpetrator of child sexual abuse, then looked at male unmarried perpetrators, and then at males who were not perpetrators.

Denzin (1989b) approaches the problem through *multiple exemplars.* The issue, he says, is not so much "analysis" as interpretive *synthesis.* After "deconstructing" prior conceptions of a particular phenomenon (e.g., "the alcoholic self"), you collect multiple instances (cases) and then "bracket" them, inspecting them carefully for essential elements or components: The elements are then rebuilt into an ordered whole and put back into the natural social context.

Many researchers approach cross-case comparison by forming *types or families.* You inspect cases in a set to see whether they fall into clusters or groups that share certain patterns or configurations. For example, Gouldner (1958), as reported in Lofland and Lofland (1984), sorted college teachers into "cosmopolitans," with subtypes of "outsider" and "empire-builder," and "locals," which included "dedicated," "true bureaucrat," "homeguard," and "elders" as subtypes.

Sometimes the clusters can be ordered or sorted along some dimensions. For example, Morse and Bottorff (1992) found that 61 lactating mothers fell into four groups; those who could or could not express milk, and those whose attitudes varied (easy, or a hassle). The meaning of the experience was fundamentally different for each type of mother.

Families or types can be formed from quite complex configurations, as we'll see in Chapter 8, where we sort cases into similar scenarios on the basis of each case's causal network.

Researchers usually assume that the cases at hand are more or less comparable, structured in similar ways. Noblit and Hare (1988) suggest a strategy they call "meta-ethnography," which makes no such assumptions. The approach is one of *synthesizing* interpretations across cases, even if

Table 7.1
Hypothetical Qualitative Data Set for a Case:
Influence Actions and Responses, Case #005, Nynke van der Molen

INFLUENCE ACTIONS AND RESPONSES (PRE-DECISION)
(illustrative entries)

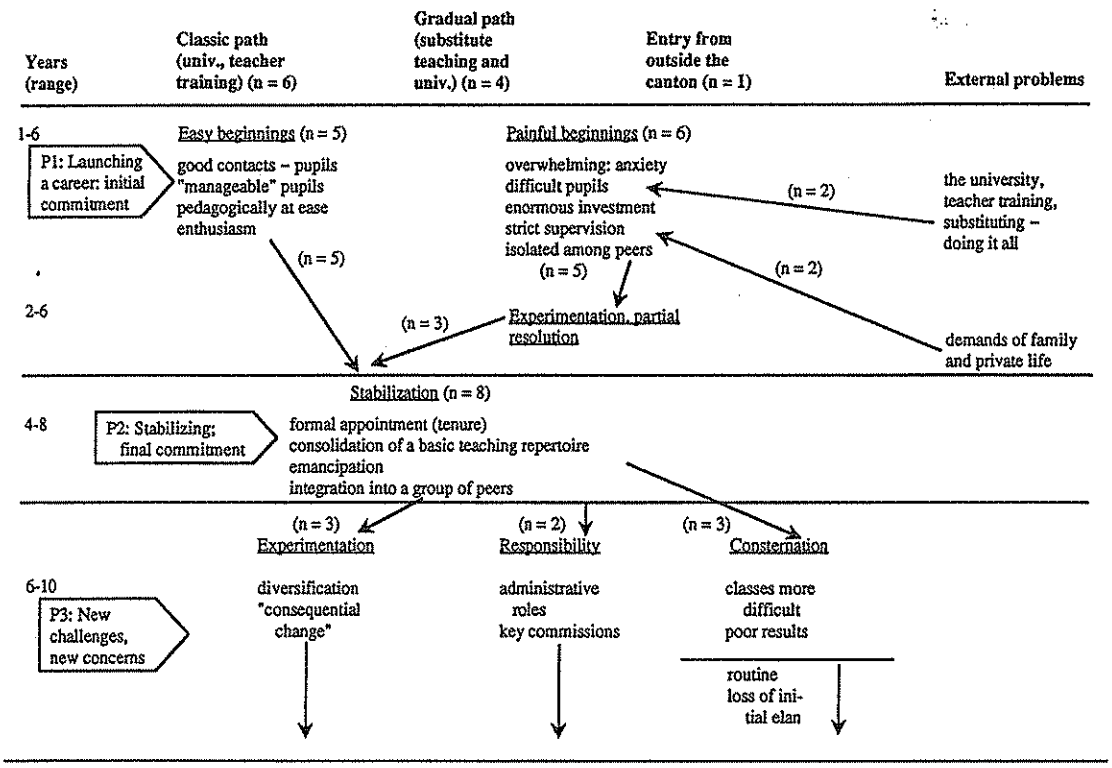
	Time 1: summer 89	Time 2: 1st term 89-90	Time 3: 2nd term, 89-90
Parents	A F says again that she should stay with the florist shop. R N. doesn't argue, [says work is "boring."]	M talks about regrets in her own life. But says it's too late to go back to social work. Seems bitter. Unknown.	M shows N. her scrap-book from social work school. N. really interested, says understood M as a young person for first time.
Siblings	A R	Brother proposes trip to Provence by motorcycle N. interested, uncertain. "My mother would be in a panic."	
Friends	A Trip to Paris with Leonie. R N. says she really loves travel, France in particular. "Leonie is my closest friend."	Leonie announces that she will go to university, study biology. N.: "That's her decision. I'll make my own when I'm ready."	
Teachers	A Math: tells N. she's bright, can do the work, makes silly mistakes. R [Shrugs].	Biol: assigned frog dissection task to Leonie & N. N. was "revolted," but enjoyed work with Leonie Got a grade of 8.	
Self	A R	"Sometimes I wish it would go away" [decision about university.] Seemed relieved to say it.	Begins part-time work at local stable. "I really, really, like it. And I get to ride a lot.

they were conducted by different researchers with different assumptions. Their substrategies include making "reciprocal translations" (can the findings of one study be used to predict those of another, and vice versa?), "refutational syntheses" (looking at apparently contradictory cross-case findings), and "lines-of-argument syntheses" (building a general interpretation grounded in the findings of separate studies).[3]

Variable-oriented strategies. Researchers often look for *themes* that cut across cases. Case dynamics as such are bypassed or underplayed. For example, Pearsol (1985) looked at interviews about sex equity programs with 25 teachers. After careful inductive coding (both descriptive and interpretive), he located recurring themes such as "concern for students," "activist view of change," and "barriers to innovation." (Later he also sorted the teachers into six types based on the configuration of themes.)

Often a key variable may not be wholly clear to you without cross-case analysis. The needed strategy might be called *pattern clarification.* For example, Eisenhardt (1989a) found evidence for the construct of "power centralization" by looking at data on CEO behavior in 10 microcomputer firms: Her matrix display included adjectives describing decision style, quantitative measures, and specific quoted examples, such as "Thou shall not hire without

Figure 7.3
Composite Sequence Analysis: Career Trajectory Data for 11 Cases (Huberman, 1989)



Presidential approval" (see section C below for more). She was, however, careful to look at the evidence within each case, not just aggregate superficially.

Mixed strategies. It's possible, and usually desirable, to combine or integrate case-oriented and variable-oriented approaches.

At a number of points in this and the next chapter, we suggest a strategy that might be called *stacking comparable cases.* You write up each of a series of cases, using a more or less standard set of variables (with leeway for uniqueness as it emerges). Then you use matrices and other displays to analyze each case in depth. After each case is well understood (the cross-cutting variables may evolve and change during this process), you "stack" the case-level displays in a "meta-matrix," which is then further condensed, permitting systematic comparison. A further example of this approach appears in Eisenhardt (1989b), who studied the dynamics of rapid organizational decision making in the microcomputer industry; see Chapter 8, section C for an illustration. Although Eisenhardt calls this a "replication" strategy, it isn't really done sequentially, but with all cases at once.

A phenomenological approach described by Fischer and Wertz (1975) might be called *interactive synthesis* (the term is ours). In a study of the meaning of being a crime victim, they wrote individual case synopses, then a cross-case narrative with themes, then did a "general condensation," then cycled back to the case synopses to see how the condensation was exemplified there, and then wrote a "general psychological structure" describing the temporal process of victimization (for more detail see Chapter 4, section I).

Ragin (1987) has written an enormously clarifying book on approaches to comparative—that is, cross-case—analysis. He points out that although combining or alternating case-oriented and variable-oriented strategies is useful, it is even more effective to think of a *synthesized* approach, which could "allow analysis of parts in a way that does not obscure wholes . . . and compare wholes [cases] as configurations of parts." His solution is *Boolean analysis,* which involves arraying binary data on several variables for each of several cases in a "truth table," and then identifying alternative patterns of "multiple conjunctural causation." The techniques are not difficult, and he provides many specific examples. The QCA and AQUAD programs (see Appendix) support this method.

Abbott (1992b), too, has interesting ideas on the problem of cross-case analysis. He shows how sheer variable-oriented analyses "lose the complexity and narrative order" of the cases, fragment them, and cannot follow the "plot" of each case. The issue, he suggests, is how to do cross-case analysis that preserves narrative sequence, "a social science expressed in terms of typical stories. . . . We need generic models for types of causality and for 'narrative steps' " (p. 76). The general strategy is one of evolving and testing *generic narrative models* for such issues as organizational birth, merger, division, and death.

One strategy that moves in this direction appears in the work of Huberman (1986, 1989) and Gladwin (1989). It might be loosely labeled *composite sequence analysis.* Huberman, looking at Swiss teachers' life histories, was analyzing a subsample of 11 teachers with 5-10 years of experience. It's easiest to see the analysis graphically (Figure 7.3). We can see that the 5 teachers with "easy beginnings" during the *launching a career* stage all moved to *stabilization,* while of the 6 "painful beginners," 3 "stabilized" partly through "experimentation." Of the 8 who "stabilized," all faced *new challenges and concerns:* "experimentation," new "responsibilities," and "consternation." The path of any particular individual can be traced through the cross-case display, but we also begin to see something like the "typical stories" that Abbott is calling for. The method is described more fully below in section E.

Gladwin's (1989) "decision tree modeling" approach describes the factors in individual decision making (e.g., freshmen deciding to buy a contract to eat breakfast in the cafeteria). Once a decision tree is constructed for each individual, you can do a composite tree that looks something like Huberman's display, and test an a priori model on it. Strictly speaking, these are not "sequences," but supposedly causal factors that are part of a logical model (see section C for a detailed description).

We've described these strategies generally, and may have left you with feelings of half-understood complexity and bewilderment. Let's turn to more detailed descriptions of how specific methods work, keeping in mind that any particular cross-case analytic method will always be embedded in a larger sequence of analytic moves.

We review five methods for cross-case analysis and eight supplementary methods in our usual format.

We begin with partially ordered displays, describing the *unordered meta-matrix,* a way of bringing together basic information from several cases into one big chart in the "stacking" style. Then we consider conceptually ordered displays: *content-analytic summary tables,* the *construct table,* and others useful in clarifying variables. Displays also can be ordered by arranging cases from high to low on some important variable; here we discuss the *case-ordered meta-matrix, scatterplots,* and others.

Finally, we describe time-ordered displays, including the *time-ordered meta-matrix* that illuminates data from several cases in relation to chronology, and *composite sequence analysis,* a process-oriented approach showing how different cases take diverse pathways through a generally-framed flow of events and states.

These displays are rich and complex, but they can be adapted easily to studies where the cases are a few individuals. They do not require large budgets or extended project length. Look for the essence of the display; if it looks promising, figure how to use it for your purposes. Focus on the doable, while keeping as much richness as you can.

B. Partially Ordered Displays

The first cut at cross-case analysis is often exploratory, to see what the general territory looks like. As with a single case, it's useful to start with displays that have some—but not too much—internal order. Don't impose conceptual, chronological, or other shape too quickly.

Partially Ordered Meta-Matrix

Analysis Problem

Cross-case analysis multiplies the data set by the number of single cases.[4] If each case produces several hundred pages of field notes and ancillary material, no matter how well you understand things at the case level, you are faced with the specter of overload and unmanageability when you try to draw meaningful cross-case conclusions deeper than, "Well, there are these similarities and these differences. . . ."

Perhaps, to be "shamelessly eclectic," as Rossman and Wilson (1991) cheerfully put it, we can borrow some of the logic of quantitative data analysts faced with too many numbers. Typically, they make all of the data comparable by standardizing them into the same metric. Then they reduce the overall amount of data by assigning data to fewer distinct sets: scores, indices, scales, distributions, and binaries with the evocative name of "dummy variables." Now you have fewer, condensed data, grouped and standardized; you can begin to see what the data look like—how numerous, how shapely, how distinct. Looking at distributions for these measures may suggest that some number sets need to be reorganized. This done, the quantitative analyst is ready to examine the relationships between these number sets, first pair by pair, then in combination.

Our problem is a good bit more complex. But perhaps some useful ideas are here. Cross-case data also need to be made comparable via common codes, common displays of commonly coded data segments, and common reporting formats for each case. Codes, displays, and reporting formats are all data-reductive devices for condensing hundreds of pages of text into workable, intellectually coherent units—tables or figures and associated analytic text.

Suppose you have comparably condensed data for each case in displays like those in preceding chapters. The problem is how to assemble them *coherently* and *in one place*—the major criteria for a good display. This is the first deep dive into cross-case analysis. It's also a critical one, because it will collapse the cross-case data set into partitions and clusters that may be used for subsequent analyses; this is how you'll first see the cross-case data. Refinements in the clustering and grouping of data may be made later, but for now you are making a commitment to a certain way of looking at what's there, making sense of it, and trying to work out the next most likely analytic steps.

Brief Description

Meta-matrices are master charts assembling descriptive data from each of several cases in a standard format. The simplest form is a juxtaposition—a stacking-up—of all of the single-case displays on one very large sheet or wall chart. The basic principle is *inclusion* of all relevant (condensed) data. We came to call such charts, affectionately, "monster-dogs."

From there, you usually move to *partition* the data further (divide it in new ways) and *cluster* data that fall together so that contrasts between sets of cases on variables of interest can come clearer. These partitioned and clustered meta-matrices are progressively more refined, usually requiring further transformations of case-level data into short quotes, summarizing phrases, ratings, and symbols.

All that sounds rather abstract, so we need an illustration. The basic thing to keep in mind is that a partially ordered monster-dog that carries all of the information from all of the cases can be gradually refined, summarized, and reduced through partitioning and clustering so that it becomes more ordered. Later we'll see how meta-matrices

Table 7.2
Case-Level Display for Partially Ordered Meta-Matrix (format)

Users	Feelings/ Concerns	How Innovation Looked	What Was User Doing Most?	Problems
1.				
2.				
3.				
4.				
5., etc.				

can be ordered conceptually (according to key variables), by cases (strong to weak on some variable), or by time (early to late).

Illustration

Let's run through the steps of an example.

Creating the display format for each case. Cross-case analysis depends on coherent within-case information. In the example shown here, the analyst was trying to answer several questions about the experience of teachers in using a new educational practice: (a) How did it look to the user? (b) What was the user doing? (c) Which feelings and concerns were paramount? (d) Which problems were looming large? How did all of this work over a long period of time? The empty case-level display might look like Table 7.2.

Entering the data. This is familiar: You sort the reduced case data into the four analytic categories, user by user, in separate but commonly formatted matrices, one per case.

Table 7.3 shows the display with real-life data entered for one case (for one time period). As we've seen before, cell entries are telegraph-style phrases boiling down coded chunks from the case data. Some selection and grouping has been done, too. The analyst has selected feelings, problems, and so on according to some explicit decision rules and has clustered discrete but similar items (e.g., under "Problems," the presence of several overloading tasks equals the problem of too limited time).

Each cell entry includes several words/phrases; this is the first thing the cross-case analyst will try to shrink.

Building the partially ordered meta-matrix and entering further-reduced data. You are now sitting in front of N case-level displays—comparable ones. If there are no more than a dozen cases or so, and if the displays are no more complex than these examples, the data can probably

Table 7.3
Case-Level Display for Partially Ordered Meta-Matrix:
Users' Second Year of Implementation at Lido

User	Feelings/ Concerns	How Innovation Looked	What Was User Doing Most?	Problems
Vance	More comfortable with style of teaching and with having kids outside	Still useful, giving good direction & helpful ideas, activities	Working through materials Giving, participating in env'l educ workshops Working with community Off-campus site work	Time too limited for tasks to be done
Drew	Concern with growing number of non-achievers in forestry/ecology class	Too discovery-oriented for kids without biology basics; lecture style more appropriate	Adapting materials & lessons to growing non-achiever population Off-campus site work	Dealing with more non-achievers successfully
Carroll	Excitement with new activities, expanding science program	Same as first year	Working with community Giving, participating in env'l educ workshops Off-campus site work	Over-extended activity commitment

all go on one composite monster-dog sheet. With more cases, or with more-complex case-level displays, an intermediate step of further condensation will need to be done. In either case, remember that you'll also have several paragraphs of analytic text clarifying and deepening the entries in the display for each case, which can help in condensation.

The partially ordered meta-matrix will be simply a stack of the case-level charts. For example, the first two columns would appear as shown in Table 7.4. Getting all of the column entries for all of the cases onto that matrix means it is going to be a big one. Maybe they should be called "mega-matrices" rather than monster-dogs. Big sheets of paper and a bare wall to stick them on are the usual ingredients for cooking up a manageable monster-dog.

The work is much easier with appropriate software; some word processors can produce matrices like these, and more-specialized programs such as AQUAD, Orbis, or NUDIST (see Appendix) are very helpful.

Usually you need to distill case-level cell entries somewhat as the work proceeds. Thus, for example, the filled-in chart for Lido in Table 7.3 might appear as in Table 7.5. We are down to fewer, case-typifying words and some symbols. Feelings/concerns have been dichotomized into positive and negative ones, and each entry has a type of feeling or concern set at about a middle range of abstraction (e.g., "comfortable," with a qualification) and often a short, category-grounding phrase (e.g., "underachievers assigned") to keep the analyst aware of what specifically was going on locally.

Stripping down an unordered meta-matrix to less information than this on the first iteration is looking for trouble. Without the qualifying and grounding task, you can get quickly into comparing different species (of, say, "comfort") with no way of seeing the noncomparability. When the analysis founders, or when you need truly representative illustrations, there is no easy way to shift back to the

Table 7.4
Partially Ordered Meta-Matrix: User Feelings/ Concerns and Other Variables (format)

Sites/Users	Feelings/Concerns	. . .
Lido 1.		
2.		
3.		
Masepa 1.		
2.		
3., etc.		

Table 7.5
Partially Ordered Meta-Matrix: User Feelings/ Concerns and Other Variables (Lido data)

Sites/Users	Feelings/Concerns	
Lido 1. Vance	+ Comfortable with new style, norms (lower control)	
2. Drew	- Shift in target public (underachievers assigned)	
3. Carroll	+ New, expanding (program, activities)	

next lower level of analysis. You have to go all the way back to the case-level displays. Doing that takes time and opens the door to a more sinister bias—looking only at *part* of the data, not at the full set (all relevant data from all the cases).

The meta-matrix has assembled and pruned down multiple tables to one big one that has considerably condensed the original data. It lets you see all of the data in one (large) place. It's used at an early stage of cross-case analysis; the analytic categories (feelings, etc.) have been carried over from the case-level displays, and individual entries (e.g., concerns over a growing number of nonachievers moving into the program) have become instances of more general analytic types (shift in target public).

From now on the analyses and matrices will depend more on what the data have to say and on the kinds of relationships you are most interested in. There is often a sequence of new meta-matrices.

Let's look briefly at two such extensions—one involving within-category sorting (tactics: **partitioning** and **clustering**—see Chapter 10, section A) and the other involving the tactic of cross-category **clustering.**

Within-category sorting. Let's assume you want to do some finer-grained work on the later "concerns" of users during project execution. This work would involve the inspection of each meta-matrix for that period—the one for "second year," for "third year," and so on. In doing such a scan, you would look for two sets of things: (a) trends: concerns that change or remain constant over time, and (b) types of concerns. A new display is being created by **partitioning** the data (in this case, by time and concerns) and by **clustering** the data (into different types of concerns). The matrix format appears in Table 7.6.

You are looking in particular for distinct *types* of concerns—ones that can be classified or catalogued—and any variation in the types of concerns over *time.* Under each year might be different entries (checks for presence/absence of a concern; ratings of "primary," "secondary," concerns, etc.). You might decide that "concerns" could be partitioned into institutional and individual categories. And you might wonder whether there were more institutional than individual concerns during later implementation (tactic: **making if-then tests**).

Table 7.6
Time-Ordered Meta-Matrix (format)

Type of concern	Year 1	Year 2	Year 3	Year 4
1.				
2.				
3., etc.				

As a next step, you could create a *summary table* of individual and institutional concerns (**partitioning** the data according to some predefined variables) and see at which of the 12 cases they were considered to be primary during later implementation. During this process you can do some empirical **clustering** within each of the two broad categories ("relational problems," "motivational problems," etc.). It might look like Table 7.7.

Here is much useful descriptive information. For example, some cases have multiple troubles (e.g., Banestown, Perry-Parkdale), others only one kind (Masepa), and two sites of the 12 have escaped unscathed (tactic: **counting**). Also it looks as if individual concerns are mostly nontechnical, and institutional concerns emphasize continuation (lower priority, poor functioning, delivering on promises) (tactic: **noting patterns, themes**).

Such a finding leads you back to other parts of the cross-case chart to test hunches (**making if-then tests**). In this case the hunch might be that projects with high technical mastery have just as many institutional concerns as those that do not.

It also suggests next steps in the analysis: Let's check whether institutional concerns actually lead to cutbacks or discontinuations in the next time period. A simple 2×2 matrix sorting case-level data can help answer that question and suggest where to look next, moving beyond a descriptive analysis to more explanatory ones.

Table 7.7 is a simple way to look at the data. Fancier displays are possible, such as one including different levels of gravity for each concern, one with the sources of the concern, one with the consequences, or one with all of these things.

Table 7.7
Summary Table:
Individual and Institutional Concerns During Later Implementation

Type of Concern/Item	Sites at Which Item Mentioned
Individual Concerns	
Relational problems-friction among project staff	Banestown, Perry-Parkdale
Motivational problems (discouragement, "distaste" for the practice)	Calston, Lido, Masepa, Dun Hollow, Proville, Tindale
Stamina, exhaustion, excessive demands of the project	Masepa, Carson, Plummet
Institutional Concerns	
Lower institutional or district-level priority of project (with attendant lessening of rewards)	Lido, Plummet
Poor overall functioning (project as a whole, or a component)	Perry-Parkdale, Dun Hollow, Proville
Resistance, obstruction, lack of support by other staff not on project	Banestown, Perry-Parkdale, Carson
Worry whether project will "deliver" on its promises or objectives	Banestown
Continuation the following year(s)	Banestown, Calston, Lido, Perry-Parkdale, Plummet

Across-category clustering. Now let's look at a summary table that reaches across and takes apart analytic categories. In this example the analyst wanted to characterize the more technical aspects of teachers' later use of the innovation. This process involved working with the last two columns shown in Table 7.2—what the user was doing and the problems encountered in doing it.

The procedure was straightforward. The analyst read down the last two columns of Table 7.5 (the filled-in meta-matrix), looking at each of the 12 cases, noting what users were "doing most" and the "problems encountered" in doing it. The rule of thumb was to look for clusters involving at least two or three cases.

Assembling the "doing most" responses, the analyst tried to catalogue them by looking for an inclusive gerund, something that would yield the underlying *activity* (e.g., "refining"). This step produced six distinct types of activity into which all instances could fit without a shoehorn. The analyst then connected the "problems encountered" to each of the activity types by (a) **clustering** problems into a lesser number of problem types and then (b) taking the modal problem type (the one mentioned by most informants) when there were more than one. The researcher then copied down some short quotes and summarizing phrases and set up Table 7.8.

Note what the cross-category clustering has accomplished. It is a variable-oriented display, but we have gone well beyond the fragmented, case-by-case account of users' later experience as shown in Table 7.3 and its following displays, and have gotten closer to a focused and integrative cross-case depiction that does not read abstractly and retains a lot of the case-specific flavor.

Reading across the rows gives a good functional sense of what, for example, "reaching up" is and what real-time problems in doing it look like. Reading down the columns gives a cumulative picture of what "later implementation" is technically all about.

Variations

Many alternative paths may be taken from the basic partially ordered meta-matrix to variously ordered meta-matrices. Sometimes, however, the first meta-matrices are sufficient for producing summary material for a report: See the "analysis tables" described by Smith and Robbins (1982) and Lee et al. (1981), which display components

Table 7.8
Clustered Summary Table: Problems Stemming From Major Tasks of Later Implementation at Field Sites

Task/Activity	Problem type	Illustrations
Reaching up	Difficulty in mastering more complex components	"Some parts just aren't working. They're harder than the first ones I did" (Teacher, Masepa) "I still can't monitor the children's progress...there are too many pupils to keep up with." (Teacher, Calston)
Improving, debugging	Solving specific problems, often connected to poor program design	"The bookkeeping is awful." (Teacher, Banestown). Paperwork (Perry-Parkdale, Carson). Scheduling problems (Banestown, Calston, Carson and Proville). Pupils disturbing, interrupting each other (Perry-Parkdale). "Materials [that are] not right for this age." (Dun Hollow)
Refining	Differentiating, tailoring	"needs tailoring ... the time isn't there" (Teacher, Lido) Inability, despite repeated attempts, to break down groups into individual exercises. (lab teachers at Banestown)
Integrating	Combining new project with regular curriculum	"Will this destroy my own curriculum? How can I fit it in?" (Teacher, Burton) Lack of time to do both old and new (Calston, Dun Hollow, Carson)
Adapting	Accommodating the project to new inputs	"The course is all floaters now ... I get the burn- outs, who sit in a daze and are hard to reach." (Teacher, Lido) Changes in released time, budget decreases (Lido); new testing requirements (Banestown)
Extending	Using components elsewhere, reorganizing practice, bringing in other segments	Trying to bring back former "stuff that worked" (Masepa); feeling "bored," looking for more stimulating materials (Tindale)

supporting parental involvement in various federal programs, sorted by case.

Advice

Don't move from case-level data directly to an ordered meta-matrix (as in the following sections). Look carefully first at the multiple-case, partially ordered information. Jumping to ordered meta-matrices is like going straight from marginals to regressions without any intermediate cross-tabbing or scatterplotting. A good grounding in the first meta-matrix helps you avoid costly blind alleys.

If your cases are less complex than ours, so will your meta-matrices be. The "stacking" approach is actually less difficult than you might think.

As we'll see, it usually takes several alternative cuts before a workable and productive set of partitions and/or clusters emerges. For more on **partitioning** and **clustering,** see Chapter 10, section A.

Time Required

Creating a partially ordered meta-matrix like this one can take about 3 or 4 hours; subsequent ordering exercises like those shown, 1 or 2 hours each. Writing associated analytic text is quicker—an hour or less. Creating the basic meta-matrix and assembling the data take the most time. Of course, much depends on the number of cases, the number of variables you are looking at, and whether the case data matrices are fully comparable and easily accessible.

Word-processing, word-retrieving, and text-base managing applications (see Appendix) can speed the mechanics of basic meta-matrix assembly and the processes of clustering and partitioning. Time savings depend on your having stored the case data in easily retrievable form—and on having some coherent, data-based ideas about clustering and partitioning. There is no real way to shortcut the *thinking* time that is needed. The process of understanding a cross-case data set, even for a few variables, is one of making gradual sense. It cannot be hurried.

Table 7.9
Content-Analytic Summary Table: The Content of Organization Changes

Type of Change	TRANSITORY CHANGES Within the innovation	TRANSITORY CHANGES In the organization	DURABLE CHANGES Within the innovation	DURABLE CHANGES In the organization
Structure	Addition of project directors Evolving departments Creation and dropping of advisory committee	Change in locus of control, funding, and supervision	Shift in student input (5) Reorganization: more department interdependence Advisory committee Team teaching	Innovation itself (5): remedial lab, alt. school work experience program, English curriculum, accountability system None (5) Scheduling (4) Expansion of innovation to new users (3) Space for program (2) Creation of coordinator role Creation of management committee Creation of in-service committee Summer-school version of program added Increased team teaching Creation of helping teacher role
Procedures	Active, coercive student recruiting Leadership providing support (emotional, material) Leader "consultation" with teachers Non-users help users with testing Aides, parents help with instruction Increased community contact	Innovation lapsed, is discontinued Field testing of materials Teachers exchange books Other instruction displaced	Tighter supervision, control (2) Student "batching" Routinization of forms/procedures Charismatic leader more distant Fewer staff meetings Reduced student time in program Staff reassignment Selective use of materials	Innovation itself (5): pupil screening, referral/intake, teaming, pullout procedures More paperwork (2) Added student transportation (2) Teacher load reduction Freeing time: substitutes, early dismissal Dropping letter grades More coordination among teachers More teacher discussion of individual students Loss of teacher control over student assignment
Climate	Conflict (4): in teams between users, users-aides, departments More cohesiveness (2), pioneer spirit, esprit User resentment (2) Tension, fear, mistrust	User-non-user conflict (2) User resentment of remedial lab Ambiguity about management committee Expectations for more upward influence, violated, lower morale Resistance by principals, then support	Discouragement, burnout, loss of interest (3) Collaborative, help-giving climate (2) Cohesiveness (2) Less cohesiveness Overload Development of friendships/enmities	Resentment re. paperwork (2) Wariness about innovation More openness to innovation Norm change: flexibility, colleagueship, el.-sec. interaction None: climate still good None: climate still bad More administrative awareness of the innovation

*Each item represents a change occurring at one site. Changes occurring in more than one site have numbers appended.

C. Conceptually Ordered Displays

As when we're looking at single cases, multicase displays may be centrally organized by variables or concepts.

Content-Analytic Summary Table

Sometimes you may want to focus primarily on the *content* of a meta-matrix, without reference to which cases it came from.

One of the first tasks in moving from a single-case to a cross-case analysis is determining how many cases share similar characteristics. As we've seen, in the simplest form, you take the original matrix used for the single cases and generate a meta-matrix that gets all of the (reduced) data in. When the same characteristic appears in more than one case, this can be noted on the matrix. But it often happens that, later on in the analysis, a new way of arranging the cases looks promising—one that will help you understand the structure of the data across all cases.

So the task is to imagine a matrix display that best captures the dimensions you are interested in, and arranges all of the pertinent data in readily analyzable form. Table 7.9 is an example of a content-analytic summary table from the school improvement study.

You can see at a glance multiple-case activity in any one of the cells by noting the numbers in parentheses. You also can tell how many kinds of organizational changes are within any one type by noting the number of entries in a cell. For example, "transitory changes in structure" within the innovation (upper left cell) are not frequent and occur in no more than one case, while "transitory climate changes within the innovation" (two cells down) are more diverse and show recurrence across several cases (tactics: **making contrasts, comparisons; counting**).

But the picture is somewhat different for durable changes. It looks as if "structural changes in the organization" (upper cell at far right) are often nonexistent, and are limited to the innovation itself or express themselves in scheduling changes alone. Even so, there are many more durable than transitory changes.

Note: This sort of tabulation deliberately drops the case identification of data. The aim is to be more conceptual, seeing main trends across the cases. You may, of course, attach an initial to each data bit to let you get back to the case if you wish.

For more examples of content-analytic summary tables, see Boxes 8.1 and 8.2, p. 212.

Box 7.1
Substructed Variable:
Extent of Organizational Change

Is the innovation in place?		Have organizational changes occurred beyond the innovation itself? Yes	Few or None
	Yes	Carson Masepa (1)	Plummet Perry-Parkdale Tindale Banestown Astoria (2)
	Partially or on limited basis		Calston Lido Burton (3)
	No		Dun Hollow Proville (4)

Substructing a Variable

Often, when you're trying to clarify a general variable, especially a variable you're using as an outcome along which to sort cases, the data seem confusing or intractable. Perhaps the variable has more than one dimension; the cases do not seem to sort themselves out simply. The technique of *substruction* helps here. It's a way of locating underlying dimensions systematically.

In Box 7.1 we show how an analyst substructed the issue of "organizational change," looking at 12 cases. The problem was that, in some sites, the innovation was poorly implemented to begin with, and couldn't be expected to induce organizational change. Also, in some sites the innovation itself was an organizational change, and in others it was not (it was simply a small-scale change limited to a classroom). So the analyst made a 3 × 2 table to incorporate these dimensions and then sorted out the cases. Looking at it this way, you can see that the ordering of cases on the extent of organizational change should go from category 1 down through 4.

The idea of substructing, originally developed by Lazarsfeld (see Lazarsfeld, Pasanella, & Rosenberg, 1972), is discussed further as "typologizing" by Lofland and Lofland (1984, pp. 93-96); see also Strauss and Corbin (1990, pp. 69-72) on "dimensionalizing."

Construct Table

Although you may have a general idea in advance about the properties of some major variable, such as "innovation stabilization" or "assistance," such variables do not usually come clear until real case data have been explored in some depth. Cross-case construct tables are an excellent way to understand a core concept because the way the variable plays out in different contexts illuminates its essential nature.

A good example appears in Eisenhardt (1989a). She was trying to clarify the concept of *power centralization* in organizations. Table 7.10 presents data from eight electronics firms, focusing on the functioning of their chief executive officers.

It includes qualitative judgments of CEO decision style, questionnaire data, organizational information, and representative, position-attributed examples. The process of assembling the information and the gradual clarification of the power centralization concept was iterative; data helped clarify the construct and suggested the need for other data to provide further grounding. Table 7.10 both summarizes the final information, and succeeds in ordering the cases for subsequent analyses (see Table 7.15).

Incidentally, Eisenhardt's discussion of the steps in building theory from case study research, drawing on six other multicase studies of organizations, is very useful; it covers early conceptualization, case selection, instrumentation, data collection, searching for patterns, and shaping hypotheses.

Note that the ordering is not a mechanical, automatic process; look at the positions of Cowboy and Neutron. Although Neutron's story decision style is authoritarian, the case still ranks lower than Cowboy's "authoritarian/consensus" style, probably because of the lower Neutron power distance measure and the CEO decision description for Cowboy as "strong, power boss."

Table 7.10
Construct Table:
Example of Tabulated Evidence for a Power Centralization Construct (Eisenhardt, 1989a)

FIRM	CEO DECISION DESCRIPTION	CEO POWER SCORE	CEO POWER DISTANCE [a]	CEO DOMINATED FUNCTIONS	STORY DECISION STYLE [b]	EXAMPLES [c]
First	Strong Volatile Dogmatic	9.6	3.5	Mkt, R&D, Ops, Fin	Authoritarian	Geoff (Chairman) is THE decision maker. He runs the whole show. (VP, Marketing)
Alpha	Impatient Parental Tunes You Out	9.6	3.8	Mkt, R&D, Ops, Fin	Authoritarian	Thou shalt not hire w/o Presidential approval. Thou shalt not promote w/o Presidential approval. Thou shalt not explore new markets w/o Presidential approval. (VP, Operations)
Cowboy	Strong Power Boss Master Strategist	9.1	3.1	Mkt, R&D, Fin	Authoritarian Consensus	The tone of meetings would change depending upon whether he was in the room. If he'd leave the room, discussion would spread out, go off the wall. It got back on focus when he came back. (Director of Marketing)
Neutron	Organized Analytic	9.1	2.3	Mkt, Ops, Fin	Authoritarian	If there is a decision to make, I will make it. (President)
Omicron	Easy Going Easy to Work With	8.4	1.2	Fin	Consensus	Bill (prior CEO) was a suppressor of ideas. Jim is more open. (VP, Mfg.)
Promise	People-Oriented Pragmatic	8.9	1.3	Ops, Fin	Consensus	(My philosophy is) to make quick decisions involving as many people as possible. (President)
Forefront	Aggressive Team Player	8.3	1.2	None	Consensus	Art depends on picking good people and letting them operate. (VP, Sales)
Zap	Consensus-Style People-Oriented	7.5	0.3	Fin	Consultative	It's very open. We're successful most of the time in building consensus. (VP, Engineering)

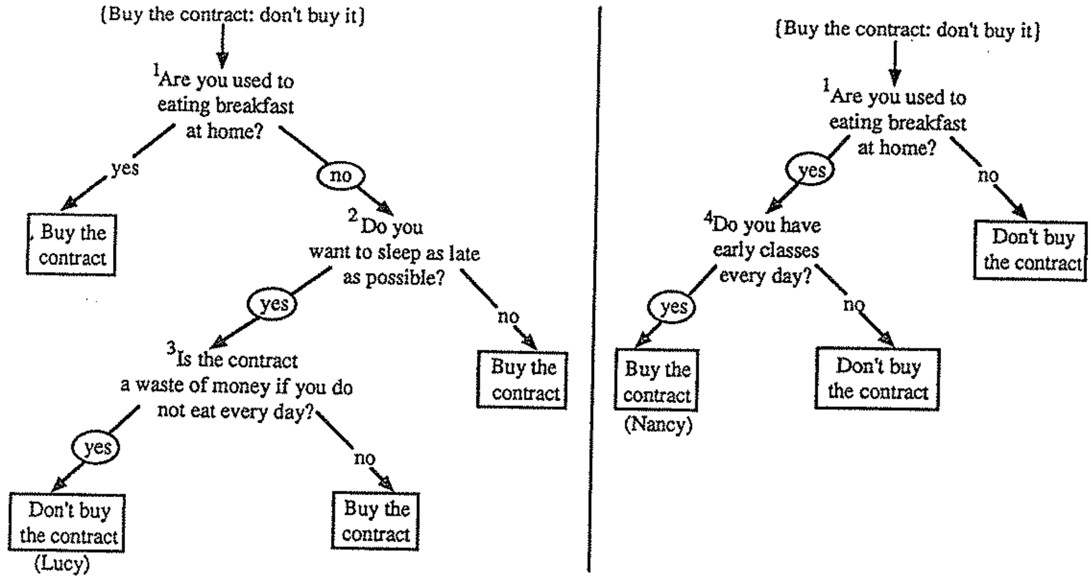
[a] Difference between CEO power score and score of next most powerful executive.
[b] Authoritarian–Decisions made either by CEO alone or in consultation with only one person.
Consultative–Decisions made by CEO in consultation with either most of or all of the team.
Consensus–Decisions made by entire team in a group format.
[c] Individual in parentheses is the source of the quotation.

Decision Tree Modeling

Sometimes you want to know how a basic process plays out across a number of cases, and to build a general model of just how it works.

One such process is decision making and related actions by individuals. How do high school students decide to attend university? What is involved in moving from soft to hard drug use? Will Third World farmers keep using a

Figure 7.4
Case Examples:
Decisions About the Breakfast Contract by Two Students (Gladwin, 1989)



fertilizer after the subsidy is removed? Why do women's support groups disband?

Gladwin (1989) describes decision tree modeling as a way to ground a description of real-world decisions and actions coherently by using multiple cases. Figure 7.4 shows a decision tree for each of two freshman students who were interviewed about why they had or had not bought a contract to eat breakfast in the college cafeteria.

The top of the display shows the *choice alternatives*, the items in the middle the *decision criteria*, and the bottom the *decision outcome*. The tree is a network display that condenses what the person actually said (e.g., Lucy said she chose not to buy a contract because she had not been eating breakfast since the fifth grade. She prefers to sleep as late as possible in the morning and thought that missing a paid breakfast on a regular basis would be a waste of money).

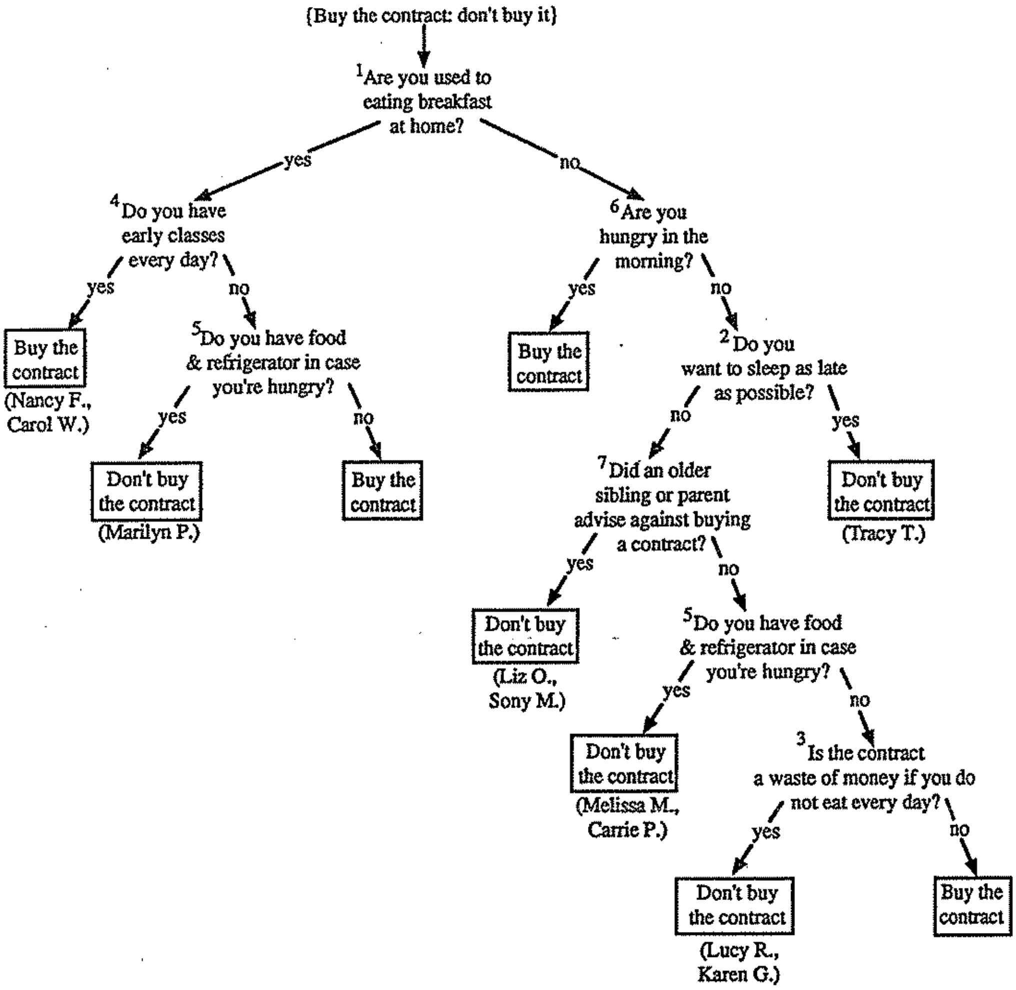
The procedure can be done iteratively: The criteria mentioned by the first person (Lucy) are used in the interview with the next person (Nancy), and so on. Gradually a composite model is built. Or a full set of individual decision trees can be done and then generalized into a composite. Figure 7.5 shows a first-cut composite model drawn from 10 cases. It was assembled intuitively; Figure 7.6 shows a more logically derived version that includes the idea of constraints (the "unless" items).

With either approach, it's essential to preserve the validity of each separate decision model; each person's path must be represented accurately, even though some more-idiosyncratic "emic" features may need to be generalized to a more abstract "etic" version. This individual focus illustrates what Abbott (1992b) says about the need to follow the "plot" of each case, preserving complexity and narrative order, and avoiding a sheerly variable-oriented "summary." Note, however, that the decision tree approach is stronger on *logical* order, rather than narrative order as such.

The next step is to test the composite model through interviews with a *new* population, using all of the decision criteria. Gladwin (1989) is very helpful on the concrete steps (and pitfalls) involved in interviewing and tree construction, and shows how to handle multiple decisions.

Note that this sort of display lends itself well to improving practice (breakfast contract procedures) or recasting policy (should the college have contracts at all?).

Figure 7.5
Composite Model 1:
The Freshman's Decision to Buy or Not Buy a Breakfast Contract (Gladwin, 1989)



D. Case-Ordered Displays

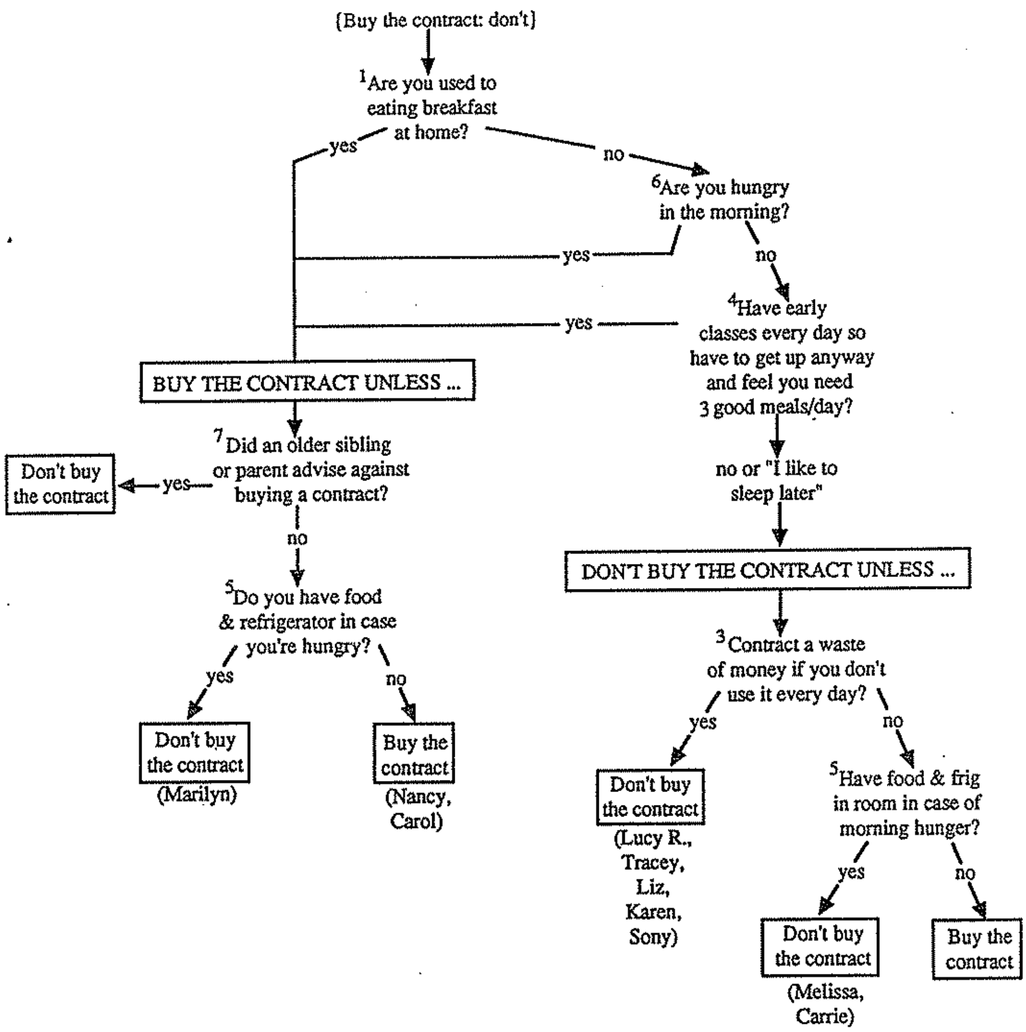
Case-ordered displays array data case by case, but the cases are ordered according to some variable of interest, so that you can easily see the differences among high, medium, and low cases. It's a powerful way to understand differences across cases.

Case-Ordered Descriptive Meta-Matrix

Analysis Problem

A partially ordered meta-matrix like the one in section B is a good start, but it usually leaves you hungering for more understanding and more focus. What is going on? What are the real patterns of more and less X in the cases?

Figure 7.6
Composite Model 2:
The Freshman's Decision to Buy or Not Buy a Breakfast Contract (Gladwin, 1989)



Why do we see more X at one site or case and less of it in another?

Brief Description

A case-ordered descriptive matrix contains first-level descriptive data from all cases, but the cases are *ordered* according to the main variable being examined. Thus it coherently arrays the basic data for a major variable, across all cases.

Illustration

In our school improvement study, we were interested in a variable we called "student impact." We had interviewed not only students themselves but also teachers, parents, administrators, counselors—and had looked at formal evaluation data as well. How could we assemble this information in one place and understand the differences among cases that showed high, medium, or low student impact?

Our interviewing and data retrieval had focused not only on "direct" student outcomes (e.g., improved reading test scores) but also on what we called "meta-level" or more general outcomes (e.g., increased interest in school activities) and even on "side effects" (e.g., "zest in experimental program leads to alienation from regular classes"). We also wanted to track both positive (good, desirable) and negative effects of the innovation on students.

Building the display. How can the cases be ordered? You can look at the relevant sections of each case report and note the general level of student impact the writer of the case report had claimed. You place the cases in rough order, and then you reskim the text in each case report to see whether the first impression is justified or whether the ordering should be changed.

This first reading also suggests ideas for building the meta-matrix. Table 7.11 presents a first possibility. The idea here would be to read the case report sections on student outcomes and to enter a code in each cell (U for user, A for administrator, E for evaluator, S for student, P for parent, etc.). If the impact was seen as strong, the letter would be underlined; if uncertain, a question mark would be added; if the evidence was conflicting, an X would be added.

But this display is not quite right. It abstracts and lumps the outcomes too grossly into categories such as "achievement," throwing away a lot of rich information on just what the outcomes were. And on reflection, we can realize that outcomes might just bear some relation to desired *objectives.* Somehow they should be in the display, too.

Table 7.11
Case-Ordered Meta-Matrix:
Format for Student Impact Data (version 1)

		Direct effects		Meta-level & side effects	
		Positive	Negative	Positive	Negative
CASE 1	Achievement				
(highest impact)	Attitudes				
	Behavior				
CASE 2 (next)	Achievement				
	Attitudes				
	Behavior				
etc.					

Table 7.12
Case-Ordered Meta-Matrix:
Format for Student Impact Data (version 2)

		Direct effects		Meta-level & side effects	
CASES	Program objectives	Positive	Negative	Positive	Negative
CASE 1 (highest impact)					
CASE 2 (next)					
etc.					

At such a point it's good to return to the case reports. What do they really look like, and what kind of display might accommodate them most easily? A reading showed us that the data were of variable robustness—ranging from detailed and thorough achievement data, interviews, and questionnaire responses to nearly unsupported opinion. Also some programs aimed high, shooting for everything from student achievement to self-concept, while others were more focused, hoping only, say, for the improvement of reading skills.

That range suggests the usefulness of a display more like Table 7.12. The idea here is to show the objectives, and to enter data in the form of phrases or sentences in the appropriate cells, marking them with code letters (U, A, E) to show the data source. In that way we keep more of the raw data, and we can see how outcomes are related to objectives.

Entering the data. Now we can return to the relevant section of each case report, look for coded material on direct and meta-level outcomes, and summarize the material in a set of brief sentences or phrases, one for each distinct outcome. We can arrange the data in the "program objectives" and the "direct effects" columns in parallel fashion to make comparisons easier. Skills and achievement goals might be placed first, followed by affective/attitudinal and other outcomes. For each sentence or phrase entered, the source is marked (A for administration, U for user, etc.).

After the data are all assembled into the matrix, review again: Are the cases correctly ordered from high to low student impact? In this case some criteria for a "high" rating are (1) Is the program achieving most of its aims? (2) Is it also achieving other positive meta-level and side effects? and (3) Are these judgments corroborated, either through repeated responses from one role through cross-role agreement or evaluation data? Weakness on criteria 1 and 2 especially should lead to lower ratings of student impact; weakness on criterion 3 may also cast doubt on a case's placement. Such a check usually suggests some reordering of cases. (Within each category, such as "high" impact, also try to order the cases from "highest" to "least of the high.")

Table 7.13 shows how this sort of data display worked out. The rows show two cases (Perry-Parkdale, Masepa) where student impact was high, two where it was moderate (Carson, Calston), one moderate to low, and two low. (For simplicity, we show an excerpted table, with only 6 of our 12 cases.)

To clarify the last two columns: We defined meta-level outcomes as congruent with the program's purposes, but affecting more-general aspects of students' functioning. Note the reading program at Calston: It led to the direct outcome of improved reading skills, but by providing interesting materials and opportunities for independent student work, it may have induced the meta-level outcome of increased student self-direction.

Side effects are more unintended: Note the alienation of Perry-Parkdale students, who liked their program's concreteness and relevance, but thereby came to dislike their regular high school's courses and activities.

Drawing conclusions. (Try your hand at analysis before reading further.) At first glance Table 7.13 looks daunting, overloading. But first try a "squint analysis." Where in the table do data look dense or sparse?

We can see right away, looking across rows, that positive effects are more frequent than negative ones (tactic: **making contrasts, comparisons**). Another squint shows us that cases where many negative effects are noted (e.g., Perry-Parkdale, Masepa) are cases where much is being attempted (in terms of the program objectives column). Apparently, large efforts are more likely to spin off negative effects along with the positive ones (tactic: **noting relations between variables**).

Looking at the two right-hand columns shows significant meta-level and side effects in nearly all cases (tactic: **noting patterns, themes**); in only the lowest impact case (Burton) do we see none at all. Perhaps this pattern means that the claims for meta-level changes in moderate-impact cases are suspect. Maybe so—they are less likely to have repeated mention (underlines) and have fewer instances of multiple-role confirmation (tactic: **using extreme cases**).

You can **contrast/compare** the projects with externally developed innovations and with locally developed ones. In the full matrix, approximately equal proportions fall in the high, moderate, and low categories; we can conclude that program sponsorship makes no difference in student impact.

Table 7.13 can also be used to assess the types of effect in high-impact cases (achievement, skills, attitudes, behavior) through a detailed analysis of objectives and positive effects. We can answer the question of who reports impact (administrators, users, etc.) and whether this differs in high- and low-impact cases.

Usually a large case-ordered meta-matrix like this one is not fully comprehensible in the first cut, although it is a perfectly fine "reference book" for keeping relevant data in one accessible place. Normally you need next-step matrices—streamlined, boiled-down, sometimes regrouped versions—as well.

Let's see what a next-step table might look like. The analysis just made suggests two likely avenues. First, we could compare cases according to the *content of outcomes,* such as knowledge, skills, and attitudes, keeping the distinction between *direct* and *other* types of outcomes. Another possibility is to compare *degrees of achieving objectives* across cases: What proportion of the objectives for each project were actually achieved, and to what extent?

A next-step, summarizing table can help in at least three ways: (a) providing a clearer check on how you actually made the ratings of "high" "moderate," and "low" student impact, (b) allowing a finer-grained comparative analysis of the outcomes, and (c) giving the reader a more digestible visual summary than Table 7.13. The summarizing chart may be the only table the general reader will see; Table 7.13 could be consigned to an appendix or included only in reports for technical readers who want to examine the material carefully.

Table 7.14 is a cut at the second option: a summary of the extent to which each case attained its objectives. The format is in the mode of a *Consumer Reports* rating, to make rapid case evaluation and comparison easier.

This is a radical condensation of Table 7.13; we are down to a set of symbols and some qualifying phrases. But it is easy enough to find your way rapidly back to the denser matrix if something ambiguous or puzzling crops up on this one. The reader now has an accessible overview of impact levels within and across projects. The analytic text accompanying such a display needs to provide contextual information, to compensate for the highly standardized, disembodied nature of these data; you pay a price for easy-to-read, comparable cross-case summarization.

In the subcolumns under "overall rating" the analyst made a differentiation that wasn't in Table 7.13. "Program-immediate" objectives are more concrete, attainable, and content-bound, such as "mastery of reading skills" or "knowledge of government processes." "Transfer" objectives are also tied to the program objectives, but are more general and usually harder to achieve. For example, the objectives at Carson call for "clearer career interests" (a program-immediate objective) and also for "improved self-concept as a learner" and "more internal locus of control," which are "transfer" objectives—they go beyond first-level mastery of the project content to a more gener-

Table 7.13

Case-Ordered Descriptive Meta-Matrix (excerpt):
Program Objectives and Student Impact (direct, meta-level, and side effects)

CASES	Program objectives	Direct outcome: Positive	Direct outcome: Negative	Meta-level outcomes and side effects: Positive	Meta-level outcomes and side effects: Negative
High impact					
Perry-Parkdale (E)	Basic skills (reading, math, communication, etc.) Life skills (critical thinking, citizenship) Life competencies (credit, health, etc.) Job-finding skills Entry-level work skills, career training Decision-making skills Self-responsibility and initiative Matching interests/ abilities to job information Understanding of others Know socioeconomic trends: career knowledge Sex role awareness Better parent-child communication	Same or better on basic skills scores U, better math (12th grade girls) E P Specific job skills S E P U Career planning and choice: aids goalless Job exploration attitudes positive E How interests fit jobs, what to look for in job E: identity, explore career options C A Experience with world of work: headstart C U S Career, occupational knowledge E Preparation for real world of work Experience in working with adults: P communicating with them better U	Less exposure to academic courses C	Personal development, identity, U E P Better communication with adults U S Self-confidence U E P S Increased attendance U Responsibility, motivation U C E P Staying in school (lower-level students) C A U Help goalless students A Improved peer relations U Getting permanent job S Parent-child communication P	Some students "get lost" again after a while U Some bide time, don't "deliver" U Program more effective with girls U Alienation from regular school activities C A S Irresponsibility, "can't handle freedom" P U
Masepa (E)	Improvement in full range of language arts skills (see at right) More on-task behavior Improved discipline	Increased skills: vocabulary U A P, spelling U A, phonetics Ux, punctuation U, reading comprehension Ux E, reading decoding U, grammar U, written expression U Low achievers more productive Ux	Retention levels not good U Too little diversity Student fatigue U	Concentration, study skills Fewer discipline problems U A More attentive to errors U Better academic self-concept U More enjoyment, enthusiasm U A Work with less supervision, more responsible A U Less time lose, more time on task U A E Rapid students progress faster E	Some lagging, failing mastery tests U Boredom U
Moderate impact					
Carson (L)	Increased achievement Clearer career interests Friendliness Improved self-concept as a learner More internal locus of control	Career knowledge U A C Work on new interests U		Achievement composite (use of resources) E Better classroom attitude U Attitude to school E U Self-concept as learner (HS) F Friendliness E Social studies (elem) Self-understanding E	Little effect on achievement U
Calston (E)	Basic reading skills, plus use of reading as tool for learning: literature appreciation: reading for enjoyment	Increased criterion referenced test scores A U More reading skills Ux		Learning to work independently self-managing U A Self-motivation A: children like independent work U Children proud to come to school A Children not competing A Reduced discipline problems A	
Moderate to low impact					
Dun Hollow (L)	Eliminate stereotyping of Eskimos Knowledge of Eskimo history, customs, current life Create positive image of Eskimo culture	Some learning of culture, reduction of stereotypes U A More reading skills Ux	Information too detailed U		Some students behind in language arts U
Low impact					
Burton (E)	Knowledge & practical skills re political/ government/legal processes (voter education, state government, individual rights)	Concept learning, by being in different roles U Experience in active learning approach A	No effects discernible U		

Legend for sources of data: U = User, A = administrator, C = counselor; P = parent, E = evaluator, S = student
Underlined letters refer to mention by at least two persons
x = presence of dissenting or conflicting opinion
(E) externally-developed innovation
(L) locally-developed innovation

Table 7.14
Case-Ordered Descriptive Meta-Matrix, Next-Step Version (excerpt):
Program Objectives and Student Impact

SITES	Achievement of objectives: no.	Achievement of objectives: rating	Comments	OVER-ALL RATINGS OF OUTCOMES #: Direct outcomes: Program-immediate	Direct outcomes: Transfer	Other outcomes: Meta	Other outcomes: Negative
High impact							
Perry-Parkdale * (E)	1	◕	math only	●	◕	●	◕
	2	○	fuzzy objective				
	3	◕					
	4	●					
	5	●					
	6	●					
	7	◕	less well for girls				
	8	●					
	9	◔	indirect evidence for some aspects				
	10	●					
	11	○					
	12	◔	only 1 respondent				
Masepa * (E)	1	●	low retention for a few	●	●	●	◕
	2	●	boredom for a few				
	3	●					
	4	●					
Moderate impact							
Carson * (L)	1	◔	1 respondent, some counter-indices	◕	◕	●	○
	2	●					
	3	●					
	4	◕	high school only				
	5	●					
Calston * (E)	1	●		●	○	●	○
	2	◔	conflicting assessments				
	3	○					
Moderate/low impact							
Dun Hollow (L)	1	◕	reduction, not elimination	◕/◔	○	○	◕
	2	◔	1 respondent, global claim				
	3	○					
Low impact							
Burton (E)	1	○	reduction, not elimination	○	N/A	○	○
	2	○	1 respondent, global claim				

* robust evidence (external evaluator and/or standardized test scores)

● strongly present
◕ partly present
◔ weakly present
○ absent

\# uses modal rating of achievement within site's objectives of this type

N/A not applicable

(E) externally developed innovation
(L) locally developed innovation

alized outcome. Note that "transfer" ratings are never higher than "program-immediate" ratings in any of the cases—another indication that they are harder to attain.

As often in qualitative analysis, this distinction came clear only as the analyst was trying to boil down Table 7.13 and noticed not only different content *types* of program objectives but also different *levels*, and that any rating system would have to take that into account. This example illustrates how next-step matrices can beget more inclusive, conceptually crisper findings as you work through successive iterations.

This next-step chart also surfaces something that was less clear in Table 7.13: Should the analyst place a case higher in the ordering when its project had many goals but uneven achievement (like Perry-Parkdale) than one with fewer goals and more consistent achievement (like Masepa)? The weighting can go either way, but the more focused and boiled-down display forces you to decide, and to make the conceptual basis of the decision explicit.

The decision rules for data transformation and entry also must be explicit. How did we get from Table 7.13 to Table 7.14? The decision rules were as follows: Include as "strongly present" only those assessments of outcome made by at least two respondents or by a formal evaluator and containing no strong qualifications or contradictions. Ratings of "absent" mean that nothing in the "outcome" cells of Table 7.13 corresponds to the program objective, or that there were no "other" outcomes (meta, negative). Intermediate ratings must have been justified in the "comments" column of Table 7.14.

Variations

Variations depend on the basic data structure. A first-cut case-ordered descriptive matrix can be big. We sometimes have used up to 20 columns in trying to sort out a wide range of variables that might bear on the main "ordering" variable. Next-step, boiled-down, redesigned matrices are nearly always needed for good analysis.

Don't assume that the readers of the final report should automatically see the first monster-dog. They should see *some* display—not just be presented with your conclusions—but one that makes sense and permits them to assess the credibility of what you are saying.

The matrix in our example could be improved by sorting detailed types of outcomes into sub-rows for each column (e.g., achievement, general knowledge, skills, attitudes, behavior). That would enable cleaner conclusion drawing about types of outcome (and would give more information than the originally proposed Table 7.11).

For less complex databases, it often helps to enter actual direct quotes from case reports, rather than constructed sentences or phrases. They give an even more direct grounding in the data for the analyst and the reader.

Advice

A case-ordered descriptive matrix is usually a fundamental next step in understanding what's going on across cases. Assuming that you have a good basis for ordering the cases, it's far more powerful than a partially ordered matrix: Patterns can be seen for high, medium, and low cases, and the beginnings of explanations can emerge. "Aha" experiences are likely.

The matrix forces you to be grounded in the case data and does not allow escape into vagueness or premature abstraction. Once the matrix is formed, it provides a good check on whether the case ordering is correct.

Expect to make several cuts at the format of the matrix before you settle on the one best for your purposes. The columns have to follow directly, of course, from the form that case-level data are in, but there will be many ways of displaying the data—each, as we have said, with trade-offs.

Don't move too quickly to a matrix that includes only cell entries such as "adequate," "weak," "confusing," "widespread," or similar judgments. Keep as much raw data visible as you can without running into excessive size or complexity.

Recheck the ordering of cases at several points as you proceed. Ask a colleague to confirm or argue with your ordering. A valid ordering is crucial to the use of this method (see also the method of summed indices just following).

During analysis, first do a general squint. What does the matrix look like—where is it dense, where empty? Draw conclusions and write them out. Then look down particular columns, **comparing/contrasting** what things look like for high, medium, and low cases. It may be useful to look at more than one column at a time to **note relations between variables.**

Keep in mind the need for next-step versions of your first display, especially with a complex matrix.

Time Required

For 12 cases like ours, a descriptive case-ordered matrix of this sort can take a long working day. That includes about 2 hours to read the relevant sections of the case reports and another 4 or 5 hours to create and fill in the matrix. Drawing conclusions and writing text (and, if necessary, reordering the display) are usually rapid—1 or 2 hours. If the data are messy and diverse (as they were in this example), the most difficult portion of the work will be deciding on a reasonable display format; if they are cleaner and simpler, formatting will be easy.

Generally speaking, time spent on adequate case-ordered descriptive matrices is well worth the investment because so much later analysis rests on them.

Ordering Cases Through Summed Indices

How can you be sure that the cases in a matrix are ordered in a valid way? Here we show a systematic method for deciding on the ordering of cases. The analyst wanted

Box 7.2
Summed Indices:
Reported Changes in User Practice and Perception

DIMENSIONS OF CHANGE ARISING FROM IMPLEMENTATION

SITES	Daily routines	Repertoire	Relation-ships	Under-standings	Self-Efficacy	Transfer	Basic Constructs, Attitudes	Magnitude of change
Masepa (E)	X	X	X	X	X	X	X	High
Plummet (L)	X	X	X	X	(X)	N/A	X	High
Banestown (E)	X	X	X	X	X	(X)		Mod
Tindale (L)	X	X		X	X	X		Mod
Carson (L)	X	(X)	X	X	X	N/A		Mod
Perry-Parkdale (E)	X	(X)	X	(X)	(X)	N/A	(X)	Mod
Calston (E)	X	X	(X)	(X)	(X)		(X)	Mod/low
Lido (E)	X	X		(X)			(X)	Mod/low
Astoria (E)	X	X	(X)	X				Low/none
Burton (E)		X						Low/none
Dun Hollow (L)	X-0	X-0					(X)	Low/none
Proville (L)	X-0				N/A	N/A		Low/none

(E) = externally developed innovation
(L) = locally developed innovation
X = change claimed unambiguously by several informants
(X) = change claimed unambiguously by only one informant
X-0 = initial change, then reversion to initial practice
N/A = not appropriate/applicable
Blank = no unambiguous changes cited

to order the 12 cases according to how much change had occurred in users. He identified a series of types of user change and put them in a rough conceptual order (Box 7.2 above) from minimal/trivial/short term in nature (daily routines, repertoire) to substantial (transfer to other tasks, basic constructs and attitudes). The legend shows the meaning of the cell entries.

To order the cases, the analyst began with a rough estimated ordering, entered the data for each case, and then rearranged the rows (and sometimes columns) until the systematic order shown appeared, with Masepa at the top (all indices present and verified by several informants) to Proville (one weak change present, then reversion). It's similar to Guttman scaling (see Robson, 1993, pp. 261-264).

The final ratings given each case are shown at the right. Note the "cut points" (e.g., between "mod" and "mod/low") the analyst chose to define the various ratings.

The analyst naturally wondered whether sites with locally developed innovations (marked with L) would have more user change than those using externally developed (E) programs. It is clear that the answer is no (tactic: **making an if-then test**) because Ls and Es are spread all the way through the ordering of cases.

This method is time-consuming but very thorough and explicit in ordering cases; it's good when much depends on the validity of the case ordering.

Two-Variable Case-Ordered Matrix

The case-ordered matrices we have been looking at focus on one variable. When you're interested in two major variables, the form can be adapted easily. If you can order the cases carefully on a well-known variable, then the columns of the matrix can include several aspects of a less well-known variable. In Box 7.3 the well-known variable is users' *practice stabilization.*

The analyst, ordering the 12 cases by that variable, wanted to know how it was related to several aspects of *local continuation* of the practice, such as "users' attitudes," "likelihood of continued use," and so on.

This display proved very rich. Try your hand at drawing some conclusions from it. In our study the analyst wrote five pages of text from it, including these excerpts:

> Were the stabilized sites more likely to continue than the poorly stabilized ones? Yes, at the extremes, but not at all in the middle range (tactic: **noting relations between variables**).
>
> Column 2 tells us that in 7 of the 12 cases, users wanted—mostly or unequivocally—to continue (tactic: **counting**). But at only 4 of these sites were their wishes likely to be fulfilled. So positive user attitudes may enhance, but do not deliver, a high likelihood of continuation (tactic: **noting relations between variables**). On the other hand, users *not* liking the innovation looks like a slightly better predictor. Still, it seems that teacher preferences are not decisive, at least when it comes to insuring continued use at a site.
>
> In the meatiest column, 4, the prime factors are listed in roughly estimated order of magnitude. They come directly from users' responses, pooled. Looking at *uncertain* continuation sites, it seems that we can expect less continuation when most of the people doing it or administering it do not like it, when other key actors do not support it, or when there is heavy external turbulence. If we look for one overarching contributor, it would clearly be lack of sustained *administrator support,* as seen explicitly, and through the medium of budget cuts (tactic: **factoring**).
>
> At the five sites where continuation was highly likely, there was also explicit evidence of administrative support: administrative fiat and administrative codification. . . . So we have direct or inferred user and administrative endorsement *together* with administrative fiat—an unbeatable combination of muscle and commitment (tactics: **using extreme cases, noting relations between variables**).

This analysis also suggested some next steps: a need to "unpack" the various causes of stabilization more clearly and to analyze the dynamics of the institutionalization process—the next issue beyond simple "continuation."

This method is very useful in studying the relationships between two variables thought to be associated, but where the direction of causality is unknown or ambiguous.

For another example of a two-variable case-ordered matrix, see Box 7.4, drawn from Eisenhardt (1989b). The researcher first used a construct table, as in section C, to sort the eight microcomputer firms she was studying according to the speed of making strategic decisions, from the firm Zap (fastest) to Presidential (slowest). The less-understood variable here was number of decision *alternatives* considered. Many researchers have believed that more alternatives mean slower decision making. Was that so?

This two-variable display lets us see some interesting things: Not only do fast-deciding firms explore more alternatives, but they also do so simultaneously. The slower firms explore fewer alternatives and do so sequentially, exploring a new alternative only when the previous one is no longer feasible. The display let Eisenhardt build a better theory: Multiple, simultaneously considered alternatives make for *faster* decisions because they (a) build confidence through comparison, (b) reduce the escalation of commitment to any one option, (c) provide quickly available fallback positions, and (d) encourage broad rather than deep analysis. Now we are far beyond the conventional wisdom that multiple alternatives slow things down.

Contrast Table

When you're trying to understand the meaning of a general variable, perhaps an important outcome for a study, and how it plays itself out across different cases, the contrast table is a useful exploratory device. You take a few "exemplary" cases where the variable is present in high or low form, and contrast several attributes of the basic variable (Box 7.5). Here the analyst tried to "unpack" the idea of how users changed as a result of trying out an innovation, choosing three representative cases and pulling out six relevant aspects of user change, such as "energy investment," that he had noticed in reading case reports.

The exercise aids conceptualization of what user change is composed of and how it works. For example, it looks as if the "negative" case was originally more like the high-change case and then regressed (do **comparisons** across rows to see this).

Note: You have to read through all or most cases before you know which cases are "exemplars." This process also helps you locate the attributes.

Contrast tables often point toward useful variables for a predictor-outcome matrix (see Chapter 8, section C). For example, in this chart, high "pervasiveness of change"

Box 7.3

Two-Variable Case-Ordered Matrix:
Relationships Between User Practice Stabilization and Local Continuation

Extent of practice stabilization/ Cases	1. End of project year	2. Users' attitudes to continuation*	3. Likelihood of continued use* (Same level or better)	4. Prime factors contributing to high/low likelihood of continuation
High Stabilization				
ASTORIA(E)	1	Positive	High	Project mandated Heavy local transformation for good fit
TINDALE (L)	3	Mostly positive	High	Local mandate well enforced Procedures codified User satisfaction
Moderate-High Stabilization				
CALSTON (E)	2	Mixed*	Low	Budget crisis-staff cuts, reassignments
PERRY-PARKDALE (E)	3	Mostly positive	Uncertain	Staff turnover Uncertain funding
LIDO (E)	4	Mixed+	Uncertain	Lower administrative support Lower priority (new facility now available) Users' discouragement
Moderate Stabilization				
BURTON (E)	1	Positive	High	Parts of project written into curriculum Heavy local transformation, good user fit
BANESTOWN (E)	2	Positive	Uncertain	Budget crisis Staff reduced, reassigned
MASEPA (E)	3	Mixed	High	Project mandated Strong logistical support Improved pupil performance
CARSON (L)	3	Mostly positive	High	Procedures codified, routinized Project mandated Widespread local support
PLUMMET (L)	4	Positive	Uncertain	Likely staff turnover Lower district support
Low Stabilization				
DUN HOLLOW (L)	3	Negative	Low	User + principal dissatisfaction No strong local advocate
PROVILLE (L)	4	Negative	Nil	Other central office priorities; no advocate Project discontinued User and principal dissatisfaction

* Researcher assessment, usually pooled from interview data and case report tables
\+ Some wanting to continue, others not

(E) = externally developed innovation
(L) = locally developed innovation

Box 7.4
Two-Variable Case-Ordered Matrix:
Speed of Decision Making and Alternatives Considered (Eisenhardt, 1989b)

Firm	*Decision*	*Number of Alternatives*	*Alternatives*	*Timing*
Zap	Alliance	4	Alliance Public offering Bank loans Venture capital	Simultaneous
Forefront	New product	3	New product Extension of existing product Status quo	Simultaneous
Promise	Strategy	3	Status quo Major strategic shift into new markets and products Minor strategic shift to capitalize on sales opportunities	Simultaneous
Triumph	Strategy	4	Refine current strategy Sell firm's technology Liquidate firm Major strategic shift	Simultaneous
	New product	2	Low-end product Moderate to high-end product	Simultaneous
Omicron	Strategy	2	Major strategic shift Better management of sales and manufacturing	Simultaneous
	Strategy	2	Major strategic shift in distribution Major strategic shift in product and market	Simultaneous
Neutron	Alliance	2	In-house development Allliance	Sequential
Alpha	New product	2	IBM-compatible product Interface product	Sequential
Presidential	New product	2	VLSI* product with U.S. partner Licensed product with Japanese partner	Sequential

* VLSI = very large scale integrated circuit.

might be causing high amounts of user change. Finally, contrast tables are a useful teaching device for readers of the final report.

Scatterplot

Analysis Problem

Many of the cross-case displays discussed up to now involve showing all of the cases on common dimensions so you can see how the cases line up. Scanning the data in matrix form can tell you something about how they should be further partitioned and clustered. But the matrix logic is one of creating categories or cutoff points, and then seeing into which bin the cases fall.

So matrices throw away much useful information on how close or far apart the cases are on dimensions of interest—how they cluster, and how clusters relate to each other. We need something a bit more *spatial*.

It's possible to plot each of the cases on two or more axes to see exactly how close to or distant from one another they are. This process is close to a scatterplot logic, as used to make sense of correlation coefficients, or to a "vector"

Box 7.5
Contrast Table: Exemplary Cases Showing Different Degrees of User Change

ASPECTS OF USER CHANGE	MASEPA HIGH CHANGE	BURTON LOW CHANGE	DUN HOLLOW NEGATIVE CHANGE
1. Start-up discrepancy from usual practice	High discrepancy	Low discrepancy	Moderate discrepancy
2. Pervasiveness of change	High—all facets	Low—repertoire only	Low/moderate—routines and attitudes
3. Technical mastery	Slow in coming	Rapid	Slow, then rapid
4. Energy investment	Very high	Low	High, then low
5. Negative changes reported	Some	None	Many
6. Stretched—pushed beyond voluntary change	Yes—well beyond	No	Yes—at the start

logic as used in statistical factor analysis. How can qualitative data be displayed in analogous ways?

Brief Description

Scatterplots are figures that display data from all cases on two or more dimensions of interest that you think are related to each other. Data from cases are scaled carefully, and the cases positioned in the space formed by respective "axes," so that similarities and contrasts among the cases can be seen. The principle is bivariate (or multivariate) analysis of relationships in coordinate-defined space.

Illustration

In the school improvement study, we were struck by how much users "bent" the innovation during project execution. Few projects came out looking like the developer's model. Gradually we saw that most of these alterations were made with the explicit or implicit approval of local administrators, who granted "latitude" to users to make appropriate changes.

Because, in most cases, administrators had leaned hard on users to adopt the new practices, we wondered whether some kind of implicit bargain was at work here. Perhaps administrators were allowing users to modify the project locally as a palliative to their feelings about being pressured to adopt. If that hypothesis were borne out, a clear relationship should be found between levels of pressure and levels of latitude (tactic: **making if-then tests**).

This relationship had not been probed in each of the case studies, but only began to emerge during cross-case analysis. So the analyst had to form comparable overall judgments of "pressure" and "latitude" in the original case-level data, and then find a way of arraying them that tested the relationship.

Building the display and entering data. The analyst scaled each case on a "pressure to adopt" continuum and on a "latitude" continuum, then created two axes and plotted each case on them (Figure 7.7). As he went back to the case reports to do this, he noted that, for some cases, latitude or pressure differed between early and later implementation. That finding could be accommodated on the display by showing two values for such a case, with an arrow indicating the direction of change.

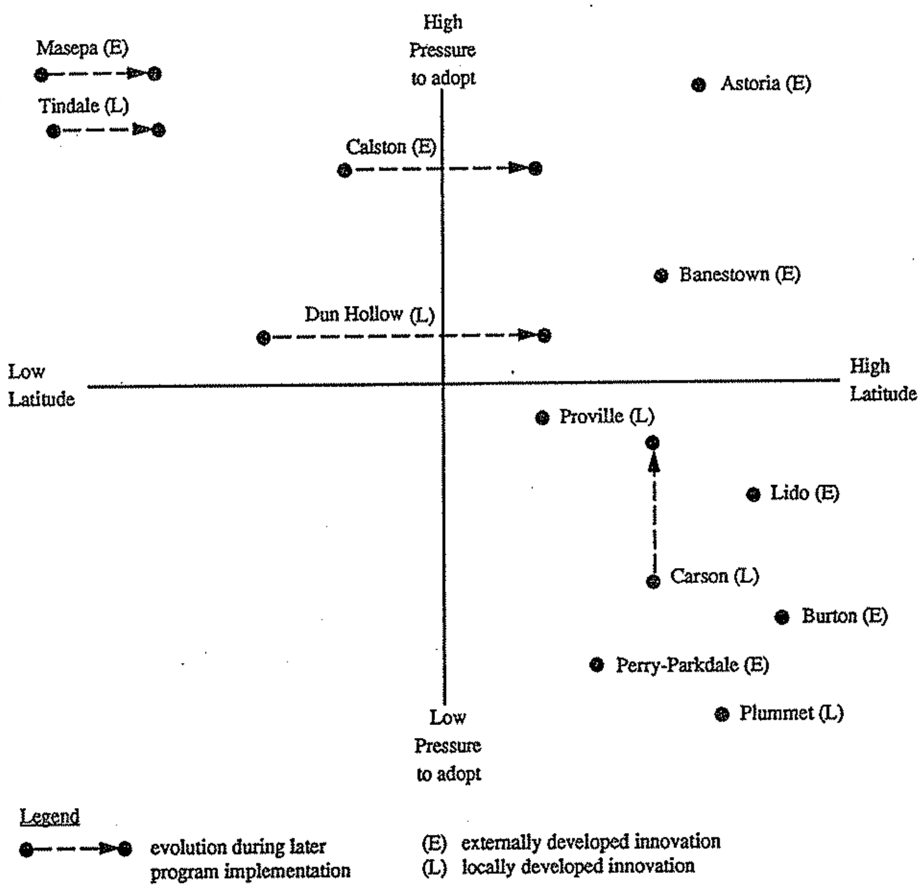
Drawing conclusions. (Try your hand first.) It looks as if the hypothesis does not pan out; if it had, most of the cases would have clustered in the upper right quadrant (high pressure, high latitude), and we have only a single case there (Astoria). In fact, it looks as if the opposite hypothesis is being borne out (tactic: **noting relations between variables**). Administrators who pressure people to adopt keep up the pressure in the form of low latitude; where adoption pressure is weaker, users get more latitude.

Note, however, the interesting drift in four of the six high-pressure cases (tactic: **noting patterns, themes**): Latitude was slightly higher during later implementation. So maybe once the innovation has started in, administrators relent a bit, and there is something to our bargaining hypothesis after all. But the negotiation comes later than we had thought.

Going back to the case reports helped clarify things further. For example, the analyst found that only in Astoria (tactic: **checking the meaning of outliers**) was an explicit initial bargain with payoffs struck between administrators and users. And plenty of data supported an administrative message like this one to users, notably in Dun Hollow, Calston, Masepa, and Tindale: "You're going to do this. Start out doing it like the developer wants. Later, you can make changes."

And the moderate pressure at Proville was accompanied by a central office bargaining stance toward principals that

Figure 7.7
Scatterplot: Relationship of Pressure to Adopt and Degree of Latitude Given to Users at 12 Sites



sounded like this: "We want to try this, and we want at least token collaboration from you."

Variations

Displays like Figure 7.7 can carry still more freight. They can easily accommodate more cases. Case data also can be differentiated (e.g., by using separate symbols for measures of administrators and users in each case). With some care (better, the services of a good graphics program) it is possible to plot three dimensions, but graphic clarity is important; otherwise interpretations will be difficult. The scatterplot can also be simplified back down to a contingency table format. See Box 7.6 for an illustration.

Advice

Scatterplots are very useful when you are in an exploratory mode, or are trying out a hypothesis, and need a way to "see" all of the cases in two-dimensional space. Plotting the cases spatially is also a good way to make more precise determinations about which cases form clusters.

Because figures like these call for a careful "scale score," rather than an impressionistic rating, you have to do finer-grained work than the case-level researcher probably did; you may need a review of the appropriate section in the case report. Techniques such as summed indices (Box 7.2) can be used. If careful scaling is not possible, and if cases cannot be precisely distinguished from each other on a variable, then a contingency table is better.

As usual, do not settle for conclusions drawn from the display alone. You need to clarify and confirm (or qualify) the conclusion through reference back to case reports.

Finally watch out for mistaken causal attributions. Maybe variable B is causing A, rather than the opposite you were assuming. Or a third variable is fooling you.

Time Required

Figures such as these are relatively easy to construct and fill in. If you have data in an easily scalable form, they can be scaled and entered for a dozen cases in an hour or so. If not, scaling can be time-consuming (as in Box 7.2). But if the variables of concern have already been scaled or categorized for some other analysis, then we are talking about 10-15 minutes for a procedure that is usually quite illuminating.

Drawing conclusions and taking a confirmatory look at case reports and writing text may occupy another hour. More cases, more time, naturally.

E. Time-Ordered Displays

As we've said, life is chronology, and it makes a lot of sense to create cross-case displays that take account of temporal sequences.

Time-Ordered Meta-Matrix

Analysis Problem

In comparing a number of cases, you often want to know about events that have occurred over a period of time, especially those events that are indicators of some underlying process or flow. How can data from several cases be displayed in a way that respects chronology?

Brief Description

A *time-ordered meta-matrix* has columns organized sequentially by time period; the rows need not be ordered, but can have the cases in arbitrary (perhaps alphabetical) order. Again, the basic principle is chronology.

Illustration

In our school improvement study, we came to be much interested in the question of *job mobility*—how the key people who were connected with an innovation changed jobs during the implementation period. Our look at adoption processes had already shown us that many people (up to 40%) liked the idea of adopting the innovation because it "might" help their job prospects. The question is, What actually happened afterward? We suspected (see the memo on career patterns in Chapter 4, section D) that people involved with an innovation might do one of several things: strengthen their status in their existing job; move into the innovative project when they had been out of it; stay in the project but move to a new job across (same status level), up, or down; or move out of the project—across, up, or down. Or they might actually move out of education altogether. And we were interested in attributed causality: Was a person's move really stimulated or induced by experience with the innovation, or would he or she have moved anyway?

Table 7.15
Time-Ordered Meta-Matrix (preliminary format)

SITES	Time 1 (early)	Time 2 (during implementation)	Time 3 (later)
Site A			
Site B			
etc.			

Building the display. How could you format a useful display to answer these questions? Table 7.15 presents a prototype setup. In each cell, you would enter examples of job mobility that had occurred in that case at that time period, with a descriptor showing what kind of mobility it was (in/out, up/down/across, etc.). Looking across rows would show the total range of job mobility examples at specific sites; looking down columns would show mobility in particular time periods.

But the display is too simpleminded. First, it does not show whose mobility we are talking about—everyone at the site, or only key people? Second, it does not include anything about attributed causality—whether the job changes were said to have occurred because of the innovative project, or would have occurred for that person anyway. Third, the time periods are defined only vaguely.

Thinking about these issues led us to the display shown in Table 7.16. It focuses on key people (with the rationale that such people are more likely to show innovation-caused effects, if they exist) and defines their number specifically, so that we can see how many of them moved. It also defines the three time periods with a key phrase (e.g., "creation/initiation of project"). And it has a column for a judgment about whether the job mobility for each person was innovation-caused or not. The last column requires an overall judgment of the amount of mobility in the case as a whole.

Table 7.16
Time-Ordered Meta-Matrix: Job Mobility at Sites

SITES	No. key actors involved+	T_1 - Creation/ initiation of project	Innovation-related?	T_2 - Early implementation (1st year)	Innovation related?	T_3 - Later implementation	Innovation-related?	Overall mobility rating for site*
Astoria (E)	5	1 cent. office (C.O.) admin. strengthens new role	√	1 bldg. admin. moves up 1 C.O. admin. moves out and over	√ √	N/A	N/A	Moderate
Banes-town (E)	6	1 aide moves in & up to teacher status 1 teacher moves back in 1 teacher prepares for move up 1 C.O. admin. positioned for move up	√√ √√ √√ √√			1 teacher moved back down 1 teacher moved over to desirable post	0 0	High
Burton (E)	9	1 C.O. admin. strengthens role	√			N/A	N/A	Nil
Calston (E)	7	1 teacher prepares for move up 1 C.O. admin. extends authority 1 bldg. admin. strengthens position	√√ √ √			1 C.O. admin. moved out and down 1 teacher (librarian) moves out and over 1 teacher in process of moving out	0 √ √	Moderate
Lido (E)	5					1 bldg. admin. moves out and up	√	Low
Masepa (E)	13			1 teacher moves up to supervisory role	√√	1 teacher moves up to supervisory role 1 C.O. admin. retires 1 proj. admin. moves out	√√ 0 0	Low/ moderate
Perry-Parkdale (E)	10	3 teachers move in from less desirable jobs 1 teacher moves in and up to admin. post	√√ √√	1 C.O. admin. moves out & (tries to go) up	√	1 teacher moves out & over 1 teacher moves out & back down 1 proj. admin. moves out and up 2 bldg. admins. move out, 1 moves up	√√ 0 √√ 0	High
Carson (L)	27	2-3 teachers move out 2-3 teachers move in 1 bldg. admin. strengthens role	√√ √√ √√	1 teacher moves out 1 bldg. admin. moves out	0 0	1 teacher (prog. coord.) moves out 1 bldg. admin. moves out and over 1 bldg. admin. moves up 1 teacher. moves in & up to prog. coord. role, then moves out & up	0 √ 0 √√	Moderate
Dun Hollow (L)	7	1 regional admin. extends role	√√			N/A	N/A	Nil
Plummet (L)	6	1 teacher moves in and up to admin. role 5 teachers move in	√√ √√	4 teachers move up to supervisory role	√√	1 bldg. admin. moves out 1 teacher moves out	√√ √√	Moderate/ high
Proville (L)	14	1 teacher moves in and up to admin. post	√√	1 C.O. admin. moves up 1 tchr moves in and up to admin. post	√√ √√	1 C.O. admin. moves out and up 1 teacher moves in and up to proj. admin. role	√√ √√	High
Tindale (L)	14			1 teacher moves up to bldg. admin. post	√√	1 bldg. admin. retires 1 C.O. admin. moves out and up 3 bldg. admins. move up	0 √√ √√	High

+ People in key roles connected to the project for whom job mobility was possible

* Researcher estimate based on proportion of actual job moves to total of key actors, with higher weighting for moves by people immediately responsible for, or crucial to, the innovation (ex: project director).

Legend
(E) = externally developed innovation
(L) = locally developed innovation
√√ = Yes, clearly
√ = possibly, in part
0 = no
N/A = not applicable

The cases are not ordered, but we did want to compare the possible job mobility in two different program types (externally developed and locally developed), so we separated them.

Entering the data. To fill this matrix, we read through the relevant sections of the case reports, first defining people in "key" positions (advocates of the innovation, supporters, project leaders, principals, active users) for whom job movement was *possible.* Then we located all instances of job change (including, if present, clear instances of "job strengthening"). For each instance, we entered the old job title, and characterized the nature of the change (in/out; up/down/across). We made a judgment, based on the case report data, about whether the move was *clearly* innovation related, possibly so, or not at all.

After all data were entered for all cases, we could look across each row and make a general rating of the overall amount of job mobility for that case. The easiest way was to look at the proportion of moves to the total possible (decision rule in this case: 0-35% = low; 36-70% = moderate; 71-100% = high). Extra weight was given if unusually central "key people" (e.g., project directors, very active principals) were the ones moving. We made provisional ratings for all cases and then reviewed them again to see whether they still seemed correct.

Drawing conclusions. A simple first step is **counting;** we can see 63 moves, of which 52, or 83%, are innovation related. The 63 moves occurred in an eligible population of 123; thus a rather substantial 51% of key people who could move did so.

Mobility obviously differs by time period, as we look down different columns (tactic: **making contrasts, comparisons**). Job moves are highest at the very outset, when the project creates new roles and teachers move into administrative/supervisory slots. We can also see that moves are frequent during later implementation, when administrators move up or out, and teachers associated with the project take their places.

By tabulating the direction of moves, we can see that 12 teachers moved up, as did 10 administrators, totaling 35% of the innovation-related job shifts. So innovation is not an automatic ticket to advancement.

We can explore the 11 people whose mobility was *not* innovation related (tactic: **using extreme cases**); here the analyst has to go back to the case reports to find out what happened. It developed that people were either already planning to leave or retire, or that sudden environmental turbulence (e.g., massive budget cuts) ejected people from their jobs.

We can also see that locally developed projects, except for Dun Hollow, have substantial mobility (37 moves in 5 sites), while externally developed projects range more widely and average lower (29 moves in 7 sites). A step back to the case reports showed that most locally developed projects tended to require new coordinative and supervisory roles. Also 3 of the 5 were relatively large scale projects, whereas externally developed projects tended to be "add-ons" or "drop-ins" with fewer demands.

Variations

This matrix could be differentiated with subcolumns for different roles—teachers, project administrators, building and central office staff—which might make for easier analysis.

Of course, time-ordered meta-matrices can be designed for a wide range of events, in addition to the one focused on here (innovation-connected job moves). In the general area of innovation in organizations (not only schools), other events include assistance efforts (e.g., workshops, training); modifications in the innovation; extension of the innovation to new users; changes in school structure or procedures; and specific managerial interventions to aid implementation. It is only important that (a) a reasonably clear class of events is specified, related to some general variable of interest, and (b) the available data identify the period when the event occurred.

We should not be too harsh about defining "events" as the only candidates for such a display. In principle, you can also do a time-ordered meta-matrix on "states," more fuzzily defined. For example, we might want to show the states of innovation users' feelings and concerns (confusion, overload, anxiety, satisfaction, pride, commitment, etc.) when they first encountered the innovation, and then during early use and during later, more stabilized use. If the times of such states can be determined validly, then good analysis is possible.

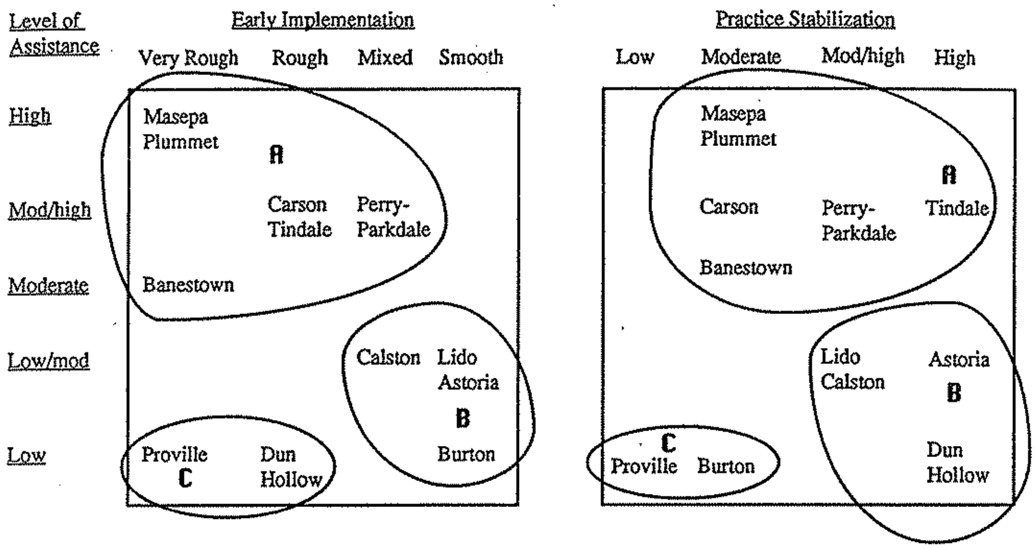
We have stressed ordering of the matrix by time, but no law says such a matrix could not be case-ordered as well. If so, you could see more clearly what the patterns of movement were in high-mobility cases and how they differed in low-mobility cases.

Advice

This sort of display is easy to do, and works best when you want to examine a reasonably specific class of time-related events (or states) across cases. For looking at the general flow of events within a case, see the methods of event listing and event-state networks (Chapter 5, section C).

Specify clearly the events (or states) being examined, so that good comparability exists across cases. Be sure you have identified sensible and comparable time periods across cases, and that you can locate unambiguously the particular event or state in a particular time period.

Box 7.6
Scatterplot Over Time: The Consequences of Assistance



Use content-analytic summary tables (Table 7.9) to pull together and look at data from the display if it is complicated, or if the totals in a row or column are not easily visible.

Time Required

The display illustrated here took about 3 hours for reading the case reports and making a rough-cut matrix, then another 3 hours for entering the data in a refined matrix. Conclusion drawing and writing occupied another hour, though more could have been spent (in this case the analyst was eager to move on to understanding the causes of mobility and could not wait to dig into them more deeply). Generally speaking, if you spend adequate time in specifying the events and developing codes (like up/down/across, in this case), the analysis can go quickly; if the events are defined blurrily, extracting meaning will be much more time-consuming.

Scatterplots Over Time

Scatterplots can be useful when they display similar variables in cases over two or more time periods. In this example, the analyst was trying to find out whether the amount of assistance given to people executing an innovation aided (a) early implementation smoothness and (b) later "stabilization" of the practice. He was looking for a display that would line up all 12 cases on these three dimensions. The best bet looked like Box 7.6. The analyst had already sorted cases into the categories noted, so plotting was quick. Once again, try your hand at drawing and writing some conclusions.

We can see at a glance that the sheer amount of assistance does not predict early smoothness or later stabilization. But there are families of cases (tactic: **clustering**). The analyst drew lines around what looked like reasonable families and then thought about what they meant.

In the first time period, Family A was cases with high assistance and rough implementation, while B was cases with low assistance and smooth implementation. Why should the relationship go that way—seemingly backward from expectations? The analyst suddenly realized that a third factor was in play (tactic: **finding intervening variables**). The Family A cases were all attempting large innovations (see Box 8.3). So, he reasoned, even large amounts of early assistance are unable to achieve smoothness if the innovation is big. The Family B cases do not need much assistance because their innovations are small.

Family C is a different kettle of fish: Their innovations are large but enjoyed little assistance. The analyst labeled this family "assistance incompetence."

Turning to the scatterplot in the second time period, the analyst noted that Family A stayed the same: High assistance eventually paid off in moderate to high stabilization for large-scale innovations. Family B, with almost the same membership as before, achieved, with minimal assistance, good stabilization for its innovations. The weak assistance in Family C probably hindered eventual stabilization.

Note that you always need to cycle back to the case reports. For example, why did the weakly assisted, large-scale innovation stabilize in Dun Hollow (tactic: **checking the meaning of outliers**)? It developed that teachers themselves helped each other, and did considerable rewriting and adaptation of the curriculum materials. We see here that anomalies and deviant cases always need careful follow-up in the case-level data. They must be accounted for satisfactorily, and usually strengthen conclusions. (The original meaning of "the exception proves the rule" is actually "the exception *tests* the rule.")

Composite Sequence Analysis

How can you extract typical "stories," "plots," or "scenarios" that a number of cases share—without forcing them or destroying meaningful sequences?

This illustration comes from a life-cycle study of 160 teachers (Huberman, 1989, 1993) that we've already mentioned. Huberman was interested in what he called career "trajectories" of teachers. He was working from 5-hour interviews with secondary-level teachers, all of whom had divided their careers into phases or stages. They themselves provided the "names" or "themes" for each phase, rather than responding to a stimulus list.

First, the material for each case was condensed onto a 25-page protocol, with an identical format for each informant. For each person's career trajectory, the sequence of themes was mapped, with the "features" of each theme added in. An initial theme labeled "drowning," for example, was often characterized by features such as difficulties with keeping discipline, exhaustion, unpredictability, feelings of professional inadequacy, overload, and the like.

How could you begin to group the cases, phase by phase, to see any "families" of trajectories? The first step was to group cases *within* phases.

The decision rules for grouping cases within a given phase were (a) the theme itself had to be denotatively or connotatively identical and (b) at least two "features" had to be the same. For example, one informant might name a "stabilizing" phase (see Figure 7.3, repeated on page 205), and another a "settling down" phase, while both teachers were at the same point in their sequence and each used two of the same features to describe that phase (e.g., gaining tenure, having a basic instructional repertoire).

Graphically, this procedure amounts to an "overlay" of individual trajectories, although it is clear that seen up close, each career is qualitatively distinct. But we do get, as Wittgenstein put it so well, "family resemblances," with some intersections, along with the distinctive properties.

Let's illustrate these intersections by a look at Figure 7.3. It depicts 11 teachers, one of 16 subgroups; here we have women with 5-10 years of teaching experience at the lower secondary level.

At the left is the range of years of teaching experience within phases, and a series of general phases defined by the researchers (P1, P2, P3) in light of the phase names given by informants (e.g., *Launching a career*). Within each general phase are subgroups. For example, of the 11 women, 5 describe their initial phase in terms of "easy beginnings" (tactic: **clustering**). The common features are noted under the theme. Six women evoke "painful beginnings," again with a list of features, and 4 mention external problems (parallel work at the university and teacher training; demands of family and private life). Through experimentation and trials and errors, these problems are attenuated for 3 women.

The second phase, *stabilization,* is common to nearly all. It is, in effect, one of those archetypical stages in the teaching career. Note, however, that 3 of the original cohort of 11 have not followed the sequence, but rather have taken another path (not shown on the chart). In that sense, the sequential analysis gets progressively more stringent. It groups only informants who have attached identical or analogous themes to the same sequence of phases.

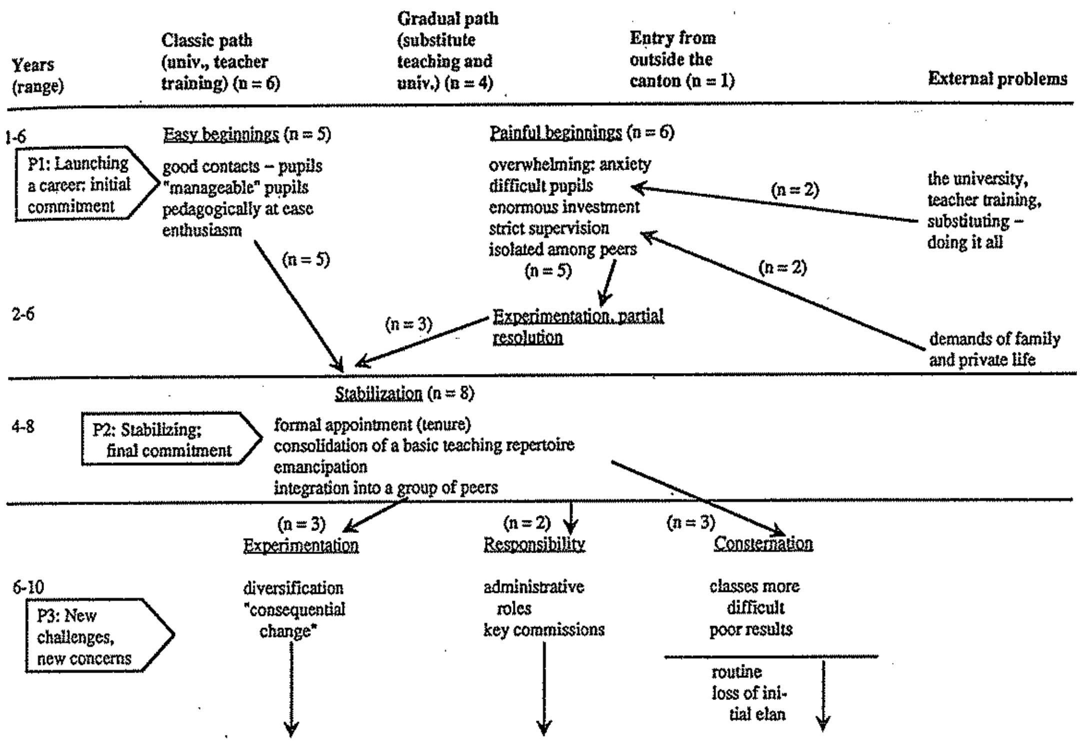
Figure 7.3 shows a third phase for this cohort, *new challenges,* in which the paths diverge still more: "experimentation" (3 teachers), "responsibility" (2 teachers), and "consternation" (3 teachers). Once again, these thematic names are provided by the informants.

A network like this one is already condensed to its qualitative limits. You might sensibly stop here, so as not to distort or simplify the texture and order of individual cases (Abbott, 1992b). In this study, however, the researchers were taking a variable-oriented approach, while still trying for a narrative ordering of the data. They were hunting for correspondences across the full set of 160 informants. By reviewing 16 charts like the one just shown, they came out with four modal sequences. Here is the first one, called "Recovered harmony":

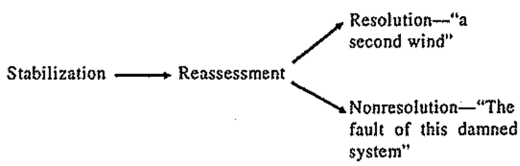
Painful beginnings ⟶ Stabilization ⟶ Experimentation

This scenario accounts for 17% of the population. Recall, too, that it was a prime sequence in the chart shown

Figure 7.3 (Repeated)
Composite Sequence Analysis: Career Trajectory Data for 11 Cases (Huberman, 1989)



above. Let's look at one more sequence, which the researchers called "Reassessment:"



Here we have 25% of the sample, by no means the majority, but a strong cluster: a clutch of people describing their career progression in these terms. Taken together, the four scenarios located account for 65% of the sample. Much has been lost in the condensation of 25 pages of interview data to a bare-bones sequence like this one.

Of course, you can stop the exercise at any point: after the transcriptions, after the 25-page protocol, after the 16 career-sequence charts. We'd argue that, in conceptual terms, it is this last scenario-building exercise that tells us most graphically how the case narratives cluster—in this instance, how the teaching career plays out for different subgroups of its members. And we have successfully linked a case-oriented and a variable-oriented approach, protecting case sequences and gaining the conceptual power that cut-across variables can bring.

This method also shows us more about network displays: They can handle narrative complexity more easily than matrices can.

Summary Comments

Looking across cases deepens our understanding and can increase generalizability. But cross-case analysis is tricky. Simply summarizing superficially across some themes or main variables by itself tells us little. We have to look carefully at the complex configuration of processes within each case, understand the local dynamics, before we can begin to see patterning of variables that transcends particular cases. Combining "process" and "variable" approaches is needed.

We've described *partially ordered* matrices that "stack up" the data from case-level displays, useful in first steps. *Conceptually ordered* displays, both matrices and networks, help in clarifying key variables and how they are patterned in cases. *Case-ordered* displays of both types deepen understanding of how variables and processes work, by sorting cases on key dimensions. *Time-ordered* matrices and networks can produce "generic narratives" that retain the "plot line" of individual cases, but show us principles that cut across them.

In the next chapter we turn again to the issue of explanation: How can cross-case displays help?

Notes

1. One of the best discussions of generalizability issues is Schofield's (1990); she discusses it in terms of not only "what is" (current reality) but also "what may be" (trends, changes) and "what could be" (the ideal, the exceptional).

2. Runkel's (1990) distinction is between "casting nets" and "testing specimens." He calls the former the "method of relative frequencies": Count cases, do statistics, look for clusters of conditions and actions, "locate densities of types of behavior." "Testing specimens" is case-oriented, looking for features of behavior that are invariant within and across cases, understanding how an organism functions in its environment. It requires a careful look at case-internal dynamics, such as feedback loops and "circular causation."

We remind the reader here of Maxwell and Miller's (1992) discussion of "paradigmatic" and "syntagmatic" modes of analysis, which parallels the variable-oriented/case-oriented distinction.

3. McPhee (1990) has a thoughtful discussion of three types of cross-case comparison: (a) when there are common data points across all cases, (b) when there is a similar dependent variable but diverse explanations of it, and (c) when there are radically different, incommensurable images (both outcomes and explanations diverge). He suggests ways of diagnosing cases to plan workable comparisons for each of these situations.

4. There are no clear rules of thumb for the number of cases that can be managed in qualitative analysis without simply regressing to numbers. We believe that once past 15 or so cases, serious gaps in understanding of local dynamics begin appearing. But note that successful cross-case qualitative studies have been carried out with case *N*s of up to 57 (see A. G. Smith & Robbins, 1982). One useful strategy involves first analyses on a representative subsample of cases and then replicating these (see Chapter 10, section B) on the remainder. But we need to learn much more about the possibilities and limitations of cross-case analysis with large numbers. The QCA program's Boolean methods can handle many cases, dealing with case-specific configurations of up to 12 variables at a time (see Appendix).

8

Cross-Case Displays: Ordering and Explaining

We've been discussing approaches to developing coherent descriptions of *what* and *how* things happen, across a number of cases. In this chapter we turn to questions of *why.* Multiple cases are extraordinarily helpful in both generating explanations, and testing them systematically. In a real sense, they are our best resource for advancing our theories about the way the world works.

A. Explaining Through Comparative Analysis

We need not reiterate the ideas about explanation and causality presented in Chapter 6. But we do want to outline what we see as the main issue to be confronted, and a series of important working principles that must be kept in mind when you are seeking to build or strengthen explanations through analysis of multiple cases.

The Main Issue

Given the idea that explanations, including explanations of causality, in human affairs are not monolithic but always involve a complex network of conditions and effects, the key problem is how to draw well-founded conclusions from *multiple* networks. Each network is in some senses unique, a "narrative" of what happened over time. Yet we need to find "generic narrative models," in Abbott's (1992b) terms. We need a theory that explains what is happening—but a theory that does not forcibly smooth the diversity in front of us, but rather uses it fully to develop and test a well-grounded set of explanations.

Working Principles

How to get there? In this field everyone is still learning. But we can suggest some principles that seem well founded in good current practice. They should, we believe, be kept firmly in mind during the explanatory analysis of multiple cases.

Understand the case. It is crucial to have understood the dynamics of each particular case *before* proceeding to cross-case explanations. Without that, superficiality sets in.

Avoid aggregation. Cases cannot simply be idly lumped, summarizing "similarities and differences" on some common variables of interest. Here we can cite Mishler (1986), who, speaking of survey research, notes that when "summary scores [are] aggregated across separate . . . [cases] each response is a fragment removed from the social and psychological context. . . . When these responses are assembled . . . the results are artificial aggregates that have no direct representation in the real world of communities, social institutions, families, or persons" (p. 26).

Preserve case configurations. The network of conditions—causes, effects, outcomes, and their temporal sequence—within each case must be protected during analysis. This is the up side of Mishler's advice. Such networks may need clarification and simplification, but they should not be decomposed into their parts. There will always be multiple causes of multiple effects, some of which become further causes of later effects. We need to see the whole picture.

Combine variable-oriented and case-oriented strategies. Nevertheless, as we've hinted before, good explanations usually involve cycling back and forth between, or synthesizing, strategies aimed at understanding *case* dynamics and at seeing the effect of key *variables.* We remind the reader here of Ragin's (1987) comparative method, which does this; you can analyze multiple cases, using key variables, preserving their configuration case by case.[1] "Story" approaches need to be married to "concept" approaches.

Inquire into deviant cases. The cases that do not fit your emerging explanations are your friends. They surprise you, confront you, and require you to rethink, expand, and revise your theories. They must not be simply cast into outer darkness as "errors," "inapplicable exceptions," or "irrelevancies."

Look for typologies, case families. If within-case configurations are preserved, it will often turn out that some "families" share certain scenarios, or "plots." These typologies are not simplistic, on the basis of a single variable such as localism or cosmopolitanism, but have similar *configurations.* Ragin (1993) gives a simple example from a study of the adoption of universal pensions in 71 countries. He found two basic families—both of which had universal pensions. Family 1 included ethnically diverse countries that had had at least 5 years of Left or Center-Left rule. Family 2 included ethnically homogeneous countries that had had at least 10 years of sustained economic growth. (Other "families" of countries did not have universal pensions.) Section F below also provides examples of shared "scenarios."

Avoid forcing. The important thing here, as implied above, is not to push your cases too quickly or abruptly into an explanation or into families. The temptation, as McPhee (1990) notes, is to assume that all of your cases will have a standard set of independent variables (causes) and dependent variables (outcomes)—and that all you have to do is treat each case as a fully comparable set of data points (as in the example in Figure 7.1, p. 173). In fact, you may find several routes to the same outcome (as in the Ragin example above). Or you may find that different cases have different routes to *different* outcomes.

We all have our preconceptions, our pet theories about what is happening. The risk is taking them for granted, imposing these willy-nilly, missing the inductive grounding that is needed. We do not hold as strongly as Glaser (1992) that all predefined concepts should be abjured in favor of solely "grounded theory" emergence.[2] As we've suggested, variable-oriented strategies can be useful—if they're well integrated with case-oriented ones.

These principles are naturally abstract. To see how they play out, we need to look at specific methods, with concrete examples. Below we describe five main methods and four supplementary ones. We deal first with the *case-ordered effects matrix* and then the *case-ordered predictor-outcome matrix,* with some variants. Then we describe the *variable-by-variable matrix* and finally turn to *causal models* and *causal networks.* These display names may sound confusing and abstruse. But as we noted before, the labels we give to types of displays are for convenience and clarity, not for reification. Look at what the display does for you in terms of the principles above, not at its name.

B. Case-Ordered Effects Matrix

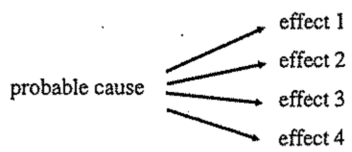
Analysis Problem

Before we come to complex networks of causal conditions and outcomes, let's clarify a simpler version. Causal displays—for single or multiple cases—tend to be implicitly modeled like this:

cause 1 →
cause 2 → outcome
cause 3 →

Another kind of analysis problem is that of understanding the *effects* of a particular variable, seen as a stimulus, an antecedent, or a probable cause. We look at this in Chapter 5 for the single-case situation.

Effects displays, examining the diverse results occurring from a single major variable, are structured like this:



So when you have several cases where an important or salient "cause" is expected to have a variety of results, the question is how to display relevant data, to see how the effects play out—across an array of cases that have a greater (or smaller) amount of the basic cause. Usually you also want to see whether the effects are "sortable" into categories.

Brief Description

A *case-ordered effects matrix* sorts the cases by degrees of the major cause being studied, and shows the diverse effects for each case. The effects may be clustered or categorized to help understanding. Again the focus is on *outcomes,* dependent variables.

Illustration

In our school improvement study, one of the main variables of interest was *assistance* provided to users of an innovation. We knew that assistance might be of various *types* (e.g., special training, supplying resources, giving emotional support); we also knew that some assistance was *event-linked* (e.g., a brief workshop) and that other assistance was *ongoing* (e.g., a principal's troubleshooting assistance given from day to day). And we expected that the assistance might have *short-run effects* (e.g., feeling clearer about the innovation) and *long-run consequences* (e.g., improved skill in using it). The problem here is how to sort out all of these features in a manageable display.

Building the display. A first cut at a possible cross-case display for ongoing assistance could look like Table 8.1. This display could be repeated for "event-linked assistance." Would this be workable? A first step is to read through the relevant sections/displays in the original case reports to see whether the information there can be sorted easily into the proposed display. (We do not advise going back to the originally coded data; those have been processed already into the case report.)

Doing this first step also helps generate a list of possible categories for short-run effects (e.g., "increased understanding," "more motivation," "reduced isolation") and long-run effects (e.g., "getting funds," "increased legitimacy of the program," "routinization of the innovation").

When you've done this step for many/most of the cases, the list of specific effects can be subsumed into fewer,

Table 8.1
Case-Ordered Effects Matrix
(preliminary format)

		EFFECTS OF ONGOING ASSISTANCE		
Cases		Types	Short-run effects	Long-run consequences
high	1.			
assis-	2.			
tance	3.			
to	4.			
low	5.			

more general bins (tactic: **subsuming particulars into the general**—see Chapter 10, section A). In this instance, both short- and long-run effects could be sorted into three general kinds: those effects on the *innovation* and its use, those on *individuals,* and those on the *organization.* So a possible refigured format could look like Table 8.2:

Table 8.2
Case-Ordered Effects Matrix
(version 2 of format)

		SHORT-RUN EFFECTS CASES high..............low					LONG-RUN CONSEQUENCES CASES high..............low				
EFFECTS	on the innovation										
	on users										
	on the organization										

The idea here is to enter specific types of effects (a phrase, usually) in appropriate cells. Returning to the case reports (and/or the lists of effects already generated) will let us see whether this format will work. It depends on how many types of effects there are. If there are only a few, the matrix will not be too unwieldy. But if there are a lot, we will have to do better—especially if we want to include the differentiation between "ongoing" and "event-linked" assistance effects. We can also see that having two submatrices for short- and long-term effects may get confusing.

If we suppose a very wide range of effects, the next iteration of the format could look like Table 8.3. This seems more workable. Remember that a preliminary list of categories already had been developed for the headings. We can show the whole range of detailed effects and enter information in each cell. For each case, we can show that a specific effect did or didn't occur, and possibly even whether it was a long-run or short-run effect. Let's see.

Table 8.3
Case-Ordered Effects Matrix
(version 3 of format)

CASES, by assistance provided

	Substantial						Moderate			Low		
Effects on the innovation												
1.												
2.												
3....												
Effects on individuals												
1.												
2.												
3....												
Effects on the organization												
1.												
2.												
3.												
4....												

Entering the data. Reading from a particular case report, we can locate the effects noted and list them in the rows. Then we can scan the case reports, and enter a check in the appropriate column for each effect noted. But what about the problems of short-run and long-run effects, and event-linked and ongoing assistance? We have several options. One is to enter O for ongoing and E for event-linked assistance, plus 1 if the effect was short-run and 2 if it was long-run. That would provide maximum differentiation in the display—data would not be smoothed over or lost—without complicating it excessively.

Table 8.4 shows the complete overall display with all data entered. It looks terribly complicated at first glance—but the *data* are rich and complex and need a complex display. Remember the counsel of Tufte (1990, p. 37), the ace of graphic presentation: "To clarify, add detail." The aim is to see core themes and patterns back of the detail.

Another refinement developed as the analyst pondered over categories. The effects of assistance for the *innovation* could be roughly temporally ordered, with initial or "front-end" effects, such as "planning, developing," and "sense of goals" arrayed at the top, and later on "back-end" effects, such as "program expansion" and "program dissemination," coming toward the bottom. A similar ordering could be done for effects on *individuals.* But the analyst found that the categories for effects on the *organization* could not be sorted by temporal order. (There is little conceptual basis for assuming that "conflict resolution" or "trust" or "improved problem solving" would necessarily come early or late during implementation.)

Naturally, as the data are entered case by case, a few new categories appear and should be put in the temporal order that seems appropriate. For example, in Table 8.4, "routinization" clearly should go toward the bottom of the list of individual effects, but not as far down as "continuation."

Drawing conclusions. Try your hand: What do you see in the display? A good first step is a rapid "squint analysis." Look at the big picture.

For example, we can see by scanning down and across blocks of rows that all cases, even those with nearly no assistance, show assistance effects on the *innovation* and (even more numerously) on the *individual. Organizational* effects are least frequent and occur in all but two sites, Calston and Burton (tactic: **making contrasts, comparisons**).

A squint across blocks of columns shows us quickly that (a) high-assistance cases have a wider *range* of effects than low-assistance ones (tactics: **making contrasts, comparisons; counting**) and (b) low-assistance cases are less likely to experience assistance effects that occur *late* during implementation (e.g, "program dissemination").

The next steps in analysis, given the complexity of the table, often require using a content-analytic summary table (as in Chapter 7, section C), radically collapsing the data from Table 8.4 to confirm or correct the squint impressions and comparing the content of effects in high- and low-assistance cases (Box 8.1).

The analyst also wanted to find out where negative effects of assistance were occurring, and added the minus marks shown, one for each instance of a negative effect. Using the summary table, he concluded that negative outcomes of assistance were more frequent in low-assistance cases, and had their effects mostly on individuals.

Any given matrix usually can generate several different summary tables. Often you do not want to push data into numbers. Another one produced from Table 8.4 appears in Box 8.2. The analyst extracted the specific effects of assistance on individuals that occurred in at least four of six cases, in both the high-assistance and low-assistance partitions.

Looking at these data, the analyst concluded:

> The recurring emphasis across cases on understanding and ownership is consistent with Fullan's (1982) view that problems of *meaning* are the central issues as users approach an innovation. For high-assistance cases, the effects noted also involve anxiety reduction/reassurance, strong support, and repertoire enlargement accompanied by problem solving. The latter two are consistent with our earlier finding that sustained assistance is more typical in high-assistance cases.

The analyst could have noted something else interesting (tactic: **following up surprises**): Why do low cases and not high ones show increasing competence and confidence? A look back at Table 8.4 shows us that, for four low sites, there was ongoing assistance but only at Time 2 or Time 1. As we might expect, the three high sites showing competence/confidence increases experience them at *both* Time 1 and Time 2, and include *event-linked* assistance as

Table 8.4
Case-Ordered Effects Matrix: Effects of Ongoing and Event-Linked Assistance

EFFECTS OF ASSISTANCE	Substantial assistance provided						Initial assistance, then minimal			Nearly no assistance		
On the innovation and its use	Masepa	Plummet	Carson	Tindale	Perry-Parkdale	Banes-town	Lido	Astoria	Calston	Dun Hollow	Proville	Burton
Planning, developing		02 E2	012	012 E2	01	E1	012	01		01	02	
Validating, confirming			012	01 E1	E12			E2				
Sense of goals, direction, priority		012-E1 2-	02	01 E2		02					01	
Increasing legitimacy of innovation		E12				E2			E1			
Obtaining funds			01		01 E2						02 E12	
Aiding start-up, launching	01	012	02		E1	01 E1			02		02	
Preparing/adding materials		02	02	E1		012	02	01	012-	01-2		
Providing general frame/model				01		E12	01					
Aiding good-quality implementation		02	02-E1				02	01-E2	E2	01-2	02-	
Maintaining program	012		02	02		E2						
Program regulating, managing	012		012			012						
Protecting, saving program		012			01-	E12	02					
Program adaptation, alteration	012	E2	02-E2-	E2-	E2	E1	02	012		E1-2-		01
Program strengthening		E2				02						
Program expansion	01					E2	02				012	
Program continuation (district)		02		E2		*		E2				
Program evaluation			01	02	012 E2							
Program dissemination			E2			012						
USER ASSESSMENT time 1 time 2	++ *	++ +	+ to ± ±	+ ±	+ to ± +	+ + to ++	+ to ++ *	+ to ++ 0 to +	0 to - +	± ±	+ to ± + to -	+ *
On individuals												
Reducing anxiety, reassuring	02		012 E12	01 E2	01 E12	E1			02 E2			01
Reducing ambiguity, uncertainty			01-E1						E2	01-		
Increasing understanding, coherence, concepts			012 E2	012 E1	012 E12	E1	01	01		E1-	E1 2	
Reducing resistance			02-E2-	E2-		01			E2		01-E1-2-	01
Reducing frustration, resentment		01-E2-		E1-	01			E2-		E1-2-	E1-2-	
Increasing optimism, hopefulness		E1										
Stimulating, motivating	01		E1							01		
Catharsis, blowing off steam				01								
Increasing interest, ownership	012	E2	012 E2	E2	01 E2	012 E2		012 E12	E2	01-E1	01-	012
Feeling supported, encouraged, backed	012	01 2-	01		012	012	01	01	01	E1	01-	
Feeling pressured, policed	01			012 E1						02		
Increasing autonomy		02	02	012-	02							
Mobilizing energy			02 E1			01				E2		
Saving energy, reduction of tasks	01			012				012	02	01		
Enlarging repertoire	01	E2	01 E1-		E2	01						01
Solving short-run problems	012		01	01 E12		012 E12			E12	01-		
Increasing satisfaction, enjoyment			01 E2							01-E1-		
Teaching/refining skills	01	02			01					02		
Routinization			01	02 E2	E2							
Increasing competence, mastery, confidence			012 E1	02 E2	02 E12		01		02	02		02
Continuation (individual level)						02						
USER ASSESSMENT time 1 time 2	++ *	± +	+ to ± + to ±	+ ± to-	+ to ++ + to ++	+ to ++ +	+ to ++ *	+ to ++ ± to -	+ to ± +	+ to - ±to -	± to - ± to -	0 to + *
On the organization												
Conflict resolution		01	02								01	
Reducing isolation, aiding linkage			E2	01 E2		01	01			01-		
Increasing cohesiveness, trust	02	01 E2	E12	02 ± E12	01			E2				
Increasing collaboration		02		E2	02							
More innovative climate		02										
Improved problem solving		E1			E2							
Increasing morale			01									
Establishing implementation "team"			012-E2	E2		E12	01					
Building assistance infrastructure			012 E12		E2	E12						
Coordination, improved organization		01		E1		02						
Lower dependency on assisters			012-									
Built links to external environment		012 E1 2	01									
USER ASSESSMENT time 1 time 2	++ *	++ +	+ to ± +	+ + to ±	+ to ± +	+ to ++ ++	+ *	+ +	NA NA	+ to 0 NA	± NA	NA *

O = ongoing assistance
E = event-linked assistance effect
– = effect was negative, or opposite

1 = short-run assistance effect
2 = long-run assistance consequence
* effects of event-linked assistance not assessed
NA = not applicable

User Assesment:
++ very effective
\+ effective
± mixed
0 neutral
– ineffective

Box 8.1
Content-Analytic Summary Table: Short- and Long-Term Effects of Assistance, by Assistance Domain

EFFECTS ON...	FOR HIGH-ASSISTANCE CASES (N=6)				FOR LOW-ASSISTANCE CASES (N=6)			
	Ongoing		Event-linked		Ongoing		Event-linked	
	Short run	Long run	Short run	Long run	Short run	Long run	Short run	Long run
The innovation and its use	25 –	24 – – –	12	20 – –	11 –	16 – –	3 –	5 –
Individuals	39 – –	28 – – – –	17 – –	22 – – –	19 – – – – – – –	10 –	10 – – – – –	12 – – – –
The organization	12	11 – –	9	13	4 –	0	0	0

– = negative or undesired effect (included in numerical totals)

well. That finding helps us see that it's not so much of a surprise. But what about those other three high sites? We will need to go back to the case reports.

So content-analytic summary tables, even restricted to counting, can be helpful in understanding a complex display or in verifying a general impression. They can be produced quickly with paper and pencil. They also can be very helpful to the reader during presentation of results.

Variations

Most case-ordered effects matrices are far simpler than this illustration, which involves not only temporally ordered effects in three major categories but also complicated cell entries (short- and long-term effects, event-linked and ongoing assistance). Effects matrices can be simplified by "unpacking" them (e.g., in this case, doing separate matrices for event-linked and ongoing assistance), by making simpler cell entries (e.g., checks), or by having fewer effect categories (though the risk is getting them too abstract and general).

Box 8.2
Content-Analytic Summary Table: Assistance Effects in High- and Low-Assistance Cases

High-Assistance Cases	Low-Assistance Cases
reducing anxiety	
increasing understanding	increasing understanding
increasing ownership	increasing ownership
feeling supported	
enlarging repertoire	
solving problems	
	increasing competence, confidence

Advice

Case-ordered effects matrices are especially useful when you anticipate a wide diversity of effects from a general cause—itself varying across cases. They fall just about at the fuzzy boundary between cross-case description and explanation, providing leads for theory building and testing. They help you avoid aggregation, keeping case data distinct, and let you look at deviant cases.

Their complexity is a natural result of effects diversity. The use of content-analytic summary tables helps verify or disconfirm impressions from the larger display. But the summary table's conclusions usually need to be rechecked against the original display. For example, low-assistance cases (Box 8.1) seem to have more negative effects on individuals. But a recheck with Table 8.4 shows that most of these effects occurred in only two cases (Proville and Dun Hollow), not in cases that have minimal later assistance (Lido, Astoria, Calston).

As with other case-ordered displays, everything depends on the accuracy with which the cases are ordered. For example, in Table 8.4 we must be fairly sure that Masepa has very substantial assistance, and Tindale less, though still of the "substantial" variety, and that Proville has "nearly no" assistance, but more than Burton. Otherwise conclusions about the effects of varying amounts of assistance will be shaky.

And as with any display, several successively revised formats are nearly always needed. As you are trying each version, and even while you are entering data in the supposedly "final" format, stay alert to simple ways of adding other data. For example, adding the minus marks for negative effects was easy to do and did not involve extensive recoding or alteration of the final matrix.

It's usually easier to show effects by rows rather than columns, simply because there is more space to write headings there. If the aim is to show the full, detailed range of effects, resist the temptation to collapse categories.

Time Required

This example actually took 4 hours for creating the basic display (alternating with reading and categorizing case report materials), 5½ hours to enter data from all 12 cases, and 2½ hours for analysis and write-up (6 pages). Once the data are in one place, analysis can go fairly quickly. Summary tables like the ones shown typically require only a few minutes each to produce.

C. Case-Ordered Predictor-Outcome Matrix

Analysis Problem

As we saw in Chapter 7, meta-matrices are the basic building blocks for cross-case analysis. They get the case-level data into an increasingly economical display that lets you work with the full set of cases on one or several variables at a time.

Partially ordered cross-case matrices are the basis for the important data-formatting, data-standardizing, and data-reducing functions that support descriptive analysis. Case-ordered descriptive matrices take us a further step toward understanding patterns.

But how can we get from the descriptive to a more inferential level of analysis? For example, the chart in Box 8.1 answers a descriptive question: Do high-assistance-giving sites have more and different effects than low-assistance-giving sites? It looks as if they do. But there is another, implicit question: Is more assistance likely to lead to or *produce* more effects (on the innovation, individuals, and the organization) than less assistance? In other words, are levels of assistance good *predictors* of these outcomes?

As we've seen, the two-variable case-ordered matrix shown in Box 7.3 (p. 196) is a way to begin asking predictor-outcome questions. The primary function of that matrix was to see whether the levels of stabilization of an innovation were associated with various indices of its continued use. But that type of display doesn't quite get us to a good explanation.

The next step is to get into a *multivariate* prediction mode, taking those predictor or antecedent variables that we have good reason to believe are contributing to outcomes, and assessing their separate and combined effects. How to do this?

Brief Description

Predictor-outcome matrices array cases on a main *outcome* or criterion variable, and provide data for each case on the main *antecedent* variables that you think may be the most important contributors to the outcome. The basic principle behind the matrix is *explanatory,* rather than purely descriptive; we want to see whether these antecedents predict or account for the criterion variable.

Illustration

Let's work our way through a cross-case adventure that includes two types of case-ordered, predictor-outcome matrices, logging the successive analytic steps.

Asking the prediction question. In the school improvement study, we found that some cases had an easy, smooth time during early use and that others had a rougher time.

What accounted for this difference? In other words, which prime factors, coming before or during early use, were associated with more or less smoothness?

Selecting the predictors. In this study, we were especially interested in a set of predictors we called "preparedness factors." For example, prior training in the use of the innovation was an important preparedness factor. Sites that were better prepared through prior training would have an easier time of it.

The individual case reports were equivocal; some facets of preparedness seemed to have made a difference at some sites, but a lot of things were happening during early use that could have added to or swamped the effects of higher or lower levels of initial preparedness.

So we needed to see in one place what was going on in all of our cases. We chose to begin with the "preparedness" factors as predictors because we already had seven of them, some broken down by role type (user, principal, central office administrator). Having a manageable number of predictors is an important start-up criterion. So is the *logic* of the relationship; few things seemed likely to be more important for smooth early use than strong preparedness. And so are *conceptual* and *empirical* considerations. Our way of construing the innovation process included the concept of *preparedness;* prior studies had shown it to be a strong determinant of successful early use, and many informants in this particular study were saying the same thing.

Scaling the outcome and the predictors. Here we come to an instance of how useful it can be to convert qualitative information into quantitative data. Ordering degrees of the outcome variable (relative smoothness of early use) is a fairly straightforward operation. You can begin by asking informants to estimate it, or by having the field researcher make an estimate in the case report. Had we been more foresightful, we probably would have done one of these things. But, like most qualitative researchers, we preferred not to foreclose too early on the choice and definition of criterion variables. We took relevant data from the individual case reports and did a kind of primitive scaling of them.

Respondents had been asked to list the problems encountered in the course of initial use of the innovation, and the field researcher commented routinely on the gravity of these problems. By counting the *number* of problems and the *gravity* ascribed to each, the analyst estimated the relative smoothness and could order the cases from "smooth" to "rough." This procedure was close to the "summed indices" technique shown in Box 7.2 (p. 194).

Cross-case analysts often face this need to "scale"—or otherwise standardize—single-case data. Doing that

Table 8.5

Case-Ordered Predictor-Outcome Matrix:
Degree of Preparedness as Related to Ease of Early Implementation

	Commitment			Understanding									
Ease of early use,[a] by sites	Users	Building Principal	Central Office Admin.	Users	Building Principal	Central Office Admin.	Resources/ Materials	Skills	Training	Preparedness score	Group Median	Ongoing aid/in-service	Building-level support
Smooth early use													
Astoria (E)	√√	√√	√√	√√	√√	√√	√√	√√	√	17	16	√	√
Burton (E)	√	√	$\underline{\surd\surd}_F$	√	0	$\underline{\surd\surd}_F$	$\underline{\surd\surd}_F$	√	√√	15		√	√
Mostly smooth													
Lido (E)	$\underline{\surd\surd}_F$	√√	?	√	√	0	$\underline{\surd\surd}_F$	√	√√	12	12	0	$\underline{\surd\surd}_F$
Mixed[b]													
Calston (E)	√√	√√	$\underline{\surd\surd}_F$	√	√	√√	$\underline{0}_B$	$\underline{\surd}_F$	√√	14	13.5	$\underline{\surd}_B$	√√
Perry-Parkdale (E)	$\underline{\surd\surd}_F$	0	√√	√√	0	√√	√√	$\underline{\surd}_B$	√√	13		√	√
Rough													
–Banestown (E)	$\underline{\surd\surd}_F$	$\underline{\surd}$	$\underline{\surd\surd}_F$	√	0	√	$\underline{\surd}_B$	$\underline{\surd}_B$	$\underline{\surd}_B$	9		$\underline{\surd}_F$	√√
–Masepa (E)	√√	√	$\underline{\surd\surd}_F$	$\underline{0}_B$	0	√	$\underline{0}_B$	$\underline{0}_B$	√	5		$\underline{\surd\surd}_F$	√
Carson (L)	$\underline{\surd\surd}_F$	$\underline{\surd\surd}_F$	√√	$\underline{\surd}_B$	$\underline{\surd}$	√	√	$\underline{\surd}_B$	$\underline{\surd}_B$	11	11	$\underline{\surd\surd}_F$	√√
Dun Hollow (L)	√	√	√	√√	√√	√√	$\underline{\surd}_B$	$\underline{\surd\surd}_F$	√	13		√	√
–Plummet (L)	$\underline{\surd\surd}_F$	$\underline{\surd\surd}_F$	√√	$\underline{\surd}$	$\underline{\surd}_B$	√	$\underline{\surd}_B$	$\underline{\surd}_F$	0	12		√	√√
–Proville (L)	$\underline{\surd}_B$	$\underline{\surd}_B$	$\underline{\surd}_F$	$\underline{\surd}$	√	$(\surd)_B$	$\underline{\surd}$	√	√	8		√	√
Tindale (L)	$\underline{0}_B$	$\underline{\surd\surd}_F$	$\underline{\surd\surd}_{F/B}$	$\underline{\surd}$	√√	√	$\underline{\surd\surd}_{F/B}$	√	√√	13		$\underline{\surd\surd}_{F/B}$	√√

(E) externally developed innovation
(L) locally developed innovation
[a] field researcher judgment from users' responses and/or from observation of practice in use
[b] smooth for some users, rough for others
underline signifies field researcher estimate that factor was decisive in affecting ease of early use

F = factor facilitated early use
B = factor was barrier to successful early use
√√ fully in place
√ partly in place
0 mostly absent, missing
? missing data

\# Computed in the following way:
√√ = 2 pts.
√ = 1 pt.
0 = 0 pts.
F = +1 pt.
B = -1 pt.

means *transforming* the single-case data from a nominal or categorical variable to a dichotomous or continuous one. This is not an operation to be taken lightly. Rather, it should be a self-conscious procedure with clear decision rules.

In Box 7.2 (p. 194), for example, the analyst specified *how* the magnitude of users' "change in practice and perceptions" would be defined by looking at a range of indicators, and *how* an assignment of "high," "moderate," or "low" for each case would be made. He entered the data for all of the cases before determining cutoff points and making the assignment. Survey research has most of these procedures routinized, but the qualitative researcher has to define decision rules and march the data through them to be sure the ordering is not being skewed in the direction of a favored insight or arresting empirical cluster.

Scaling the "preparedness" predictors was easier. Each case report included a "preparedness" checklist matrix (see Table 5.2, p. 107) that could be converted directly into an ordinal scale running from "factor not in place" to "factor fully in place" (the checklist matrix already had ordered judgments included). Now we are ready for the case-ordered predictor-outcome matrix.

Building the matrix and entering data. We have a clear list of predictors, and the construction of the matrix (Table 8.5) is straightforward. The analyst works case by case, reviewing the case matrices and forming a judgment of the degree to which each of the predictors was in place. We may need to look at the associated analytic text as well, especially in judging whether the predictor was decisive in affecting early implementation (see *underline* explanation in legend.) The text is also often helpful in aiding another judgment: whether the factor was a facilitator or a barrier (note that the analyst *does not take this for granted,* but looks for specific evidence of such near-causal attribution). A second analyst should verify or disconfirm such judgments.

Drawing first conclusions. A first look can center on the first nine columns, ending with the ratings on "training"; these are the main "preparedness" indices. Looking down the columns, because the 12 cases are ordered by ease of early use, yields a lot of useful information (tactics: **seeing patterns, themes; making contrasts, comparisons**). For example, most sites are only partly prepared; sites with externally developed innovations are, without exception, better prepared than those with locally developed ones (Hmmmm . . . Wonder why that's happening? . . . We may be on to something.). We can also see that most sites are well prepared on the "commitment" factor, but few on the "skills" factor.

Testing the prediction. What predicts implementation smoothness? Try your hand, starting with a squint analysis.

If "preparedness" were related linearly to "ease of early use," the double check marks should turn progressively to single check marks and then to zeros as we read down the columns. Perversely enough, this does not seem to be happening for all columns, although there is some thinning out of check marks as we go from smooth starters to rough starters. Part of the reason the prediction is not working out is limited variation in some of the predictors (e.g., central office administrators' commitment); all or most cases have the same rating, or the scale is truncated (no zero ratings). Scaled matrices such as these live or die by the variance of their scales. Without variation, the matrix is out of business for other than descriptive purposes. Finding what *does not* predict is not much help.

Another test of the prediction can be made by converting the checks and zeros to numbers. As we've noted, the tactic of **counting** is not irrelevant in qualitative research. This is done in columns 10 and 11 of Table 8.5. The legend at the bottom right also shows that some **weighting of the evidence** has been done when facilitators or barriers were clearly present.

The "preparedness score" and "group medians" do show a relationship, but it is certainly not cleanly linear. For example, preparedness seems a bit better in "mixed" than in "mostly smooth" cases. Of course the missing data in Lido lower its score artifactually. We can also note the key difference between Lido and its "mixed" partners Calston and Perry-Parkdale: absence of central office understanding. Furthermore, we can see some moderately prepared but "rough" sites (Dun Hollow, Tindale, Plummet).

That finding led the analyst to go back to the report text and to fine-tune the case ordering, breaking out scores for "rough starters" and "very rough starters" (Banestown, Masepa, Plummet, Proville). The first group median was 13, and the second one 8.5. That somewhat strengthened the linearity of the relationship between preparedness and smoothness of early use. But it still wasn't very clear. And the analyst looked at two more variables (see the last two columns of the table) to determine whether these two conditions were actually present during early use (Note **surprise** here: More in-service is occurring in "rough" sites. You would think it would help smoothness. To be followed up....).

So the numbers help. They make it easier to manipulate the data in a case-ordered matrix when you're testing a hypothesis. They also make it easier for you to see what is there. Finally, they keep you honest; you have to look at all of the data on the chart and use the same "computing" rules for each cell.

All of these steps put insights, hunches, strong desires, and conceptual preferences to a more stringent test. But notice that the numbers are not *alone* in Table 8.5. The judgments that led to the numbers are still present on the chart (fully in place or not, facilitative or not, factor seen as decisive, etc.), and you can get back easily to the original case displays and text—or even, if needed, the coded field notes—when puzzling results turn up. Also the numbers and weights are, thankfully, primitive ones, which keeps you from veering off into fancier data manipulations, which would almost surely be tepid or misleading.

We're making some analytic progress here. We can see, for instance, that some preparedness conditions matter more than others. For example, looking at the underlined cell entries, along with the Bs and Fs, shows that user "commitment" matters more than user "understanding" (tactic: **making contrasts, comparisons**). Also a scan of the Bs and Fs shows that being well prepared does not help smoothness of initial use as much as being poorly geared up hurts it.

And for some cases, few or none of the factors are underlined. Astoria, for instance, has the highest preparedness score, but no underlines. *Something else* must also be influencing smoothness of early use—and, for the smooth cases, influencing it more than levels of preparedness. A lot can be wrung from a case-ordered predictor-outcome matrix—including the idea that our explanation is not working well. The tactic we are coming up on is **finding intervening variables**—additional predictors that are in play.

Strengthening the prediction. What does the analyst do now? Go fishing for more predictors, look for the "something else." The choice can be made on logical, conceptual, and/or empirical grounds. In qualitative research, fishing is mostly an empirical recreation. What did the informants say was helping or hurting early use? How did the case-level analyst make sense of the "problems encountered" information?

So, back to individual case reports. If they have been formatted in the same way, that search can be done in a few hours of reviewing the subsections and the displays.

Here the analyst derived five more variables at a low level of inference. He then did a new predictor-outcome

Table 8.6

Case-Ordered Predictor-Outcome Matrix:
Additional Factors Related to Early Implementation

Ease of early use, by cases	Users (1st generation) volunteered or pressured	Actual classroom/ organizational fit	Actual degree of practice change	Latitude for making changes	Actual size/scope of innovation +
Smooth early use [a]					
Astoria (E)	constrained	$\underline{\text{good}}_F$	$\underline{\text{minor}}_F$	$\underline{\text{high}}_F$	small
Burton (E)	$\underline{\text{pressured}}_F$	$\underline{\text{good}}_F$	$\underline{\text{minor}}_F$	$\underline{\text{high}}_F$	$\underline{\text{small}}_F$
Mostly smooth					
Lido (E)	$\underline{\text{pressured}}_F$	$\underline{\text{moderate}}_B$	$\underline{\text{moderate}}_B$	$\underline{\text{high}}_F$	small/moderate
Mixed [b]					
Calston (E)	pressured	$\underline{\text{poor}}_B$	$\underline{\text{moderate}}_B$	$\underline{\text{mod./high}}_B$	small
Perry-Parkdale (E)	volunteered	moderate	$\underline{\text{moderate}}_B$	$\underline{\text{high}}_F$	moderate
Rough					
Banestown (E)	pressured	moderate	$\underline{\text{major}}_B$	$\underline{\text{high}}$	small/moderate
Masepa (E)	volunteered	good	$\underline{\text{major}}_B$	$\underline{\text{low}}_B$	large
Carson (L)	$\underline{\text{volunteered}}_F$	moderate	$\underline{\text{major}}_B$	$\underline{\text{high/mod.}}_F$	$\underline{\text{large}}_B$
Dun Hollow (L)	volunteered	$\underline{\text{poor}}_B$	$\underline{\text{minor}}_F$	moderate	$\underline{\text{small}}_F$
Plummet (L)	$\underline{\text{volunteered}}_F$	$\underline{\text{poor}}_B$	$\underline{\text{mod-major}}_B$	$\underline{\text{high}}_F$	$\underline{\text{large}}_B$
Proville (L)	$\underline{\text{pressured}}_B$	moderate	minor	high	$\underline{\text{moderate}}_B$*
Tindale (L)	$\underline{\text{pressured}}_B$	$\underline{\text{moderate}}_B$	$\underline{\text{major}}_B$	$\underline{\text{low}}_B$	moderate/large

(E) externally-developed innovation
(L) locally-developed innovation
[a] field researcher judgment from users' responses and/or from observation of practice in use
[b] smooth for some users, rough for others
$\underline{\text{underline}}$ signifies researcher estimate that factor was decisive in affecting ease of early use
F = factor facilitated early use
B = factor was a barrier to successful early use
* substantial role changes, but for limited populations
+ as contrasted with pre-implementation size/scope

matrix (Table 8.6). As before, all of the predictors are scaled, either dichotomously or continuously. The conclusion-drawing work is also similar: looking for covariation in the scales and in the underlines, Fs and Bs (tactic: **noting relations between variables**). We are taking a variable-oriented approach, hoping at the same time to get a better sense of the "story" in each case.

What's here? Scan down the columns of Table 8.6 one by one; do you see a progression in any of them, related to the progression of cases from smooth to rough?

The columns that pay off most are "actual degree of practice change" (minor change helps, major change hurts), "latitude for making changes" (higher latitude helps), and "actual size/scope of innovation" (smaller ones have an easier time). So it looks as if smaller scale ventures with a lesser leap from past practice to this innovation and some latitude to make changes (probably downsizing ones) is a combination accounting, in part, for relative ease of use. Aha! The "smooth" sites did not need much preparation because they were not taking on very much, whereas the "rough" sites were making ambitious changes that inadequate readiness could bedevil.

Notice something important here. We have shifted from the mainly *variable-oriented* approach of Table 8.5 to a *case-oriented* look. We're finding *configurations* of variables that show us families of cases.

The analyst begins to piece together this mosaic by looking for possible explanatory configurations, and cycling back to the case-level reports to see whether the emerging picture makes sense. Here the analyst saw one cluster of cases and variables (small practice change, high latitude, and small-sized innovation) with its countervailing cluster (large practice change, lower latitude, and large innovation size). He then could look back at the case reports to see whether such animals actually existed, and whether the hypothesis about size of attempted changes affecting early smoothness was plausible in each case or not (tactic: **replicating a finding**).

Variations

Predictor-outcome matrices are close—perhaps too close—analogues to statistical prediction procedures. They are something of a statistical albatross: no zero-order or regression coefficients, no algorithms for entering or deleting predictors, no error terms. You could, of course, use simple nonparametric analyses to test the trends in the matrix.

But it's much more helpful, and humbling, to look at the empirical validity of a prediction by confronting a supposedly "elegant" pattern of covariation with the grounded local patterns actually observed in the case reports. If no case has a configuration that fits the prediction cluster, that cluster is hollow and the prediction is baseless.

Another way to go is to switch to an analytic induction logic (as in Chapter 6, section D, on causal networks, and in section E of this chapter, on causal models). That switch moves us away from the domain of probabilities (looking for good post hoc predictions) and toward finding case-specific causes or determinants (providing explanations for a given outcome) that can be tested through successive **replication.** Such induction is sometimes easier—both to do and to explain to others—when the predictor-outcome matrix has cell entries in symbolic, stripped-down form, like the *Consumer Reports*-type entries in Table 8.7, drawn from Lee et al. (1981, p. B48). The 16 cases are ordered by degrees of parental involvement; we can see at a glance that the key contributing factors for parent involvement are "DAC has specified authority," "parent coordinator role," "DAC training," and "powerful person."

You can also build matrices that go from antecedents to outcomes to later consequences of those outcomes—three rather than two steps. Box 8.3 shows an example.

Advice

Use a case-ordered predictor-outcome matrix when you want to see how several contributing factors function *together* in relation to different levels of a criterion measure. It's variable oriented, but keeps the configuration of variables for each case.

Begin work with a manageable number of candidate predictors chosen on the basis of good sense, conceptual plausibility, and empirical groundedness. More than a dozen at a time is a formula for overload. Don't let predictors balloon. A pure fishing expedition is never helpful; the conceptual framework or the data from the cases should be providing promising clues from the outset.

Be very careful when transforming case-level text into cross-case scales. Summed indices, using a quasi-Guttman approach, as in Box 7.2, p. 194, are a good, bias-deflating device. Record and report the decision rules you used to do the scaling.

Work with the matrix *descriptively* before you work with the relationships between the variables; that is, go down each column first (tactic: **noting patterns, themes**) and then start **comparing, contrasting** columns. If there is too little variation in the matrix, be content with what you can extract descriptively, or focus solely on the variables that *do* have some range.

Test out promising patterns with simple computations. Think about the quality of the field data and consider **weighting** some of them over others. But never let the computations get unhinged from the raw data. Keep the words and the numbers or scales together throughout the analysis.

Expect that one matrix will lead to another and that the next one will be smaller and better informed as a result of the first one.

Test your "best and final" prediction against the individual case reports to be sure the configuration does actually occur locally.

Time Required

How long it takes for case-ordered predictor-outcome matrices like these depends on how many cases you have, how many predictors you load into the matrix, and how scale-ready and scale-worthy each one is in the single-case reports. The greater the number of cases, predictors, and transformations needed to get each predictor standardized, the longer the process. Time also depends on how good the matrix is—that is, how well-informed your choice of candidate variables was.

Table 8.5 is at the middle-high range on all of these counts; for 12 cases, it took about 3 hours to assemble and another 3 hours to analyze. Table 8.6 took longer (about 4 hours) to assemble because the choice of predictors was less obvious and the data transformations greater for all of the predictors. As usual, the intermixed activity of writing analytic text—itself a clarifying and focusing device—went quickly; it occupied, all told, under an hour for each matrix.

Predictor-Outcome Consequences Matrix

So far in this chapter, we've seen how cross-case matrices can be used to explore and test relationships between two, then among several, variables. It's also possible to do more directly explanatory work with cross-case matrices, moving beyond probabilistic predictions to more causal (deterministic) analyses. We can hook a chain of predictors to some intermediate outcome, and then show the consequences of that outcome—the outcome of the outcome. Three steps, rather than two, are involved. Such a display is especially good when you are looking at an intermediate

Table 8.7

Case-Ordered Predictor-Outcome Matrix:
Factors Contributing to Parent Involvement (Lee, Kessling, & Melaragno, 1981, p. B48)

	No DAC	No Involvement							Minor Involvement					Major Involvement		
	Mt. View	Roller	King Edward	Brisbane	Benjamin Co.	Maple	Summer Place	Meadow-lands	Plains	Stadium	Bonnet Co.	Kings-town	Clete-ville	Redlands	Compass	Johns Co.
State Guidelines Exist, Are Implemented			√					√			√	√		√	√	√
DAC Has Specified Authority														√	√	√
Parent Coordinator Role	○	○	◒	○	○	○	○	●	●	◒	◒	●	◒	●	●	●
Staff Attitude: Parental Involvement = Support	√	√	√	√	√			√			√	√				
Parent Attitude: Satisfied with Project/ Professional Make Decisions	√			√	√	√	√	√				√	√			
Decisions Reserved for High-Level Administrators		√	√			√			√			√	√			
DAC Training		⬓	⬓	□	□	⬓	□	□	■	■	⬓	□	□	■+	■	■+
Powerful Person		△	△	△	△	△	△		◭	△	◭	△	△	▲	▲	▲

○ = no PC ◒ = PC dominates DAC ● = PC supports DAC

□ = none ⬓ = Title I ■ = group processes ■ + = group process + Title I

△ = professional ◭ = shared, parent and professional ▲ = parent

SOURCE: Drawn from Lee et al, (1981), p. B48).

Box 8.3

Predictor-Outcome Consequences Matrix: Antecedents and Consequences of Assistance

	ANTECEDENT CONDITIONS							ASSISTANCE	CONSEQUENCES	
SITES/ Scale of Assistance	Actual size/scope of innov.	Required practice change	Actual classroom/ organizational fit	Implementation requirements #	Scale of funding	Central office commitment to change	Admin latitude	OVER-ALL PRESENCE	Smoothness/ Roughness of early implementation	Practice-stabilization (later implementation)
Substantial assistance										
Masepa (E)	large	major	mod/good	12	$30-50K	high	low	HIGH	very rough	mod
Plummet (L)	large	mod/major	good/poor*	12	$300K	high	high	HIGH	very rough	mod
Carson (L)	large	major	mod/good	12	$96K	high	mod	MOD/HIGH	rough	mod
Tindale (L)	large/mod	major	mod	12	$87K	high	low	MOD/HIGH	rough	high
Perry-Parkdale (E)	mod	mod/major	mod	10	$300K@	mod	high	MOD/HIGH	mixed	mod/high
Banestown (E)	small/mod	major	mod	10	$5.6K	high	high	MOD	very rough	mod
Initial assistance, then minimal										
Lido (E)	small	mod	mod	7	$6.1K	low	high	LOW/MOD	mostly smooth	mod/high
Astoria (E)	small	minor	good	3	None	high	high	LOW/MOD	smooth	high
Calston (E)	small	mod	poor	9	None	mod/high	mod/high	LOW/MOD	mixed	mod/high
Nearly none										
Dun Hollow (L)	small	minor	poor	7	None	low	mod	LOW	rough	low
Proville (L)	mod	minor	mod	7	$180K	high/low	high	LOW	very rough	low
Burton (E)	small	minor	good	3	$3.1K	mod/high	high	LOW	smooth	mod

(E) externally developed innovation
(L) locally developed innovation
* good at district level, poor for students.
\# weighted sum of three variables at left, scaled 1-5.
@ local innovation funds, to adopt + disseminate externally developed innovation

or intervening variable (in this case, assistance) that is not construed as a "final" outcome.

In Box 8.3 the first aim is to predict the degree of assistance provided at the 12 field sites (which are ordered in the left-hand column). To do this, the analyst assembled the most likely predictors emerging from the individual cases and from preceding cross-case analyses.

The first three columns deal with variables that are bundled together and are given a numerical weight signifying "strength of implementation requirements."

Then we have three more antecedents (funding scale, central office commitment, and administrative latitude). The level of assistance is simply repeated in the next column for clarity.

But the display can help with another objective: looking beyond assistance to its consequences during early implementation (next to last column) and during later implementation (last column). In other words, although the degree of assistance was the *dependent* variable for the earlier analysis, it is, in turn, the chief *predictor* of the two later consequences.

The analyst is going beyond a simple double prediction here, and is also looking to see whether the predictors that best accounted for the degree of assistance make sense—when combined with assistance level—to lead to different levels of later outcome. So we have the rudiments of a causal chain appearing in the matrix. We are going beyond a simple "configuration" in each case to understanding, if we can, the actual causal *mechanisms* involved.

The same general analysis strategies as those for a regular predictor-outcome matrix can be used here. Because of the three-step aspect, expect to spend more time in analysis.

D. Variable-by-Variable Matrix

Analysis Problem

Sometimes, after cases have been analyzed carefully, you begin to see that a few major variables are in play, underlying many specific phenomena. Normally you would like to know how these variables are connected—without going to a simplified scatterplot mode, as in Chapter 7. The problem is how to keep the richness of specific indicators of the main variables, so as to see in more depth what the core relationships are.

Brief Description

A *variable-by-variable matrix* has two main variables in its rows and columns; specific indicators of each variable are *ordered* by intensity. The cell entries, rather than being data bits, are *case names*, so that we can see which cases (themselves also ordered on a variable) have included a specific type of interaction of the two main variables.

Illustration

That description sounds quite arcane, but this example shows the simplicity of the display. In a study of urban high school reform (Louis & Miles, 1990), one of the questions of interest was how schools coped with the inevitable problems encountered during the change process.

Building the display. Much conceptual work already had been done (Miles, 1986). The researcher had clustered and categorized the types of *problem* reported, and ordered them from most to least "tractable" or solvable. Problems such as "program process" and "program content" were relatively easy to deal with, and problems stemming from the school's physical setting or available financial resources were judged as more difficult to resolve, on the basis of prior studies of educational reform. The first ordering was adjudicated with other members of the research team.

The researcher also developed, from the five high school cases at hand, a list of 23 *coping strategies:* What do you do when faced with a problem? The strategies, such as "delay," "avoidance," "people-shuffling," "vision-building," and "role redesign," were sorted into nine stylistic clusters, such as "do nothing," "do it the usual way," "do it harder," and "redesign the system"; these clusters then were ordered according to a major shallow-deep dimension, one end of which was "shallow, soft, informal, less penetrating" and the other "deeper, more structurally oriented, deliberate, and person-changing." Here, too, the first cut required colleague feedback and revision.

Once the conceptual work was done, the matrix design was easy: Rows were coping strategies, and problem types were columns. What to do about cell entries? They could be initials denoting the five cases. Two "A" cases had the most successful implementation, one "B" case was moderately successful, and two "C" cases were least successful. The display appears as Table 8.8.

Entering the data. Data on problems and coping had been collected in all sites. In a cross-case analysis meeting of project researchers, instances of problems and the associated coping strategies used were drawn from case report displays, field notes, and researcher judgments. The problem was sorted into the appropriate column, and the strategy used was identified by the researchers most knowledgeable through their fieldwork and case analysis. If more than one strategy had been used for the same problem, more than one cell entry was made. Interresearcher differences were resolved through discussion and reference back to displays and case data. In most cases the problems and coping strategies were very familiar and salient. When the researcher was less certain about the problem-coping connection, "?" was added to the cell entry. Altogether, 76 problem-coping pairs were entered for the five sites.

Drawing conclusions. This is a new kind of display. Take a close look to see possible patterns in it, and jot down your conclusions.

The analyst, looking at the top part of the table, wrote:

> Most of the instances of "shallow" coping (none at all, temporizing, using normal routines) appear in the two less-successful sites, Caruso and Chester (C1 and C2) (tactic: **noting patterns, themes**).
>
> Second, moving to the lower region of [the table] it is quite striking to note that the deeper managing and coping strategies (improving capacities of persons and the system, re-staffing, and role/organizational redesign) are used mostly by A-level (more-successful) sites. Except for the new orchestration structure and the monitoring strategy in C1, such strategies are mostly absent in the C-level sites. The B-level (moderately successful) site employs a few "deep" strategies, but fewer than the A-level sites.

Then the analyst saw another **pattern,** a bit more complex. He focused on the A-level sites and wrote:

> The A-level sites actually use a very wide range of strategies, *except* for "none" and "avoidance/delay." The feeling is one of actively doing whatever the problem seems to require, but having a repertoire that includes vigorous managing and capacity-building interventions where they are needed . . . [they] show "deeper" coping for the more difficult problems.

To check the meaning of these findings further, the researcher returned to the case reports, to a section where the analyst had described the typical results of problem-coping efforts, and made a summary table, shown in Figure 8.1 (tactic: **replicating a finding**).

The analyst saw nothing to contradict the general findings from Table 8.8.

Table 8.8
Variable-by-Variable Matrix: Coping Strategies and Problems, by Case

STYLE	Coping Strategies	Program process	Program content	Target pop.	Skill lacks	Attitudes	Crises	Competing demands	Low control	Physical setting	Resources
-	Can't determine	B	A2						C1		
DO NOTHING	None	C2			C1	C2			C2		
TEMPORIZE	Delaying/avoiding		C1			C2					
DO IT THE USUAL WAY	Short-run coping	C1 C2	C1							C1	C1
	Using existing meetings/roles	A2 B C1 C2	C2								C1 A2
	Action-taking	C1									
	People-shuffling		C1?			A1	C1 C2				C1
EASE OFF	Program modification		B	A1 B C1	B?	B		B			
DO IT HARDER	Symbolic support					A1 C1 C2		A1			
	Rewards/incentives					A1 A2 B C2					
	Negotiating										C2
	Pressuring/requiring					B C2					
BUILD PERSONAL CAPACITY	Person-changing	A2		A1 A2 B?	A1						
BUILD SYSTEM CAPACITY	New orchestration structure	A2 C1	C1		A2						A2 C1
	New interaction arenas			B							
	Vision-building/sharing		A1			A1 A2 B					
	Monitoring		C1								
	Rolling planning					A2					
	Ongoing assistance		A1								A1
ADD NEW PEOPLE	Re-staffing					A2 B					
REDESIGN THE SYSTEM	Increasing control							A1	B?		A1?
	Empowering	A1 A2 B?				A1 A2					
	Role redesign										A1
	Organization redesign										A1 A2

Legend:
A1 Agassiz [more successful]
A2 Alameda
B Burroughs [moderately successful]
C1 Caruso [less successful]
C2 Chester
? Items with less researcher certainty

Variations

The time dimension can be added by identifying cell entries as occurring earlier or later during some time period. (In this case, interestingly enough, case C1, which did a modest amount of deeper coping, moved up 2 years later to "B" level, according to researcher judgments.)

Each cell entry for this matrix could also include a code indicating how "successful" the entry turned out to be. (In this particular display, did the problem recur, lead to new problems, or was it effectively solved?)

If cases are not ordered on a major variable, this type of display can help you cluster them into types or families.

Advice

Enjoy the process! This display and its analysis led to delighted Aha! feelings in the research team.

Careful *conceptual* ordering of the two main variables is critical for this display, or else the empirical results will be dubious. It works best if there are specific indicators, as shown here, and the possibility exists of identifying clear-cut, eventlike instances of interaction between pairs of the indicators. Without these, you run the risk of forcing findings via vague judgments.

It also helps to link the conceptualization of the main variables with prior work. In this case, work on individual coping mechanisms (Roth & Cohen, 1986) proved very helpful in further interpretation of the findings. For example, the ability of the "A" sites to tailor coping depth to the problem's difficulty was paralleled by the finding that healthy individuals often use shallow, avoidance-oriented coping (e.g., denial, postponement) for trivial problems, while using deeper, approach-oriented coping for serious ones.

Time Required

Creating the display in this example went rather quickly: The categories for problem type and problem coping al-

Figure 8.1
Summary Table:
Typical Consequences of Coping, by Case

Agassiz (A1):	Most problems have been acknowledged, addressed, and solved. Main consequence is an increase in individual and collective self-confidence (both teachers and students). More downtown (central office) respect for school.
Alameda (A2):	Reduced anxiety, built more constructive norms. Better mesh with staff development, met teacher needs. New roles and empowerment meant that principal could be reassigned [without disruption of program].
Burroughs (B):	Increased respect for principal's authority. Staff schism: people either big fans or opponents. Little investment by new staff in program. As trained staff moved to other schools, new staff did not know or have skill in program.
Caruso (C1):	Problems slightly ameliorated, but will recur.
Chester (C2):	Moderate coordination of the curriculum alignment part of the program, but of the remainder not. Problems will continue.

ready had been developed in near-final form for an interim report. The researchers understood the five cases well, and their ordering on overall outcome was defined clearly. The format of the display was proposed and settled on during a few minutes of a staff meeting, and in less than an hour the researchers most closely familiar with each case reported examples of problems and coping strategies from their case reports, sorting them into the appropriate cells. Initial conclusions followed in another few minutes, and the first analytic text took another hour. During writing, there were a few revisions of the order of the indicators, and revised conclusions; another hour.

Naturally, if the indicators of each main variable are unclear or poorly ordered, or if the case ordering is not firm, much more time will be required.

E. Causal Models

Analysis Problem

Carefully ordered meta-matrices, like predictor-outcome matrices, time-ordered matrices, or effects matrices, can begin to tell us a lot about what goes with what. Variable by variable, we can understand that X comes before Y, and more of X goes with more of Y; that less of Q goes with more of Y; and perhaps that J looks unrelated to Y—unless, maybe, you take varying levels of K into account. This is all rather atomistic, though. Two problems remain.

The first is beginning to transcend mere "association" (the weasel word of quantitative researchers) and to arrive at something like a judgment that variable X not only precedes Y in time but also looks coherently connected to it: If X goes up for some reason, we would expect Y to go up or down, as well.

The second problem is how to go beyond mere list-making (X, Q, and J with K controlled are all predictors of Y) to something like an integrated set of relationships among the variables: in short, a *model*. In practice, these problems usually need to be solved together, and they require a network rather than a matrix display.[3]

We have discussed (Chapter 6, sections A and D) the practical problems of assessing local causality and of forming a sensible, data-grounded causal network. The question at hand is this: Given multiple-case data in ordered matrix form and a set of variable-oriented conclusions, how can you integrate those findings into a general model of the variables involved that specifies causal connections clearly?

Brief Description

A *causal model* is a network of variables with causal connections among them, drawn from multiple-case analyses. Although empirically grounded, it is essentially a higher order effort to derive a testable set of propositions about the complete network of variables and interrelationships. The principle is one of theory building.

Illustration

In the school improvement study, we asked the question, "In what way do users of an innovation change their 'practice' after using the innovation—and what determines greater or lesser amounts of practice change?" Because causal modeling is usually a second-order activity drawing on other displays, we depart here from our usual format.

Preliminary display: Outcomes. When you're thinking about causal modeling, you usually have to clarify the basic dependent variable. Is it monolithic, or does it have several parts? Can some display of data from all cases help clarify this?

One line of analysis goes like this. Divide the dependent or outcome variable into aspects or components; make a rough ordering of the cases by degree of presence of the overall variable; and see whether the components have variance—that is, are differentially present. For example, a component present only in "high" cases is presumably "deeper," harder to achieve than a component that appears in nearly all cases.

Box 7.2, repeated here, shows how this line works. We can see that the "user practice change" outcome is not monolithic. Some components of it, such as "basic constructs, attitudes" and "transfer" (of learnings to new situations), appear only in high-change cases, while others, such as "daily routines" and "repertoire," appear in nearly all cases, regardless of how much user practice change occurred (tactic: **making contrasts, comparisons**).

Box 7.2 (Repeated)
Summed Indices:
Reported Changes in User Practice and Perception

DIMENSIONS OF CHANGE ARISING FROM IMPLEMENTATION

SITES	Daily routines	Repertoire	Relation-ships	Under-standings	Self-Efficacy	Transfer	Basic Constructs, Attitudes	Magnitude of change
Masepa (E)	X	X	X	X	X	X	X	High
Plummet (L)	X	X	X	X	(X)	N/A	X	High
Banestown (E)	X	X	X	X	X	(X)		Mod
Tindale (L)	X	X		X	X	X		Mod
Carson (L)	X	(X)	X	X	X	N/A		Mod
Perry-Parkdale (E)	X	(X)	X	(X)	(X)	N/A	(X)	Mod
Calston (E)	X	X	(X)	(X)	(X)		(X)	Mod/low
Lido (E)	X	X		(X)			(X)	Mod/low
Astoria (E)	X	X	(X)	X				Low/none
Burton (E)		X						Low/none
Dun Hollow (L)	X-0	X-0					(X)	Low/none
Proville (L)	X-0				N/A	N/A		Low/none

(E) = externally developed innovation
(L) = locally developed innovation
X = change claimed unambiguously by several informants
(X) = change claimed unambiguously by only one informant
X-0 = initial change, then reversion to initial practice
N/A = not appropriate/applicable
Blank = no unambiguous changes cited

Thus we have "unbundled" the outcome, finding that some aspects of it change more easily than others. This process is interactive: The various dimensions came from inductive examination of teachers' reports of their changes after using the innovation; the analyst tried to sort them out in a rough order of "depth," meanwhile sorting out the cases in a rough order of "overall impact." Both orderings shifted; rows and columns were transposed several times until the analyst came out with the result.

Preliminary display: Predictors and outcomes. Next we need to look at the question of which predictor variables might be associated with the outcome. For that we need to build a case-ordered predictor-outcome meta-matrix.

The analyst in this case chose to order the cases in terms of overall intensity of the outcome (holding in abeyance, for the moment, the idea that suboutcomes seemed to be differentially achievable) and picked a number of predictors that seemed to have been important on a case-by-case basis in inducing user practice change (Table 8.9). As usual, we suggest that you scan the chart, draw some preliminary conclusions about the weight of the different predictors, and jot them down.

In this instance the analyst concluded that "required practice change," "project size/scope," and "classroom/organizational fit" were all associated with degree of user practice change (tactic: **making contrasts, comparisons**). This conclusion also involved another tactic: **checking the**

Table 8.9
Predictor-Outcome Matrix:
Predictors of Magnitude of User Practice Change

EARLY IMPLEMENTATION REQUIREMENTS

Magnitude of change, by sites	Required practice change *	Project size/scope	Classroom/ organizational fit	Index of early implementation requirements @	General attitude during implementation	Administrative Pressure Direct: Strong-arming	Indirect: exhortin reinforci
High Change							
Masepa (E)	major	large	mod/good	14	+	high	high
Plummet (L)	mod-major	large	good/poor #	12	+	low	high
Moderate Change							
Banestown (E)	major	small/mod	moderate	10	+	mod	high
Tindale (L)	major	large/mod	moderate	12	+	high	high
Carson (L)	major	large	moderate	13	+	low	high
Perry-Parkdale (E)	mod-major	mod	moderate	10	+	low	low/mo
Moderate-Low Change							
Calston (E)	moderate	small	poor	9	+	mod	mod
Lido (E)	moderate	small	moderate	7	+	low	mod
Small-No Change							
Burton (E)	minor	small	good	3	+	low	mod
Dun Hollow (L)	minor	small	poor	7	-	mod	low
Proville (L)	minor	moderate	moderate	7	-	mod	mod
Astoria (E)	minor	small	good	3	+	low	low

(E) externally developed innovation
(L) locally developed innovation

* Discrepancy between users' customary instructional practices and those required to implement the innovation at the time of initial use

Good in the district, poor for needs of incoming students

@sum of 5-poi scales for fir 3 columns.

meaning of outliers. For example, the analyst noted that the moderate-sized program at Proville was simply sabotaged by dissatisfied users, so there was no opportunity for user practice change to occur. At Calston, another outlier, the small (institutionally speaking) project was nevertheless an ambitious one for the individual teachers involved.

The analyst also noted that "general attitude during implementation" was a poor predictor because it was largely positive, except in cases where discontinuations took place (Dun Hollow, Proville).

Finally the analyst saw that "administrative pressure" of a direct sort was, by itself, not a very good predictor. But if it was combined with indirect pressure (exhortation, reinforcing), the consequences were clear: A carrot-and-stick approach was associated with greater user practice change (tactic: **finding intervening variables**).

Building the causal model. Now you begin reflecting: How can the associations noted in these two displays be integrated into a meaningful explanatory model? So far, the analysis has been mostly variable-oriented. Let's apply some process-oriented rules of thumb.

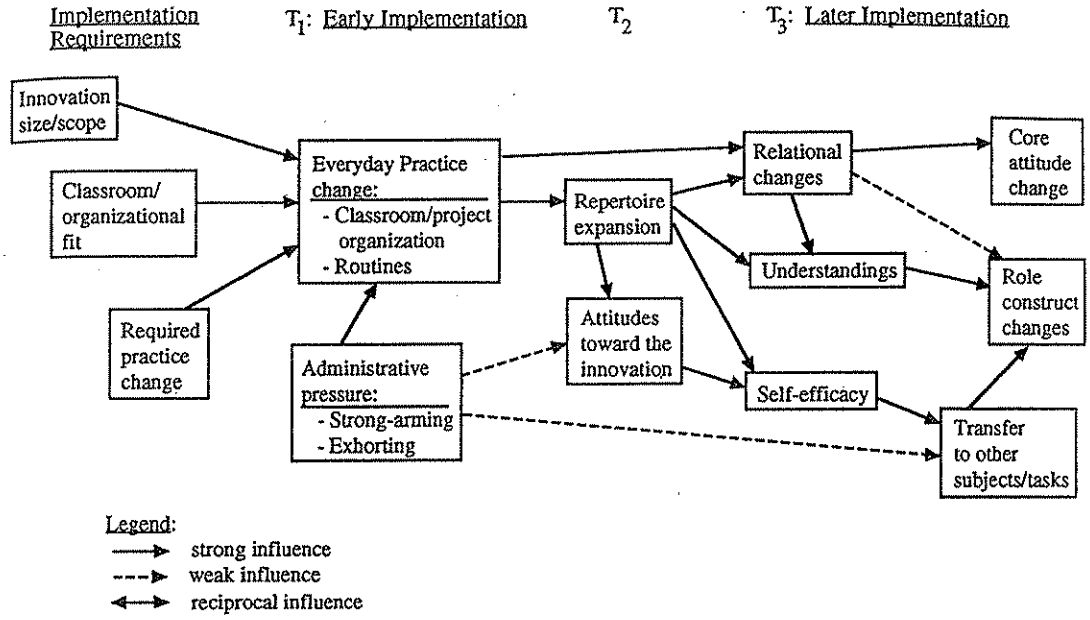
One rule of thumb is to order the model *temporally.* So the question is, Which variables of those found relevant occur first in time, which occur along the way during implementation, and which might be seen as early and later outcomes?

The second rule of thumb is to consider which variables might reasonably be expected to have a direct *impact* on other variables, both preceding them in time and having a plausible direct connection. This rule means thinking of the mechanisms that could be involved.

The third rule is to *check case informants' explanations.* What causal linkages do they claim are present? Consult the field notes.

The fourth rule is to consider what available *research and theory* have to say about causal connections. In this case, past studies of implementation and individual

Figure 8.2
Causal Flowchart Tracing User Practice Changes



learning theory might suggest relationships among variables.

Using these rules, you have to noodle for a while. As with other network displays, it helps to put variables on cards, move them around into various configurations, and look at the connections that seem sensible. If you have some user-friendly graphical software, such as MacDraw, Inspiration, or Meta Design, use that. (Network programs that work directly from your codes, such as ATLAS/ti, or programs that let you test propositions, such as HyperRESEARCH, are also very helpful.)

And because the process is a creative, synthesizing one with a good dose of serendipity thrown in, it definitely helps to have a colleague look at the first few versions.

Figure 8.2 shows the model that finally emerged from this process. Let's look at some of its aspects. First, the analyst concludes that the innovation's "implementation requirements" are logically prior to what happens when it is implemented. So he places *innovation size/scope, classroom/organizational fit,* and *required practice change* as antecedents of some aspects of user practice change.

But which aspects? From Table 8.9 we can suggest that the easiest (hence possibly earliest) things to change are *classroom routines.* The analyst, reflecting a moment and reviewing case data, also realizes that such routines are being changed, not by magic, but by the form of organization of the project and the way it is carried out in classrooms. So those are drawn in as the immediate "user practice" changes during early implementation. The model is growing.

Where should *administrative pressure* fit in the model? It occurs mainly during early implementation, the field notes show, and the analyst reasons that it probably has a direct effect on classroom and project organization and users' routines: The pressure and exhortation serve to keep the immediate short-run practice changes in place. Prior implementation research tends to support this linkage (see Fullan, 1982; Fullan & Stiegelbauer, 1991).

The analyst also postulates that the pressure has weaker, later effects as well (dotted lines): It encourages positive *attitudes to the innovation* itself (through the exhortation and reinforcement) and makes it more likely that users will *transfer* their learnings to other subjects and teaching tasks. (The analyst is assuming a mechanism: The pressure keeps user learnings salient and visible so that transfer to new situations is more likely. Assumptions like these must (a) be made explicit as the work proceeds and (b) be checked with the data and your colleagues.)

Once the initial user changes in practice have taken place, what is a reasonable model that would get us to the other, "deeper" outcomes noted in Figure 8.2? It almost surely is not a linear chain, but a network. We have to consider how succeeding types of changes in teachers might influence and feed each other. Here again the four rules of thumb apply. The analyst also has the outcome orderings from Box 7.2 as an empirical start.

The analyst says that the project's *routines*, given some teacher experience with them, lead to *repertoire expansion* (teachers have new things they know how to do). Expansion of repertoire is satisfying (increased skills usually feel good), so we might expect increased positive *attitudes to the innovation*. Checking back with the original case report and the field notes gives (a) clear examples supporting those causal links and (b) no examples of disconfirmation (tactic: **making if-then tests**).

These two practice changes have further consequences. An expanded repertoire leads to *relational changes* (e.g., more willingness to share one's new knowledge with other teachers); in some cases, the analyst notes, the innovation's routines themselves (e.g., team teaching or work-experience programs) bring people into new working relationships. But the repertoire expansion also deepens teachers' *understandings*—of classroom dynamics, of their own and pupils' roles, and, in some cases, of school and district dynamics. Finally both repertoire expansion and positive attitudes to the innovation lead to increased *self-efficacy* (put schematically: I am more skilled in doing this good new thing, therefore I feel good about myself professionally). These formulations, once again, need to be checked through cycling back to Box 7.2 and Table 8.9—and to the analyst's knowledge of other research.

How do we get to the final, "deepest" set of outcomes? First, the analyst postulates that after a good deal of implementation, *core attitude* changes (e.g., "I learned to let go of the child," "I was learning to be flexible," "How to trust students when you have no control over them") come essentially from the teacher's working relations with peers, as well as with students themselves.

Second, the analyst proposes that *role construct* changes (e.g., reconceptualizing "structured" teaching as productive, rather than authoritarian) come essentially from basic understandings, and from the successful experience of transfer to other subjects and tasks. (The assumption is that transfer has a "cosmopolitanizing," "generalizing," and possibly "reinforcing" aspect.) *Transfer,* itself, comes mainly from self-efficacy, in this model: The better I feel about my competence, the more likely I am to try my new ideas and practices in other aspects of my work.

Thus endeth the explication. Another analyst could well have come out with a somewhat different model. The rules of thumb may be weighted differently, or they may turn up alternative, equally compelling accounts. For example, you might postulate that core attitude changes are unlikely if no change has occurred in self-efficacy: They cannot flow solely from repertoire changes and administrative pressure. But this is not a major change in the model; independent analysts using the same rules of thumb, the same database, and assumptions that are not wildly different will be likely to come up with a similar causal picture.

You should subject the "final" version to verification, most easily through the depredations of colleagues (including earlier analysts, if any), who can help you clarify the assumptions you are making and suggest alternative views (tactic: **checking out rival explanations**).

And, given a causal model with which you are reasonably satisfied intellectually, return once again to the written-up field notes for evidence of disconfirmation or needed revision.

Variations

It sometimes helps to make submodels that are wholly linear ($W \rightarrow X \rightarrow Y \rightarrow Z$) as a simplifying strategy (see discussion of "causal chains" below). It's also useful to consider a backward mapping approach: You begin with final outcomes and reason back along the causal byways (Which variable, in principle, would have to be changed in order to induce this change?).

This particular outcome variable (user practice change) had many parts, which could be seen plausibly as influencing each other. If you are interested primarily in processes, and if the outcome variable is a straightforward, simple one (e.g., percentage of teachers using an innovation in a school), then the modeling will naturally ramify on process issues.

Advice

Expect to do several versions of the model. Use a simple technology (cards or software) that permits flexibility in rearrangement. Get advice and argument from colleagues. Don't close up too fast.

Return repeatedly to field notes to check, test, and extend the model. It usually helps to cut back to the preceding matrices, with a good look at **outliers** and **extreme cases.** When a specific, well-understood case does not fit the model, CHANGE THE MODEL to accommodate that information, rather than try to explain away the "inconvenient" information.

It often happens that you feel blocked, unable to understand how variable A could lead to variable C in any plausible way. For example, in the school improvement study, we started a causal model that led from "size of project funding" to "organizational change." (More expensive projects were also those that transformed the local school and district.). That reminds us of a half-truth from the Reagan era: "You can't change [. . . x . . .] by throwing money at it."

So what does money actually do? There must be one or more other variables in the picture that it affects, and they,

Box 8.4
Causal Chain: Illustration

Individualization/pupil centeredness → pupil self-pacing, self-organization → inability to monitor (in class, off campus) → development of trust → letting go → changed self-image as a teacher

in turn, affect the outcome. In our example these turned out to be features of implementation, such as administrative support (which was greater for more expensive projects). Use the tactic of **finding intervening variables.**

The classical way of testing a model is to apply it to fresh data, new cases. If you have such data easily at hand, wonderful. You also might make a braver and riskier choice that involves a basic design decision at the beginning of the project: holding out one or more cases from the initial analysis, without analyzing them. The causal model now can be tested cleanly (tactic: **replicating a finding**). This approach is usually difficult or unfeasible if your cases are few in number and complex, as was the case in our school improvement study. You cannot avoid partial analysis during data collection itself, even if you resolutely decide otherwise. But if your cases are large in number and less complex (e.g., interviews with individual students about their choice of careers), it is much more feasible to have a "hold-out" sample for model verifying.

Time Required

If the preliminary display work has been completed and a list of strong candidate variables for the model assembled, a causal model of the size shown here can be put together quickly in its first version: an hour or less. But the dictum here comes from Norman Mailer (1959): "Do not understand me too quickly." Discussing the model with colleagues, cycling back to the displays and the field notes, and revising may take another 3 or 4 hours.

Causal Chains

During the early stages of causal modeling, it helps to make simplifying assumptions about what leads to what, placing causes and effects in a linear chain. Box 8.4 shows an analyst's effort to understand how teachers using an innovation that stressed individualization and pupil-centeredness came to experience a change in themselves they called "letting go."

Such a chain helps the analyst lay out explicitly what may be causing certain phenomena. Although the chain does represent a simplification, that very simplification carries with it the seeds of a fuller explanation. As the analyst remarked here:

> We have left out some steps. For example, pupil-centeredness also leads to relational closeness and to a better grasp of individual ability levels and emotional states. Trust then develops as a result of the bond between staff and pupils—"trust violation" was a common leitmotiv at many sites. There is something like an implicit contract between parties that pupils going off campus will not "betray" their teachers in return for pastoral care . . . [or that] in classroom-bound projects pupils will complete their work and do the requisite exercises or mastery tests in return for individual help with difficult concepts or operations.

Such chains can be used to study less constructive processes. Box 8.5 presents one chain explaining why an initially ambitious innovation eventually was discontinued, with energy expenditures dwindling steadily.

Box 8.5
Causal Chain: Illustration

Initial difficulties → annoyance → resignation (to redesigning it, investing energy in it) → guilt (over the decision to cut back) → futility (of spending time on it)

Causal chains are useful just because they require little elaboration or textual explanation. They are a rapid, simple way to communicate with colleagues (and final readers) about the meaning of—the mechanisms underlying—a process. They must ordinarily be elaborated and linked with other chains to form causal models, as in this section, or to build within-case causal networks.

F. Causal Networks—Cross-Case Analysis

Analysis Problem

In Chapter 6, section D, we show how to make an inclusive, explanatory analysis of single-case data by using causal network analysis. Typically a causal analysis is a late, inferential—and powerful—step, building from pattern coding, tests of bivariate relationships, conceptual clustering, and predictor-outcome analysis.

Two questions naturally arise: Can we get *meaningful* inferences from loading case-level causal analyses into a cross-case procedure—and can it be *managed?* Can you really juggle a dozen such networks, each containing up to 30-35 variables that are put together in case-specific ways? The answer to both questions is yes.

First, doing a cross-case analysis with a core list of variables found to have significance across several cases is a powerful way to move from case-specific explanations to findings that bridge to the discovery or reinforcement of constructs. At the same time, we can identify and account for local particularities. We are going from local causality to clusters or "families" of cases sharing important attributes. As Maxwell (n. d.) argues forcefully, "causal explanation is a legitimate goal of qualitative research, and one for which qualitative methods have some unique advantages" (p. 1).

As for the manageability of such an analysis, we've found that the basic operations are very similar to those used in single-case causal networking. Such an analysis, as with single-case analysis, is the ultimate step after the researcher has built from partially ordered to case-ordered meta-matrices and scatterplots, then to multiple-case effects matrices and predictor-outcome matrices. In brief, it can be done.

Brief Description

Cross-case causal networking is a comparative analysis of all cases in a sample, using variables estimated to be the most influential in accounting for the outcome or criterion. You look at each outcome measure and examine, for each case, the stream of variables leading to or "determining" that outcome. Streams that are similar or identical across cases, and that differ in some consistent way from other streams are extracted and interpreted. The basic principle is that of developing one or more meta-networks that respect the individual case networks from which they have been derived (Huberman & Miles, 1989).

Illustration

Because cross-case causal networking involves manipulating several sets of boxes and arrows at the same time, it can be confusing to explain. So let's break down the process into successive analytic steps. We'll work from a familiar example, and take an outcome variable that lends itself easily to cross-case analysis.

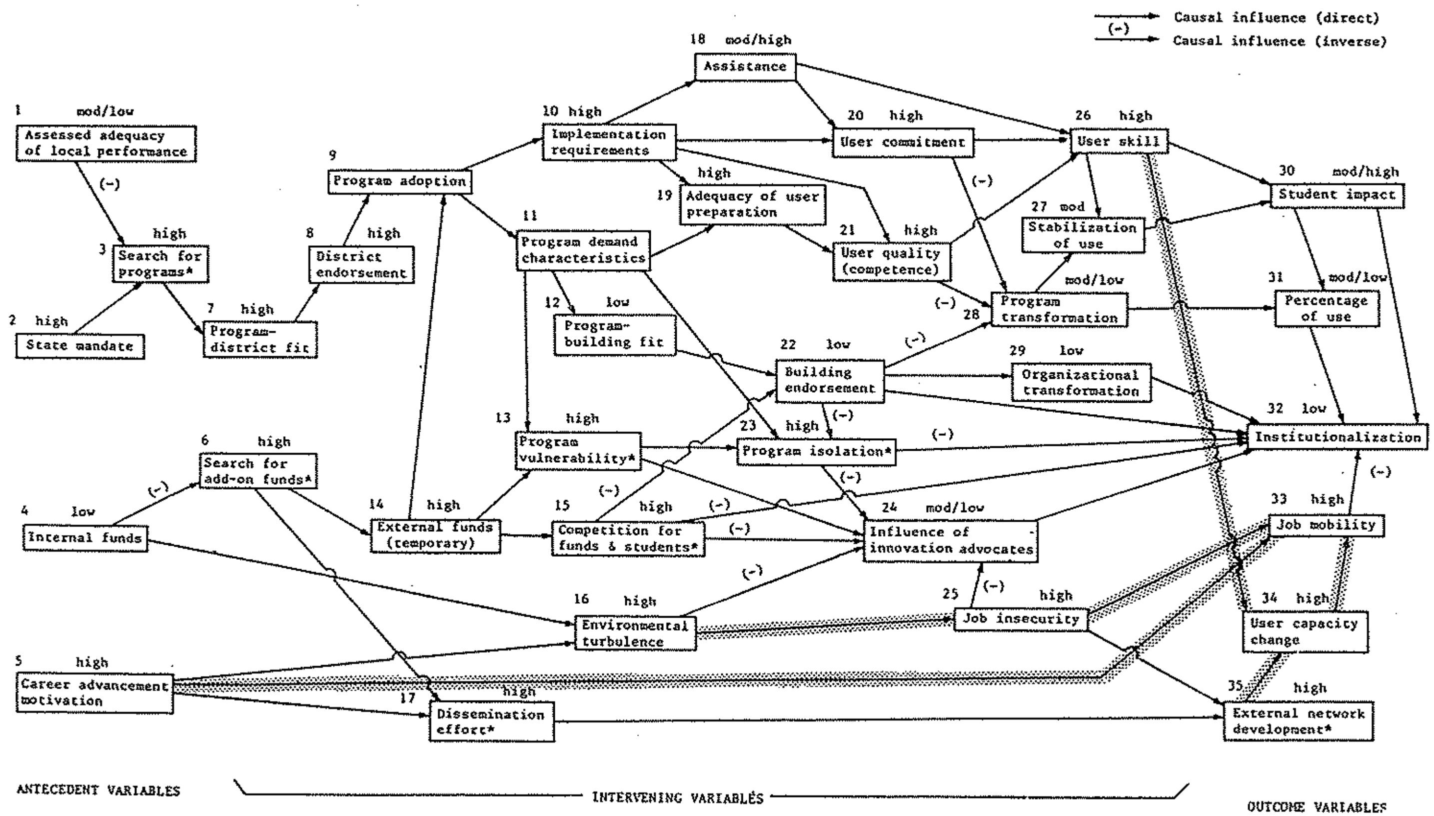
Step 1: Assembling the causal networks. We will assume that you have the basic building blocks discussed in Chapter 6—that is, the list of core variables (see Figure 6.3, p. 157) and the causal network (see Figure 6.1, p. 154), along with the narrative (see Figure 6.6, p. 161), for each of the cases. In our school improvement study, for example, we had 12 such networks and narratives, incorporating a core list of some 35 variables, together with a half-dozen case-specific variables. It helps to have the networks up on a wall or another display surface.

Step 2: Isolating the causal "streams" for each case that lead to the dependent variable being analyzed. This is the major task. Let's assume, for purposes of illustration, that you want to see how an outcome called "job mobility" is determined in a study of people's role and career shifts. We begin by looking at the predictors of that outcome for one case. We start with an exhibit we have seen already, the causal network for the Perry-Parkdale case (Figure 6.1, p. 154). Figure 8.3 shows the network, marked up to help our analysis. The "job mobility" box is at the far right, midway in the list of outcome variables (box 33). It is rated "high"; there was a lot of shifting of staff out of the project, to different roles within the district, to similar roles in another district, or out of education altogether. (To see how this rating was made for Perry-Parkdale, take a look at Table 7.16, p. 201.)

The job mobility box has three arrows leading to it from boxes 25, 34, and 5. Two of those boxes, in turn, have others leading to them along what we have called a "stream." The streams leading to "job mobility" are easy to see if you use a highlighter marker to draw backward from box 33 to preceding boxes. It helps to go only two steps back from the outcome measure. The boxes reached in that way can be thought of as "immediate" causes of the outcome; earlier boxes on the causal stream are "remote" causes.

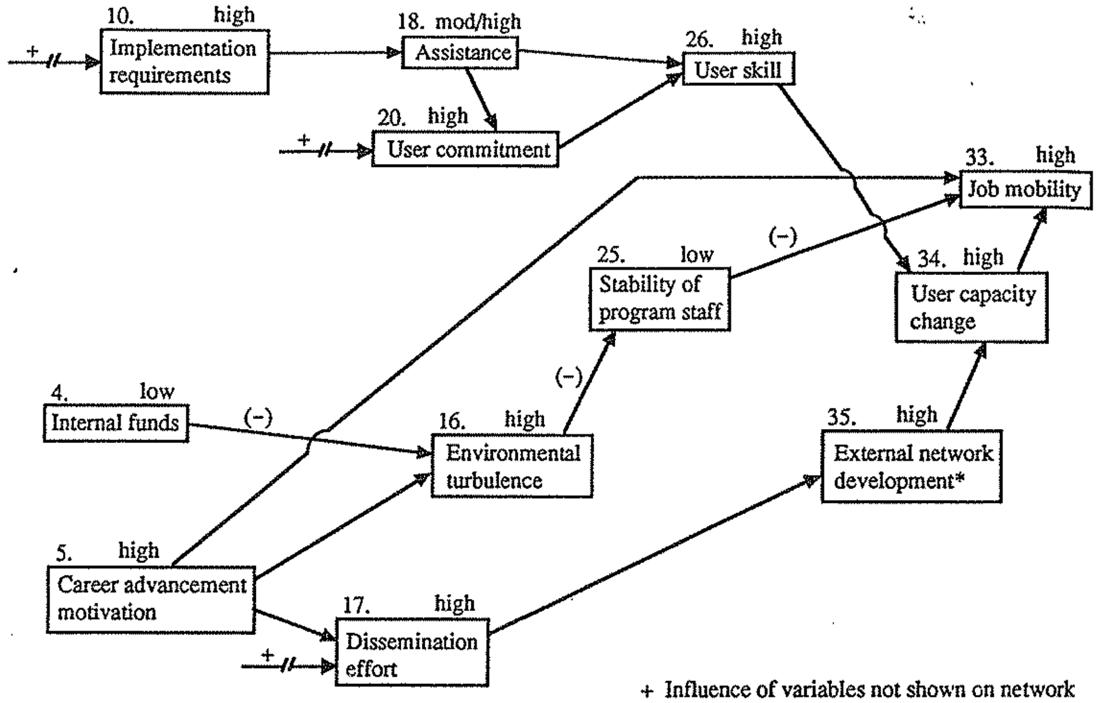
Working from the bottom of the display, let's look at the streams. You go from 5 and 17 (remote causes) on through 35 and 34 to 33 (job mobility). The next is from 5 directly

Figure 8.3
Causal Network for Perry-Parkdale CARED Program
(immediate causes of job mobility marked)



*Variable specific to this site

Figure 8.4
Subnetwork: Variable Streams Leading to High Job Mobility, Perry-Parkdale Case



\+ Influence of variables not shown on network
* Case-specific variable
(-) Inverse causal influence

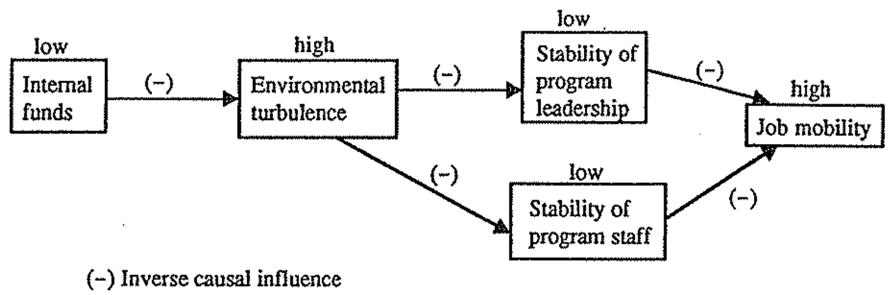
to 33. The third stream is 5, 16, 25, 33. The fourth stream is closely connected to the third: 4(remote), 16, 25, 33. And the fifth stream, taking a reasonable starting point just after program adoption, runs from the remote causes 10 through 18 to 20, and then to the immediate causes 26 and 34 to 33.

To make the analysis easier to follow, let's extract the streams from the causal network, as shown in Figure 8.4. We've pulled out the *subnetwork* of variables leading to the outcome of job mobility. To understand the five streams within that subnetwork, we have two tools. First, we can read across the stream to see what is happening—what the theme or logical succession is. To do that, we have the variable labels and their ratings. For instance, take the substream 5—17—35. The message here is that key actors at the field case were interested in getting promoted (5), which led to a more energetic effort to disseminate (17) the innovation they had developed to others outside their home district (35), where, presumably, some interesting jobs might be found.

To be sure that such an interpretation is plausible, we have the second tool: the causal network narrative (see Figure 6.6, p. 161) the narrative for the case discussed here). The narrative nails down the context, shows the temporal and causal relationships mapped on the network, and explains why the variables are chained as they are.

Let's take up each stream in Figure 8.4. The stream 4—16—25—33 looks ominous: low funds, high turbulence, high job insecurity, low stability of program staff, high job mobility. Reading the narrative confirms our impression that this is a "casualty" scenario: Low funding led to local uncertainties about continuing the project, which caused project staff to be shifted to other jobs. We can clump another stream into the same scenario: 5—16—25—33. It adds the "career advancement motivation" variable and thereby strengthens the "casualty" notion; people did not get promoted or reassigned to desired job slots, even though they had hoped to. But the stream 5—33 qualifies that; some people did get where they wanted via the project. For now, let's call this one an "opportunism" stream.

Figure 8.5
Subnetwork for Job Mobility, Calston Case



Others also made desirable job shifts, but in what looks like a more socially redeeming way. This is shown on the fourth stream: 5—17—35—34—33. The dissemination effort may be fed by career advancement motives, but the process of spreading the project to other districts develops training and consulting skills (capacity changes) that then are used on the next job. Reading the narrative to get a clearer sense of what is happening here yields a "career crystallization" theme: People doing the disseminating realize they want to go on doing this kind of work, rather than return to their former jobs.

Now for the last stream: 10—18 & 20—26—34—33. First, how did we know to *begin* that stream at variable 10 rather than, say, at variable 1 or variable 2, where the stream "really" began? Here, again, we can use our two tools. The logical theme in this stream has to do with project mastery, so you begin with the setting for that mastery process—that is, just after "program adoption" (variable 9 on the full network). And the narrative confirms that hypothesis. The stream itself is upbeat: Local staff take on a stiff project (high implementation requirements), receive decent assistance, develop strong commitment, master the project (high user skill), develop new capacities, and move to desirable new jobs. Let's call this stream "success-driven advancement."[4]

Step 3: Matching the variable streams to other cases with the same outcome. This step can be carried out in two ways, depending on your preferences. The first, case-oriented way, involves "pattern matching" (Campbell, 1966, 1975) on a case-by-case basis; we outline that approach here. The second way, an initially more "macro" approach for those who would like to see how specific *variables* perform across all cases, involves the creation of an *antecedents matrix* (see Box 8.6) showing immediate and remote causal variables in all of the subnetworks for the complete set of cases.

The issue in the "pattern matching" approach is discovering whether a pattern found in one case is replicated in other ones as well, suggesting a common scenario. Are the same patterns of core variables involved? Are the ratings (high, moderate, low) the same? Let's take another high "job mobility" case, extracting the streams leading to that outcome. Figure 8.5 presents the result for the Calston case. This one is easier to contend with.

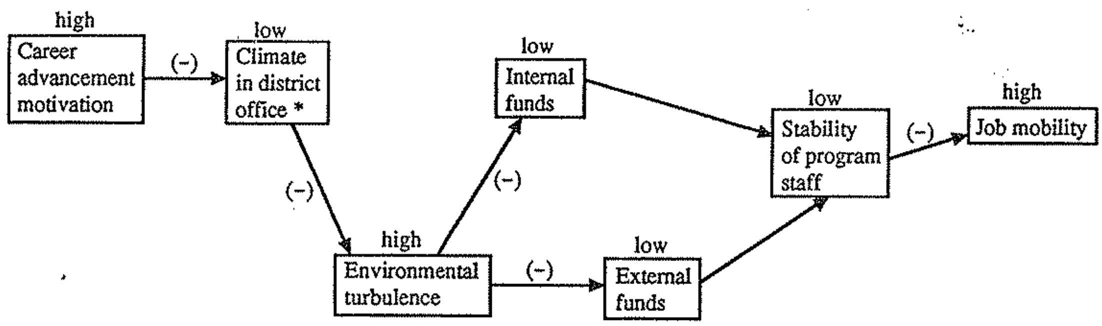
Only two streams lead to the "job mobility" variable, and both highlight the same scenario. If you look back at the "casualty" stream for the Perry-Parkdale case (Figure 8.4), three of the same variables (internal funds, environmental turbulence, stability of program staff) are here. Better still, they are in the same sequence and have identical ratings.

Let's try another case. Figure 8.6 shows the extracted portion for Banestown. This is clearly another "casualty" case. If we take out the case-specific variable ("climate in district office") and the "external funds" variable, we have the same variables, identically rated, as in Perry-Parkdale, although not in the same order; here the "internal funds" variable comes after, not before, the "environmental turbulence" variable. But reading the case narrative produces the same, sad scenario of dashed career hopes; the shift in sequence means little.

Let's look at one more example (Figure 8.7). By now the reader probably can dissect the network more rapidly. Three arrows lead to high "job mobility," but so do five possible streams. The first one—variables 20—18—26—36—37—should be familiar; it is the "success-driven advancement" stream we saw at Perry-Parkdale, at least in the *immediate* causal variables, those closest to the criterion measure: skill, capacity change. The sequence is the same, and the ratings are identical.

Let's try another stream, say, 27—31 & 38—37. This looks like the "opportunist" scenario at Perry-Parkdale, with some more fleshing out (the "stability variables" are

Figure 8.6
Subnetwork for Job Mobility, Banestown Case



* Case-specific variable
(-) Inverse causal influence

the intervening processes at work). Finally, let's take a third stream: 39—38—37. This is the same sequence as the final three predictors in the Calston "casualty" scenario, which itself is virtually identical to the Perry-Parkdale "casualty" scenario. To confirm this, the analyst rereads the causal network narratives for the three cases.

At the end of this analysis, the 40-odd streams for the 12 cases fell into the four families we identified along the way: casualty scenarios, opportunism scenarios, success-driven advancement scenarios, and career crystallization scenarios.

Moreover, the analyst can move up an analytical notch and cluster *cases,* rather than individual streams. For example, both Perry-Parkdale and Plummet have successful and unsuccessful job mobility scenarios; some people get what they are after, and others do not. This also obtained for other cases; we could group them at the case level as "win-lose" scenarios.

Step 4: Verifying the scenarios for similar and contrasting outcomes. It's time to confirm the emerging scenarios. Here are some decision rules you can use to decide whether two streams belong in the same scenario or, to take Campbell's term, have "matched" patterns. We've used seven such rules, of which five are illustrated above:

1. All (or all but one or two) of the core predictor variables on the stream are the same.
2. The most immediate predictor variables—the two or three closest to the outcome measure—are the same, and are in the same sequence.
3. The common predictors have the same ratings (high, moderate, low).
4. The outcome theme is the same (e.g., "casualty," "success-driven advancement").
5. The narrative confirms the similarity or identity of the outcome theme derived from the stream of variables in the network.

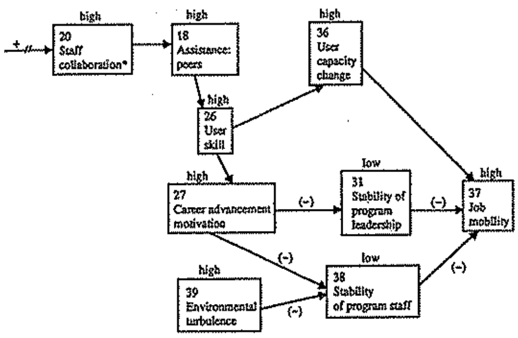
Let's note in passing that the analyst is using three tactics to determine similarity of pattern. First is simple **counting,** which can be done rapidly with the aid of antecedents matrices (Box 8.6): number of identical predictors, number of similar sequences, identity of ratings (rules 1-3). Second is matching of outcome themes from a stream of predictors—in short, getting a plausibly similar **pattern or theme** from two sets of four or five variable flows (rules 1-4). Third is **comparison** of the outcome themes with the narratives—again meaning matching units of two different sorts (rule 5). So we have both somewhat "arithmetical" and more structural or "qualitative" forms of analysis operating together.[5]

The two last decision rules round out the picture:

6. The outcome themes are different (or absent) in cases with a differently rated outcome variable (e.g., high or low job mobility).
7. In these differently-rated cases, the predictor variables closest to the outcome variable are different or, if the same, are rated differently.

We used these last rules to complete and test the analysis (not shown here). They bring home the point that, in a sample of cases, we have not only comparability but also *variability.* There are moderate and low job mobility cases, and we have to be sure their causal flows are either differ-

Figure 8.7
Subnetwork for Job Mobility, Plummet Case



* Case-specific variable
\+ Influence of variables not shown on network
(-) Inverse causal influence

ent or contrary to the ones in the high job mobility cases. Ideally the low cases would have some of the same *variables* in the same or similar *sequences* as for the high cases, but the *ratings* would be different. For example, high internal funds, combined with low environmental turbulence and high staff stability, would lead to low job mobility—the opposite of the casualty scenario, but with the same variables in the same order. That would strengthen the analysis and increase its explanatory power. We now can claim with more confidence that job mobility is caused by a combination of funding, turbulence, and staff stability, because both high and low mobility result from the same interaction of these variables.

Antecedents Matrix

Sometimes it's easy to get lost in the welter of stream-by-stream matching across cases. You need to get a general fix on the variables that seem to be leading to the outcome in question.

An antecedents matrix is ordered by the outcome variable, and shows all of the variables that seem to be at work in inducing changes in it. The easiest way to proceed, after the subnetworks have been identified as in step 2, is to scan through them for the variables that appear on, say, at least one-third of the subnetworks. You may want to add variables that are conceptually important as well. Box 8.6 shows an example of an antecedents matrix for the outcome variable "student impact." Note that the variables are *clustered* conceptually under general headings to make understanding easier.

An "immediate" cause is considered to be not more than two steps back from the outcome. A "remote" cause is still on a stream connected to the outcome, but farther back. A remote or immediate cause should have its strength (high, moderate, or low) entered. Some variables will either not be on a direct stream, or not be on the network at all; leave those cells blank.

During analysis it helps to take each column, one by one, and summarize the finding. (Try this and see what you find in the matrix.)

For example, we can note that "user commitment" is high and immediately causal in three of the five high-impact cases, and a remote cause in the other two; is a high or moderate remote cause in the moderate-impact cases; is not causal at all in moderate/low cases; and is low in low-impact cases. So user commitment, even though it is often a remote cause, has a real bearing on eventual student impact. After all of the variables have been reviewed and the analytic text written, it helps to make a general summary. Here is one for Box 8.6:

> It appears that high student impact is more likely when user commitment is high, when program transformation is either held to a minimum or is a corrective to an initially overambitious start, and when strong assistance takes place, resulting in practice mastery and stabilization of use. The presence of administrative pressure is a remote cause of student impact—but only, it appears, when it translates into willingness to restrict program transformation, to supply strong assistance, or do both.

What we have here is an overview statement—which could be used, in fact, to construct a preliminary cross-case causal model (see section E). We now have an idea of what variables are most powerful antecedents of the outcome. But such a model would only be a smoothed-down, averaged version of what really happens case by case. You now need to begin the process of looking for "families" of cases that share the same scenario.

To do this, you can make case-by-case **comparisons** across the columns, looking for similar **patterns**. It helps to start with cases that have the same general level of outcome (e.g., the high-impact ones). When the matrix suggests a plausible pair, examine the original subnetworks to see whether they and their component streams are a reasonable match. You are using the same pattern-matching operations we've just described, and now have gotten to the same result—a set of coherent scenarios explaining what happened in the cases. As before, they need to be verified (step 4).

The antecedents matrix is an easy way to see the big picture without getting swept away in myriad specific streams and going under for the third time. The main thing to remember is that the matrix takes apart the causal flows,

Box 8.6
Antecedents Matrix: Antecedents of Student Impact

	PRESSURE	COMMITMENT	DEMANDS/FIT						ASSISTANCE	STABILIZATION	
SITES	Admin. pressure	User commitment	Innovation size/scope	Implementation requirements	Program-school fit	Program transformation	Innovation change scenario	Organization transformation	Assistance	Practice mastery	Stabilization of use
High impact											
Perry-Parkdale (E)		high →	mod	high ⇢	low ⇢	mod →	overreaching		mod/high →	high →	mod/high →
Plummet (L)	high ⇢	high →	large	high ⇢		high ⇢	overreaching	high	high →	high →	high →
Tindale (L)	high ⇢	high →	mod/large	high ⇢		low →	enforcing	high	high →	high →	
Masepa (E)	high ⇢	high →	large	high ⇢	low ⇢	low →	enforcing		high →	high →	
Banestown (E)		high →	small/mod	low ⇢			salvaging		high ⇢	high →	
Moderate impact											
Carson (L)		high ⇢	large	high →	low →	high →	overreaching	mod/high →	mod/low →		mod →
Calston (E)		high ⇢	small	low/mod ⇢	high ⇢	mod →	refitting		low (admin.) high (self) →	high →	high →
Lido (E)		low ⇢	small/mod				refitting		high ⇢	high ⇢	high →
Low to moderate impact											
Dun Hollow (L)			small			low →	refitting			high ⇢	
Astoria (E)			large	high →	low →	high →	(none assigned)				
Low impact											
Proville (L)	high ⇢	low ⇢	mod		low ⇢	high →	salvaging		low →		low →
Burton (E)		high →	small/mod	low ⇢	low →	high →	refitting		low ⇢	low ⇢	low →

Legend: → = immediate cause ⇢ = remote cause. Blank spaces indicate not a cause, or not on network.
(E) externally-developed innovation (L) locally-developed innovation

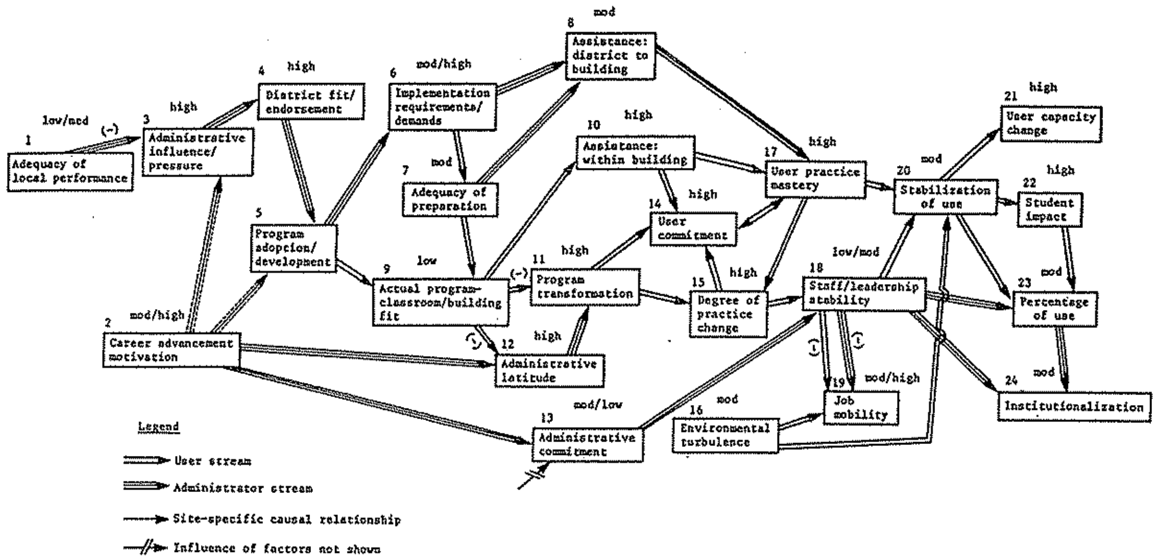
and genuinely powerful conclusions cannot be drawn from it. You must always go back to the actual local flow of causality and sort networks into families, not individual variables, out of sequence. Doing an antecedents matrix adds to analysis time, but it often saves you from the overload and incoherence that are likely when you are doing stream-by-stream analysis, especially with more than a few cases.

Variations

We consider cross-case causal networking to be in its adolescence, and expect others to improve on it as they work out alternative techniques of drawing conclusions (see the work of Leithwood et al., 1991, Box 6.5, p. 163). Plenty of questions still remain. For instance, it isn't clear whether you would get more than one or two replications of a causal stream with an N of only 5 or 6 cases, or whether these procedures are cost-effective with an N of more than 15-20 cases. The cross-case configural analyses possible with the programs QCA and AQUAD (see Appendix) are helpful with large numbers of cases, although they can't handle the multiple pathways that network displays give us, and are restricted to 12 variables at a time.

The next obvious threshold is cross-case analysis of several outcomes at once. We have tried our hand at this with reasonable success. For example, using the same analytic procedures and decision rules, we derived four multiple-outcome scenarios for the 12 cases in the school improvement study. Figure 8.8 is a causal network grouping four cases into one scenario (with a total of five outcome variables). Here the analyst, discarding case-specific variables, could map consistent flows across the four cases. They were defined as a "family" sharing a common scenario, because they had high student impact and user capacity change but only moderate percentage of use and institutionalization. (The consistent mapping is achieved, in part, by accepting correspondences such as "moderate/low"

Figure 8.8
Scenario 2:
Moderate/High Outcomes From High Mastery and Low Settledness



rather than a harsher criterion such as full agreement.) The critical factors in the scenario appear to be high user mastery accompanied by only moderate stabilization of use. Note that the analyst also found that causal flows were somewhat different for users and administrators, suggesting that role-specific causal networks could be generated as well.

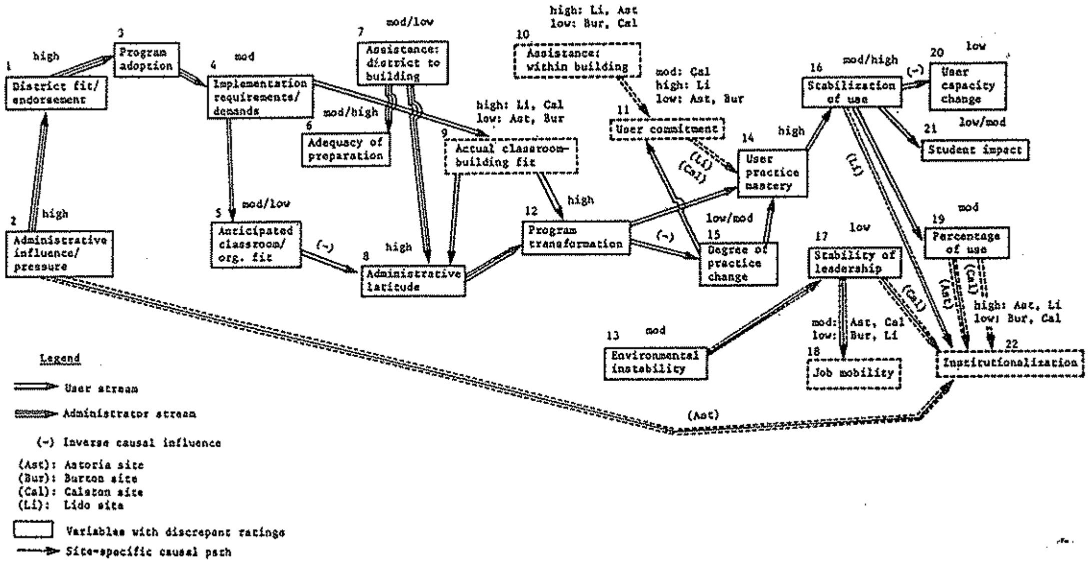
Not all multiple-outcome scenarios come out so nicely. In Figure 8.9 four cases are grouped in one causal network with five outcome variables. As before, there are some distinct causal streams for users and administrators. But in several instances, either the ratings or the paths are discrepant among the set of four cases. For example, the variable boxes surrounded by broken lines (e.g., 9, 10, 11) have discrepant between-case ratings. Also some of the paths are case-specific (e.g., those linking variables 19—22, 11—14, 17—18). Still, the overall scenario is plausible (moderate to low outcomes are likely when administrators allow users a lot of latitude (8) to transform the innovation (12) to make it "easier to use"). Even though the innovation is used stably (16), its "downsizing" results in less capacity change (20) and student impact (21), along with lower percentage of use (19) and institutionalization (22).

What about the between-case discrepancies on this multiple-outcome network? On balance, they are far fewer than the correspondences. Furthermore, the discrepancies can be shown directly on the network itself without overloading the display—and the reader can get back easily to the original case network to understand more.

These are ambitious variations. How about more modest ones? You can think of causal networking as a rapid, rough-cut technique for finding out how well you understand your data or, with a research team, the degree of agreement about the core variables in play. This exercise is quite useful about three-quarters of the way into the study, just before the final analyses.

Let's take the lone researcher first. You sit down and say: What do I think the main causal stream looks like in my data? A half-hour later, you have sketched the hypothesized stream. A good half-day later, you have surveyed the available case material—pattern codes, interim summaries, memos, display-related text—and know how plausible this emerging model is and how it could be confirmed and strengthened. Because you usually have another shot at the field, your candidate causal stream can be put to the test empirically. This procedure is easier to manage when you've begun with or evolved a conceptual framework. You're taking a core "stream" of predictors, moderator variables and outcomes, and seeing whether it aligns with the data collected and analyzed up to now. It's a worked-out cross-case version of the causal fragments we discussed in Chapter 6, section D.

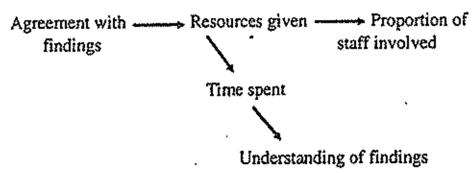
Figure 8.9
Scenario 3:
Moderate/Low Outcomes From Program Blunting/Downsizing



Now for a multiple-researcher example. Huberman and Gather-Thurler (1991) studied the dissemination of findings from 11 research projects to 23 "target publics" in the domain of vocational training. In trying to understand whether these organizations were likely to be receptive (or not) to the research findings, a set of key variables began to emerge:

time devoted to the study of the research findings
resources devoted to the study (time, consultants)
proportion of staff knowledgeable about the study
agreement of study findings with espoused organizational goals
understanding of study findings

Each of three researchers drew and explicated a causal stream by using these variables, and the individual streams were compared. Each version was slightly different, but there were several intersections. The discovery of those intersections drove the next—and final—wave of data gathering, to determine whether and exactly how they were borne out. Incidentally, here is the stream with the most commonalities across cases:



Advice

Cross-case causal network analysis, using a simple set of inferential procedures, can yield explanatory accounts that make good conceptual and empirical sense. They are a good example of linking variables and processes. Things get more unwieldy when more than one dependent variable is cranked into the analysis, but the use of the same decision rules can produce equally compelling findings.

Cross-case causal network analysis is feasible mainly because you include attention to the micro-analytical level, such as individual "paths" or "streams." It is important to see the big picture, but the more general insights gained from an antecedents matrix, as in Box 8.6, ultimately have to be run back through the lens of case-specific streams. You can't lose sight of the web of *local causality*.

An important corollary, however: This exercise rests on prior cross-case analyses. If the causal network findings are inconsistent with, say, the case-ordered predictor-outcome analyses or the case-ordered effects matrices, something is wrong. The variable-oriented cross-case analyses should lead inferentially to the more encompassing pattern matching of cross-case causal network analysis. The network analysis has to make sense when you check back to the matrix work.

A few more specific bits of advice:

Always keep the network narratives close at hand. Case accounts with the same set of core variables on a similarly sequenced stream can actually mean different things.

Don't try for a theme or scenario name until you have looked at several cases. A stream usually has a stray predictor or two, and you need to wait until these strays can be separated out from those variables appearing consistently, which will give the name to the stream.

Be cautious about giving tacit *weights* to individual network variables. Presumably the ones with more arrows leading to them are more significant, but a stream-level analysis cancels this out. In fact, a variable with too many arrows leading to it becomes a nuisance, because it multiplies the possible number of streams. Unlike path-analytic models, our networks have no regression coefficients to show how important any substream (two or more connected variables) is. We do, however, have the *narrative*, which usually tells the reader which factors count most and which relationships are especially meaningful. So, once again, the analyst should keep the analytic text and the displays together, moving back and forth between them.

It's a good idea to have a second person replicate the analysis by using the same networks, narratives, and core list of variables. It's easy to slip into believing that streams from two cases are the "same"; a colleague may well note an obstinate exception ignored in the first run.

If a cross-case scenario doesn't square with the case network narrative or the case report material, the scenario needs to be redone. The scenario is your last product, and it rests on your earlier work. Don't let the scenarios get reified and disconnected from the analyses they were based on.

If the person doing the cross-case analysis did not do the case-level causal network, the cross-case account should be fed back to the single-case analyst to verify that his or her case has not been misrepresented.

Remember, finally, that you're moving from the locally causal to the translocally causal, in order to derive a wide-ranging, more generalizable explanation. That explanation has to make not only local or context-informed sense but also *theoretical* sense. It should fit with an extant or emerging account of why phenomena like these, seen across multiple instances, are likely to come together as they did.

Time Required

Although it may look forbidding at first, cross-case causal network analysis turns out to be pleasurable and not as time-consuming as you would think. For example, the *single-outcome* "job mobility" analysis discussed here took about a working day from the time the analyst began work with the 12 networks and narratives. Steps 2 and 3 (isolating and matching streams across cases with similar outcome levels) took about 4 hours. Step 4 (double-checking the decision rules, looking at cases with contrasting outcomes) took about 3 hours. Revising to final text finished the day.

The alternative strategy using an antecedents matrix takes more time. Marking the networks, creating the matrix for 12 cases might take 3 or 4 hours, drawing conclusions and writing analytic text another hour, and then moving back to the clustering of cases into common scenarios, extracting prototypical subnetworks, and writing final analytic text another 5 or 6 hours.

Doing cross-case network analysis is enjoyable; you're making sense of things in an integrative way, with the "aha" rate never being zero. The price of the enjoyment, though, is a certain obsessive care. You have to take the time to (a) exhaustively account for the full set of cases and (b) back out of certain dead ends. Dead ends are streams that make no sense and need to be broken up. Even though they seem common across cases, they do not square with the narratives. You have to tell yourself that this is a necessary part of an analytic process that is bound to lead somewhere—ultimately.

The full set of cases has to be accounted for using several scenarios—a far more stringent criterion than in probabilistic analyses. If, let's say, the predictor ratings or sequences for networks with low levels of an outcome measure, as contrasted with the high cases, do not meet decision rules 6 and 7 (consistent contrasts), you must find out why. The question can usually be resolved in the narratives, but it takes time.

Taking on *multiple-outcome* networks is labor-intensive, to put it gently. Figures 8.8 and 8.9 took roughly 2 days each to derive. Plan on interspersing this work with physical exercise, or else you may become seriously disoriented with friends and relatives. Also expect to have dreams with unceasing arrays of boxes and arrows that never quite interconnect. . . .

Summary Comments

We've been looking at how to derive good explanations, better theories, by working across a set of cases. The working principles are to understand your cases well, avoiding aggregating or smoothing them, and keeping the local case configuration of variables intact. Both variable-oriented and process-oriented strategies are needed; they complement each other. Matrices tend to work more easily when the focus is on variables; networks give the process view more fully. Cases can often be sorted into explanatory "families" sharing common scenarios. But it's important to look carefully at deviant cases and not to force cross-case explanations.

By our count, we have described something like 27 matrix types in this book. In the next chapter, we move back of this diversity to offer some general suggestions about matrix building and use.

Notes

1. The best introductory overview of qualitative comparative analysis appears in Ragin (1993); computer programs that accomplish it include QCA and AQUAD (see Appendix). Note, however, that the method is not strong in arraying variables temporally, so it falls short of linking "concepts" to "stories."

2. For a thorough discussion of what conceptually driven "forcing" means, see Glaser's (1992) sharply critical discussion of Strauss's (1988) and Strauss and Corbin's (1990) views of grounded theory. He clarifies his own strongly inductive, emergent approach as the true version of grounded theory development, and suggests that Strauss's later work should be called "full conceptual description."

3. The best case for causal modeling with qualitative data has been made by Maxwell (1984, n.d.); see also House (1991) and Salomon (1991).

The literature on various forms of *quantitative* causal modeling is large, and goes back 30 years. Some useful treatments include Blalock's early classic (1964) and later overviews (Asher, 1983; Blalock, 1971); James, Mulaik, and Brett's (1982) material on confirmatory analysis, with a practical training emphasis; Berry's (1984) text on nonrecursive modeling; Fornell's (1982) review of "second-generation" approaches to quantitative causal modeling (including LISREL, PLS, etc.); and Bollen and Long's (1993) edited collection.

Developments in quantitative causal modeling have done much to help researchers manage and understand multivariable data in a coherent, non-self-deluding way. Much of this literature, however, deals with technical problems of measurement and data manipulation, and little of it deals with actual model building, or how you move from a list of "associated" variables to an integrated, causally coherent picture (see Davis, 1985, for help on this).

Excessive preoccupation with technical problems can lead to final models that are quantitatively sound but make no sense, leading to stretched, forced explanations. The strength of qualitative causal models is that they are well grounded in local temporal reality, with all its complexities. And they can always be run back against events in the field case for verification.

As noted earlier, "system dynamics" approaches, modeled by computer, are especially useful because they can incorporate feedback loops and study model behavior over an extended time period. See Gaynor and Clauset (1985), Clauset (1989), and Clauset and Gaynor (1992) for some informative examples in education.

4. Note that, for simplicity, we have omitted another associated substream in Figure 8.3 (the full network): 10—19—21—26, thence to 34 and 33. Early user preparation and selection, producing competent users (even without assistance) was also a factor in the level of user skill attained.

5. Much of the inferential "pattern matching" that qualitative analysts do within and across cases involves the kind of simple **counting** and **contrasting** in rules 1-3 above: noting that some things always happen together, that others happen more or less often in similar sequences, that there are several instances of X and only one or two of non-X, and so on. Most qualitative analyses include, implicitly or explicitly, such estimates of frequency, intensity, and covariation. In fact, you cannot get to rule 4 above (comparing outcome themes) without the preliminary counts and comparisons of rules 1-3. So there is plenty of simple computing in qualitative work, just as there is constant inductive sense-making in quantitative work (e.g., as in factor analysis or cluster analysis procedures).

9

Matrix Displays: Some Rules of Thumb

Matrices are so pervasive throughout this book that we would like to offer some summarizing suggestions for building and using them. Not everyone loves matrices—and not everyone thinks visually. But as we argue in Chapter 5, section A, displaying your reduced data in a systematic way has immense consequences for your understanding. It requires you to think about your research questions and what portions of your data are needed to answer them; it requires you to make full analyses, ignoring no relevant information; and it focuses and organizes your information coherently. These advantages are repeated when you include displays in a final report; the reader can re-create your intellectual journey with some confidence.

We noted earlier that displays fall generally into two main categories: matrices and networks. *Matrices* essentially involve the crossing of two or more main dimensions or variables (often with subvariables) to see how they interact. They lend themselves well to a variable-oriented analysis style and can be expanded to a more holistic case-oriented style.

Networks are not dimensional in the same sense, but involve a series of nodes connected by links. They lend themselves well to a case-oriented, syntagmatic approach that re-creates the "plot" of events over time, as well as showing the complex interaction of variables. They give us the kinds of narratives that tend to get chopped up analytically in matrices.[1]

Advice on network displays is harder to synthesize; we encourage you to consult the following "Advice" sections for specific kinds of networks:

Chapter 5, section B (*context chart*); section C (*event-state network*)

Chapter 6, section D (*causal network*)

Chapter 7, section E (*composite sequence analysis*)

Chapter 8, section E (*causal models*); section F (*cross-case causal networks*)

This chapter aims to be user-friendly and is more than a simple reprise. For those whose style is to avoid displays and deal with qualitative texts as wholes, it is a practical overview of what matrices may be able to do for you. For readers who like using displays, and matrices in particular, it's a summary of how you can use them effectively. In turn, we offer ideas on building matrices, on entering data, and on drawing conclusions. The focus is on modest rules of thumb, grounded in our experience and that of colleagues.[2] We also try a small experiment: talking about displays using text alone!

A. Building Matrix Displays

We begin with some hortatory remarks. Matrix construction is normally interesting, easy, and satisfying. It is not arcane, forbidding, or obsessive, but rather an enjoyable problem-solving activity.

Most of the matrix formats described in this book were conceived, literally, in a few minutes. Developing them and working them through was a longer process. But we have repeatedly had the experience, during training workshops using this book, of formulating a research question and some data types and then asking participants to "make a matrix" that would answer the question. The average participant does the task in 15-20 minutes, with a good deal of pleasure—and surprise at how easy it was.

Furthermore that exercise usually produces in a group of, say, 25 people, at least 20 discriminably different matrix designs, each with different types of data entries. *Each* can be used to answer the research questions; *each* has certain advantages and certain limitations (e.g., the variations on a checklist matrix shown in Chapter 5, section B).

That diversity underscores a basic point. There are no fixed canons for constructing a matrix. Rather, matrix construction is a creative—yet systematic—task that furthers your understanding of the substance and meaning of your database, even before you begin entering information. Thus the issue is not whether you are building a "correct" matrix, but whether it is a *helpful* one that will give you reasonable answers to the questions you are asking—or suggest promising new ways to lay out the data to get answers. At a deeper level, the message of this book is not "Use these matrices," but "Think *display,* and invent formats that will serve you best."

We also reemphasize here that the names we gave our illustrative matrices were for convenient description, not a means of reification. Who cares whether it is a "time-oriented meta-matrix" or a "predictor-outcome-consequences matrix"? The issue is what a matrix you have built does for your *understanding* of your data.

Matrix Elements

Here's a simple checklist—aspects of matrices about which you can make choices during the design process. Essentially we're considering how to *partition* the data.

1. *Descriptive versus explanatory intent.* Are you essentially trying to lay out data to see "what's there," or do you want to generate some explanations about why things happen as they do? The latter intent usually requires more careful attention to various forms of ordering.

2. *Partially ordered versus well-ordered.* Are you essentially placing data in rows and columns that represent descriptive categories, one by one, or are those categories ordered in some way—by strength or intensity of a variable, by time, by roles of participants, by cases that have different levels of some key variable? Once past initial description, some form of ordering is typically very helpful.

3. *Time-ordered versus not.* Ordering matrices by time is a special case of the above. If a matrix is time-ordered, it enables the analysis of flow, sequences, perhaps cycles and chronologies, and maybe causes and effects.

4. *Categories of variables.* Which types of rows and column are possible? The set of possibilities is almost infinite. To illustrate, here are two brief lists adapted from Lofland and Lofland (1984) and Bogdan and Biklen (1992). For concreteness, we use examples from health care settings. One list refers to the size of the social unit being examined:

individuals	Jane Hughes, Luis Garcia
roles	patients, nurses, doctors, administrators
relationships, pairs	patient-spouse, nurse-resident
groups	intensive care unit staff, accounting dept., surgical team
settings (places or locales with sites)	operating room, emergency room, cafeteria
sites as wholes	Good Samaritan Hospital, Downtown HMO

Another list refers to what social units are doing, in effect:

specific acts, behaviors (what people do or say)	diagnostic questions, answers, listening, information giving, eye contact
events (marked-off happenings or occurrences)	admission, operation, discharge
activities (regularly occurring, connected sets of behavior)	grand rounds, lab tests, billing
strategies (activities aimed toward some goal)	rehabilitation plan, nutrition counseling, radiation treatment
meanings, perspectives (how people construe events)	patient and physician views of HIV-positive finding, the experience of labor
states (general conditions)	staff morale, patient uncertainty, critical condition
processes (ongoing flows, phases, changes over time)	wound healing, convalescence, triage, decision making, social support, bereavement

Once you're clear about the *types* of rows and columns, you usually must make further partitioning decisions. For example, if you are working with "roles" as columns, are head nurses shown separately, lumped with nurses, or with administrators? Are you sorting doctors by specialty and/or by status (head, attending, resident)?

5. *Two-way, three-way, N-way.* The simplest matrices, as in quantitative data display, are organized in two dimensions. You have a choice to move to more complexity if the data demand it. Subdividing each column in a parallel fashion for each column permits a "three-way" matrix—even though the data still are displayed in two dimensions. Subdividing rows in the same way permits a four-way analysis. More complexity is always possible, but going further than four-way tables probably means you should be breaking out submatrices for more clarity.

6. *Cell entries.* You always have choices about the level and type of data to be entered. For example, you can include:

direct quotes, extracts from written-up field notes	"Surgeons are basically macho types—the best ones, I mean. They have to be, but they can't help it."
summaries, paraphrases, or abstracts	Patient advocate seems increasingly distrusted. Rules for residents' time use largely ignored.
researcher explanations	Patient uncertainty about diagnosis is a function of life-threatening conditions, and use of denial as a buffer.
ratings or summarized judgments	Risk-reduction behavior after cardiovascular bypass: worse, no change, improved, much improved.
combinations of the above	RR: *improved.* Diet ("I even tried broccoli."), some exercise (walks 20-30 min. daily); smoking (reduced, is considering nicotine patch).

7. *Single-case versus multiple-case data.* Are you focusing on describing/explaining phenomena attached to a single case in context (an individual, a family, a group, an organization, a community, as discussed in Chapter 2, section C) or to several cases from which reasonably comparable data have been collected? If the latter, rows or columns will usually have to be devoted to cases (though, as we saw in Chapter 8, section D, cases can be cell entries).

Rules of Thumb for Matrix Building

Given these choices, what can we say informally about the best and easiest ways to build matrix displays? We state these briskly, as softly given advice rather than as harsh dicta.

1. Look at your research question(s) and key variables, and think of the data that are or will be available. Sketch the matrix roughly.
2. Get a colleague to look at your format, to help you detect the assumptions you are making, and to suggest alternative ways to display your data.
3. Set up the revised matrix by using your word processor if possible. Aim for a display that will go on one large sheet, even if that sheet covers a wall. You have to be able to see it all at once.
4. Don't try to include more than a dozen or so variables in rows or columns; five or six is more like it. If you are drawn to a design (e.g., like Table 8.4, p. 211) with larger numbers of variables, plan to cluster or partition them. In effect, regroup the matrix into "streams" or adjacent "families."
5. Expect that your preliminary format(s) will be iterated several times during early data entry—and often later on as well. You'll always get better ideas about how the matrix should be set up.
6. If the matrix is an ordered one, expect to transpose rows and columns for a while until you have a satisfactory version. Most word processors can do this quite easily.
7. Always stay open to the idea of adding new rows or columns, even late in your analysis operations.
8. Keep rows and columns fine-grained enough to accommodate meaningful differentiations in the data, but not so fine as to bury you under indiscriminate detail.
9. Keep in mind that any particular research question may require a *series* of matrices; for example, an initial partially ordered descriptive matrix, leading to a small summary table, then to variously ordered, more boiled-down matrices. Think ahead to this possibility but allow new matrix forms to emerge as the analysis proceeds. Trust us: they invariably do.

B. Entering Matrix Data

The choice of data for entry into matrix cells, and the operations involved in doing it, are critical issues in qualitative data analysis. *The conclusions drawn from a matrix can never be better than the quality of the data entered.* A completed matrix may look coherent, plausible, and fascinating, but if the data were poorly collected in the first place or were entered in a hasty, ill-partitioned, or vague way, the conclusions are suspect.

Rules of Thumb for Data Entry

1. Be clear about the level of the data you want to enter. Will you be entering relatively "thick" descriptions (Den-

zin, 1989b; Geertz, 1973) that render the context well, show intentions and meanings, and show how things developed over time?[3] Will you include direct quotes, close-up detail? Will you make a short summary or paraphrase? Will you make more general summary judgments or global ratings?

When in doubt, make the entries "thicker." The usefulness of any data display is greater when "data density" is higher (Tufte, 1983, 1986, 1990; Wainer, 1982). We requote Tufte: "To clarify, add detail." More information is better than less. Too-thin cell entries keep you away from the meaning of the data.

2. Remember, however, that even a dense matrix is displaying only a very small percentage of the available data. That part will inevitably be "thinner" than the original description. There is always a great deal of selection and condensation from the mass of field notes. Be aware of how you make that selection and how you boil data down. You are not throwing away your field notes—you can always get back to the full material—but you should *notice* what you are doing.

3. Use codes to locate key material. Entering these data is much easier with a word processor or software having multiple screens; they permit you to retrieve coded chunks to one screen or region and to select/edit/reduce them on another. Some software is specially designed to generate and display matrix entries (e.g., AQUAD, NUD-IST; see Appendix).

It's important to know where these retrieved chunks came from in the written-up field notes; they should include some form of "source tag," so you can easily get back to them in context if need be.

4. Because, as we just noted, you can never retain full "thickness" of the data you are displaying, everything depends on your knowing—and others' knowing—just how you did the data selection and condensation.

So keep an explicit record of the "decision rules" you followed in selecting data chunks for entry (e.g., the extent of agreement among respondents or data types, the intensity of respondent feelings, researcher confidence in the data, or the basis for making judgments or ratings). You will need this for your report. The record should be generated as a *log* as you proceed—otherwise you may delude yourself retrospectively, forget how you did it, or shift your decision rules during the process (see Chapter 10, section D for specific self-documentation suggestions).

5. The decision rules and other aspects of the matrix should be explained clearly in a legend that will always be associated with the matrix, whether you, a colleague, or the final reader is looking at it.

6. When data are missing, ambiguous, or were not asked for from certain respondents, show this explicitly in the matrix.

7. Don't lock up your format until very late in the process. Entering data tests the adequacy, realism, and helpfulness of the matrix format. Keep revising it as needed.

8. The basic strength of qualitative matrices lies in their inclusion of *text.* But you can also be open to using numbers, either direct quantities, or judgments in the form of ratings or scales (decision rules on the latter must be clear).

If you do use numbers, *keep words together with the numbers* in the matrix to clarify, support, and deepen their meaning.

9. It pays to take unusual care when scaling any variable on which you will order cases in a multicase matrix. Unclear ordering will enfeeble all subsequent analyses. The same holds true for any major dependent variable—for example, the outcome variable in a predictor-outcome matrix. If the outcome is poorly scaled, then all of the predictor analyses will be founded on sand.

10. Early in the game, get a colleague to review a filled-in matrix, along with your decision rules and written-up field notes, to check the procedural adequacy of your work. This step will go best if you have kept a thorough log of the data entry process as noted under point 4. Such audits are time-consuming (carefully done, they take about half the time an original data entry and analysis process took). But used *selectively,* they are an important check on the "confirmability" of the procedures you used.

C. Drawing Conclusions From Matrix Data

The test of any matrix is what it helps you understand—and how valid that understanding is. As we indicated in preceding chapters through use of **boldface,** there are many specific "tactics" for conclusion drawing and verification. Each has its specific advantages and pitfalls, which we discuss in Chapter 10. Here we content ourselves with some general advice.

Rules of Thumb for Conclusion Drawing

1. It's always helpful to start with a quick scan—a "squint analysis" down rows and across columns to see what jumps out. Then verify, revise, or disconfirm that impression through a more careful review.

2. Any given matrix will always have multiple tactics used on it. The most frequent, basic tactics we have seen used for drawing first conclusions are **noting patterns, themes; making contrasts, comparisons; clustering;** and **counting.**

3. As conclusions form in your mind, always *write text* explaining them. Make your conclusions explicit. The process of writing inevitably leads to reformulation, added clarity, and ideas for further analysis. Writing is itself a form of analysis. We discuss the interaction of displays and analytic text in Chapter 5 (see Figure 5.4, p. 101). Displays never speak for themselves—either to you or to the reader; associated text is always needed.

4. For initial descriptive matrices, which are often large and complex because of the need to "get it all in," use content-analytic summary tables (Chapter 7, section C) to clarify your understanding. Then check back with the larger matrix to be sure you have not oversimplified or distorted your conclusions.

5. First conclusions almost always need to be checked back against written-up field notes. If a conclusion does not ring true at the "ground level" when you try it out there, it needs revision. The very existence of matrices can sometimes lead you to unjustified feelings of certainty about conclusions. Look at the raw data to guard against this.

6. More generally, any early conclusion typically needs confirmation, checking, *verification.* Is it really right? The most frequently used tactics we've noted are **following up surprises; triangulating; making if-then tests;** and **checking out rival explanations.** Matrix displays are also useful when you are seeking confirmation through **feedback from informants.** More on these in Chapter 10.

7. In looking at case-ordered matrices, make a first sweep through the data for *each case,* one at a time, to be sure your descriptive understanding is clear at the case level before you try to understand cross-case patterns.

8. In writing semifinal text explaining your (checked, confirmed) conclusions, include specific illustrations from written-up field notes. In doing this, avoid the temptation to sprinkle the text with vivid or interesting examples to spice up the narrative. Rather, look for genuinely representative *exemplars* of the conclusions you are presenting. If you cannot find them, something is wrong with the conclusions—revise.

9. Remember that analysis usually has to go beyond descriptive summation and reach to explanation. Checking the conclusions against the data is only part of the meaning-establishment task. The other part is clarifying the *conceptual* import of those conclusions, how they tie into your or someone else's theory of social behavior. Matrix analyses that yield verifiable but meaning-poor conclusions are of little use to anyone.

10. Document your conclusion-drawing procedures (see Chapter 10, section D, Table 10.2 [p. 283] and Figure 10.8 [p. 285] for specific suggestions) and ask for an occasional review from a colleague, especially during your early work. Reviews at this level help *you* now, and help out a future researcher who may want to extend or replicate your work. Save full-scale "audits" of your work for high-stakes projects or for aspects of your study that you consider crucial.

11. In developing text for the final report, think carefully about the data the reader will need. In many cases the complete matrix you used should be presented, along with your analytic text, to the reader, who can thereby follow and test/confirm your conclusions. Your findings are not taken on trust.

The tradition of presenting basic data is deeply ingrained in reports of quantitative data analysis, so much so that it would be unthinkable for a researcher to present conclusions without data tables, or at least without reference to working documents containing them. We believe the same norms should apply to qualitative researchers. Some data displays should be a normal part of reporting conclusions. The typical final report, as Carney (1990) notes, normally will contain a mixture of full-scale narrative text where thoroughly "thick" description is needed, displays, and associated analytic text.

As Krathwohl (1991) suggests, the reader can thus "reconstruct how an analysis developed, check the translation fidelity of constructs and the logical validity of conclusions."

In some instances, of course, a summary table or boiled-down version may suffice. And in a few others—which ought to be the rarest, in our view—you may conclude that text with illustrations will be sufficient. Where basic matrices are not presented, you owe the reader a clear explanation of the display and analysis methods used to get to the text.

In any case, let's remember what the reader needs from displays. We can hardly do better than quote Tufte (1986): "What we are seeking in graphic and tabular displays of information is the clear portrayal of complexity. Not the complication of the simple; rather the task of the designer is to give visual access to the subtle and the difficult—that is, the revelation of the complex" (p. 80).[4]

We won't try to summarize this summary chapter, but want to point forward to the next one. In it we deal in some depth with the various tactics we have identified along the way for drawing and verifying conclusions. We also explore the issue of "goodness" of conclusions, and methods for documenting analysis operations.

Notes

1. We found the discussion in Maxwell and Miller (1992) very helpful in clarifying these two broad families of display.

2. For added ideas on "matrix thinking" as a way to aid conceptualization and research planning, see Hage (1972) and Patton (1990, Chap. 8). Carney (1990) also has useful suggestions, though he limits himself to matrices with numerical entries reduced from text.

3. Denzin's (1989b) discussion of "thick description" is especially helpful; he points out how it lays the basis for "thick interpretation." And he argues that some of Geertz's supposedly "thick" descriptions (e.g., of Balinese cockfights) are thin, analytic, and observer-oriented. Thick description is sometimes superficially equated with rich detail; the issue is more, as Maxwell (1992b) points out, the *meaningfulness* of detail in the local cultural context.

4. Although Tufte (1986) was speaking here of the display of quantitative information, his ideas are generic, as are those in Wainer's (1982) ingenious piece "How to Display Data Badly." See also Tufte's classic, marvelous books on display (1983, 1990).

10

Making Good Sense

DRAWING AND VERIFYING CONCLUSIONS

This is a big chapter, at the core of the book. The displays we've described are tools for analysis. But how do you use them? As you work with any display, there's always a flow of specific analysis *tactics*—ways of drawing and verifying conclusions that you use during the process. These tactics have been named in **boldface** in preceding chapters. Now we discuss each in detail. Following that, we turn to the question of the goodness—the *quality*—of conclusions you reach. The chapter concludes with a look at the *documentation* of qualitative analyses.

A. Tactics for Generating Meaning

In this section we discuss 13 specific tactics for drawing meaning from a particular configuration of data in a display. Usually we describe the general analysis situation being faced, explain the tactic, and then give one or more examples, often referring back to previous sections (where the tactic was noted in **boldface**). We also refer to others' work for examples.[1]

If we can muster advice, we present it, too. Our approach is brisk; the test of these tactics comes in the using. We also make suggestions for practicing the tactic with some specific displays in prior chapters. Note: A given display usually invites the use of *other* tactics, in addition to the one suggested.

People are meaning-finders; they can very quickly make sense of the most chaotic events. Our equilibrium depends on such skills: We keep the world consistent and predictable by organizing and interpreting it. The critical question is whether the meanings you find in qualitative data are valid, repeatable, and right. The following section (B) discusses tactics for testing or confirming meanings, avoiding bias, and assuring the quality of conclusions.

First, here is a quick overview of tactics for generating meaning, numbered from 1 to 13. They are arranged roughly from the descriptive to the explanatory, and from the concrete to the more conceptual and abstract.

Noting patterns, themes (1), **seeing plausibility** (2), and **clustering** (3) help the analyst see "what goes with what." **Making metaphors** (4), like the preceding three tactics, is a way to achieve more integration among diverse pieces of data. **Counting** (5) is also a familiar way to see "what's there."

Making contrasts/comparisons (6) is a pervasive tactic that sharpens understanding. Differentiation sometimes is needed, too, as in **partitioning variables** (7).

We also need tactics for seeing things and their relationships more abstractly. These include **subsuming particulars into the general** (8); **factoring** (9), an analogue of a familiar quantitative technique; **noting relations between variables** (10); and **finding intervening variables** (11).

Finally, how can we systematically assemble a coherent understanding of data? The tactics discussed are **building a logical chain of evidence** (12) and **making conceptual/theoretical coherence** (13).

1. Noting Patterns, Themes

When you're working with text or less well organized displays, you often note recurring patterns, themes, or "gestalts," which pull together many separate pieces of data. Something "jumps out" at you, suddenly makes sense. We discuss this under "pattern coding" (Chapter 4, section C).

Some examples of patterns from our school improvement study:

The frequent citing of a "miracle case" (a failing student who was rejuvenated by the innovation) as either an explanation or a justification for the project

"Deep coping" as a problem-solving style in a high school staff group

The use of "administrative latitude"—freedom to alter an innovation in return for trying it at all

What kinds of patterns can there be? As usual, we can expect patterns of *variables* involving similarities and differences among categories, and patterns of *processes* involving connections in time and space within a context.

Pattern finding can be very productive when the number of cases and/or the data overload is severe. Stearns et al. (1980), for example, studied 22 school sites (see Chapter 4, section D). Field-workers generated unstructured "pattern" statements of findings from more than one site; these were reduced gradually to a few hundred "propositions," ordered into 21 categories.

The human mind finds patterns so quickly and easily that it needs no how-to advice. Patterns just "happen," almost too quickly. We rapidly make "stories" from a range of data (Goleman, 1992). The important thing, rather, is to be able to (a) see added evidence of the same pattern ("recurring regularities," as Guba, 1978, puts it) and (b) remain open to disconfirming evidence when it appears. As Ross and Lepper (1980) point out, beliefs (in this case in the existence of a pattern) are remarkably resistant to new evidence. As we'll see, patterns need to be subjected to *skepticism*—your own or that of others—and to conceptual and empirical testing (Does it really make conceptual sense? Do we find it elsewhere in the data where it was expected? Are there any counterexamples?) before they represent useful knowledge. See Yin (1984, pp. 104-108) for further ideas on matching empirically found patterns to prior predictions.

Suggestions for practice. Look at Figure 5.5 to see what patterns jump out for you. Write some analytic text with your conclusions. Try it: You'll like it. Then compare with the text following the figure. The same approach can be used with Figures 5.1 and 5.2; Box 5.8; and Tables 5.2, 5.7, and 6.1. Look for a pattern in Figure 7.7. Other practice possibilities are in Tables 8.5 and 8.8 (one that especially rewards pattern-finding efforts). To locate any of the displays mentioned, see the List of Figures, Tables, and Boxes (page viii).

A reminder: Any given display usually will invite (and reward) your use of *other* tactics as you proceed.

2. Seeing Plausibility

The faithful field-worker Boswell (1791) reports that Dr. Johnson said, "Patriotism is the last refuge of a scoundrel." With good reason: A noble sentiment can easily be exploited for other purposes. There is a crude parallelism to the idea of "plausibility" as a last-refuge tactic for drawing conclusions.

It often happens during analysis that a conclusion is plausible, "makes good sense," "fits." If a colleague asks you how you came to the conclusion or what you based it on, the initial answer is something like, "I don't really know. . . . It just feels right." Many scientific discoveries initially appeared to their authors in this guise; the history of science is full of global, intuitive understandings that, after laborious verification, proved to be true. So plausibility, with intuition as the underlying basis, is not to be sneered at.

But people are meaning-finders, even in the most genuinely chaotic data sets. Patterns can be found even in random data, as the activities of numerologically obsessed people show. So plausibility can easily become the refuge of, if not scoundrels, at least analysts who are too ready to jump to conclusions.

During documentation of our own analysis efforts (as reported in section D), we often found ourselves giving the "plausibility" basis for conclusions we drew, particularly in the early stages of analysis. Nearly always, it turned out, "plausibility" was an initial impression that needed further checking through other tactics. Plausibility in this sense was a sort of *pointer,* drawing the analyst's attention to a conclusion that looked reasonable and sensible on the face of it—but what was the real basis involved?

Here's a brief illustration. An analyst in our school improvement study was trying to order 12 school cases on "percentage of use" of the innovation—the number of teachers in a school or district eligible to use an innovation (e.g., a reading program) who were actually using it. On the face of it, this looks simple: the ratio of two numbers. The cases could be sorted into a few categories: full, moderate, minimal percentage of use. But as the analyst pro-

Table 10.1
Case-Ordered Meta-Matrix: Percentage of Use, by Cases

		In the building				In the district					
Percentage of Use	CASES	Years of use	# users in bldg.	Eligible users in bldg.	% of use	Years of use	# users in dist.	Eligible users in dist.	% of use	Eligibility criteria	Remarks
Substantial	Carson (L)	3	20	20	100	3	42	42	100	All regular teachers in district (1 elem., 1 HS).	Mandated as of Fall '79.
	Masepa (E)	3	9	11	82	4	36	43	84	All teachers, grades 3-7, 6 schools.	Mandated as of April, 1980.
	Tindale (L)	4	29	36?	80?	4	48	60?	80?	English, math and science teachers for lower track students, 2 HSs. 60= max. of teachers ever using innovation.	Not clear what maximum student population is.
Full in buildings, less in district	Plummet (L)	4	25	25	100	4	N/A	N/A	?	Innovation is complete school; all present staff eligible.	Unknown whether school deals with all of potential target population (delinquents, 11 high schools).
	Perry-Parkdale (E)	3	6	6	100	3	N/A	N/A	?	Innovation is self-contained program; all present staff eligible.	Program accommodates 3% of total juniors and seniors from 2 HSs; unknown what eligible population is.
	Banestown (E)	$1^1/_2$	3	3	100	$1^1/_2$	10?	DK	?	Innovation is remedial lab; all lab staff eligible.	Used in 5 schools of large county system.
Moderate to full for specialized population	Astoria (E)	1	5	5	100	2	DK	DK	100?	All 1st-grade teachers and aides (92 schools).	Program mandated as of Fall '78; actual percentage of use throughout district unknown.
	Calston (E)	2	2	2?	100?	4	4	4--50	100 to 8	Narrow definition; all intermediate teachers in schools with an interested principal (N=2).	Percentage of use for all 25 schools is 8%.
	Lido (E)	4	3	5	60	4	3	5	60 to 8	High school science teachers, in one high school.	In principle, all 36 HS teachers are eligible since innovation is interdisciplinary. Percentage of use would thus be 8%.
Minimal	Burton (E)	1	1	5	20	1	3	20?	15?	All social studies teachers in 4 high schools.	Use defined as "experimental" during first year.
	Dun Hollow (L)	$2^1/_2$	2	13?	15?	$2^1/_2$	3	84	3?	All primary teachers–grades 1-3 in 7 elementary schools	Use defined as "field testing."
	Proville (L)	3	0	11 to 36?	0	3	0	44 to 144	0	44 vocational education teachers in 4 HSs and/or 100 classified personnel.	A discontinuation as of Spring '79.

? = missing or inconclusive data

ceeded, life looked more complicated. His documentation comment was: "The categories are OK, but they don't form an ordinal scale. Much depends on the size or scale of what is being attempted in the first place." This statement was based, the analyst said, on "plausibility. . . . Seems clear on the face of it."

The next step was to get more systematic. The analyst went through several iterations of a matrix, ending with Table 10.1. As he did this, the analyst also realized that percentage of use in the *district* and in the *school* were quite different things, and he clustered cases accordingly.

We can see in Table 10.1 that the initial "plausibility" basis had to be supplemented by some tactics we explain further below: **clustering** (e.g., realizing that Astoria, Calston, and Lido all were dealing with a specialized population of students) and **partitioning a variable** (building-level and district-level *percentage of use*).

The final four clusters of schools fall in an order that has a clear basis (substantial use in building and in district; full in building, less in district; moderate to full, but for a specialized, hence less demanding, population; and clearly minimal use). The analyst also can order the cases within each category.

So the moral is: Trust your "plausibility" intuitions, but don't fall in love with them. Subject the preliminary conclusions to other tactics of conclusion drawing and verification.

Incidentally, a somewhat more trustworthy tactic involves noticing *lack* of plausibility. When a conclusion someone is advancing "just doesn't make sense," it's a bit safer to rule it out. But not completely safe. Counterintuitive or puzzling findings sometimes can be extraordinarily stimulating and rich, so they should be allowed their day in the sun, too. (We discuss the tactic of **following up surprises** shortly.)

Most conclusions drawn during analysis are substantive, based on the content. But the analyst is constantly drawing *procedural* conclusions along the way as well: to

Box 10.1
Clusters of Information-Processing Capacities

scan, skim, look into, pick over, inspect, dip into, flip through, browse, glance into, leaf through, glean,

itemize, list, review,

select, edit, single out, choose, screen, filter, skip,

highlight, focus, chunk, isolate,

condense, reduce, boil down, abstract, summarize, synopsize, approximate,

discriminate, distinguish, winnow wheat from chaff, separate the sheep from the goats,

merge, harmonize, synthesize, integrate, blend, average, smooth, lump,

group, cluster, aggregate, pair,

categorize, classify, sort, pigeonhole,

organize, catalog, outline, structure,

refine, idealize

transpose two rows in a matrix; to add or discard a column; to collapse the data into a summary table; to change a decision rule for data entry. Our experience is that "plausibility" is often a reasonable guide for such decisions, and that laboring to try other tactics or to "verify" the procedural decision is unnecessary. Of course, it's important, as we've said many times, to log and report the procedural decisions made (the final matrix, the operative decision rules, etc.).

Suggestions for practice. Scan Table 5.7 and jot down what some plausible conclusions seem to be. Then how would you check them out with other tactics? Then compare with the following text. Working with a data set of your own may be helpful here: What looks like a first-cut "plausible" finding? And how could you check it out further?

3. Clustering

In daily life, we're constantly clumping things into classes, categories, bins: Things that do not move around but grow are called "plants"; things that move around and have babies are called "animals"; things that move around, have four wheels, have an engine run by fossil fuels, and carry people are called "automobiles." Most categories require other categories to define them: "wheel," "engine," "babies."

That sounds a bit more conceptually formal than the way clustering usually happens. Take, for example, this quote from Tufte (1990):

> We thrive in information-thick worlds because of our marvelous and everyday capacities to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthesize, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, refine, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, sort, pick over, group, pigeonhole, integrate, blend, average, filter, lump, skip, smooth, chunk, inspect, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, list, glean, synopsize, winnow wheat from chaff, and separate the sheep from the goats. (p. 50)

As you scan this lovely list, do you begin to form some categories—things that seem to fall together? Try your hand at a few before you read further.

With this list entered in a word processor, one of us produced the clusters in Box 10.1 in 6 minutes.

The next step often is to figure out what to call clusters. The first one in Box 10.1 might be labeled SCANNING or EXPLORING; the next ORDERING or REVIEWING; the next SELECTING, and so on. Perhaps your clusters and cluster names would differ somewhat from these. The point is that we all have cognitive frames that let us clump things rapidly. In this instance the analyst had an implicit "sequence" frame that led him to think that clusters at the top of the list might come earlier than those that followed in the sense-making process.

As we note in our discussion of coding (Chapter 4, section B), the qualitative analyst is looking to see, as

Le Compte and Goetz (1983) put it, "What things are like each other? Which things go together and which do not?" The categories or classes used may be preexisting (for school districts: urban, suburban, rural) or may emerge from the data ("controlling," "directive," "facilitative," and "neglectful"), as found in Berman and Weiler's study of districts in California school improvement projects (Degener, 1983). Typically, as Bulmer (1979) points out, they emerge from an interaction of theory and data.[2]

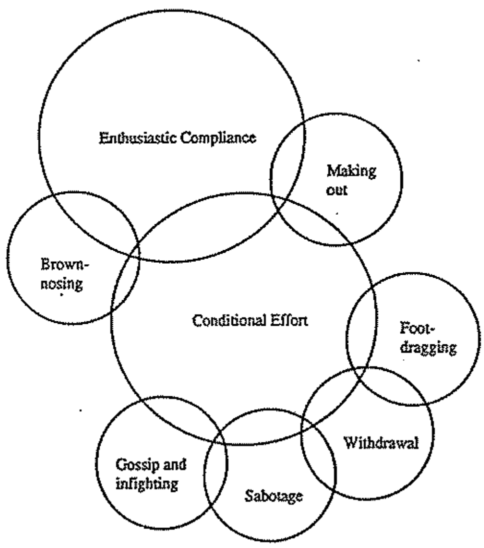
Clustering is a tactic that can be applied at many levels to qualitative data: at the level of events or acts, of individual actors, of processes, of settings/locales, of sites or cases as wholes. In all instances, we're trying to understand a phenomenon better by *grouping* and then *conceptualizing* objects that have similar patterns or characteristics. Here are some illustrations.

Davis (1959), quoted in Lofland (1971), was studying the *acts* of cab drivers who were interested in receiving a larger tip. They clustered this way:

fumbling in the making of change
giving the passenger a hard-luck story
making fictitious charges for services
providing a concerted show of fast, fancy driving
displaying extraordinary courtesy

Another example of clusters of *acts* comes from Hodson (Box 10.2). It usefully shows us that clusters are not always mutually exclusive and may overlap.

Box 10.2
Overlapping Clusters:
Behavioral Modes at the Workplace
(Hodson, 1991)



Focusing on key *actors,* we asked teachers to draw pictures of how their high school looked to them—which people and groups were involved. This task led to categories of teachers such as "goners," "boy coaches . . . girl coaches," and the "old new guard" (see Chapter 5, section B).

At the level of *processes,* we clustered the activities involved in coping with the problems of later implementation of an innovation (Table 7.8): reaching up, improving, debugging, refining, integrating, adapting, and extending. These clusters took a good deal of summarizing and reworking before they came clear.

Another example comes from our look at teachers' and administrators' job mobility, which clumped rather easily into these categories: moving in, moving out, moving up, moving in and up, moving out and up, and moving over.

It's also possible to cluster *settings* or locales. For example, in schools, we might sort places where people interact into these clusters:

formal instructional (classroom, gym)
informal instructional (library, club room)
formal adult work (meeting room, office)
informal adult association (teacher lunchroom, restroom, corridor)
mixed (cafeteria, playground)

Finally, as we've seen repeatedly, it is possible to sort *sites* as complex cases into meaningful clusters. For example, in Box 7.6 we show how 12 cases were sorted into three "families" according to the level of assistance provided and roughness/smoothness of early implementation, then the later degree of practice stabilization. Or, in Chapter 8, section F, using causal networks, we clustered 12 school cases into four scenarios: "casualties," "success-driven advancement," "opportunism," and "career crystallization."

We can see from these examples that "clustering" is a general name given to the process of inductively forming categories, and the iterative sorting of things—events, actors, processes, settings, sites—into those categories. Where lower level, less complex things are being sorted (events, actors, etc.), the clustering tactic typically relies on aggregation and comparison ("What things are like each other/unlike each other?") and is naturally closely interwoven with the creation and use of codes, both at the first level (Chapter 4, section B) and the pattern-coding level (section C).

As the analyst works at clustering processes, settings, and whole cases, the clustering operations become more and more complex and extended—just as sorting things into "animals" and "plants" is a (perhaps deceptively) simpler task than sorting various kinds of wheeled machines (automobiles, trucks, golf carts, airplanes, ski-lift gondolas, typewriter tables, and floor polishers) into sensible clusters.

The typical problem in making clusters at these more complex levels is that the entities being clustered have many attributes that may be relevant to the clustering task. A simple way to proceed is the "case-by-attribute matrix." Listing cases as rows and attributes as columns lets you see the whole picture. By inspecting the columns, you can find which attributes are critical in differentiating cases. Then "families" of cases can be formed by rearranging the rows of the matrix, a task most word processors can do quickly. (See also the QCA program, Appendix.) Each family shares the same set of critical attributes. Miles, Farrar, and Neufeld (1983) have an example.

Clustering also can be seen as a process of moving to higher levels of abstraction (see **subsuming particulars into the general**). Figure 10.1 illustrates this with a content-analytic "dendrogram" display for representing clusters, from Krippendorff (1980a). It emerged from computer clustering of groupings made by 28 people sorting 300 advertising appeals. Human analysts are not so elegant, but the principle holds.

As this example suggests, the qualitative analyst need not assume that clustering techniques have to be completely self-invented. There is a long tradition of content-analytic techniques dealing directly with issues of coding, unitizing, and clustering qualitative data that can be very helpful, from Holsti (1968, 1969) and Berelson (1971) to Krippendorff (1980b) and Weber (1990). The best-organized comprehensive review we have seen is by Carley (1990).[3]

Clusters, like the results of other conclusion-drawing tactics, must be held lightly in the analyst's mind; you have to ward off premature closure. And they need to be verified, as well. For example, does a cluster **"outlier"** or **extreme case** really belong to cluster A, or more appropriately to cluster B? Are the data being used fully **representative** of the universe of events, actors, or settings being studied?

Suggestions for practice. If you haven't already tried your hand at clustering the Tufte data above, have a go. Also look for patterns in Table 5.7 and see whether they can be clustered; compare with Table 5.8. In Table 8.6 try clustering the cases according to shared patterns, and compare with the following text.

4. Making Metaphors

The notion that research should focus only on matter-of-fact, literal—even actuarial—description, with cautious, later ventures into interpretation and meaning, is responsible for much intellectual poverty and misery.

That sentence itself is stuffed (a metaphor) with metaphors (cautious ventures, the number-driven obsessions of actuaries, poverty, misery). Metaphors, seen as one major type of trope or literary device, involve comparing two things via their similarities and ignoring their differences.[4] As Morgan (1980) notes, calling a boxer "a tiger in the ring" evokes fierceness, grace, and power—and ignores striped fur, fangs, and four-leggedness. Metaphors are thus a "partial abstraction."

As such, they have an immense and central place in the development of theory. Lakoff (1987) argues that people can only grasp abstract ideas by mapping them on to more concrete ones. Gertner and Grudin (1985) show how psychological theory has evolved over nine decades from using animistic to systemlike metaphors.

A. Miller (1986) points to the centrality of metaphor in the physical sciences in clarifying arguments and explaining puzzles. Geertz (1983) comments that metaphors used to explain society (which he calls "the instruments of reasoning") have shifted from "elaborate mechanisms or quasi-organisms" to those based on "serious games or sidewalk dramas."

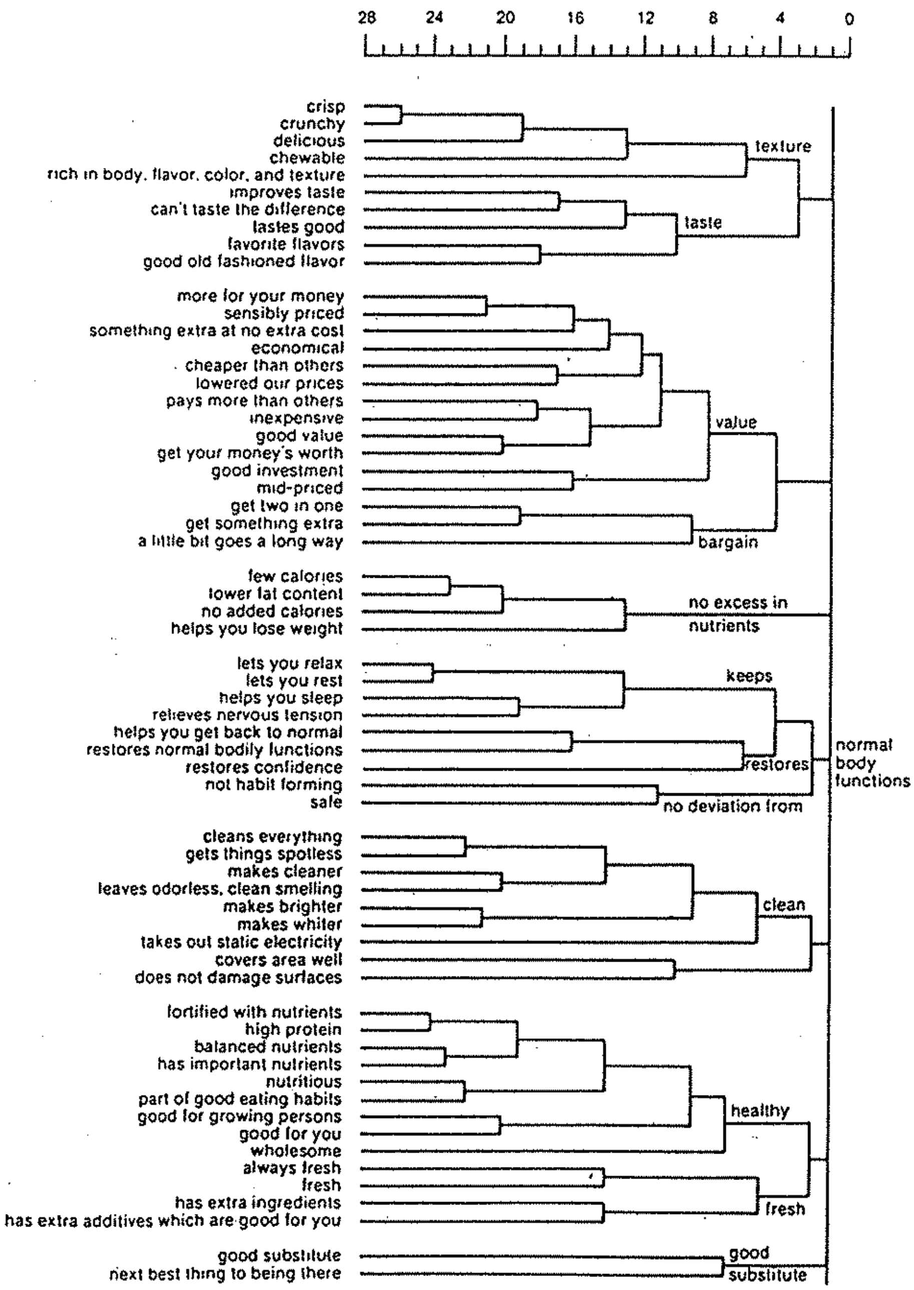
The people we study use metaphors constantly as a way of making sense of their experience. We do the same thing as we examine our data. The issue, perhaps, is not *whether* to use metaphor as an analysis tactic, but to be *aware* of how we—and the people we study—use it.

For instance, suppose you found yourself referring to "the empty nest" when you were looking at an interview with someone who has grown children. You are, in effect, making allusions to an important environmental setting ("nest"), the idea of nurturance aching for a newly absent (but grown-up) object, and the acknowledgment that nurturance to the point of nest-leaving has taken a good deal of time. But you may also be assuming that the nest itself is of little value and will be abandoned—and not considering the idea that the nest may be refilled with a new brood.

So the *richness and complexity* of metaphors are useful. The "empty nest" metaphor leads us farther than a single variable, such as "mother's separation anxiety," would. Our metaphor lets us see new theoretical possibilities (maybe if socialization for independence is weak, the child will regress).

What else is true of metaphors? They are *data-reducing devices,* taking several particulars and making a single generality of them. For instance the "scapegoat" metaphor pulls together into one package facts about group norms,

Figure 10.1
Illustration of Clustering, Using Dendrogram Method (Krippendorff, 1980a)



treatment of deviants, social rituals, and social rationalizations. This ability is not to be sneezed at. Qualitative researchers looking at mountains of field note write-ups are grateful for any device that will reduce the bulk without locking out multiple avenues for analysis.

Metaphors are also *pattern-making devices*—for ourselves and for the people we study. For example, in the school improvement study, we found at one site that the remedial learning room was something like an "oasis" for the pupils sent there for part of each day. (A teacher used the word spontaneously, and we began to see the pattern.) The metaphor "oasis" pulls together separate bits of information: The larger school is harsh (like a desert); not only can students rest in the remedial room, but they also can get sustenance (learning); some resources are very abundant there (like water in an oasis); and so on. Such metaphors also help place the pattern noted in the larger context (in this case, the harsh, resource-thin school).

Metaphors are also excellent *decentering devices.* You step back from the welter of observations and conversations at the field site and say, "What's going on here?" Because metaphors will not let you simply describe or denote a phenomenon, you have to move up a notch to a more inferential or analytical level. The remedial learning room does not look like an oasis, and most people are not actually describing it that way, nor is anyone behaving literally like an exhausted Bedouin under a date palm.

Finally metaphors or analogies are ways of *connecting findings to theory.* The "oasis" metaphor makes you think of how institutions develop compensating mechanisms to reduce the stress they put on role occupants, or of how they nurture as well as isolate deviants. Or you start considering social control mechanisms more generally. As many authors note (Lakoff & Johnson, 1980; Miller & Fredericks, 1988; Weick, 1989), metaphoric thinking effectively unites reason and imagination.

The metaphor is halfway from the empirical facts to the conceptual *significance* of those facts; it gets you up and over the particulars en route to the basic social processes that give meaning to those particulars. For instance, Glaser (1978) advises the field researcher struggling to make sense of social phenomena to attach metaphorical gerunds to them (e.g., servicing, bargaining, becoming). In doing that, you're shifting from facts to *processes,* and those processes are likely to account for the phenomena being studied at the most inferential level.[5]

Now a few words of advice for metaphor makers.

1. Stay aware of the metaphors you—and people in your study—are using, often with only partial awareness. Dig into them for implicit meanings, for more explicit exploration and testing.

2. Looking for overarching metaphors too early in the study is dangerous. It distracts you from fieldwork, and it leads to hasty judgments that clamp down on the meaning of what you are studying. You get all too quickly wedded to your metaphor; it sounds good, other people resonate, it makes you feel insightful. You start to look around less, and you project the metaphor on things that are, at best, remotely related to it. For big or inclusive metaphors, better to wait until, say, two-thirds of the way through data collection, when you have a strong body of information to draw from, and still some time to test the validity of the image.

3. Being cognitively playful helps generate metaphors. You say to yourself, "What's the gerund here?" or "If I only had two words to describe an important feature at this site, what would they be?" or "What does it feel like?" The trick is to move from the denotative to the connotative. As users of synectics methods know, it helps to shift into another domain. Instead of the social or the personal, go to biological, mechanical, or spatial domains to find useful metaphors.

4. Interaction helps. Groups stimulate their members' thinking by increasing the inputs, bringing in ideas from a new angle, making you decenter from your cognitive ruts (cf. Wicker, 1985), and creating a contagiously playful thinking environment.

5. Know when to stop pressing the metaphor for its juice. When the oasis starts to have camels, camel drivers, a bazaar, and a howling sandstorm, you know you're forcing things. Use it as long as it's fruitful, and don't overmetaphorize. Remember that the two things compared in a metaphor *always* have differences.

Suggestions for practice. Scan Table 5.7 and see what metaphors spring to mind: It could have several, in addition to the one noted by the researchers.

With one or more colleagues, look at a single-case explanatory display, such as Tables 6.1 or 6.2, and independently generate metaphors that organize what you see; compare. Try this with cross-case displays, such as Box 8.2, Table 8.6 (try a metaphor that names the families of cases), or Figures 8.4, 8.5, 8.6, and 8.7 (what metaphors would *you* use to characterize the streams?).

5. Counting

Some readers interpreted the first edition of this book as an assertion that qualitative data should always be "forced" into a quantitative mold. That particular dead horse (speaking of metaphor . . .) should not be beaten any more, other than by inviting the reader to look again at what we said in Chapter 3, section A, and at what follows here.

In qualitative research, numbers tend to get ignored. After all, the hallmark of qualitative research is that it goes beyond *how much* there is of something to tell us about its essential *qualities.*

However, a lot of counting goes on in the background when judgments of qualities are being made. When we identify a theme or a pattern, we're isolating something that (a) happens a number of times and (b) consistently happens in a specific way. The "number of times" and "consistency" judgments are based on counting. When we make a generalization, we amass a swarm of particulars and decide, almost unconsciously, which particulars are there *more often, matter more* than others, *go together,* and so on. When we say something is "important" or "significant" or "recurrent," we have come to that estimate, in part, by making counts, comparisons, and weights.[6]

So it's important in qualitative research to know (a) that we are sometimes counting and (b) when it is a good idea to work self-consciously with frequencies, and when it's not.[7]

There are three good reasons to resort to numbers: to see rapidly what you have in a large batch of data; to verify a hunch or hypothesis; and to keep yourself analytically honest, protecting against bias.

Seeing what you have. Numbers, we noted earlier, are more economical and manipulable than words; you "see" the general drift of the data more easily and rapidly by looking at distributions. For instance, in the school improvement study, we asked informants why they were using the new school practices we were studying. We got a mass of answers from several informants at each of 12 field sites. It seemed that many people were saying they had been pushed, more or less gently, into these projects, rather than diving in voluntarily. To see more clearly, we did a content analysis of the responses, totaled them, and derived Figure 10.2.

Figure 10.2
Reasons Given for Adoption by Users

Reasons/Motives	Number of Respondents Mentioning Item (N = 56)
Administrative pressure, constraint	35
Improves classroom practice (new resources, relative advantage over current practice)	16
Novelty value, challenge	10
Social (usually peer influence)	9*
Opportunity to shape projects	5
Professional growth	5
Gives better working conditions	3
Solves problems	2
Provides extra money	1
Total	86

* Seven mentions from one site

It turns out that 62% of the respondents mentioned pressure and constraint. And, counterintuitively, very few of the practices were adopted to solve problems. There also seemed to be a general "professional development/capacity enhancement" theme (challenge, shaping projects, professional growth). Seeing that theme, gauging the importance of the "constraint" motive, and noting the infrequent problem-solving incentive were all helpful. We saw the overall trends, got some new leads, and saw some unexpected differences. All these findings helped in the subsequent nonquantitative analysis. Even within a single case, that kind of exercise would have been a useful one.

Verifying a hypothesis. Table 8.5 is probably the best example of verifying. We reasoned that good preparation was the key to smooth initial use, so we created and computed a preparation "index" and set it against an estimate of smoothness of early use. Except at the extremes, we were wrong. But, by doing the counts, we saw *which* cases we were wrong about, and *why* this appeared to be the case. We set aside the numbers and followed up those leads.[8]

Keeping yourself analytically honest. We had expected from the start that careers would be important in the school improvement projects we studied. The more data we got, the more it seemed that "innovating" was a vehicle for moving up, in, over, or out (seldom down). The finding seemed important, potentially controversial, and might have been a result of our expectation. So (Table 7.16) we actually counted the number of job moves (63 for 12 sites) and estimated how many could be attributed to the innovation (83% were). Afterward we felt far more comfortable about the claims we were making. For example, it seemed that only 35% of the job-related shifts were upward ones, contrary to our early impression.

As qualitative researchers we work to some extent by insight and intuition. We have moments of illumination. Things "come together." The problem is that we could be wrong. A near-library of research evidence shows that people (researchers included) habitually tend to *overweight* facts they believe in or depend on, to *ignore or forget* data not going in the direction of their reasoning, and to *"see" confirming instances* far more easily than disconfirming instances (Nisbett & Ross, 1980). We do this by differentially weighting information, and by looking at part of the data, not all of them. Doing qualitative analysis of all data with the aid of numbers is a good way of testing for possible bias, and seeing how robust our insights are.

Suggestions for practice. Explore Tables 5.9 and 5.15 and Box 5.7 to see what counting can do for your understanding. Tables 7.9 and 8.4, along with Box 8.6, are rather underanalyzed examples that will repay a careful counting approach. Table 7.16 also has a number of unanalyzed aspects that deserve this sort of exploration.

6. Making Contrasts/Comparisons

Although comparisons are supposedly odious, they are what we do naturally and quickly when faced with any life experience, including looking at a qualitative data display. How does X differ from Y? Comparison is a time-honored, classic way to test a conclusion; we draw a contrast or make a comparison between two sets of things—persons, roles, activities, cases as a whole—that are known to differ in some other important respect. This is the "method of differences," which goes back to Aristotle, if not further. (The contrast between experimental and control groups was not invented by R. A. Fisher.) A few examples from our work:

- When we looked at the *variable* of preparedness (Table 5.2), the comparison showed that administrators were enthusiastic about the innovation and understood it well, but that users were bewildered—a picture that fit with their roles: Administrators press for adoption; users have to do the actual implementation work.
- Comparing *cases* that had many negative effects of implementation with those that had few (Table 7.13) showed us that high-negative cases were also ones with demanding innovations. Comparing cases with externally versus locally developed innovations showed no differences in final student impact, a finding that fit with prior expectation.
- Contrast tables (Box 7.5) comparing *cases* on amounts of user change made it clear that "change pervasiveness" might be causing user change.
- Predictor-outcome matrices array cases by high and low *outcomes,* and use that leverage to examine the impact of possible predictors. In our example the comparison was between roughness and smoothness of implementation, showing which predictors were present in smooth cases but not in rough ones (Tables 8.5, 8.6).
- Comparing job mobility during the middle, early, and later portions of projects showed (Table 7.16) the effects of *timing* and project development; more people moved in at the beginning and out toward the end.

Some advice:

1. Mindless comparisons are useless. The trick is to be sure that the comparisons being made are the right ones and that they make sense.

2. The results of a comparison should themselves be compared with what *else* we know about the roles, persons, groups, activities, or cases being compared.

3. Take a moment before you display a comparison, and think, "How big must a difference be before it makes a difference?" and "How do I think I know that?" You do not have a statistical significance test to fall back on. The *practical* significance is what you need to assess.

Suggestions for practice. Do a quick comparative cut at Boxes 7.5 and 8.2. What meanings do you draw from comparison?

For any of the following tables, draw some conclusions and write them down: 5.2, 5.9 (especially juicy), 5.12, 7.13, 7.16, 8.4, 8.5 (many comparisons possible), 8.6, and 8.9. Then compare with the following text.

7. Partitioning Variables

"Westward the course of empire takes its way." So wrote George Berkeley in 1752 in *On the Prospect of Planting Arts and Learning in America.* To Berkeley the main direction was obvious. It sometimes seems equally "obvious" that the course of qualitative data analysis is toward integration—ever and ever greater linkage among variables, expressed at a more and more powerful level.

But just as there was plenty of colonial push to the south and to the east, there are many times when *differentiation* is more important than integration. You must say, Stop! That wonderful variable is not really one variable, but two, or maybe even three. You need the courage to question what might be called "premature parsimony."

Partitioning variables can occur at many points during analysis. At the stage of initial conceptualization, it pays to "unbundle" variables rather than assume a monolithic simplicity. For example, the checklist matrix shown in Table 5.2 splits the general variable of "preparedness" to carry out an innovation into 10 subvariables or components, ranging from states of the user ("commitment," "understanding," "skills") to the availability of materials and actions taken by administrators ("time allocation," "inservice").

When coding schemes are being developed and elaborated, partitioning is often useful. For example (Chapter 4, section B), our study originally had a code TR-ORG, used for instances of change in the school as an organization as a result of innovation use. Initial fieldwork showed us that this variable should be split to distinguish among organizational changes that were *practices* (e.g., such as staffing, scheduling, planning, use of resources) and an equally important, "softer" variable, organizational *climate* (norms, interpersonal relationships, power, social networks).

When you are designing matrix formats, variable partitioning is also useful; more differentiation lets you see differences that might otherwise be blurred or buried. For example, in Box 6.2 the analyst wanted to study and explain the effects of various types of assistance to teachers. He realized very early that the variable "effects of assistance" should be separated into *short-run* effects (the user's immediate "state" after receiving assistance) and *longer-run* effects (what the user was able or unable to do as a result of the assistance).

As conclusion drawing proceeds, you'll often realize that a variable needs to be divided. In Chapter 7, section B, where the issue was the feelings and concerns of innovation users, the analyst was struck by the fact that some concerns, as expected, were *individual* ones (discouragement, fatigue, or friction with others), and many other concerns were *institutional* in nature (changes in district priorities, poor overall functioning of the project, or uncertainties about project continuation). This realization led to the summary tabulation in Table 7.7.

Sometimes creating a two-variable matrix, as in Box 7.3, helps clarify whether partitioning a variable will be illuminating. Here the analyst was struggling with the relationship between user "practice stabilization" and the "continuation" of the innovation locally. The first look at continuation emphasized users' attitudes to continuation as the indicator. That made sense: A stabilized innovation ought to evoke more positive attitudes on the part of its users. But the relationship was quite weak. Users from almost all sites had positive attitudes to continuation; only in very low-stabilized sites were negative attitudes found. This finding suggested splitting the "continuation" variable into two parts: users' *attitudes* toward continuation and their estimates of the *probability* of continuation.

Adding a column to the matrix led the analyst to pay dirt. The probability variable illuminated the relationship between stabilization and continuation more clearly than the attitude one, because it took into account factors beyond individuals, such as district support, staff turnover, and funding difficulties.

When is variable-partitioning a good tactic? The first answer: Divide variables in early stages (conceptualizing, coding) to avoid monolithism and data blurring. The second answer: Partition a variable when it is not relating as well to another variable as your conceptual framework (or other available data) has led you to expect.

Finally, we should say that variable-partitioning is not a virtue in itself. Extreme differentiation can lead to complexity and atomization, poor mapping of events and processes. When you divide a variable, it should be in the service of finding coherent, integrated descriptions and explanations.

Suggestions for practice. Table 5.2 has the main variable of preparedness now maximally differentiated. Try recasting the rows to make *fewer* variables (in effect, you are clustering the parts of an overdivided variable).

How might you divide the variables in Table 7.7 any further?

Consider a favorite predictor or outcome in your own study. Find a colleague who will propose ways of dividing it; do the same yourself and compare.

8. Subsuming Particulars Into the General

Clustering involves clumping together things that "go together" by using single or multiple dimensions. That process is often an intuitive, first-level process corresponding to ordinary coding (Chapter 4, section B).

A related tactic is to ask, "What is this specific thing an instance of? Does it belong to a more general class?" This tactic is like the "pattern coding" we outline in Chapter 4, section C. You are taking a step up, trying to locate the immediate act, event, actor, or activity in a more abstractly defined class. That class may have been predefined, or it may have emerged as a result of memoing (Chapter 4, section D).

For example, LeCompte (1975), looking at a classroom, noted that the following events fell into a cluster:

- Teacher vacuums room twice daily.
- Students cannot leave in the afternoon until they wash their desks.
- Teacher refers to personal hygiene when reprimanding students.
- Teacher says, "Children who come to school looking like that [soiled shirt] you just can't expect to do the same kind of work as the others."

At first glance this list is simply a cluster of behaviors dealing with cleanliness. But LeCompte also noticed that children's statements about the cleanliness issue tended to parallel the teacher's. Furthermore, most of the items (a) recurred and (b) had a strong regulatory emphasis. She concluded that the general class here was "cleanliness norm"—a shared set of standards about appropriate and inappropriate behavior.

Glaser (1978) uses this tactic during his "constant comparative method," looking for "basic social processes," such as negotiation or bargaining, that are a more general class into which specific behaviors (e.g., bickering, arguing, refusing, offering, soothing) can be subsumed.

In our school improvement study, we noted specific statements made by teachers and administrators, such as:

- If you want to depart from the guide, ask me and also tell me why you want to do it and how it will fulfill the guide's objectives.
- The basic philosophy is there, but the use of (the innovation) is flexible, and doesn't require use of all units.
- In this program you're like a robot . . . but I learned that if I wanted to change something I would just go ahead and do it. . . . I learned to cut corners and do it just as well.

These statements can be subsumed into a more general class: the presence of high or low *administrative latitude* given to teachers to adapt or alter an innovation, a variable that turned out to be very important in explaining the amount of adaptation that occurred.

The process of moving up a step on the abstraction ladder is not a mechanical or automatic process. For example, all three statements above could be seen as an instance of the more general class "adherence to objectives," if you attend to such phrases as "the guide's objectives," "basic philosophy," and "do it just as well." Here the analyst was convinced, through the presence of many other statements, that the question of administrative action restricting or permitting latitude was the more important general class.

Subsuming particulars into more general classes is a conceptual and theoretical activity (Glaser, 1978) in which you shuttle back and forth between first-level data and more general categories that evolve and develop through successive iterations until the category is "saturated" (new data do not add to the meaning of the general category).

Arbitrary movement up the abstraction ladder gets you nowhere. Suppose you observed a teacher writing her name on the chalkboard on the first day of school. That specific action can be subsumed in a larger class of "written communication," then in a larger class of "information transmission," and then in a still larger class of "human action." That is a sort of taxonomic classification without useful meaning, however.

Suggestion for practice. Consider how you might subsume that action (teacher writes name on chalkboard) into a general class.

Depending on your assumptions and the purposes of the study you were doing, the action could be classified usefully as an instance of "thoroughness of preparation," "norm setting," "system launching," "reassurance/anxiety reduction," "control through legitimacy," or "institutional perpetuation." You probably proposed still others.

You cannot decide in a vacuum which of these classes is "right" or "best." There must be a clear linkage to the study's conceptual framework and research questions. And you need to move repeatedly back down the abstraction ladder, as Korzybski (1933) always counseled, to "find the referent": the concrete instance alluded to when a phrase such as "institutional perpetuation" is used in an analysis.

Other suggestions. Do some subsuming of particulars in Table 5.7 (Hint: There may be several classes useful for different particulars) and then check Table 5.8 and the text. Try your hand at recasting Table 8.2: What alternative classes of effects would you propose?

9. Factoring

"Factoring" comes from factor analysis, a statistical technique for representing a large number of measured variables in terms of a smaller number of unobserved, usually hypothetical, variables. These second-order variables ("factors") may be largely uncorrelated, or may have some "communality," overlapping with each other. In either case it is possible to identify general themes, giving a name to the statistically identified factor or factor cluster. What is the qualitative researcher's analogue?

Most of the tactics we've discussed are designed to do two things: (a) reduce the bulk of data and (b) find patterns in them. Clustering, making metaphors, and moving up the abstraction ladder are all pattern-forcing exercises. The task essentially is saying to yourself, "I have a mountain of information here. Which bits go together?" When you derive a pattern code (Chapter 4, section C), you really are hypothesizing that some disparate facts or words do something in common or are something in common. What they do or are is the "factor," and the process by which you generate it is "factoring." Time for an illustration.

Here are some data from a study of interorganizational arrangements linking a university or college with a set of surrounding schools (Havelock, Cox, Huberman, & Levinson, 1983). At one successful site in a Midwestern state, we collected some study-relevant characteristics of the state college. Most were coded COLL-CHARS, so we could easily yank them out and list them. Here is the list:

service is a central objective
few contacts with state-level agencies
small-scale projects
low publishing rate by staff
little research activity
numerous outreach activities to surrounding schools
concern for following up on preservice training with in-service training
college staff active on community councils
use of area teachers as resource people in activities
majority of within-state college staff
few cross-university discipline contacts

Suggestion for practice. How can you "factor" such a list? Try scanning these items to see what factor might underlie them.

The analyst, in the course of successive scans, identified the thematic commonalities shown in italics.

- service is a central objective: *activist, client-centered*
- few contacts with state-level agencies: *localism* (not looking beyond own region)
- small-scale projects: *client-centered* (operating on scale of community)
- low publishing rate by staff: *service over academic orientation*
- little research activity: *service over academic orientation*
- numerous outreach activities to local schools: *service* again; *localism*, too
- concern for following up on preservice training with in-service training: *localism* (keeping connected to graduates in local area)
- college staff active on community councils: *service, localism* again
- use of area teachers as resource people in activities: *activism, nonacademic* orientation again
- majority of within-state college staff: *localism* again
- few cross-university discipline contacts: *localism*, also *non-academic orientation*

The same themes recur: activism, service orientation, client-centeredness, nonacademic emphasis, localism. We settled on a double-barreled theme we called "localism/activism." All of the characteristics fit into this theme without a conceptual shoehorn. From these 11 particulars, which most of the people we observed or interviewed had in common, we derived a general characteristic.

We also noted that the large state university, another site in the same state, had different characteristics. It, too, was activist, but it was *not* localist. Its principal client was the state, not the city where the university was located. There was more cross-university contact, and publishing was higher. So perhaps we had a bipolar factor in our study, "localism versus cosmopolitanism," with the effect of activism "controlled," analogically speaking. Now we are factoring at a slightly higher level of abstraction by using multiple cases to do a sort of "second-order factoring." It gets us to still fewer overarching themes or constructs that subsume bigger chunks of the data. We get there by asking the question, "What is there a lot of in one place that there is little of in another—and are they comparable things?"

Then comes the more consequential question, "Do these contrasts make any meaningful difference, or are they essentially decorative?" The factors have to contribute to our understanding of the case or of its underlying dynamics. Otherwise they are no more useful than the big, gift-wrapped boxes that unpack into a succession of smaller but equally empty gift-wrapped boxes, leaving us at the end with a shapeless heap of ribbon and cardboard.

In this case the localism/cosmopolitanism factor, when associated with the activism factor, was very helpful in explaining why some school-university arrangements were very successful and others limped along.

Other suggestions. Look at the text examples after Table 6.1. Can you infer a factor underlying them? Also look at column 4 of Box 7.3. What *general* factor can you see that is of a higher order than the specific "contributing factors" listed there?

10. Noting Relations Between Variables

The idea of relations between variables is prefigured in our discussion of conceptual frameworks (Chapter 2, section A), most easily depicted as sets of boxes and arrows; boxes are variables, and the arrows show relationships between them. Once you are reasonably clear about what variables may be in play in a situation, the natural next query is, How do they relate to each other?

What sort of relations can we envision between variables A and B? A variable is something that varies. Thus we might have:

1. A+, B+ (both are high, or both low at the same time)
2. A+, B– (A is high while B is low, or vice versa)
3. A↑, B↑ (A has increased, and B has increased)
4. A↑, B↓ (A has increased, and B has decreased)
5. A↑, then B↑ (A increased first, then B increased)
6. A↑, then B↑, then A↑ (A increased, then B increased, then A increased some more)

(These don't cover all of the possible permutations, of course.) Relationship 1 is a direct association: Both variables are high (or low) at the same time. For variables that are "all or nothing," this relationship can be read as follows: When A is present, B is also present, or both may be absent.

Relationship 2 is "inverse." With relationship 3, we are noting that changes have occurred recently in A and in B in the same direction; 4 is the inverse. No claims are necessarily being made that the changes are linked; they are just present.

In relationship 5, we verge toward causality: A changed, then B changed (and—not shown—there is a reasonable belief that A "could" have caused B). If A is an evening of heavy drinking and B is a headache the next morning, there is a presumptive connection. But little connection is likely—in most cases—if B is a morning headache and A is the announcement of the new city budget. (Still, if the headache belongs to the mayor, maybe . . .)

Finally, in relationship 6 we see a mutual ("nonrecursive") relation: Change in A leads to a subsequent change in B and then to a subsequent change in A. The strength of these associations can vary: We can have decisive, strong, clear relationships—or feeble, weak, ambiguous ones. And, as Davis (1985) reminds us, relations can be direct, like those above, or indirect, mediated through other variables via "ripple effects."

The basic analysis tactic here involves trying to discover what sort of relationship—if any—exists between two (or more) variables. The important thing to keep in mind is that we are talking about *variables*, concepts, not necessarily specific acts or behaviors.

Even when we are focusing on specific events, usually underlying or more general variables are involved. The event of an evening of heavy drinking and the event of the morning headache do not quite affect each other directly. All sorts of variables are at work: the presence of certain esters in the beverage involved and the body's ability to metabolize alcohol, the amount consumed, the time intervening, and so on. So we concur with Glaser (1978), who says that the researcher is "making theoretical statements about the relationship between concepts, rather than writing descriptive statements about people."

How are relationships detected? We argue in this book that matrix displays are an especially economical way to see them: Data bearing on two or more variables can be arrayed for systematic inspection, and conclusions drawn. Network displays help us look at more complex configurations, and show the temporal dimension more clearly.

For an illustration, let's use Table 5.12. What is the relationship between the "centrality" of an innovation to a person's interests and the initial attitude held toward it? Are people more favorable to a "central" innovation than one that is more superficial in their life space? Or maybe, might we find that people are more negative when an innovation looms large in their scheme of things?

Scanning down the columns of Table 5.12 for "centrality" and "initial attitude toward program," we indeed can see a relationship, but not a tight one. For users the centrality is always high and initial attitudes are neutral or favorable; only one is unfavorable. And it looks as if early apprehension is at work with the "neutral" teachers.

We can't know what the consequences of moderate or low centrality are for teachers—because we have no such cases. There is more of a range of centrality for administrators, but no clear pattern in terms of attitudes. So we have only modest support for the idea of a positive relationship. The analyst might want to go back to the field and hunt for some low-centrality teachers to fill out the picture. Pending that, we can turn to other possible associations with positive attitude—such as "career relevance," in Table 5.12. There the range is wider, and it looks as if the relationship may be clearer: For the teachers (and administrators) who indicate no career relevance, the attitudes are neutral or unfavorable; where career relevance grows (e.g., for K. Weelling), attitude becomes more positive; and for two of three teachers who indicate high career relevance, initial attitude is positive.

People tend to think in causal terms. The risk in trying to understand relationships between two variables is jumping too rapidly to the conclusion that A "causes" B, rather than that A happens to be high and B happens to be high. Here it helps to shift to verification tactics (in section B following), such as proposing and **checking out rival explanations, ruling out spurious relations,** or **using extreme cases.**

Drawing in skeptical colleagues to use one or more of these tactics can be very useful. One friend of ours says that any causal statement made about a social situation should be reversed immediately to see whether it looks truer that way:

"The students are late to class because they hate the teacher." (resistance driven by dislike)

"The students hate the teacher because they are late to class." (lateness, caused by other reasons, leads to dislike—perhaps mediated by the teacher's reactions to tardiness)

That example may sound a little fanciful, but the reversal exercise is useful. In our school improvement study, we considered this conventional statement: "Teacher involvement and commitment lead to more effort in using the innovation." Then we considered the reverse: "High teacher effort leads to teacher involvement and commitment."

That made good theoretical sense in terms of cognitive dissonance theory. And we had seen several examples of cases where early strong teacher effort led to later increases in commitment.

Suggestions for practice. Look at Table 8.5; what do you conclude about the relation between user commitment and ease of early implementation, and between user understanding and ease? (Hint: Use the Fs and Bs, as well as the check marks.) Then try your hand at the relationships between ease and the predictor variables in Table 8.6.

Using the scatterplot in Figure 7.7, what conclusions do you draw about the relation between administrative pressure and latitude? Box 7.3 is also a fruitful practice display.

11. Finding Intervening Variables

It often happens during analysis that two variables that "ought" to go together according to your conceptual expectations, or your early understanding of events in the case have only a tepid or inconclusive relation. Another puzzle is the case of two variables that do go together, but without making much sense. The analyst cannot quite figure out *why* they go together.

In both of these conditions, looking for other variables that may be in the picture is a useful tactic. Perhaps a third variable Q is confusing, depressing, or elevating the relationship between A and B, so if you "controlled" for Q, the relationship between A and B would become clearer:

$$\overset{\text{Q}}{\text{A} — ? — \text{B}}$$

Or perhaps the third variable actually fills out a reasonable chain, mediating or linking between A and B:

$$\text{A} \rightarrow \text{Q} \rightarrow \text{B}$$

Let's take the latter instance first. In Chapter 8, section E, we reported a finding: Schools adopting innovations bearing large associated funding changed more than those adopting less well funded innovations. That finding leaves a great deal unexplained. Why should it be that a well-funded innovation "induces" more organizational change?

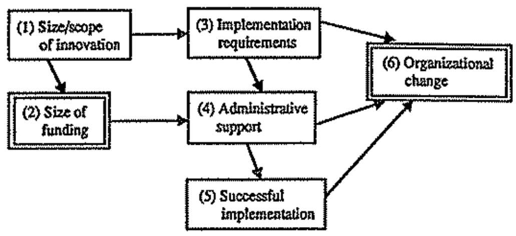
In this case the analyst created a case-ordered matrix of other possible correlates of organizational change, such as "environmental pressure," "problem-solving orientation," "implementation requirements," and "administrative support." A careful scan showed that the original relation (Figure 10.3) could be understood much more realistically when several other variables entered the picture (Figure 10.4). Here we see that "size of funding" is part of a web of other variables.

Larger innovations (1) carry more funds with them (2). The funds increase the support administrators give (4), but so do the heavier implementation requirements (3) of larger innovations. Organizational change (6) comes from at least three sources: the direct requirements of the implementation itself (3), administrative support (4), and the degree to which implementation is successful (5). As we can see, administrative support is a very central intervening variable.

In this case the effort to clarify a plausible but puzzling relationship led to a much clearer—if more complex—formulation. Simpler cases of finding intervening variables also exist. For example, we found that *administrative pressure* to use an innovation was associated with its eventual *institutionalization*. That finding looked puzzling until we discovered that cases with administrative pressure were also those where *organizational changes* had occurred—and these, in turn, supported institutionalization.

Figure 10.3
Two-Variable Relationship

Figure 10.4
Two-Variable Relationship With Intervening Variables



Now let's turn back briefly to the other sort of puzzle: where the relationship between A and B is tepid when it "ought" to be hot. The tactic here is also to examine a series of other candidate variables that may be "depressing" or confusing the relation between A and B.

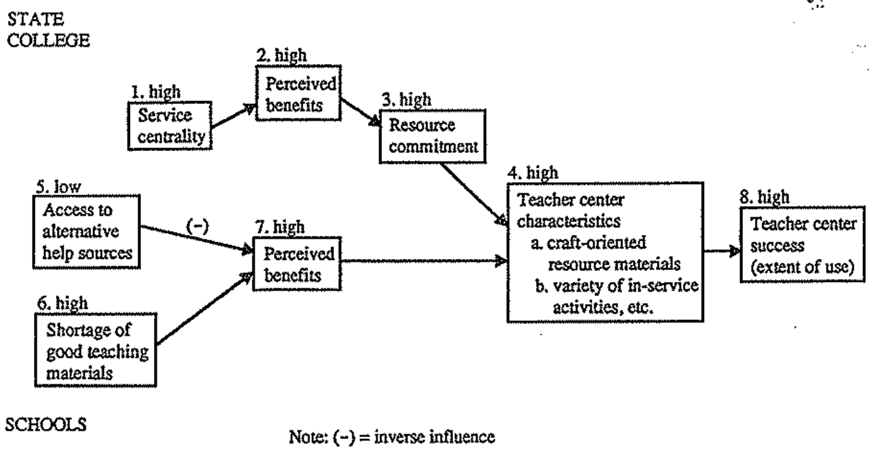
Suggestions for practice. Check Table 8.5 to get a feel for "tepidness": Preparedness and smoothness of early implementation do not relate clearly at all. What else might be in the picture?

(In this case the analyst went back to the written-up field notes and located a series of five other variables that might affect smoothness; these were displayed in a case-ordered predictor-outcome matrix [Table 8.6]. Presto! It came very clear that the main issue making for smoothness/roughness was the size of the innovation. Even if people were well prepared, if the innovation was large and demanding, early implementation was rough.)

Finding intervening variables is easiest with multiple examples of the two-variable relationship to look at, contrast, and compare. In our illustrations, these were cases. But the same principle can be used to examine multiple instances of actors, settings, or events.

Other displays useful for doing detective work on intervening variables are Box 7.5 (first look at Box 8.3 and make a note of the cases that are trying large-size innovations); Box 8.3, which shows you an intervening variable (assistance) directly; and Table 8.9 (focus on direct administrative pressure's relation to user practice change).

Figure 10.5
Example of a Chain of Evidence Supporting an Observed Outcome



12. Building a Logical Chain of Evidence

We've been talking about **noting patterns, making metaphors, clustering, subsuming particulars into the general,** and **factoring.** With these tactics, discrete bits of information come together to make a more economical whole that, analytically speaking, is more than the sum of its parts. How do you actually accomplish this? Is there some kind of heuristic or algorithm you can use?

In the study of interorganizational arrangements between schools and universities, we happened upon one especially successful case. It was a "teacher center" connected to a rural state college and undertaking a variety of in-service training activities for schools within a radius of some 60 miles.

We developed a logical chain of factors that could be leading to success, as seen from the state college side and from the school side (Figure 10.5). The logical chain of evidence goes like this. The state college might regard service and outreach activities as very central (1). Because that is, in fact, so, we would expect college staff to see the benefits (2) of a teacher center as high (they did). That should, in turn, lead to higher resource commitment (3) to the center; such commitment was found in the form of money and staff. Looking at the school side, we found few other opportunities for in-service help (5), and a shortage of good teaching materials (6); both of these should lead to high perceived benefits (7) from using the center—if the center did, in fact, give good in-service help and provide new materials. As it turned out, the high resource commitment did permit that; teacher center assets (4) and extent of use (8) were high.

This example illustrates how to build an evidential chain. Some minimal conditions have to be met. *Several* informants with *different* roles have to emphasize the factors independently and *indicate the causal links,* directly or indirectly (e.g., "We didn't have any other facility to go to to find out about new materials, so the center looked good" = the link between 5 and 7). You have to *verify* the logical predictions and the claims (e.g., the actual funds committed, the lack of alternative resource sources, the activities actually undertaken). *Countervailing evidence* has to be accounted for.

How does building a chain of evidence differ from the "causal network" method we describe in Chapter 6, section D? This approach is more tactically, specifically oriented. Building a chain of evidence requires painstaking verification *at each step:* Does this really happen, and what would we logically predict as a consequence—and does that consequence appear in the data? We are stringing together a series of **if-then** tactics: "If that were true, we should find X. We do find X. Therefore . . ."[9]

Furthermore the *relationships have to make sense;* there must be a logical basis for the claim that "perceived college benefits" lead to "resource commitment." The *chain must be complete;* the stream from antecedents to outcomes should have no gaps. For instance, in Figure 10.5, the link between 3 and 4 is not so obvious. The college might have committed resources and come up with another model or a center with different characteristics. Committing resources does not translate automatically into, say, craft-oriented resource materials. Something is missing in the logical chain and needs to be filled in.

You construct this evidential trail gradually, getting an initial sense of the main factors, plotting the logical relationships tentatively, testing them against the yield from the next wave of data collection, and modifying and refining them into a new explanatory map, which then gets tested against new cases and instances. This is the classic procedure of analytic induction. It has been used in qualitative research to make a case for the necessary and sufficient causes of social behavior (see Lindesmith's [1947, 1968] celebrated studies of opiate addiction; see also Manning [1982] for more on this method).

At its most powerful, the method uses two interlocking cycles. One is called "enumerative induction," in which you collect a number and variety of instances all going in the same direction. The second is called "eliminative induction," in which you test your hypothesis against alternatives and look carefully for qualifications that bound the generality of the conclusion. When qualitative researchers invoke "progressive focusing," they are talking about enumerative induction; when they get into "constant comparisons" and "structural corroborations," they are switching into a more eliminative inductive mode of work. The "modus operandi" logic used in several professions as a troubleshooting device—for forensic pathologists, garage mechanics, clinicians, detectives, classroom teachers—is a good example of a back-and-forth cycling between enumerative and eliminative induction.

The "decision tree" methods outlined by Gladwin (1989) can be helpful here, because they focus not so much on specific event sequences, but on the logical consequences of decisions. See also Strauss and Corbin's (1990) ideas on "tracing conditional paths."

Suggestions for practice. Look at Box 8.3, which arrays a series of variables relating to assistance and, in turn, its consequences. Try your hand at constructing a logical chain of these and any other variables you believe are logically needed. The problem here is that you are short on the concrete data needed to test each step. You can, however, partially test your chain within one of the cases by using the data in the matrix. Then try **replicating** the model with another case.

Table 5.1 and associated text can be used for practice as well; see also Table 5.14.

13. Making Conceptual/Theoretical Coherence

When you're trying to determine what someone's behavior "means," the mental exercise involves connecting a discrete fact with other discrete facts, and then grouping these into lawful, comprehensible, and more abstract patterns. With the preceding tactics, we're moving up progressively from the empirical trenches to a more conceptual overview of the landscape. We're no longer just dealing with observables, but also with unobservables, and are connecting the two with successive layers of inferential glue.

The next step is to move from metaphors and interrelationships to *constructs* and from there to *theories.* We need to tie the findings of our study to overarching, across-more-than-one-study propositions that can account for the "how" and "why" of the phenomena under study.

In the school improvement study, we came to notice that people at some of the field sites were literally exhausting themselves in the course of using new instructional practices. These very people were also making strong claims that the practice had substantially improved reading scores or children's attitudes toward school. The interesting part was that hard data to substantiate the outcome claims were either nonexistent or gave little support for them.

These are the "facts" from which we made a pattern. Field site informants could—and did—agree with the facts, but they didn't put them together as we did. To some extent we were only able to see the pattern because things were happening otherwise at *other* sites—less investment, fewer claims, or more accurate claims. Multicase field research is very useful in providing contrast and variance.

Let's call the pattern we saw "consensual delusion"—everyone agrees that more is happening with outcomes than really is. The metaphors, in gerund form, might be "groupthinking" or "self-deluding" or "wish-fulfilling." We also could sketch out a logical flowchart like the one shown in Figure 10.5 to get a fix on how this happens at the several sites. But we are still within the confines of our study; we can't converse meaningfully with other social scientists except in the sense of reporting replication of an existing finding.

The analyst now has to ask, "Do any broader constructs put these facts together the way I am putting them together?" In principle, as we noted earlier, the gerunds should help. The first one points toward group behavior, the next one toward individual cognitive process, and the third one toward motivational dynamics.

Could we invoke the idea of *group conformity* to mistaken judgments (Asch, 1951)? No, that does not help so much because it ignores the exhaustion/hard work aspect of the empirical finding.

Better: We picked up an appropriate and powerful construct from cognitive social psychology: *effort justification* (Festinger, 1957). To justify the effort expended, you "see" more results than are objectively there. This action led us into the domain of *cognitive dissonance* and how people learn to love that for which they have suffered.

Where does this process get us? For starters, it tells us that our finding has a conceptual analogue, which lends more plausibility to the finding and to the concept, which is now empirically grounded in a new context. It also helps explain why such a pattern occurs. And it throws light on larger issues (e.g., how people, at our sites and more generally, cope with uncertainty). Finally, the construct can be trained back on our cases to explain related but puzzling phenomena. For example, we can now see why objective criteria (the test scores) are being systematically ignored, when they are easily available.

We have progressed here from the bottom up—from the field to the concepts. The steps: (a) establishing the discrete findings, (b) relating the findings to each other, (c) naming the pattern, and (d) identifying a corresponding construct. (This is like Pierce's process of "abduction" [Furman, 1990]: We are making conceptual "additions" to the observed data to make them applicable more broadly.)

As we showed earlier in the book (Chapter 2, section A), it is perfectly legitimate, and often desirable, to work from the top down—from a conceptual framework to the collection of information testing its validity. Of course, you have to stay open to the idea that the concept is inapplicable, or has to be bent or discarded when you see the data. Concepts without corresponding facts are hollow, just as facts without concepts are, literally, meaningless.

One of the best discussions we have seen of building theory from qualitative data is Eisenhardt's (1989a). She emphasizes the importance of cross-case pattern finding, sharpening constructs by careful connection to your data, and looking in the literature for constructs that *conflict* with, as well as support, your emerging findings.

How to decide whether one coherent explanation is better than another? Garfinkel (1981) has useful suggestions (e.g., better explanations are often more "structural," not too local and atomistic).

Suggestions for practice. Scan Tables 5.12 or 5.14 for what patterns you see; then consider what concepts might be applicable from your own or your colleagues' disciplines that might explain the findings on a higher level.

B. Tactics for Testing or Confirming Findings

We've spent some time on particular tactics for generating meaning—making and interpreting findings at different levels of inference. Now we need to confront the issue of validity. Qualitative analyses can be evocative, illuminating, masterful—and wrong. The story, well told as it is, does not fit the data. Reasonable colleagues double-checking the case come up with quite different findings. The interpretations of case informants do not match those of the researchers. The phenomenologist chuckles, reinforced in the idea that there is no single reality to get "right"—but cannot escape a sneaky feeling that, in fact, reasonable conclusions are out there somewhere. The psychometrician, from the polar opposite stance, wonders whether nonstatistical research is an oxymoron—but often hungers for some close-up information that would tell what some puzzling numbers "really" mean.

A decade ago, we noted (Huberman & Miles, 1983a) that qualitative researchers shared no canons, decision rules, algorithms, or even any agreed-upon heuristics to indicate whether findings were valid and procedures robust. That situation is changing, but slowly. In section C below, we review some general guidelines for judging the "goodness" of qualitative research; in this section we want to focus more on what researchers can do at the tactical, operating level to test and confirm findings. With Katz (1983), we want to address the "four Rs": representativeness, reactivity, reliability, and replicability.

First let's take a longer view of the problem. Most qualitative researchers work alone in the field. Each is a one-person research machine: defining the problem, doing the sampling, designing the instruments, collecting the information, reducing the information, analyzing it, interpreting it, writing it up. A vertical monopoly.

And when we read the reports, they are most often heavy on the "what" (the findings, the descriptions) and rather thin on the "how" (how you got to the "what"). We rarely see data displays—only the conclusions. In most cases, we don't see a procedural account of the analysis, explaining just how the researcher got from 500 pages of field notes to the main conclusions drawn. So we don't know how much confidence we can place in them. Researchers are not being cryptic or obtuse. It's just that they have a slim tradition to guide their analytic moves, and few guidelines for explaining to their colleagues what they did, and how.

The other part of the problem appears when we look at research on information processing by individuals. There is a long research tradition (Dawes, 1971; Goldberg, 1970; Meehl, 1954, 1965) showing that human judgments are consistently less accurate than statistical/actuarial ones—and even that "expert" judges can be worse than untrained ones (Taft, 1955). Oskamp (1965) found that experts felt more and more confident of their *erroneous* judgments as they got more accurate information, without modifying the initial judgments. (Think of the lone field-worker out there, trying to collect trustworthy data and make sense of it.)

Why are such errors made? There's a substantial body of research on how people attribute causality (Heider, 1944; Kelley, 1967, 1971) and form judgments (Kahnemann & Tversky, 1972; Nisbett & Ross, 1980). The general finding, baldly put, is that most people are rotten scientists, relying heavily on preexisting beliefs and making bias-ridden judgments.[10] They find patterns in random data, look at incomplete data, see what they want to see, and over-rely on others' judgments, real or assumed (Gilovich, 1991). Most people don't keep track of frequencies, make probabilistic estimates, sample representatively, make accurate deductions, or entertain alternate hypotheses. Do you, in your daily life?

What these studies don't tell us is when and how researchers behave like most people. The suspicion is that they do so rather often. For example, Kahnemann and Tversky (1972) studied the "availability heuristic": Data that are "vivid," rather than "pallid," tend to be noticed, retrieved, and used more frequently. As a result you overweight the importance and frequency of concrete, immediate inputs. One personally experienced or witnessed dramatic event "means more" than several you have read about. Think now of the field-worker who witnesses a crisis or a dramatic conflict and then makes inferences about the significance of the event—and still other inferences about what the site is like when he or she is absent.

So we need to be watchful in qualitative research about the multiple sources of analytic bias that can weaken or even invalidate our findings. Some of these biases have been identified in mainstream anthropological textbooks, along with some pointers for avoiding them (e.g., see Wax, 1971). The archetypical ones include the following:

1. the *holistic fallacy:* interpreting events as more patterned and congruent than they really are, lopping off the many loose ends of which social life is made
2. *elite bias:* overweighting data from articulate, well-informed, usually high-status informants and underrepresenting data from less articulate, lower-status ones
3. *going native:* losing your perspective or your "bracketing" ability, being co-opted into the perceptions and explanations of local informants

It's useful to note that these three biases correspond, respectively, to the three major judgmental heuristics identified by cognitive researchers: "representativeness," "availability," and "weighting." We recommend Nisbett and Ross (1980) and Gilovich (1991) as good guides to this literature.

We also draw on this literature as we discuss tactics for testing/confirming findings. The language of "confirming" and "verifying" is bolder than what we can usually achieve. But the intent is still there: How can we increase our—and our readers'—confidence in what we've found? As Lee (1991) suggests, the answer lies in linking three levels of understanding: the meanings and interpretations of our informants, our own interpretations of those meanings, and our confirmatory, theory-connected operations.

We describe 13 tactics, beginning with ones aimed at ensuring the basic quality of the data, then moving to those that check findings by examining exceptions to early patterns, and conclude with tactics that take a skeptical, demanding approach to emerging explanations.

Data quality can be assessed through **checking for representativeness** (1); **checking for researcher effects** (2) on the case, and vice versa; and **triangulating** (3) across data sources and methods. These checks also may involve **weighting the evidence** (4), deciding which kinds of data are most trustable.

Looking at "unpatterns" can tell us a lot. **Checking the meaning of outliers** (5), **using extreme cases** (6), **following up surprises** (7), and **looking for negative evidence** (8) are all tactics that test a conclusion about a "pattern" by saying what it is *not* like.

How can we really test our explanations? **Making if-then tests** (9), **ruling out spurious relations** (10), **replicating a finding** (11), and **checking out rival explanations** (12) are all ways of submitting our beautiful theories to the assault of brute facts, or to a race with someone else's beautiful theory.

Finally, a good explanation deserves attention from the very people whose behavior it is about—informants who supplied the original data. The tactic of **getting feedback from informants** (13) concludes our list.

As before, we suggest practice opportunities, often inviting you to consider your own data—which is where the confirmation crunch comes.

1. Checking for Representativeness

When we come up with a "finding" in a field study, we quickly assume it to be typical, an instance of a more general phenomenon. But is it? And, if it is, *how* representative? (One of us, a few minutes after picking up first-time foreign visitors at Kennedy airport, often asks, only half-ironically, "What do you think of America?" A first impression of an entire country, sometimes less than ironic, usually emerges—on the basis of a sample of the terminal's bustle, a parking lot, and some roadways.)

In their studies of the "representativeness heuristic" used by individuals, Tversky and Kahneman (1971) show how biased we are in moving from particulars to generalities. From one or two concrete, vivid instances, we assume dozens more are lurking in the bushes—but we do not verify whether or how many (usually fewer than we think).

People typically make a generalization, then illustrate it ("For example, my friend X . . . "), but would be hard put to come up with several more instances of a supposedly widespread occurrence. To compound the problem, people as information seekers—and as processors—are far more likely to see *confirming* instances of original beliefs or perceptions than to see disconfirming instances, even when disconfirmations are more frequent (e.g., see Edwards, 1968).

Still more ominously, Kahneman and Tversky (1972) caught mathematical psychologists in the act of making biased inferences from samples to populations. What, then, of the field-worker who, Prometheus-like, is doing all of the sampling, measuring, and inference making *de novo?*

Operating alone, without any standardized or validated instruments, the field researcher runs several risks of generalizing wrongly from specific instances. Here are some of the most common pitfalls, and their associated sources of error:

Pitfall	*Source of Error*
sampling nonrepresentative informants	overreliance on accessible and elite informants
generalizing from nonrepresentative events or activities	researcher's noncontinuous presence at the site; overweighting dramatic events
drawing inferences from nonrepresentative processes	nonrepresentative informants and events; heavy reliance on plausibility; good fit into emerging explanation; holistic bias

The first pitfall highlights the fact that you can talk only with people who can be contacted; some people are harder to contact than others. This problem in itself signals something particularistic; their accessibility may be connected with work load, lack of cooperativeness, or both.

Anthropologists have often warned of field-workers' tendency to rely too much on articulate, insightful, attractive, and intellectually responsive informants; such people often turn out to be in the local elite. For a good discussion of the atypicality of persons chosen as prime informants, see Dean, Eichorn, and Dean (1967).

The second pitfall results from the researcher's noncontinuous presence; you have to infer what is happening when you are not there. When you observe a dramatic event (a crisis, an argument), the tendency is to assume it has "built up" when you were absent, or that it "symbolizes" a more general pattern. These are plausible, but certainly not well-grounded inferences.

Finally, in looking for underlying processes explaining what you've seen and heard, you draw heavily from the people, events, and activities you have sampled. But if the samples are faulty, the explanation cannot be generalized beyond them.

"Plausibility" is the opiate of the intellectual. If an emerging account makes good logical sense and fits well with other, independently derived analyses within the same universe, you lock onto it and begin to make a stronger case for it; this action constitutes the "confirmability" bias studied by Edwards (1968) and others.

Elementary safeguards: Some illustrations. The real problem with selective sampling and abusive generalizing is that you can slide incrementally into these biases, with the first layer preparing the ground for the next. Gradually you become a prisoner of your emerging system of comprehending the case. There is no longer any possibility, cognitively speaking, of "standing back" or "reviewing critically" what you have observed up to then. What you now understand has been accumulated very gradually from within, not drawn validly from without.

So if you want to stand back and review critically, you need someone else to do it—or you must build in safeguards against self-delusion. We've already addressed the former approach (critical friends, colleagues, other field-workers, auditors), so let's illustrate some safeguards.[11]

In our study of linkage between universities and local school districts, we worried that we were drifting into the three pitfalls just listed. Each researcher covered several institutions, which meant discontinuous field visiting. We wondered if we were systematically missing people, even though the people we *did* talk to assured us that we had sampled broadly.

Our first safeguard was to look for the **outliers** (see tactic 5 below) in order to check the representativeness of those with whom we had talked. As a second safeguard, we asked for personnel rosters and randomly sampled eight informants from each one for interviewing. As it turned out, we had been in danger of sampling a biased group.

In the same study, informants often talked about one type of project, saying it was the most frequent activity sponsored by the university. We began to wonder whether informants were confusing frequency with impact; did they remember this activity type most because it seemed to have been the most practice-relevant? So we took the last two years' newsletters, extracted a calendar of activities, and content-analyzed them. Informants had, in fact, misled themselves and us by overweighting one activity type.

A final example from that study fits the third pitfall. Policy board meetings were held periodically between delegates from the university and the school districts. From our observations it appeared that university staff were dominating them; the university delegates spoke more often and longer than the school people, and most decisions went their way. Pattern: a university-dominated arrangement, an idea that fit the data we had.

To check it, we sampled more widely, observing other meetings that (a) were less formal and did not involve representatives or delegates or (b) were formal, but not as prominent as the policy board. This was salutary; our initial hypothesis got bent. In less formal, less prominent meetings, school people held their own in the decision making.

Some advice: Assume you are selectively sampling and drawing inferences from a weak or nonrepresentative sample of "cases," be they people, events, or processes. You are guilty until you prove yourself innocent—by extending the universe of study. How?

1. Increase the number of cases.
2. Look purposively for contrasting cases (negative, extreme, countervailing).
3. Sort the cases systematically (perhaps using substructing, as in Box 7.1) and fill out weakly sampled case types.
4. Sample randomly within the total universe of people and phenomena under study.

The last two procedures correspond to the "stratification" and "randomization" conventions used by experimental researchers to enhance internal validity. But while the experimental researcher uses the conventions *early,* as anticipatory controls against sampling and measurement error, the qualitative researcher typically uses them *later,* as verification devices. That use allows you to let in all the candidate people and data, so the most influential ones will have a chance of emerging. But you still have to carry the burden of proof that the patterns you ultimately pinpoint are, in fact, representative.

Suggestions for practice. Here we suggest focusing on a study you are involved in. Use the pitfalls and error sources we outlined as a checklist and consider what concrete safeguards you might initiate. Ask a colleague to help.

2. Checking for Researcher Effects

"Outsiders" to a group influence "insiders," and vice versa. So it is with the researcher who disembarks in a field setting to study the natives. You are likely, especially at the outset, to create social behavior in others that would not have occurred ordinarily.

That behavior, in turn, can lead you into biased observations and inferences, thus "confounding" (a nice term in this instance) the "natural" characteristics of the setting with the artificial effects of the researcher-native relationship. Unconfounding them is like threading your way through a hall of mirrors.

So we have two possible sources of bias here: (A) the effects of the researcher on the case and (B) the effects of the case on the researcher. Field study researchers are often less worried about bias A because they typically spend enough time on-site to become part of the local landscape. But that, of course, increases the danger of bias B: being co-opted, going native, swallowing the agreed-upon or taken-for-granted version of local events.[12]

Although we discuss these biases as they occur during site visits, they influence analysis deeply, both during and after data collection. The researcher who has gone native stays native during analysis. The researcher who has influenced the site in un-understood ways suffers unaware from that influence during analysis.

Bias A occurs when the researcher threatens or disrupts ongoing social and institutional relationships. People now have to figure out who this person is, why he or she is there, and what might be done with the information being collected. While they are figuring that out, informants typically will switch into an on-stage role or special persona, a presentation of self to the outsider. (They have other personae, of course, for fellow insiders, as Goffman, 1959, shows so well.)

Even after this preliminary dance, informants will often craft their responses to be amenable to the researcher and to protect their self-interests. For some analysts (Douglas, 1976), local informants' interests are fundamentally in conflict with those of the researcher, who might penetrate to the core of the rivalries, compromises, weaknesses, or contradictions that make up much of the basic history of the site. Insiders do not want outsiders—and sometimes other insiders—to know about such things. So the researcher, who is usually interested in uncovering precisely this kind of information, must assume that people will try to be misleading and must shift into a more investigative mode.[13]

Field research can, at bottom, be considered as an act of betrayal, no matter how well intentioned or well integrated the researcher. You make the private public and leave the locals to take the consequences.[14] (We discuss this issue further in Chapter 11.)

But that is not the only way bias A can occur. In some instances biases A and B can team up to create artifactual effects as a result of the complicity between researcher and local actors. This is the famous "experimenter" effect, studied intensively by Rosenthal (1976).

We've been caught napping several times on this one. For instance, one field site in the school improvement project was about to phase out the project we had come to see. For some mysterious reason, the phase-out decision was canceled during our time on-site. The reasoning, which we unraveled only after several more days, was that the practice *had* to be better than it appeared, because people (that is, we) had come from so far away to see it. There was also the desire to avoid a public indictment; the researcher and/or the public reading her research might convey the impression that the school had botched things.

Bias A can take still other forms. For example, local informants can implicitly or explicitly boycott the researcher, who is seen variously as a spy, a voyeur, or a pest. Or the researcher can inhibit local actors. After several days on-site and multiple interviews, people are not sure anymore how much the researcher has found out, and assume—wrongly in most cases—that the researcher knows too much. This opinion then triggers bias B: The researcher accordingly becomes more reassuring or, alternatively, moves into the investigative-adversarial mode. Both strategies are likely to affect the data being collected.

Assuming, then, that you have only a few months, weeks, or even days on-site, how can these two interlocking forms of bias be countered? Below is a short shopping list of suggestions, many of which are treated in far more detail in fieldwork-related literature (e.g., Adams & Preiss, 1960; Bogdan & Biklen, 1992; LeCompte et al., 1992; LeCompte & Preissle, with Tesch, 1993; Lofland & Lofland, 1984; Morse, 1989; Pelto & Pelto, 1978; Wax, 1971; Whyte, 1984).

A. *Avoiding biases stemming from researcher effects on the site:*

- Stay as long on-site as possible; spend some time simply hanging around, fitting into the landscape, taking a lower profile.
- Use unobtrusive measures where you can (McCall & Simmons, 1969; Webb et al., 1965).
- Make sure your intentions are unequivocal for informants: why you are there, what you are studying, how you will be collecting information, what you will do with it.
- Consider co-opting an informant—asking that person to be attentive to your influence on the site and its inhabitants.
- Do some of your interviewing off-site, in a congenial social environment (cafe, restaurant, informant's home), by way of reducing both your threat quotient and your exoticism.
- Don't inflate the potential problem; you are not really such an important presence in the lives of these people.

B. *Avoiding biases stemming from the effects of the site on the researcher:*

- Avoid the "elite" bias by spreading out your informants; include lower-status informants and people outside the focus of your study (peripheral or former actors).
- Avoid co-optation or going native by spending time away from the site; spread out site visits (see Whyte, 1943, on "temporary withdrawal").
- Be sure to include dissidents, cranks, deviants, marginals, isolates—people with different points of view from the mainstream, people less committed to tranquility and equilibrium in the setting.
- Keep thinking conceptually; translate sentimental or interpersonal thoughts into more theoretical ones.
- Consider finding an informant who agrees to provide background and historical information for you and to collect information when you are off-site (such co-optation may be more useful, in bias-reduction terms, than the information provided).
- Triangulate with several data collection methods; don't overly depend on talk, or on observation, to make sense of the setting.
- If you sense you are being misled, try to understand, and focus on why an informant would find it necessary to mislead you. Follow that trace as far upstream as you can.
- Do not casually show off how much you do know; this is a covert plea for confirmation that deludes only the person making it.
- Show your field notes to a colleague. Another researcher is often much quicker to see where and how you are being misled or co-opted.
- Keep your research questions firmly in mind; don't wander too far from them to follow alluring leads or drop them in the face of a more dramatic or momentous event.

As with all such lists, following some items gets you in trouble on others. For instance, if you have only a few days on-site, off-site interviewing may be too costly. Or you may be co-opted by the informant you are trying to co-opt.

Supposedly bias detection and removal take time. The more time you have, the more layers you can peel off the setting to get to the core explanatory factors, and the less subject you are to biases A and B. However, we take that with a grain of salt. Long exposure can just push up bias B and make bias A harder to see (cf. Whyte's [1984] quote from another researcher: "I began as a nonparticipating observer and ended up as a nonobserving participant.").

We say again that people who are discreet, savvy in the environment under study, and conceptually ecumenical are often able to get to the core of a case in a matter of days, sidestepping both types of researcher bias and coming away with good-quality data. It's possible that the methodologists demanding months or years on-site before valid data can be obtained are confusing time with competence.

Suggestions for practice. Here, too, the data displays and reports in this book do not help much. You need to examine your own study for possible A- and B-type biases, seek colleague feedback, and then invent and try out safeguards.

3. Triangulating

Much has been written about "triangulation" as a near-talismanic method of confirming findings. Stripped to its basics, triangulation is supposed to support a finding by showing that independent measures of it agree with it or, at least, do not contradict it.

It's not as simple as that, so we won't try to be brief. For one thing, if they do contradict it, and if you have only two measures, you are stuck with a deeper question: Which do you believe?[15]

First, a little history and some examples. Webb et al. (1965) seem to have coined the term: They spoke of validating a finding by subjecting it to "the onslaught of a series of imperfect measures." Before that was "concurrent validity" in the development of psychological tests, checking items against other, already validated measures by using different measuring methods (Campbell & Fiske, 1959); and the methods of "corroboration" and "multiple validation procedures" (Becker, 1958) used to ensure the dependability of a field study finding.

Here, too, listen to the historian Barbara Tuchman (1981):

> Bias in a primary source is to be expected. One allows for it and corrects it by reading another version.... Even if an event is not controversial, it will have been seen and remembered from different angles of view by different observers.... As the lion in Aesop said to the Man, "There are many statues of men slaying lions, but if only the lions were sculptors there might be quite a different set of statues." (p. 19)

There are also links to the *modus operandi* approach used by detectives, mechanics, and general practitioners. When the detective amasses fingerprints, hair samples, alibis, and eyewitness accounts, a case is being made that presumably fits one suspect far better than others; the strategy is pattern matching, using several data sources. Diagnosing engine failure or chest pain follows a similar approach. The signs presumably point to the same conclusion and/or rule out other conclusions. Note the importance of having different kinds of measurements, which provides repeated verification.

What kinds of triangulation can there be? Following Denzin's (1978) classic distinctions, we can think of triangulation by *data source* (which can include persons, times, places, etc.), by *method* (observation, interview document), by *researcher* (investigator A, B, etc.), and by *theory.*[16] To this we can add *data type* (qualitative text, recordings, quantitative).

How to choose which? The aim is to pick triangulation sources that have different biases, different strengths, so they can complement each other. Carney (1990) also suggests moving from uncontrolled to more controlled data types. In effect a new source forces the researcher to **replicate** the finding in a place where, if valid, it should reoccur.

In some senses we are *always* faced with triangulating data, whether we attend to it or not. Here is Brown's (1991) quote from V. W. Turner: "There may be a correspondence between life as lived, life as experienced, and life as told, but the anthropologist should never assume the correspondence, or fail to make the distinction" (Bruner, 1984, p. 7).

Now a few examples. Stringfield and Teddlie (1991) compared interviews with school people, researcher observations of classrooms, and child achievement data (which were unknown to the other two data sources), finding good convergence. Stringfield and Teddlie (1990) also compared their findings with those from another *study* of schools (Huberman & Miles, 1984). Warner (1991) asked children to make camcorder recordings of adult-child interactions; the camcorder operator's spontaneous comments were also recorded and added immensely to interpretive validity.

What can we expect from triangulation? We may get *corroboration;* more typically, we can "bracket" the findings, getting something like a confidence interval (Greene et al., 1989). Even then, we may only be getting "reliability" rather than validity information, as Mathison (1988) notes.

It may well happen that our different sources are *inconsistent* or even directly *conflicting.* Help! At the best, this can push us into a more complex, context-respecting set of explanations (Jick, 1979) and helpfully reduce what Mark and Shotland (1987) call "inappropriate certainty." Such results may help elaborate our findings, or even initiate a whole new line of thinking (Rossman & Wilson, 1984).

Perhaps our basic point is that triangulation is not so much a tactic as a way of life. If you self-consciously set out to collect and double-check findings, using multiple sources and modes of evidence, the verification process will largely be built into data collection as you go. In effect, triangulation is a way to get to the finding in the first place—by seeing or hearing multiple *instances* of it from different *sources* by using different *methods* and by squaring the finding with others it needs to be squared with. Analytic induction, once again.

Suggestions for practice. Consider your own study and some of its preliminary findings. Then try what Neff (1987) did: make a matrix of findings by data sources/methods/types to see how well supported they are, and to note any inconsistencies and contradictions.

Many of the displays mentioned as practice locations for **making contrasts/comparisons** (section A, tactic 6) also can be used for triangulation practice.

4. Weighting the Evidence

Any given preliminary conclusion is always based on certain data. Maybe we should use the word some historians have employed: *capta.* There are events in the real world, from which we "capture" only a partial record, in the form of raw field notes, from which we further extract only certain information in the form of write-ups, which we then call "data." There is, in turn, further condensation, selection, and transformation as these data are entered into various displays.

Some of these data are "better" than others. Fortunately you can exploit that fact beautifully in verifying conclusions. If the data on which a conclusion is based are known to be stronger, more valid than the average, then the conclusion is strengthened. Stronger data can be given more weight in the conclusion. Conversely, a conclusion based on weak or suspect data can be, at the least, held lightly and, optimally, discarded if an alternative conclusion has stronger data back of it.

Basically, there is a very large range of reasons why certain data are stronger or weaker than others—essentially the question is one of validity (Dawson, 1979, 1982; Kirk & Miller, 1986)—more specifically, "descriptive" and "interpretive" validity, as discussed by Maxwell (1992c). We cannot be encyclopedic here, but suggest a number of markers the analyst can use in deciding whether to give more weight to some data than to others. (See also the discussion of internal validity in section C below.)

First, data from some informants are "better." The informant may be articulate and reflective and may enjoy talking about events and processes. Or the informant may be knowledgeable, close to the event, action, process, or setting with which you're concerned. In our study, for example, we gave more weight to school superintendents' judgments about forward budget categories than we did to those of teachers about that topic.

Second, the circumstances of the data collection may have strengthened (or weakened) the quality of the data. Here is a partial list (see also Becker, 1970; Bogdan & Taylor, 1975; Sieber, 1976):

Stronger Data	*Weaker Data*
Collected later, or after repeated contact	Collected early, during entry
Seen or reported firsthand	Heard secondhand
Observed behavior, activities	Reports or statements
Field-worker is trusted	Field-worker is not trusted
Collected in informal setting	Collected in official or formal setting
Respondent is alone with field-worker	Respondent is in presence of others, in group setting

These differences do not always run this way. For example, early data may be stronger if you have gone native later on. Or data in official settings can tell you a lot about role expectations that might be denied in an informal interview. But judicious weighting along these general lines can help.

Third, data quality may be stronger because of a field-worker's validation efforts. These may be of several varieties:

checking for researcher effects and biases (section 2)
checking for representativeness (section 1)
getting feedback from informants (section 13)
triangulating (section 3)
looking for ulterior motives
looking for deception

The last two have not been attended to in other sections. Douglas (1976) emphasizes that regardless of the degree of trust a field-worker may believe has developed, people in field sites nearly always have some reasons for omitting, selecting, or distorting data, and also may have reasons for deceiving the field-worker (not to mention deceiving themselves). If you have entertained such a view of certain respondents, and of the data from them, and have done something to validate the data, more confidence is justified. The interested reader should consult Douglas (1976) for a wide range of specific interventions. Here are a few:

- Check against "hard facts."
- Check against alternative accounts.
- Look for "the trapdoor"—what is going on beyond the obvious.
- Share own personal material to open up the respondent.
- Assert your knowledge of "what's going on" and see whether respondent buys it.
- Summarize a state of affairs and ask the respondent to deny it.
- Name possible ulterior motives, and see respondent response (denial, acknowledgment).

Field-workers who rely mainly on "trust" may quail at such interventions or dismiss them as too intrusive. Nevertheless, Douglas makes a good case for such validating tactics when the object of investigation has good reasons for being evasive and/or self-deceiving (some of his studies have included people who had attempted suicide, clients of massage parlors, habitués of nude beaches, and police who work in emergency rooms). And even for less dramatic settings, we have found that it pays to be suspicious, to expect to be lied to sometimes, to look for respondent self-deception, and to push respondents from time to time on such matters.[17]

Two added suggestions. First, we've found it useful to keep a running log of data quality issues (often in the form of reflective or marginal remarks on the field notes; see Boxes 4.2 and 4.3), together with recurrent efforts to improve data quality in subsequent site visits.

Second, when approaching final write-up of a case analysis, it's useful to summarize your views of data quality—both for yourself and for your eventual readers. Here is an example from a case report in our school improvement study, which appeared after the researcher listed the number of interviews (46), informal talks (24), and observations (17) held during three site visits:

> The data base is probably biased toward administrators and central program personnel (3-6 interviews apiece), and may underrepresent those of normal program users, and certainly those of peripheral (and more disenchanted) people. So the information may be fuller about the ins and outs of operation as seen by key operators, and thinner on what day-to-day life in the Carson schools is like, with the IPA program as a feature in that life. . . .
>
> I have moderately good retrospective data on the first and second years of the program, except for assistance provided. Data on the current year's operations are quite full, except that I observed no actual student-parent-teacher conferences in the high school. The only key informant I missed talking to was the former IPA director, Helena Rolland. I should probably have interviewed Mark Covington, as the archetypal coach. But with these exceptions the coverage was thorough. . . .
>
> In spite of write-up delays up to several weeks, I experienced little decay from notes. Where puzzling or meaningless notes could not be reconstructed, this was marked in the field notes and noted during dictation; such indications appeared, however, for only 3-5 percent of field notes.

Editorial Comment: The researcher's confidence about "little decay" is almost surely self-deluding. The probable loss of detail will need to be compensated for by triangulation and by looking for repeat examples from the same respondent. And comments from "disenchanted" people probably should be given more weight in conclusion verification.

Suggestions for practice. Consider the database for your current study and apply the "stronger/weaker" characteristics to it as a checklist. A data accounting sheet (Boxes 4.4 and 4.5) is very useful here.

At a more micro level, examine a specific data display and mark it up with +'s, ?'s and —'s to indicate where you would give more weight to specific items. Table 8.5 shows a convention for indicating extra weight in the display itself.

5. Checking the Meaning of Outliers

Any given finding usually has exceptions. The temptation is to smooth them over, ignore them, or explain them away. But *the outlier is your friend.* A good look at the exceptions, or the ends of a distribution, can test and strengthen the basic finding. It not only tests the generality of the finding but also protects you against self-selecting biases, and may help you build a better explanation.

For example, in the school improvement study, we happened on one case where the new practice was seen by many teachers as a near-miraculous cure for local ills. Although teachers found it hard to get on top of, the project eventually led to dramatic increases in reading and composition scores. Enthusiasm was high.

Was it a "miracle cure"? To test the finding, we asked about people who either had not adopted the practice or had used it and found it wanting. After some thought, our key informants came up with one each.

Our interviews with these two people were instructive. First, we found that their reasons for not adopting were opposite to—and thereby coherent with—the reasons given by the other informants for adopting. Then we found that the dissident user had not really mastered the innovation in the way the contented users had. We already had good evidence linking technical mastery to positive results. So our findings were strengthened, and we understood far better why deviant cases were deviant.

So it came clearer that the innovation was only "miracle cure"-like if it were technically well carried out. Furthermore, these dissidents told us there were more people like them around than advocates had admitted.

We realized then that we had oversampled contented users and, in a sense, had been "sucked in" to the taken-for-granted version of events among local actors. In widening the sampling of discontented users thereafter, we got a somewhat different and more intricate picture of the case.

There are often more exceptions or deviant cases than you realize at first; you have to go looking for them. They do not come calling, nor do you usually think spontaneously of sampling for them. After all, they are inconvenient—not only hard to reach or observe but also spoilers of your artfully built, coherent version of case dynamics.

Remember, too, that outliers are not only people; they can be discrepant *cases,* atypical *settings,* unique *treatments,* or unusual *events.* You need to find the outliers, and then verify whether what is present in them is absent or different in other, more mainstream examples (see also the discussion of **using extreme cases,** section 6).

A good illustration appears in Figure 7.7, where the Astoria case proved to be an outlier. Most other cases had high pressure to adopt the innovation and gave little latitude to users, or vice versa—little pressure and lots of latitude. Only in Astoria was there high pressure and high latitude. Why should this be? A look back at the case report showed that some careful administrator-teacher bargaining had been going on: The administrator said the innovation was mandated, but agreed that it could be flexibly adapted. Astoria was also the only parochial school in our sample, suggesting that authority was less questioned. In this instance the outlier exploration gave us more confidence that the basic finding was right.

Finding outliers is easier when you have well-ordered displays. Cases, roles, settings, events, and individuals can be shown in coherent relation to others. If you are still collecting data, display what you have and go for the outliers. If things are closely clumped (no apparent outliers), consider where you might go to find some outlying persons, events, settings. And, on following up, be cognitively open to the eventuality that the exceptional cases can turn out to be prototypical ones.

Suggestions for practice. Review the illustrations of outlier analysis in Box 7.5 and Table 8.9 to get a further flavor. You nearly always have to go back to *other* data than those shown in the display—to what else you know about the case. The best practice strategy is to take a display of your own, review the main finding, look for outliers, and work on what they mean by reviewing what else you know—or by collecting new data to understand outlierdom better.

In many cases outlier analysis strengthens an original conclusion ("the exception proves the rule"). But be careful. Don't force it. Stay open to the idea that the outlier is telling you something useful and important about how your conclusion needs changing.

6. Using Extreme Cases

We've just described the use of outliers in deepening preliminary conclusions. Outliers of a certain type, which we call extreme cases, can be very useful in verifying and confirming conclusions.

Here we can use two illustrations from Sieber (1976). The first involves extreme *situations.* Sieber asks us to imagine a situation in which an educational innovation failed, and then to look at the possible antecedents of that failure. If we found that a particular case had many positive factors—such as high motivation to innovate, access to resources, and implementation skills—but that *administrative support* was lacking, we could argue that we have found the key factor responsible for the failure. This is a tactic of "holding everything else constant"—looking for the most extreme case, where there should have been success, but wasn't. Note that you need conceptual and/or empirical knowledge of the variables involved; it cannot be done in a vacuum. You are not just looking for empirical outliers, but are conceptually defining extreme cases, and looking at whether they exist.

Sometimes extreme cases don't carry as much conceptual freight as in Sieber's example. In Table 7.16 the great majority of people changed jobs as a direct result of work with an innovation. The analyst wanted to look at the 17% of people who moved, but for *other* reasons. These turned out to be local turbulence and personal, preexisting plans—two rather different reasons.

The second sort of extreme case Sieber mentions is *persons* known to have a strong bias. For example, suppose you are talking with a very conservative administrator who you know from past contact is rather defensive. You ask him why the teachers he works with are reluctant to try innovations. He answers that it's due to his own lack of cooperation and support. That answer is very persuasive because you wouldn't expect this particular administrator to make such a statement at all.

To put this another way: Look for the person in a site who would have the most to gain (or lose) by affirming or denying something, and pop the question. If you get a surprising answer (e.g., the person who has much to gain by denying the statement/question, in fact, affirms it), then you can be more confident. This maneuver requires you to have a good prior understanding of the person's typical stance and biases.

In a way, this is another style of differentially **weighting evidence** (section 4). For example, if you are interviewing enthusiastic proponents of an innovation, their comments on the innovation's warts and trouble spots should be taken quite seriously. Once again, you are conceptualizing what "extreme" means, and capitalizing on the data you find there.

Suggestions for practice. Scan Table 7.13. Look at the columns dealing with meta-level outcomes and side effects. By looking at the high-impact and then the low-impact cases, what do you conclude about the trustworthiness of the moderate-impact claims to have experienced meta-effects?

In Box 7.3, sort out the extreme cases where continuation is seen as high, and develop an explanation. Test the explanation further by looking at the lows.

7. Following Up Surprises

Surprises have more juice than outliers. When you are surprised, it usually means that something has occurred well outside the range of your expectations. You had an implicit theory of what was happening, like, "I am coming home after a hard day at the office." And—surprise! You walk into a roomful of friends who are giving you a half-birthday party 6 months in advance. So you laugh, bask, and maybe reflect on their kindness, and perhaps on the perfidy of your Significant Other, who colluded with the plot to delude you . . . in a good cause, of course. And how did all of the shopping get done without your noticing?

In qualitative analysis, the party is less important than the follow-up reflection and sleuthing. What does this event tell me about my expectations, implicit theories, taken-for-granted assumptions? And where can I go in my data to help me rebuild my theory?

In Table 5.9, for example, the analyst was surprised that principals, supposedly close to the innovative action, did not know much about the innovation. A check with the field notes showed they actually were forbidden by rule to concern themselves with curriculum.

Or in Table 6.1, which displayed much evidence that teachers were preoccupied with maintaining their own autonomy, we encountered a surprise: The issue seemed to be resolved by more collaboration and linkage with others, not less. How could that be? A careful check with the case report first tried the hypothesis that teachers gained increased power and influence through collaboration in the "Management Council." Not so, it appeared. Maybe it was a matter of shared work with others leading to new benefits, more classroom effectiveness. That fit better. So the creation of new linkage structures in the school made the maintenance of autonomy less desirable; collaboration meant gain, not loss.

Following up surprises has three aspects. You (a) reflect on the surprise to surface your violated theory, (b) consider how to revise it, and (c) look for evidence to support your revision. You may also work from (c) to (b), hunting in a sniffing mode to find new aspects of the case that possibly could lead to a new theory.

Suggestions for practice. Try Box 8.2. There's a surprise there: Sites that received *low* amounts of assistance report more confidence and competence, whereas high-assistance ones do not. What? See whether you find anything that would explain it by scrutinizing Table 8.1.

Table 8.5 shows us that, in roughly-implementing sites, in fact, *more* in-service help is being given. Why wouldn't it have helped things go more smoothly? See whether you can develop several alternative explanations here. Then go on to Table 8.6 and see whether any of them fit—or improve on—what the analyst concluded.

8. Looking for Negative Evidence

The tactic **looking for negative evidence** is easy to describe but, given people's pattern-making proclivities, not naturally come by. When a preliminary conclusion is in hand, the tactic is to say, "Do any data oppose this conclusion, or are any inconsistent with this conclusion?" This is a more extreme version of looking for **outliers** (section 5) and for **rival explanations** (section 12); you are actively seeking *dis*confirmation of what you think is true. As Ely et al. (1991) point out, negative evidence *refutes* a construct, rather than merely refining it, as outlier data often do.

Einstein is supposed to have said, "No amount of evidence can prove me right, and any amount of evidence can prove me wrong." That is so, in the abstract, but most of us act as if the converse were true. Our beautiful theories need little data to convince us of their solidity, and we are not eager to encounter the many brutal facts that could doom our frameworks.

As Glaser and Strauss (1967) remark, "There are no guidelines specifying how, and how long, to search for negative cases," so it is never quite clear what "enough" effort is. For a good case study of such an effort, see the account in Judd et al. (1991, pp. 310-314) of Cressey's (1953) classic study of embezzlers; Cressey revised his working hypotheses five times, looking for negative cases for each version in turn, using prior data, newly collected data, and the studies of others, until there was no disconfirmation. Judd et al. argue that "what makes qualitative research systematic is not standardization but negative case analysis."

Miller (n.d.) advises that "disconfirming instances should be handled with care"—discarding your original hypothesis too quickly or modifying it hastily to accommodate the negative evidence are both undesirable. Miller suggests that although one instance of negative evidence may be enough to require reconsideration, the *proportion* of negative to positive evidence should probably play a part.

Suggestions for practice. The easiest way to proceed is to commission a friendly but curmudgeonly skeptic to take a good look at the conclusion at hand, avoiding your data display and seeking data back in the written-up field notes that would effectively disconfirm your conclusion.

If such evidence is found, do your best to rejoice, and proceed to the formulation of an alternative conclusion that deals with the evidence. If such evidence cannot be marshaled in, say, half the time it took you to do the analysis, more confidence is justified.

But note what might be called the "delusional error." Absence of negative evidence can never be decisive as a confirmatory tactic. As in this exchange:

Q. Why do you have that blue ribbon on your little finger every day?

A. It's to keep elephants from following me.

Q. But there are no elephants here.

A. See? It's working.

9. Making If-Then Tests

If-then tests are the workhorse of qualitative data analysis. They are, in Kaplan's (1964) terms, " 'test hypotheses,' . . . what we very well think may be the truth of the matter, and we then organize the inquiry so as to facilitate the decision on whether the conjecture is correct." They are more focused than a generalized "working hypothesis," which supports a general analytic direction.

The classic formal statement is, "If p, then q." It is a statement of expected relationship. Assuming p to be true (an important condition), then we look to see whether q is true. If q is true, then we have a building block for understanding. We are a long way from a "law" about the relation of p and q, which requires universality, linkage to a larger theory, and nontriviality, among other conditions. But we know more than we did, and can take some next analytic steps—notably, making more if-then tests and connecting them into a general theory of what is happening.

As an example: In one study we gradually got the picture that school administrators in the district office were aggressively promoting changes in three local schools, bypassing the school principals. This looked important, because it could realign the entire power-influence network. We reasoned that, given the bypassing, a principal would have relatively little detailed understanding of the changes even though they were happening in that principal's own building. So we asked the principals to tell us about the new practices. As hypothesized, there was a good deal of hemming and hawing, even in the very small schools that principals could cover easily.

Here are some further illustrations:

Display	*IF—*	*THEN—*
Table 5.9	Director and chairs share work on curriculum.	Their views of the innovation will be similar, more than with teachers.
Table 6.1	Teachers are on the Management Team.	They'll report having more influence over school matters.
Table 7.7	The period of implementation is "later."	Institutional concerns will be more frequent than individual ones.
Box 7.2	The innovation is developed locally.	Change in users will be greater than for externally developed innovations.
Figure 7.7	Administrative pressure is high.	Latitude for users will be high.
Table 8.9 & Figure 8.2	Teachers expand their repertoire.	Their attitudes to the innovation will be more positive.

The use of the conditional future tense in the "then" statements helps remind us that we have to *look* to see whether the "then" has happened.

If-then statements are a way to formalize *"propositions"* (Chapter 4, section D) for testing. The method of generating *predictions* (Chapter 6, section E) involves linking together a large number of "if's" to a single major "then."

Suggestions for practice. Review some of the illustrations above, walking them through the display to see what the basis is for concluding that the "then" statement is true.

In your own study, try the if-then tactic on some aspect of your data about which you (a) are increasingly puzzled or blocked or (b) feel on the brink of an Aha! experience. Formulating several different if-then statements (see **rival explanations,** tactic 12) helps.

A number of computer programs, including HyperRESEARCH, ATLAS/ti, QCA, and AQUAD permit you to test propositions of an if-then sort in your data (see Appendix).

10. Ruling Out Spurious Relations

Suppose you've been able, through assorted ingenious tactics, to establish that variable A is indeed related to B. Whenever A, we see B, and vice versa. Before breathing a sigh of relief and proceeding to the next conclusion, it pays to consider that the picture you are drawing:

$$A \rightarrow B$$

may in fact be more accurately portrayed as

$$\begin{tikzcd} & X \arrow[dl] \arrow[dr] & \\ A & & B \end{tikzcd}$$

where some third factor is in play, causing both A and B to occur.

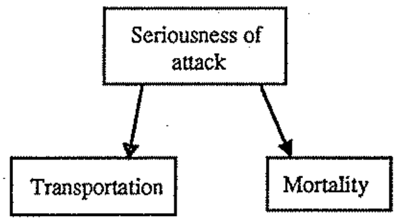
This is an old problem, which statisticians have dealt with well. We can draw a nice example from Wallis and Roberts (1956), describing a study from the *Journal of the American Medical Association.* Researchers noted that polio patients who traveled longer distances (average, 85 miles) to a hospital were more likely to die than patients who traveled little (average, 7 miles) and were more likely to die sooner (50% died within 24 hours, vs. 20%). They concluded: "The greater mortality in the transported group, occurring shortly after admission to the hospital, is a manifestation of the effect of long transportation during the acute stage of illness" (p. 285).

Wallis and Roberts suggest that a third variable may be influencing both A (transportation) and B (mortality). It is *seriousness of the initial attack.* All of the patients were seen in a certain hospital, Willard Parker, a noted center for treatment of contagious diseases. Polio patients who lived farther away were probably brought to Willard Parker only if their conditions were serious; milder cases would be treated nearer their own homes. Thus the picture develops as shown in Figure 10.6.

This interpretation can be checked out through the sort of display found in Figure 10.7, with Ns and mortality rates entered in the cells.

And, as Wallis and Roberts faithfully point out, even if the reanalysis could be done and it supported "seriousness of initial attack" as the real issue, you would have to do additional ruling out. Perhaps those coming from a distance had poorer basic health to begin with. Perhaps they came from an area where a particularly virulent strain of polio was widespread. And so on.

Figure 10.6
Possible Explanation of Spurious Relationship



Finding a candidate third variable is not always easy, especially if the original explanation "makes sense," as the transportation-mortality link seemed to at first glance. The Willard Parker researchers did think of one third variable—length of prior illness—which showed no real difference. Then they stopped, not realizing that the "Willard Parkerness" of Willard Parker was probably in the picture. Had they been able to recruit Wallis and Roberts as "friendly strangers" partway through their analysis, the story might have been different.

The moral for qualitative researchers is the same. When two variables look correlated, especially when you think they are causally associated, wait a beat and consider whether some third variable might be underlying/influencing/causing them both.[18] Use a knowledgeable but detached colleague to help in the search. Then consider new displays that will give you a clean look at such third variables and their effects.

Doing this procedure in a more than cursory way takes time, so it is worth it mainly when you have a major (but perhaps surprising) conclusion, or one on which a lot is riding, in practical terms.

Earlier we discussed the tactic of **finding intervening variables** (section 11 of part A), but there we were concerned with understanding a tepid relationship better. Here the issue is *undoing* a relationship you think looks plausible and strong.

Suggestions for practice. This is a special case of finding **rival explanations,** tactic 12. You might try your hand at the Wallis and Roberts (1956) example to see whether you can come up with any other way of ruling out the spurious relationship—and with a design for the needed display.

With your own data, the easiest way is to use one or more friendly skeptics to dismantle your favorite A-B relationship. Recall that in *case analysis meetings* (Chapter 4, section E) we found it very fruitful to invite alternative interpretations for emerging conclusions.

Figure 10.7
Display for Testing Explanation in Figure 10.6

	Local patients	Distant patients
Mild attack		
Severe attack		

11. Replicating a Finding

As we showed in section 3 on **triangulating,** findings are more dependable when they can be buttressed from several independent sources. Their validity is enhanced when they are confirmed by more than one "instrument" measuring the same thing.

Still the fact that usually one person is doing all of this measuring with homemade instruments is grounds for precaution. Once you've latched onto a hypothesis that makes powerful sense of the case, it's the dickens to get rid of it. Confirmation seems, almost magically, to come from all quarters. New interviews, observations, and documents all appear to bring verification, and to fit together coherently. Disconfirming evidence is absent or feeble. This is a heady and very dangerous time, and it usually means you are knee-deep in the "holistic fallacy": putting more logic, coherence, and meaning into events than the inherent sloppiness of social life warrants. How to protect against this?

One line of attack is to think in terms of replication, often cast as the bedrock of science. If I can reproduce the finding in a new context or in another part of my database, it is a dependable one. If someone else can reproduce it, better still.[19]

This tactic can be used in several ways. At the most elementary level, you are replicating as you collect new information from new informants, from new settings and events. New data bolster or qualify old data by testing their validity and generality.

At a notch higher in the confidence scale, you can test an emerging hypothesis in another part of the case or data set, as in the example we cited of bypassed principals and heir knowledge of new practices. Another example appears in Table 8.8 and Figure 8.1, where the analyst checked a conclusion about problem-coping behavior by looking at its consequences.

Such a test is more rigorous; it's harder to bootleg researcher bias into it. Even stiffer tests can be made by looking at multiple cases: finding a pattern in a cross-case display, and then tracking carefully through all of the cases to see whether the pattern is repeated.

Even stronger: replicate in a brand new case. For example, in another study, we came up with the hypothesis that a "localist" orientation at a college or university was more promising than a more "cosmopolitan" orientation for the college's work with surrounding school districts. To test this, we had "saved" one college in the sample for study later. We worked out what we meant by localism and cosmopolitanism (see section 9 of part A), described for the other cases just how the two variables were associated with successful collaborative projects, and then went to the new (most localist) site and tried to determine whether our hypothesis was viable. (It was, but needed some qualifying.) Others (see Stake & Easley, 1978; Yin, 1984) have used this kind of staggered replication device even more methodically.

Some brief words of advice:

1. In cross-case studies, replication is an important part of the basic data collection effort. Emerging patterns from one case need to be tested in others. They usually surface in some of the methods we described earlier: *pattern codes, memos, case analysis meetings, interim case summaries.* Be prepared, in the course of fieldwork, to do the kind of corroborative testing described here and in section 3 on **triangulation.**

2. If provisions are not made in advance for replications, they will not happen; there is inevitably too little time and too much information still to compile.

3. Doing replication at the very end of fieldwork, during final analysis and write-ups, is very difficult and less credible. To test a hypothesis in another part of the data set assumes that all of the requisite data are there for the test to be made. They usually are not, unless the researcher has made sure to collect them beforehand in anticipation of just such an exercise.

Suggestions for practice. Try your hand at using replication within Tables 8.6 or 8.7, or Boxes 7.4 or 8.6. Look for a plausible pattern within one case, then proceed to other parts of the display and see whether the pattern is repeated.

Reasonably early in your study, think forward a bit. Imagine that your findings will turn out to be truly amazing, with profound implications for your field. Then ask a friendly colleague to perform a rapid "replicability check." Based on a few hours' review of the project data base and methods, how easy or difficult would it be *for that colleague* to repeat your study?

12. Checking Out Rival Explanations

Thinking that rival explanations may account for the phenomena you have carefully studied and masterfully unraveled is a healthy exercise in self-discipline and hubris avoidance. But that thought often gets lost in the shuffle. During data collection you are often too busy making minimal sense of the welter of stimuli. Later on, you tend to get married to your emerging account and usually opt for investing the scant time left to buttress, rather than to unhorse, your explanation. Then, during data analysis, it is often too late to "test" any other explanation than the one arrived at; the data necessary for doing that in any but a cosmetic way just aren't there.

So, in qualitative research, there appears at first blush to be little of the kind of rival hypothesis-testing that Platt (1964) praises so much when he speaks of "strong inference." Platt, drawing on fields such as molecular biology, emphasizes (a) developing alternative hypotheses, (b) specifying critical experiments, the outcomes of which will exclude one or more hypotheses, (c) carrying out the experiments, and (d) recycling until a satisfactory conclusion is reached. But in most social settings, we cannot easily conduct the series of carefully controlled, elegantly scaffolded critical experiments that theoretically—but seldom practically—do away with equivocal findings from competing studies. Maybe we shouldn't worry too much about trying rival hypotheses on our data, because our rivals will cheerfully do it for us afterward. (Somehow, though, their conclusions don't match the ones we would have reached had we done the same exercise.)

Still, we think the search for rival explanations is often more thorough in qualitative research than in survey research or in most laboratory studies, and that it's relatively easy to do. The competent field researcher looks for the most plausible, empirically grounded explanation of local events from among the several competing for attention in the course of fieldwork. You are not looking for one account, forsaking all others, but for the best of several alternative accounts. This is rendered beautifully, we think, in Eco's novel *Le Nom de la Rose* (1982, pp. 311-312), in this exchange between the protagonist and his foil:

Guillaume I have a number of good hypotheses [for explaining a series of events], but no overriding fact to tell me which is the best . . .

Adso So then you must be far from finding the solution?

Guillaume I'm very close, but I don't know to which one.

Adso So you don't only have one solution?

Guillaume Adso, if that were the case, I'd be teaching theology at the Sorbonne.

Adso So in Paris they always have the right solution?

Guillaume Never, but they're very confident of their errors.

Adso What about you? Don't you make errors too?

Guillaume Often. But instead of imagining one, I try to imagine several. That way I don't become a slave of any one error in particular. [translated from the French]

We can see how checking rival explanations works in Figure 7.7. The first explanation, a hypothesis, was that high pressure to adopt an innovation will be accompanied by plenty of latitude to change it, which administrators grant in a bargain with users. A rival hypothesis turned up when the analyst cross-tabulated the data: Pressure and latitude were inversely related. The rival hypothesis needed some revision: Some bargaining was going on, but much later during the relationship; the inverse finding was essentially upheld. Note that, in this instance, the rival hypothesis was not dredged up from the depths of the analyst's mind or presented by a "stranger" colleague, but was forced on him by the data.

The trick is to hold on to several possible (rival) explanations until one of them gets increasingly more compelling as the result of more, stronger, and varied sources of evidence. Looking at it from the other end, you give each rival explanation a good chance. Is it maybe better than your main love? Do you have some biases you weren't aware of? Do you need to collect any new data?

Foreclosing too early on alternative explanations is a harbinger of bias, of what psychometricians call "systematic measurement error." You lock into a particular way of construing the case and selectively scan the environment for supporting evidence. Discounting evidence is ignored, underregistered, or handled as "exceptional"—and thus further increases your confidence in your thesis.[20]

We should also note that closing too *late* on alternative explanations builds too weak a case for the best one. It also adds great bulk to the corpus of data. So rival explanations should be looked at fairly promptly in fieldwork and sustained until they prove genuinely unviable—or prove to be better. This step should happen, if possible, before most of the fieldwork is done. The same principle applies to analysis done after fieldwork. Check out alternative explanations early, but don't iterate forever.

Suggestions for practice. It's usually difficult for anyone who has spent weeks or months coming up with one explanation to get involved seriously with another one. The idea may have to come from someone else who is not on the site and has the cognitive distance to imagine alternatives, to be a devil's advocate. You can ask a colleague (preferably from another discipline), "Here's how I see it and here's why I see it this way. Can you think of another way to look at it?" We also heartily commend the use of what Platt (1964) calls "The Question": "What could disprove your hypothesis?" or "What hypothesis does your analysis disprove?"

It also helps to fasten on to discrepant information—things that "do not fit" or are still puzzling (tactics: **looking for negative evidence, following up surprises**). The trick is not to explain them away in light of your favorite explanation—that's a piece of cake—but rather to run with them, to ask yourself what kind of alternative case these bits of information could build, and then check them out further.

A useful subtactic is: During final analysis, first check out the merits of the "next best" explanation you or others can think of as an alternative to the one you preferred at the end of fieldwork. "Next bests" have more pulling power than fanciful alternatives.

For more on rival explanations, including many fascinating problem exercises for practice, see Huck and Sandler's (1979) text.

13. Getting Feedback From Informants

One of the most logical sources of corroboration is the people you have talked with and watched. After all, an alert and observant actor in the setting is bound to know more than the researcher ever will about the realities under investigation (see Blumer, 1969). In that sense, local informants can act as judges, evaluating the major findings of a study (Denzin, 1978).

Feeding findings back to informants is a venerated, but not always executed, practice in qualitative research. It dates back at least to Malinowski's fieldwork in the 1920s and has been used in numerous field studies since then. More recently, Bronfenbrenner (1976) classified feedback to informants as a source of "phenomenological validity," and Guba (1981) built it into his repertoire of devices for assuring the "confirmability" of findings by using the sociologists' term "member checks." Other researchers (e.g., Stake, 1976) have made the case for feedback a quasi-ethical one—informants have the right to know what the researcher has found—and still others, more and more numerous, are feeding back findings because people at the site are making this a precondition for access.

Earlier (in sections D and E of Chapter 6) we showed two techniques for feedback to case informants. The first one requires the reader to comment on a short summary of findings, and then to evaluate the accuracy of a causal network with higher-inference findings. In the second, the researcher generates predictions that should play out if the findings are valid, then submits them to case informants for verification a year later.

Feedback may happen during data collection, too. When a finding begins to take shape, the researcher may check it out with new informants and/or with key informants, often called "confidants." The check-out process may be more indirect with the former than with the latter, who are often built into the study as sources of verification (e.g., Becker, Geer, Hughes, & Strauss, 1961; Lofland & Lofland, 1984; Whyte, 1943). The delicate issue here, of course, is that of introducing bias (see section 2 on **researcher effects**). Feeding things back in the course of a study may change informants' behaviors or perspectives.

Still we've seen some useful examples of thorough early feedback. In a phenomenological study, Melnick and Beaudry (1990) interviewed teachers and then, in a second interview, fed back the transcript, which had been annotated with marginal remarks on themes, and follow-up questions; the transcript became a base for shared reflection. Warner (1991) not only had children run camcorders of adult-child interaction and add their comments, but also played back the tapes for yet more comment by the children. The more emic the study, the more useful early feedback is likely to be—*if,* we would suggest, a clear, stable product is at hand, like these transcripts or tapes.

There are also good reasons for conducting feedback after final analysis. For one thing, you know more. You also know *better* what you know—are less tentative, have more supporting evidence, can illustrate it. In addition, you can get feedback at a higher level of inference: on main factors, on causal relationships, on interpretive conclusions. Finally, the feedback process can be done less haphazardly. You can lay out the findings clearly and systematically and present them to the reader for careful scrutiny and comment.

As we showed in sections D and E of Chapter 6, it is crucial that the reader be able to connect to the feedback—understand it, relate it to local experience and perceptions, *do* something with it (draw on it, cross out parts and add others, etc.). So *formatting* the feedback is crucial. Sending back an abstract, an executive summary, or the concluding chapter, without transforming it into the language of the site—which the researcher has come to learn in the course of field research—is of little value if you are seriously after verification.

Some advice:

1. If you don't plan deliberately for this exercise—setting aside the time, transforming write-ups into site-comprehensible language and formats, leaving time to incorporate the results of the exercise into your final write-up—it probably won't happen. Once again, there will be too many competing claims on your time.

2. Think carefully about displays. As with analysis, matrices and networks work much better than text alone to help local informants access the information. They will find it easier to get an overview, to see how the pieces fit together.

3. Providing information at more macroanalytical levels of inference (e.g., main factors and relationships, plus causal determinants) has to be done very carefully by working up from particulars. If this is not done, informants may discount the whole exercise because the overarching findings look abstract or incomprehensible. Or they may swallow these macro findings whole because these read so "scientifically." As the symbolic interactionists have shown convincingly (see Blumer, 1962), people do not act toward social structures or institutions or roles; they act toward *situations,* and are only likely to understand metasituational language if it is directly connected to these situations.

Still, the people we study do make maps of their reality. Those maps may not coincide with the researcher's, but if we can cast our pictures and feed them back in an informant-friendly way, they can, at the minimum, be acknowledged. Beyond this is the possibility of mutual enlightenment.

4. Counterpoint: Think very carefully before feeding back any specific incident. Will anyone's self-esteem, job chances, or standing in the organization be damaged by the report? (One of us once fed back first-draft case reports to people in five cases and was threatened with lawsuits in four of the five because of specific incidents in the reports—even though the incidents were reported accurately.) We have more to say on the ethical aspects of feedback in Chapter 11.

5. Don't expect that informants will always agree with you or with one another. If they always did, life at the site would be more conflict free than you probably found it to be. People often have widely varying perceptions of the same phenomenon. Interpretivists consider this natural, and as Denzin (1989b) notes, consider "all interpretations unfinished, provisional, and incomplete."[21]

Critical theorists emphasize the "emancipatory" aspects of distortion-free communication between researcher and researched—but supply little help on how this might actually work out. Prior work on feeding back *quantitative* information (Miles, Calder, Hornstein, Callahan, & Schiavo, 1966) suggests the importance of connecting it with sustained problem-solving efforts, a feature usually missing in the aftermath of qualitative studies.

In practice, feedback efforts, even collaboratively-run ones, often run into difficulty (King, Louth, & Wasley, 1993) and don't encourage further reflection. Let's look at some possible reasons why informants might reject the conclusions or interpretations of a field researcher. Guba and Lincoln (1981, pp. 110-111) reviewed them succinctly and well. To recapitulate: The informant is not *familiar* with the information; the informant does not *understand* it (jargon, lack of sophistication or awareness); the informant thinks the report is *biased;* the information *conflicts* with the informant's basic values, beliefs, or self-image; the information threatens the informant's *self-interest;* or this is not the way the informant *construes* or puts together the same information.

If you think these reactions are unique to "informants," look at the reactions a group of graduate student *researchers* had when Kurzman (1991) fed back his conclusions about their work; he concludes skeptically that written feedback is unwise, doesn't correct mistakes, doesn't help cooperation or empowerment, and may even cause self-censorship by the researcher.

General implication: It won't be easy. It helps to remember that data feedback is an occasion to learn more about the case, not only about your feedback.

Suggestions for practice. Early in your study, consider alternative approaches to verification through member feedback, ranging from none at all to early/intermittent to late and well-structured (case critique, verification or prediction forms, etc.); consider costs and benefits of each. For promising candidates, use several colleagues familiar with the *setting* you are studying, but not your research, to act as surrogates for real informants in tryouts of your feedback plans.

C. Standards for the Quality of Conclusions

We've reviewed 26 tactics for drawing and for verifying conclusions. How will you, or anyone else, know whether the finally emerging findings are *good?* That term has many possible definitions: possibly or probably true, reliable, valid, dependable, reasonable, confirmable, credible, useful, compelling, significant, empowering (add others of your choice). It's not enough to say that well-carried-out tactics will make for good conclusions. As Firestone (1990) notes, "The major justification for the research enterprise is that we have the time and the skills to develop approximations of the truth that have a firmer warrant than common sense" (p. 123).

As we were writing this book, we collected survey data from a wide range of qualitative researchers, asking about key analysis issues they were concerned about. The basic question of the goodness of conclusions can be approached from a "justificatory" point of view, as one of our respondents noted:

> How to really convince reductionists/positivists that naturalistic inquiry/qualitative research methods are *not* any more biased or inaccurate or imprecise than their methods? (From a specialist in recreation and leisure studies)

In this frame the issue is legitimacy. But as many researchers have pointed out, the problem of quality in qualitative studies deserves attention in its *own* terms, not just as a justification device. Other respondents said:

> It is probably an issue for all of us: when your own judgment has to be relied on for strength, significance of a finding. This is what makes the process seem subjective. (Public health researcher)

> Creation of lenses to understand my data. It is difficult to avoid seeing what you expect to see without using careful procedures. (Industrial engineer)

> If criteria for "goodness" are a false hope, how do we share/teach/defend context-bound, hermeneutically analyzed, interpretive inquiry? (Philosopher)

In this section we explore some practical standards that could help us all judge the quality of conclusions.

The battles in this domain have been extensive, and they continue. Many interpretivist researchers take the position that there is no "fact of the matter" (e.g., Schwandt, 1990) and suggest by extension that it is not really possible to specify criteria for good qualitative work—and that the effort to do so is somehow expert-centered and exclusionary, not responsive to the contingent, contextual, personally interpretive nature of any qualitative study.[22]

But the problem of quality, of trustworthiness, of authenticity of findings will not go away. The fact is that some accounts are better than others. Although we may acknowledge that "getting it all right" is an unworkable aim, we should, as Wolcott (1990a) suggests, try to "not get it all wrong." Would you be bothered if a journalist did not tell the truth about you, or if a court dismissed a case against someone who had assaulted you, saying that everyone has different interpretations?

Our view is that qualitative studies take place in a real social world, and can have real consequences in people's lives; that there is a reasonable view of "what happened" in any particular situation (including what was believed, interpreted, etc.); and that we who render accounts of it can do so well or poorly, and should not consider our work unjudgable. In other words, shared standards are worth striving for (Howe & Eisenhart, 1990; Williams, 1986).

We cannot enter here into a discussion of how "goodness criteria" flow from epistemological positions (for thoughtful discussions, see J. K. Smith, 1990, and Hammersley, 1992). Rather, we remain broadly in the "critical realist" tradition and discuss five main, somewhat overlapping, issues: the objectivity/confirmability of qualitative work; reliability/dependability/auditability; internal validity/credibility/authenticity; external validity/transferability/fittingness; and utilization/application/action orientation. (Here we are pairing traditional terms with those proposed as more viable alternatives for assessing the "trustworthiness" and "authenticity" of naturalistic research [Guba & Lincoln, 1981; Lincoln, 1990; Lincoln & Guba, 1985]).

In each section we describe the issues generally, without trying to straighten out all of the thorny problems involved. (Better people than us are still trying.) Then we propose some practical guidelines in the form of questions that can be applied to qualitative work—your own or that of others.

As in Marshall's (1990) practical list, these are not "rules" to be stiffly applied, but queries we think can be fruitfully posed when you are reflecting on the question, "How good is this piece of work?"—whether it is yours or someone else's. We cite useful sources as we go.

Objectivity/Confirmability

The basic issue here can be framed as one of relative neutrality and reasonable freedom from unacknowledged researcher biases—at the minimum, explicitness about the inevitable biases that exist. In short, do the conclusions depend on "the subjects and conditions of the inquiry," rather than on the inquirer (Guba & Lincoln, 1981)? This domain is sometimes labeled "external reliability," with emphasis on the replicability of a study by others (LeCompte & Goetz, 1982).

Relevant queries. Some useful questions to be asked of a qualitative study about this issue are:

1. Are the study's general methods and procedures described explicitly and in detail: Do we feel that we have a complete picture, including "backstage" information?
2. Can we follow the actual sequence of how data were collected, processed, condensed/transformed, and displayed for specific conclusion drawing?
3. Are the conclusions explicitly linked with exhibits of condensed/displayed data?
4. Is there a record of the study's methods and procedures, detailed enough to be followed as an "audit trail" (Schwandt & Halpern, 1988)?
5. Has the researcher been explicit and as self-aware as possible about personal assumptions, values and biases, affective states—and how they may have come into play during the study?
6. Were competing hypotheses or rival conclusions really considered? At what point in the study? Do other rival conclusions seem plausible?
7. Are study data retained and available for reanalysis by others?

Reliability/Dependability/Auditability

The underlying issue here is whether the process of the study is consistent, reasonably stable over time and across researchers and methods. We can, in effect, speak of "quality control" (Goetz & LeCompte, 1984; A. G. Smith & Robbins, 1984). Have things been done with reasonable care?

Kirk and Miller (1986) helpfully distinguish "diachronic" reliability (stability of observations over time) from "synchronic" reliability (stability in the same time frame) and point to the danger of "quixotic" reliability (what happens when multiple respondents give a monolithic, party-line answer: Remember that a broken thermometer is 100% reliable—but not very valid).

Relevant queries. What can be usefully asked in this domain?

1. Are the research questions clear, and are the features of the study design congruent with them?
2. Is the researcher's role and status within the site explicitly described?
3. Do findings show meaningful parallelism across data sources (informants, contexts, times)?
4. Are basic paradigms and analytic constructs clearly specified? (Reliability depends, in part, on its connectedness to theory.)
5. Were data collected across the full range of appropriate settings, times, respondents, and so on suggested by the research questions?
6. If multiple field-workers are involved, do they have comparable data collection protocols?
7. Were coding checks made, and did they show adequate agreement?
8. Were data quality checks made (e.g., for bias, deceit, informant knowledgeability?)
9. Do multiple observers' accounts converge, in instances, settings, or times when they might be expected to?
10. Were any forms of peer or colleague review in place?

Internal Validity/Credibility/Authenticity

Here we arrive at the crunch question: truth value. Do the findings of the study make sense? Are they credible to the people we study and to our readers? Do we have an authentic portrait of what we were looking at?

Validity itself is not monolithic; the classic, measurement-oriented view (Brewer & Hunter, 1989) differentiates face, content, convergent, discriminant, and predictive validity. More helpfully for our purposes, Maxwell's (1992b) thoughtful review distinguishes among the types of *understanding* that may emerge from a qualitative study: descriptive (what happened in specific situations); interpretive (what it meant to the people involved); theoretical (concepts, and their relationships, used to explain actions and meanings); and evaluative (judgments of the worth or value of actions and meanings). Warner (1991) also speaks of "natural" validity—the idea that the events and settings studied are uncontrived, unmodified by the researcher's presence and actions.

Furthermore, work on the "validity" of narratives emerging from interpretive studies (Connelly & Clandinin, 1990) suggests the importance of such aspects as "apparency" and "verisimilitude" (Van Maanen, 1988), "authenticity," "plausibility," and "adequacy," among others.

We should also note Kvale's (1989b) emphasis on validity as a process of checking, questioning, and theorizing, not as a strategy for establishing rule-based correspondence between our findings and the "real world." "Validation becomes the issue of choosing among competing and falsifiable explanations."

Relevant queries. Some useful possibilities are:

1. How context-rich and meaningful ("thick") are the descriptions (Denzin, 1989b; Geertz, 1973)?
2. Does the account "ring true," make sense, seem convincing or plausible, enable a "vicarious presence" for the reader?
3. Is the account rendered a comprehensive one, respecting the configuration and temporal arrangement of elements in the local context (Campbell, 1986)?
4. Did triangulation among complementary methods and data sources produce generally converging conclusions? If not, is there a coherent explanation for this (Mathison, 1988)?
5. Are the presented data well linked to the categories of prior or emerging theory? Do the measures reflect the constructs in play?
6. Are the findings internally coherent (Eisner, 1991); are concepts systematically related (Strauss & Corbin, 1990)?
7. Were rules used for confirmation of propositions, hypotheses, and so on made explicit (Miller, n. d.)?
8. Are areas of uncertainty identified? (There should be some.)
9. Was negative evidence sought for? Found? What happened then?
10. Have rival explanations been actively considered? What happened to them?
11. Have findings been replicated in other parts of the database than the one they arose from?
12. Were the conclusions considered to be accurate by original informants? If not, is there a coherent explanation for this?
13. Were any predictions made in the study, and how accurate were they?

External Validity/Transferability/Fittingness

We need to know whether the conclusions of a study have any larger import. Are they transferable to other contexts? Do they "fit" (Lincoln & Guba, 1985)? How far can they be "generalized"?

Firestone's (1993) review suggests three levels of generalization: from sample to population (less helpful for qualitative studies), analytic (theory-connected), and case-to-case transfer (see also Kennedy, 1979).

Maxwell (1992c) also speaks of "theoretical" validity, the presence of a more abstract explanation of described actions and interpreted meanings. Such explanations could be considered as "internal" validity, but they gain added power if they are connected to theoretical networks beyond the immediate study. Generalizability, he suggests, requires such connection-making, either to unstudied parts of the original case or to other cases.

Schofield (1990) usefully distinguishes generalizing to "what is" (other actual contexts), to "what may be" (sites in the forefront of some similar process) and to "what could be" (outstanding or ideal cases).

The generalizing process is far from mechanical, as Noblit and Hare (1988) note in their discussion of "meta-ethnography": It is more like translating, refuting, or synthesizing two or more studies of similar phenomena. It is careful interpretation, not just "adding up."

Relevant queries. Here we may usefully ask:

1. Are the characteristics of the original sample of persons, settings, processes (etc.) fully described enough to permit adequate comparisons with other samples?
2. Does the report examine possible threats to generalizability? Have limiting effects of sample selection, the setting, history and constructs used been discussed (LeCompte & Preissle, 1993)?
3. Is the sampling theoretically diverse enough to encourage broader applicability?
4. Does the researcher define the scope and the boundaries of reasonable generalization from the study (McGrath & Brinberg, 1983)?
5. Do the findings include enough "thick description" for readers to assess the potential transferability, appropriateness for their own settings?
6. Does a range of readers report the findings to be consistent with their own experience?
7. Are the findings congruent with, connected to, or confirmatory of prior theory?
8. Are the processes and outcomes described in conclusions generic enough to be applicable in other settings, even ones of a different nature (Becker, 1990; Bogdan & Biklen, 1992)?
9. Is the transferable *theory* from the study made explicit (Maxwell, 1992b)?
10. Have narrative sequences (plots, histories, stories) been preserved unobscured? Has a general cross-case theory using the sequences been developed (Abbott, 1992b)?
11. Does the report suggest settings where the findings could fruitfully be tested further?
12. Have the findings been replicated in other studies to assess their robustness? If not, could replication efforts be mounted easily?

Utilization/Application/Action Orientation

Even if a study's findings are "valid" and transferable, we still need to know what the study does for its participants, both researchers and researched—and for its consumers. We simply cannot avoid the question of "pragmatic validity" (Kvale, 1989a); it's an essential addition to more traditional views of "goodness."

Evaluation and policy studies in particular are supposed to lead to more intelligent action; whether or not they do, real people's lives are being affected, and large amounts of money are being spent (or misspent). As Patton (1990) notes, "The ultimate test of the credibility of an evaluation report is the response of decision-makers and information users to that report" (p. 469). Managers and consultants often rely on qualitative studies (Gummeson, 1991). Action research (Watkins, 1991) is designed to solve particular local problems through a cycle of reconnaissance, planning, action, and re-reconnaissance. Critical theorists (for good overviews see Carspecken & Apple, 1992, and Popkewitz, 1990) stress the importance of "emancipatory" research that makes people aware of inequitable or oppressed positions and empowers their corrective action.

Lincoln (1990) speaks more broadly of good qualitative research as enhancing "(a) levels of understanding and sophistication and (b) the ability of participants and stakeholders to take action during and after an inquiry and to negotiate on behalf of themselves and their own interests in the political arena."

There are also questions of ethics—Who benefits from a qualitative study, and who may be harmed? (see also Chapter 11)—and of "evaluative validity" (Maxwell, 1992b): judgments made about the worth, legitimacy, or goodness of actions or meanings.

Relevant queries. What are some fruitful probes to be made here? Our range is deliberately wide.

1. Are the findings intellectually and physically accessible to potential users? (Note Gummeson's [1991] tongue-in-cheek dictum that "a scientific report should be boring and difficult to read.")
2. Do the findings stimulate "working hypotheses" on the part of the reader as guidance for future action (Donmoyer, 1990)?
3. What is the level of usable knowledge offered? It may range from consciousness-raising and the development of insight or self-understanding to broader considerations: a theory to guide action, or policy advice. Or it may be local and specific: corrective recommendations, specific action images.
4. Do the findings have a catalyzing effect leading to specific actions (Lincoln, 1990)?
5. Do the actions taken actually help solve the local problem?
6. Have users of the findings experienced any sense of empowerment, of increased control over their lives (Lincoln & Guba, 1990)?
7. Have users of findings learned, or developed new capacities?
8. Are value-based or ethical concerns raised explicitly in the report? If not, do some exist implicitly that the researcher is not attending to?

These questions are pointers, not rules. They will get clearer as the craft of qualitative analysis advances. The way to get them clearer is to ask—and try to answer—them more frequently than we do now.

At a number of points during this discussion of goodness criteria, we alluded to the importance of careful documentation and its value as an "audit trail" that peers or other colleagues can follow. We turn to this next in more detail.

D. Documentation

The Problem

Good qualitative research, like any other research, requires careful record keeping as a way of connecting with important audiences. The first audience is the *self:* The notebooks of the molecular biologist, the industrial sociologist, or the clinical drug tester help each keep track of what was done along the way, suggest ways of improving next steps, and give reassurance about the reproducibility of the results.

The second audience is *readers* of the research reports, who need to know what was done and how, as a way of assessing the credibility of the findings. Other researchers make such judgments carefully, even obsessively. And other readers such as local informants, policymakers, managers, and the general public often raise such questions as, "Whom did you talk to, anyway?" "How do you know?" "Why are you being so negative?" and "Where did you get that recommendation?"

The third audience is a subset of the second: *other researchers* who are interested in secondary analysis of the data, want to carry out a meta-analysis or synthesis of several studies, or want to replicate the findings to strengthen or refute them.[23]

For the latter two audiences, most journals require authors of empirical studies to report on their procedures as an integral part of the article. The formats are often so familiar that the author can almost fill in the blanks when writing sections on sampling, methods, and data analysis. For quantitative studies, it's also typical to report such basic data details as means, standard deviations, derived scores and scales, marginals, zero-order correlations, and various validity and reliability coefficients.

There is, in other words, an agreed-upon set of conventions for documenting empirical research (cf. Wolfe, 1992) and for reporting it, and a corresponding set of methods for verifying the report. But these tools seem confined to statistical studies. Qualitative researchers don't have very clear alternatives to fall back on. Lofland (1974, p. 101) says it well: "Qualitative field research seems distinct in the degree to which its practitioners lack a public, shared and codified conception of how what they do is done and how what they report should be formulated."

On the face of it, this is a curious state of affairs. Although qualitative studies are rich in descriptions of settings, people, events, and processes, they often say little about how the researcher got the information, and almost nothing about how conclusions were drawn.

When procedures are left so opaque, we have only vague criteria for judging the goodness of conclusions: the "plausibility," the "coherence," or the "compellingness" of the study—all evocative but ultimately hollow terms. The researcher can always provide a plausible final account and, with careful editing, may ensure its coherence. If the writing is good, we will be won over by the undeniability and vividness of the report. But, as we saw earlier, plausible and coherent accounts can be terribly biased, and vividness lands us in the "availability" heuristic, where we overweight concrete or dramatic data.

So we have an unappealing double bind. Qualitative studies cannot be verified because researchers don't report clearly on their methodology, and they don't report clearly on their methodology because there are no shared conventions for doing so. (Epistemological reasons may be claimed as well—for instance, the idea that a qualitative study is such a person-specific, artistic, private/interpretive act that no one else can viably verify or replicate it—but that takes us far from shore just now.) What to do?

Since the first edition of this book, we have seen clear signs of movement. "Methods" sections are much more typical in journal reports of qualitative studies (though they are often discursive and not specific enough). Some texts review generally what should be in a methods section. For example, Lofland and Lofland (1984) suggest explaining the study's inception, relations with informants, private feelings, data gathering, data focusing and analysis, and retrospective learnings. Miller (n. d.) suggests covering the research questions/hypotheses, methods used, data types, and exactly how the questions/hypotheses were confirmed. Pitman and Maxwell (1992) suggest including a description of the database and how it was produced, along with methods of category development, the coding system, and data displays.

We've seen a number of thoughtful methodological reports that bring the reader close to what was done (e.g., Chesler [1987] on grounded theory work; Melnick & Beaudry [1990] and Fischer & Wertz [1975] on specific phenomenological studies; Merryfield [1990] on the auditing of extended fieldwork as a basis for case studies; Pearsol [1985] on the chronology of data analysis in a hermeneutic study; and Eisenhardt's [1989b] report of analysis operations in a multicase study). Many qualitative evaluation studies, especially multicase ones, document their analytic procedures carefully (e.g., Kell, 1990).

We've also begun to see more public critique and discussion of specific qualitative studies (e.g., Miller's [n. d.] review of how five recent qualitative studies proceeded to "confirmed" findings; Noblit's [1988] analysis of the published Cusick-Everhart debate on their studies of schooling cited in Chapter 6; Miles & Huberman [1990] on studies by Barone [1990] and Clark [1990]; and Eisenhart & Howe's [1992] review of Holland & Eisenhart's [1990] study of college women).

And we are beginning to see secondary analyses of qualitative studies: Reinharz (1993) cites three, presents another in detail, and proposes extending "qualitative data banks" like the Murray Center at Radcliffe.

But there's still plenty of ambiguity about methodological documentation and reporting. We need better descriptive information than we now are getting—though we should not get too far into methodological canons until we know what qualitative researchers are *doing* when they do qualitative research. As Mills (1959) noted many years ago: "Only by conversations in which experienced thinkers exchange information about their actual ways of working can a useful sense of method and theory be imparted" (p. 195).

We need to know many things in these conversations. Which kinds of qualitative designs do researchers actually use? How are sampling decisions actually made? (We think of the wonderful comment by Star & Strauss [1985] on this topic: "Tell the truth, no matter how embarrassing.") What does the instrumentation look like, and how does the researcher know whether it is measuring accurately what it was meant to measure? How did fieldwork actually proceed (we think of Lofland & Lofland's [1984] comment that most reports only mention "the second worst things that happened")?

Most fundamentally, how do researchers get from hundreds of pages of field notes to a final report? How are the data aggregated, reduced, partitioned, displayed, analyzed, interpreted? Until we can share clear descriptions of qualitative research procedures, we cannot talk about them intelligibly with one another—let alone set up conventions for verification.[24]

We have, to begin, then, by logging and then describing our procedures clearly enough so that others can understand them, reconstruct them, and subject them to scrutiny.

Guba (1981) and Guba and Lincoln (1981) provided a good orienting map for doing this with care. They used the metaphor of the fiscal auditor who examines the books of a company to determine whether accounts have been kept satisfactorily and whether the "bottom line" is correct. Making such an audit depends on the auditee's having documented income, outlays, and transfers. Without such an "audit trail," you cannot determine the dependability or the confirmability of the bookkeeping.

The "audit" idea was first worked out operationally by Halpern (1983), who also shows a case study, and was detailed further by Schwandt and Halpern (1988) with specific and helpful suggestions.

We note again that the first and basic audience for good documentation is the *self.* Even if no "audit" is ever intended, such devices as the researcher's notebook, methodological diary (Walker, 1981), "reflexivity journal" (Carney, 1990) recording changes in design and associated thoughts, "data analysis chronologies" (McPhee, 1990; Pearsol, 1985), or running diary (L. M. Smith, 1992a) strengthen the study *as it goes.* Secondarily, they enable easy production of the methods section. Thorough auditing, as such, is relatively rare, as far as we know. The audit metaphor is forbidding: It connotes an external, stern, obsessive expert and misses the idea that you, with close colleagues, can frequently look at documentation very fruitfully.

Illustration

We describe a "maxi-approach" to documentation of analysis operations, and also suggest more modest variations. Any documentation approach must be keyed to purposes. In our case, during the school improvement study, we were also writing the first edition of this book and needed to know just how the methods—many of them new to us or just invented—really functioned and how dependable they were. Otherwise we could hardly pass them on to others.

So over 3 years, we developed seven successive versions of a documentation form focusing on analytic operations. (Note: It does not deal with issues of sampling, instrumentation, fieldwork, etc.) The challenge was to come up with a documentation form that met several criteria: facility and rapidity of use, easy transfer into a methodological report, easy access and comprehension by a second reader, and believability/validity.

There are some clear trade-offs—for example, comprehension and believability usually mean the form cannot be tossed off rapidly, but requires some time and care.

We show here (Table 10.2) version VII of our form, a streamlined version of some earlier and (would you believe it?) more elaborate efforts. We include it here for essentially heuristic, illustrative reasons. Of course we encourage others to develop their own documentation methods. The nature of good documentation of analysis operations is something we have to discover inductively.

As we used it, the form is focused on a single research question or issue (item 1). Item 2 asks the researcher to explain, generally, what the analysis was designed to do, and to situate it in the context of other analyses. Item 3 calls for a fairly complete description (actually written as a log or diary during analysis), including the *data sets* in which the analysis was conducted, the *procedural steps,* the *decision rules* used to manage data, the *analysis operations* involved, the preliminary *conclusions* to which the analysis led, and any concluding *comments.* All of this information goes on a single sheet, so the analyst can log in the successive analytical steps and a reviewer (if any) can grasp quickly what was done.

The researcher indicates all relevant *exhibits* (materials used or developed in the course of analysis), so they can be referred to easily. For a reviewer, they should be appended.

Let's walk through the form, explaining and commenting on its use. (In practice, the form as printed here was on 11 × 17 paper.) Items 1 and 2 are orienting ones for the immediate analysis enterprise (normally, one analysis task per form). Of course a string of procedures will be involved, but if the general task changes (e.g., from summarizing the general nature of the data to defining a precise set of predictors), a new form should be used.

The *data sets* should be described clearly enough so that you, a colleague, or an auditor can find them in your files. Some may appear as exhibits (see below).

Make running notes for the section on *procedural steps* as you go; much gets lost in recollection. Then it's easy to fill in the section with a few summary phrases. Here are some examples from our work:

- Reviewed preliminary matrix format, entered data from 2 sites.
- Tried a cross-tabulation of program type by ease of implementation; abandoned after several entries.
- Reviewed meta-matrix, decided to lump Lido case with Astoria and Carson.
- Went through causal network, reviewing links, adding and subtracting links to make final version.
- Scanned Table 8.4 and Box 8.1, comparing high- and low-assistance cases.
- Wrote text based on preceding step.

(The language here is often technical because the audience was other colleagues on the project closely familiar with the methods. A more "strange" reviewer typically needs a somewhat fuller description.)

Noting the exact *decision rules* used is important, especially for any operations involved in "readying" data for entry into displays. Making them explicit as you go is very

Table 10.2

Qualitative Analysis Documentation Form

QUALITATIVE ANALYSIS DOCUMENTATION FORM

M.B. Miles, A.M. Huberman
NIE G-81-0018

Version VII
3/83

1. Research Issue being explored: ______________________ Analyst ______________ Date ____________ FORM NO. ______
2. In this analysis task, what, specifically, were you aiming to do? (Give context and a short rationale; say whether focus is exploratory or confirmatory; make the connection with earlier analyses.)
3. Description of procedures. Work sequentially, keeping a log or diary of steps as you go through the analysis. Use a second sheet if needed. If the analysis task changes substantially, use a new form, redoing items 1 and 2 above.

SPECIFIC DATA SETS IN USE*	PROCEDURAL STEPS (number each one, explain what was done, and exactly how it was done)@	DECISION RULES# followed during analysis operations	ANALYSIS OPERATIONS (enter codes) Readying data for analysis	Drawing conclusions	Confirming conclusions	CONCLUSIONS DRAWN from these specific analysis operations; give substance in brief.	RESEARCH COMMENTS, reflections, remarks on any of the preceding.

* Indicate whether single or multi-site. May include: field notes; write-ups; summaries; documents; figures; matrices; tables; tape recordings; photos; film/videotapes; others (specify).

Explicit list of actual rules used for "readying" data (clustering, sorting, scaling, etc.); may also apply to drawing/confirming conclusions

@ PROVIDE ALL RELEVANT EXHIBITS; give each a letter, describe briefly, and also give numbers of procedural steps where used.
Work notes/work sheets: ______________
Interim data displays: ______________
Final data displays: ______________
Final text written (excerpts): ______________

useful; it reduces error and can aid self-correction. Here are some examples:

- Theme was coded as present for an informant if mentioned repeatedly or with strong affect during the interview.
- Item entered in display if mentioned by more than one informant at sites with more than three informants; at sites with fewer users, entered if mentioned by one informant with no contradiction from others.
- "Smoothness of use" was a field-worker judgment drawn from the interview responses and/or from observation of the practice in use; more weight given to observation data.
- If judgments of the value of assistance sources varied in strength, the final rating gave more weight to the source closer to the user (e.g., assistance from principal, materials, and peers had more weight than assistance from external consultants).
- Rating of +/– given when implementation smooth for some users, rough for others. Otherwise ratings followed modal response.
- "Immediate causes" of the outcome were defined as two steps or less back along the causal network. "Distal causes" were on the causal stream leading to the outcome, but farther back. "Other" variables were off the stream or not on the network.
- A "strong" predictor had to be present in the majority of high-assistance sites, and absent in at least 40% of low-assistance sites.
- Cutting point between "high" and "moderate" cases made when half or less of the indicators were absent.
- When contradictory data on adoption date were found, gave more weight to those informants on-site from the beginning, and who were actively involved.

Making decision rules explicit is critical for data-readying operations but also is important for conclusion drawing and confirmation as well. A reader will usually want to know how and why you concluded that variable A was a stronger predictor of the outcome than variable B, *and* how you verified or confirmed that conclusion. Your rules may be (alas) arbitrary—but they are your rules.

For tracking *analysis operations,* we developed a code list (Figure 10.8). Scanning the list serves as a kind of prompt for the analyst, both helping to label the operations followed and suggesting other avenues that might be followed. Note that although we have clustered the codes for convenience into three general categories, many of them may fall into any of the three categories. For example, COH (conceptual and theoretical coherence), though it appears under "drawing conclusions," can also be used to confirm or test conclusions. Or SUB (subsuming data under a higher level variable) can be seen as readying data for analysis, or as a way of drawing conclusions.

Note that many of the items, along with the logic underlying their use, are tactics defined more fully in sections A and B above. Because there is no common language about analysis operations, and there is a variety of partial dialects from different disciplines, we tried to find items that would have meaning to most researchers. In any case, we need a common, agreed-upon language for describing qualitative research procedures badly enough that we should begin with an imperfect set of terms and clean them up gradually.

We've aimed to keep terms clearly defined. We include no items for noting "insights," "bracketing," "structural corroboration," and "disciplined subjectivity" because it isn't clear to us, operationally, what these terms mean—how they actually work. Other researchers may have more success on this.

The next to last column of the form asks for the researcher's substantive *conclusions* in capsule form; they do not need to be detailed, and reference can be made to the analytic text in which they appear.

The last column is open for the researcher to comment on confidence in the conclusions, reflect, make remarks, or vent any relevant feelings. This material helps clarify the meaning of the analysis episodes being reported.

We repeatedly found that indicating or appending all available *exhibits* was crucial for other readers. It is simply impossible to understand the analysis procedures followed without direct recourse to such exhibits. The final analytic text is equally important. It's the end of the audit trail.

Variations

Less complex documentation methods are not hard to find. The most straightforward are a running log of study decisions and activities, accompanied by a journal of reflections (Merryfield, 1990) and commentary.

At a more abstract level, it is often helpful to cast the description of analytic activities as a series of transformational steps. For example, Fischer and Wertz (1975), studying the experience of crime victims, started with case synopses, then went to cross-case "illustrated narratives," then to a general cross-case "condensation," and then cycled back to "exemplary case synopses" and out to a general cross-case explanatory structure (see Chapter 4, section I). Each step was well defined and documented. See also the description of Chesler's (1987) "in vivo" coding work that appears in Chapter 4.

Pearsol (1985) developed a "data analysis chronology" with more specific steps than Fischer and Wertz's. Samples are "enlivening the transcripts" while listening to tapes, "initial listing of topic discussed," "classifying teachers' conclusions about the project," and "identifying frames of reference" shared by groups of teachers, such as "traditional family values."

Figure 10.8
Code List for Analysis Operations

Readying data for analysis

TAB	tabulating coded segments
MAT	filling in matrices
CLAS	classifying, categorizing
RANK	ranking/weighting data
SUMM	summarizing phrases, generating key words
SUB	subsuming data under higher-level variable
SCAL	scaling, summing indices
COMP	computing, tabulating
SPLT	splitting one variable into two
PAR	partitioning
AGG	aggregating

Drawing conclusions

PLAUS	seeing "plausibility" only
GES	seeing a Gestalt, pattern, theme
MET	generating metaphors
CLUS	clustering
COUNT	counting/frequencies
CEN	establishing central tendencies
CONT	systematic contrasts/comparisons
FAC	establishing factors
REL	establishing relationships between variables/sets of variables
LOG	logical chain of evidence
INTV	establishing intervening/linking conditions
COH	conceptual/theoretical coherence
CAUSE	determining directional influence
TEMP	determining temporal order/temporal relationships
INF	making inferences
INF-COMP	by computations
INF-DED	by deduction
INF-IND	by induction (e.g., determining antecedents, covariates, consequences)

Confirming conclusions

REPR	checking for representativeness
RES-EFF	checking for researcher effects
BIAS-CONTR	control for bias (specify)
TRI	triangulation
TRI-DATA	from different data sources
TRI-METH	from different methods
TRI-CONC	conceptually (different theories)
TRI-RES	from different researchers
WT	weighting of evidence
OUT	use of outliers, exceptions
EXTR-SIT	extreme situation verification
EXTR-BIAS	extreme bias verification
SURP	following up surprises
EMP	empirical evidence from elsewhere
NONEG	absence of negative evidence
IF-THEN	testing if-then relationships
FALSE-REL	checking false relation due to third variable
REPL	replication
RIV	test of rival explanation
FB	corroboration from informant feedback

Eisenhardt's (1989b) documentation procedures are not described explicitly, but her "methods" section is exemplary, including general design, details of data sources, and data analysis moves such as development of case profiles, "decision stories," pairwise comparison of cases, and proposition development and iteration.

Advice

1. Be clear in your own mind *why* you are doing documentation for this study. Is it for study steering and revision, for your own personal learning, for getting feedback from a colleague or other critical friend, for "methods section" reportage, for a methodological article as such, for actual audit?

2. Any study has more or less riding on it. A high-stakes study (e.g., of an expensive but controversial program) demands more care. Remember, too, that audits always have a political dimension (Schwandt, 1989): Whose interests are being served and for what purposes?

3. Documentation detailedness also depends on your study's focus. With our interest in close depiction of analytical moves, we found that any given research question usually involved a flow of seven or eight analysis episodes, each using one of these documentation sheets. Less-micro purposes, or fewer of them, would mean less detail.

4. Work with a form of the sort we have shown usually goes faster with a stepwise procedure. First, while actually conducting the analysis, you make rough running notes; then you order them, logging in the procedural steps, decision rules, and conclusions, giving enough detail to be clear; and assemble the exhibits (tables, work sheets, text, etc.). Then, after the analysis is complete, you review the entries, cleaning them up where needed and adding analysis codes,

and then fill out the more reflective part. It's not a good idea to do a thorough reflection on your analysis while doing the analysis. You need all the energy you can spare for the analysis itself.

Whether you're using this form or a much simpler one, the *log* is crucial. The dictum is this: If it isn't on the documentation form or your original worksheets, you didn't do it. Avoid "laundering" or retrospective enlightenment. Do not let incomplete documentation forms pile up—that defeats your purposes. Do them as you go.

5. The code list we suggested can be used as cues for tactics or procedures you aren't using but could use. This flag will turn up automatically as you notice yourself using the same codes frequently; it probably means that you are relying too heavily on too few devices.

Time Required

Documenting your analytic steps methodically adds time to analysis itself. Sometimes it's fun and helpful; sometimes it isn't. Still, depending on your purposes, it can be very useful, even indispensable.

Our experience with this type of documentation varied, depending on complexity of the analyses being carried out. For an analysis that occupied 2 hours, using Form VII, we usually spent 10 minutes or so doing concurrent logging, and another 15-20 minutes in refining, coding, and reflecting on the experience. For a longer analysis task (say, 4 or 5 hours), logging took 20 minutes or so, and coding/reflection 30-40 minutes. Thus analysis-related documentation at the level shown here can be expected to require something like 20% of total analysis time.

But note: Documentation is not a separate, onerous task carried out for "someone else." It's a method of improving the immediate analysis task being carried out, advancing the sophistication of later analyses, and deepening the confidence you have in final conclusions. The value of friend, colleague, reader, replicator, meta-evaluator, or auditor responses comes later as an add-on. It may be possible to reduce documentation time with further experience or through a simpler form, but spending up to 20% is, we think, a good investment in research quality.

A number of software programs automatically develop a log of analysis operations of certain sorts, such as code construction, searches and retrievals, and proposition testing. See, for example, NUDIST, ATLAS/ti, and QCA (see Appendix). Note, too, that depending on how fully computerized your data management system is (Figure 3.3), your "audit trail" may be relatively easy to maintain. Your system may well include many normal working products: coded data, memos, displays, analytic text, and the like.[25]

How much time does a *review* by a colleague or auditor take, using this form and exhibits? Our experience is: 40-50% as much time as the original analysis. In the main, however, auditing should happen *selectively,* early in a study's analysis work, and for sections about which you want to be especially careful. A full-scale audit of an entire study usually will occur only for high-stakes projects. For example, Merryfield (1990), who was about to embark on a solo study of African schools—one that would become her dissertation—contracted with a colleague to audit her work regularly. That decision itself led to a more carefully designed study.

Summary Comments

Drawing coherent conclusions from displayed data and confirming their strength means using a series of specific tactics. We've reviewed each of 13 tactics for drawing initial meaning from a display that involve forming patterns, looking at contrasts, clarifying relationships, and building a coherent understanding. Thirteen more tactics are focused on confirming or verifying initial conclusions; they consider the quality of the data, focus on "unpatterns" and exceptions, making rigorous tests, and collecting respondent feedback.

The "goodness" of qualitative work needs careful assessment; we proposed a series of questions that researchers and readers can apply to any particular study. We concluded with suggestions for ongoing study documentation, seen mainly as a formative support.

In the next chapter we turn to exploring ethical issues, emphasizing those that bear most centrally on analysis.

Notes

1. We certainly don't consider this list of tactics exhaustive. For example, Martin (1990) describes tactics used in "deconstructing" qualitative texts that only partially overlap these. She includes, for example, **dismantling dichotomies, examining silences** (what is not said, tacitly excluded), **analyzing double entendres,** and deliberately **reconstructing** text through systematic substitutions of words or phrases.

2. The traditional philosophical position here notes the supposedly insoluble "paradox of categorization" (Scheffler, 1967, p. 14): "Observation contaminated by thought yields circular tests; observation uncontaminated by thought yields no tests at all." In spite of this paradox, categories do happen, practically speaking, in the interaction between someone thinking and a set of data. Each influences the other.

Note, too, that conventions for working with categories as variables are available (e.g., grading their intensity from high to low) and that prior theory always involves categories.

3. There are also many other quantitative approaches to clustering; for a good overview see Aldenderfer and Blashfield (1984). See also Trochim (1989a, 1989b, 1989c) for a quantitative approach to producing concept maps from qualitative statements rated by members of a group.

4. Qualitative researchers often use other tropes as well: irony (the view from the opposite, sometimes incongruous or paradoxical side), along with synecdoche (linking instances to a larger concept), and metonymy (representing a whole in terms of one or more of its parts; cf. Morgan, 1983). Noblit (1989) argues convincingly that still other literary

devices, including tragedy, comedy, satire, farce, and allegory, could well be used more widely in developing interpretations.

5. This discussion comes nowhere near the depth of understanding of metaphor that is available in treatments such as Lakoff's (1987) fascinating analysis of categorization. See also Ortony (1979), Lakoff and Johnson (1980), and Johnson (1981). Miller and Fredericks (1988) and Noblit (1989) are especially thoughtful about qualitative research implications; see also Miller (1985) on metaphors in social policy thinking.

6. Hook (1953) claims that qualitative changes are the cumulative result of variations in quantities. Beyond a certain limit of quantitative change, there is a transformation from differences of degree to differences in *kind*—as in changes of weather, emotions, structures, relationships, and the like.

7. Gherardi and Turner (1987) are helpful on when not to count: The "units" involved cannot be treated as reasonably standardized, or we have no rules for ignoring nonstandard information, or we are "reluctant to declare the variety in quality between units to be unimportant."

8. Many "mainstream" qualitative researchers resort routinely to frequency counts for verifying hypotheses. In symbolic interactionism, for example, we can cite Denzin's (1978) ideas on achieving "behavioral validity" through observations of a social phenomenon's appearance in an actor's behavioral repertoire; the greater the frequency of appearance across times and situations, the greater the behavioral validity. See also R. B. Smith (1982) on "enumerative induction" in this same tradition.

9. This step-by-step verification is close to Hempel's ideas on confirmation of hypotheses as requiring a set of sentences that are each directly confirmed by an observation statement (Fredericks & Miller, 1988).

10. Charmingly enough, Christensen-Szalanski and Beach (1984) show that a countering bias exists: Studies showing fallibility are overcited in the literature.

See Sadler (1981) for a thoughtful, empirically based discussion of 13 "information-processing limitations"; they include problems such as data overload, persistence of first impressions, inattention to unreliability in data, and overweighting of novel information. A practical checklist could be constructed easily from the article to guide analysis activities.

11. For a classic, detailed and comprehensive treatment of safeguards against self-delusion, see Becker et al. (1961).

12. For an early and good discussion of researcher effects on the case, see Schwartz and Schwartz (1955). For a thoughtful recent review, see LeCompte and Preissle (1993), who also explore issues of advocacy, role asymmetry, collaboration, boundary spanning, and ethical issues.

13. Douglas (1976) has astutely catalogued and dissected the various evasions, lies, and fronts visited on researchers by informants (see especially pp. 55-82). He cites a marvelous quote from John Leonard: "It took me a long time to discover that the key thing in acting is honesty. Once you know how to fake that, you've got it made" (p. 55).

See also Mitchell (1991), who has a typology of researcher roles: naive-sympathetic (the researcher as prey); naive-unsympathetic (the researcher as butt of jokes); informed-sympathetic (the researcher as over-intimate); and informed-unsympathetic (the researcher as self-deluding spy).

14. Writer Joan Didion once noted: "I am so small, so neurotic, and so inoffensive that people invariably forget an important point: the writer will *do you in.*"

15. On this topic one of us wrote:

> The rhetoric of triangulation, a term drawn from the field of surveying, implies that three elements of a triangle are known. . . . Where there are two data points, all we have is a measure of agreement or disagreement. . . . Real triangulation requires additional information, which may be data from an actual third source (one whose position relative to the two other sources is known); a more general theoretical explanation that subsumes the apparent disagreement; or information about the trustworthiness of the two sources drawn from other data (the assistant superintendent is known to be a liar; the survey item is unreliable, and so on). (Miles, 1982, pp. 125-126)

16. As Miller (1983) notes, triangulation by theory is quite problematic. What if two incompatible theories predict the same result, which is then found?

17. See also Van Maanen's (1979) discussion of the differences between "operational" (observed) data and "presentational" data, which involve respondents' efforts to maintain or enhance appearances. He also notes two other sources of data invalidity: respondent ignorance and taken-for-granted assumptions.

18. The best discussion we have seen on this is by Davis (1985). He reminds us that *spurious* does not mean the association between A and B is false. It is clearly there. What is spurious is the *causal* assumption being made.

19. Systematic replication of studies, in all fields, tends to be rare. The important issue is that studies, in principle, be *replicable*—well enough described that they could be rerun. As Bergman (1989) notes, this condition is often unmet even in fields such as synthetic chemistry.

20. We should remember that explanations are not loose bundles of particular facts, but interrelated mental *constructions* that hold the particulars together. When people register or comprehend a bit of information, they automatically include its *causes*, not just its particulars (Sherman & Titus, 1982). Undoing a thought (or a fact) means unbundling the whole causal stream of reasoning to which it belongs. So a few contrary facts can threaten the whole system, which is why we are loath to notice them—as lay people or as social scientists—and why we are usually unaware that we are blocking them out.

21. Denzin (1989b) says, "In a certain sense, interpretive studies hope to understand the subject better than she understands herself," citing the researcher's broader range of vision and methods of interpretation. But he concludes, "The interpretations that are developed about a subject's life, however, must be understandable to the subject. If they are not, they are unacceptable" (p. 65).

22. Even Peshkin's (1993) thoughtful piece on the "goodness of qualitative research" is limited to listing outcomes including description, interpretation, verification, and evaluation as legitimate results of qualitative studies—without addressing the question of relative goodness.

23. A rarely evoked fourth audience for documentation is other researchers or agencies who suspect scientific fraud. We rarely consider this eventuality in qualitative studies—perhaps the stakes are not high enough to invite it—but we would be naive to rule out the possibility of misconduct (see Teich & LaFollette, 1992 for a range of articles on the topic).

24. We would also mention as exhibits the well-done case accounts in Hammond's *Sociologists at Work* (1964), the methodological report by Lee et al. (1981), the five-case comparison by Firestone and Herriott (1983), and the reports of analysis episodes in Huberman and Miles (1983a). See also Turner (1988). Although Constas's (1992) effort to propose methods for assessing the development of analytical categories is a step in the right direction, it needs to be applied to specific studies.

25. Ironic note: Freedland and Carney (1992) explain in detail how computer use in (quantitative) research can result in problems with data quality control, documentation, and data retention that did not appear when everything was done on paper. For example, software that does not log data transformations obscures the various versions of measures; documentation of variables may be thin; changes in one part of a multiple data set may not appear in other related ones, and so on.

11

Ethical Issues in Analysis

This is a book about methods. Its steady focus on ways of doing things could obscure an important issue: that qualitative data analysis is more than a technical matter. We cannot focus only on the quality of the knowledge we are producing, as if its truth were all that counts. We must also consider the rightness or wrongness of our actions as qualitative researchers in relation to the people whose lives we are studying, to our colleagues, and to those who sponsor our work.

Any qualitative researcher who is not asleep ponders moral and ethical questions: Is my project really worth doing? Do people really understand what they are getting into? Am I exploiting people with my "innocent" questions? What about their privacy? Do respondents have a right to see my report? What good is anonymity if people and their colleagues can easily recognize themselves in a case study? When they do, might it hurt or damage them in some way? What do I do if I observe harmful behavior in my cases? Who will benefit and who will lose as a result of my study? Who owns the data, and who owns the report? The qualitative literature is full of rueful testimony on such questions, peppered with sentences beginning with "I never expected . . ." and "If I had only known that . . ." and "I only belatedly realized that . . ." We need to attend more to the ethics of what we are planning and doing. As Mirvis and Seashore (1982) say, "Naivete [about ethics] itself is unethical" (p. 100).

Some Framing Remarks

Our aim in this chapter is limited: to outline recurring ethical issues in qualitative studies, as these issues relate to *analysis,* and to suggest some bases for dealing with them. We cannot begin to deal with all of the issues, but we do aim to cut across some of the problems involved and raise as much ethical consciousness as we can. We suggest thoughtful treatments by others that we've found useful.

Most professions have well-defined codes of ethics. The American Psychological Association (APA, 1992) developed its code over more than a decade from cases provided by members. It includes guidelines on everything from "boundaries of competence" to "fees and financial arrangements," "plagiarism," and "sexual intimacies with current patients or clients," along with a detailed set of procedures for complaints, adjudication, and sanctions. The American Sociological Association (1989) and the American Educational Research Association (1992) have similar, less detailed codes. Such professional codes function largely to legitimate the profession involved, and to discipline the minority (usually a fraction of 1%) of members who are found to have committed grossly unethical actions. But they seem less helpful in supporting day-to-day decision making on ethical issues.

Although the American Anthropological Association (Cassell & Jacobs, 1987) has developed a handbook on ethical issues for its members, there is still no well-formulated set of ethical guidelines usable by qualitative researchers across a range of disciplines.

Why is this? There are several reasons. As May (1987) suggests, a postpositivist emphasis on multiple realities and researcher interpretation makes for a personalistic, nonsystematic approach to ethical issues. So do the emphasis on the idiosyncratic local context and the "closeness to the respondent" nature of most qualitative studies. These features tend to obscure general principles and make for situation-specific coping.

Another factor is that fieldwork and its accompanying analytic work are often quite unpredictable and situation-specific (Punch, 1986). Conceptual foci, research questions, instrumentation, and sampling all may evolve over the course of a study, making traditional ideas of informed consent or prereview of study design nearly impossible to apply. It is not surprising, as Howe and Dougherty (1992) found, that institutional review boards responsible for approving research proposals have little sense of how to deal with qualitative studies.

Nevertheless, ethical issues are floating constantly beneath the surface of what we do, and they typically receive too little public attention within and beyond the research community. As May (1987) concludes: "If we continue to gloss over the potential power of communicating our ethical questions, decisions and actions among ourselves and educating others about our work, we ultimately [will] have acted in ways that gloss over the rights of those we study and our responsibilities to them" (p. 32).

In the rest of this chapter, we examine some basic theories underlying ethical choices, and then look at a series of specific issues that often arise in qualitative research, exploring their implications for analysis. We examine some of the conflicts, dilemmas, and trade-offs involved, and conclude with some general advice.

Ethical Theories

Specific ethical issues are, explicitly or not, nested in larger theories of how we decide that an action is right, correct, or appropriate. Deyhle, Hess, and LeCompte (1992) suggest five general theories. A *teleological* theory judges actions according to primary ends, good in themselves (e.g., the presumed value of knowledge resulting from research). A *utilitarian*, pragmatic approach judges actions according to their specific consequences—benefits and costs—for various audiences: the researcher, the researched, colleagues, the public. A *deontological* view invokes one or more universal rules (e.g., Kant's categorical and practical imperatives, which boil down to: (a) Would I like this action to be applied to everyone—including me? and (b) Will I treat every person I encounter as an end, and not as a means to something I want?). A *critical theory* approach judges actions according to whether one provides direct benefits to the researched and/or becomes an advocate for them. A *covenantal* view judges actions according to whether they are congruent with specific agreements made with others in trusted relationships.

Flinders's (1992) idea of *relational* ethics emphasizes issues of attachment, caring, and respect, more than agreements made as such; he also adds the idea of an *ecological* basis for ethical decisions, emphasizing the impact of actions on a complete, interdependent system "in the broadest possible context." Both Deyhle et al. and Flinders emphasize that the ethical meaning of specific actions depends very much on the general theory one is espousing. Table 11.1, from Flinders (1992), illustrates this well.

Flinders suggests that researchers with a *utilitarian* view (often, a traditional "scientific" stance) address the recruitment of respondents via informed consent, conduct fieldwork so as to avoid harm to others, and extend this stance through protection of confidentiality in reports. A rule-based, *deontological* view gets more absolutist: Recruitment must emphasize reciprocity (both researcher and researched are to gain), fieldwork must avoid wronging others, and reports must be just, fair, and honest.

Researchers taking a *relational* view stress equal-status collaboration; researcher and researched are now more symmetrical. Fieldwork seeks to avoid imposition of any sort, and reports serve to "confirm," support, or even celebrate people who are defined as "friends."

A comprehensive *ecological* view of ethics, Flinders suggests, leads the researcher during recruitment to be sensitive to the language and meanings of the local "culture"; to avoid "detachment" in favor of being attentive to the wrong or harm one may be doing to the entire relevant environment; and to consider during report writing "how to act responsibly in making public what we have learned," with attention to the effects of implicit, multiplex, stereotypic aspects of language.

Sometimes theorists offer statements of core "principles" to guide ethical choices. In her thorough, practical guide, Sieber (1992, p. 18) speaks of:

beneficence—maximizing good outcomes for science, humanity, and the individual research participants while avoiding or minimizing unnecessary harm, risk, or wrong.

respect—protecting the autonomy of (autonomous) persons, with courtesy and respect for individuals as persons,

Table 11.1
Ethical Frameworks and Aspects of Research (Flinders, 1992)

	UTILITARIAN	DEONTOLOGICAL	RELATIONAL	ECOLOGICAL
RECRUITMENT	1. Informed Consent	1. Reciprocity	1. Collaboration	1. Cultural Sensitivity
FIELDWORK	2. Avoidance of Harm	2. Avoidance of Wrong	2. Avoidance of Imposition	2. Avoidance of Detachment
REPORTING	3. Confidentiality	3. Fairness	3. Confirmation	3. Responsive Communication

including those who are not autonomous (e.g., infants, the mentally retarded, senile persons).

justice—ensuring reasonable, nonexploitative, and carefully considered procedures and their fair administration; fair distribution of costs and benefits among persons and groups (those who bear the risks of research should be those who benefit from it).

House (1990a) also suggests three basic principles, only one overlapping with Sieber's:

mutual respect—understanding others' aims and interests, not damaging self-esteem, not condescending.

noncoercion and nonmanipulation—not using force or threats or leading others to cooperate when it is against their interests.

support for democratic values and institutions—commitment to equality and liberty, working against oppression and subjugation.

We can see that such principles are linked to basic ethical theories: Sieber's look deontological with utilitarian overtones; House's are essentially relational. They provide us with general guidelines, but we are still a good way from making specific choices.

Specific Ethical Issues

Here we identify a series of issues that typically need attention before, during, and after qualitative studies. They are ordered crudely from early to late in a study. A number of them are exemplified in the helpful casebook assembled by Mathison, Ross, and Cornett (1993), which we commend both for its directness and for the teaching suggestions included. Because ethical choices always occur in case context, we also suggest a look at the case descriptions provided by Sieber (1982), L. M. Smith (1990, 1991), Burgess (1989), Marshall and Rossman (1989), and Whyte (1984).

We outline each issue briefly, posing key questions, giving mini-examples and citations, and suggesting the *analysis*-related implications. The aim is not to provide guidelines, but to raise awareness and make ethical issues more visible to the reader, who ultimately—whether deliberately or de facto—will make relevant choices. As we'll see, ethical choices nearly always involve trade-offs, balances, compromises among competing goods and threatening bads.

Worthiness of the project. The question may sound pompous and hortatory, but the issue is not trivial: Is my contemplated study worth doing? Will it contribute in some significant way to a domain broader than my funding, my publication opportunities, my career?

And is it congruent with values important to me? The incongruence might not be as sharp as in Project Camelot (Horowitz, 1974), which recruited social scientists to study counterinsurgency in Latin America (practical translation: The funders wanted to know how to suppress popular revolts). But even an apparently innocuous issue, say a study of school "dropouts," could be decidedly incongruent with your view that bad schools are causing "pushouts."

Implications for analysis. In general, a study that is only opportunistic, without larger significance or real meaning to you, is likely to be pursued in a shallow way, with less care devoted to design and data collection. First conclusions may not be questioned; follow-up analyses with rival hypotheses may be rare. The report will be written to "look good," rather than to be right.

Value incongruence, assuming that it persists during the study, tends to enfeeble and restrict conclusions; if you believe that your work will be misused, you are unlikely to put your mind into full gear. Or worse: If you let yourself be used wittingly, you become complicitous and dishonest.

Competence boundaries. Do I (and my colleagues) have the expertise to carry out a study of good quality? Or (because researchers, both novice and experienced ones, are always exploring things they do not quite know how to do), am I prepared to study, be supervised, trained, consulted with? Is such help available?

Implications for analysis. Unacknowledged (or not understood) incompetence is, we think, responsible for a certain pattern of analytic weakness in qualitative studies: blissful (rather than deliberate) underdesign; accumulation of large amounts of poorly collected, unanalyzed data; and superficial and hasty conclusion drawing as deadlines loom. This picture often occurs when lone researchers fail to seek help from friends, colleagues, or mentors. Graduate students often understand their own inexperience, but sometimes cannot get support and help from their teachers. That is research malpractice.

Informed consent. Do the people I am studying have full information about what the study will involve? Is their "consent" to participate freely given—fully voluntary and uncoerced? Does a hierarchy of consent (e.g., children, parents, teachers, administrators) affect such decisions?

It can be argued (e.g., Eisner, 1991; Wax, 1982) that truly informed consent is impossible in qualitative studies because events in the field and the researcher's actions—such as following up new and promising leads—cannot be anticipated. The major topics of a study may well shift during its conduct (Deyhle et al., 1992).

Nevertheless the issue does not go away. We need to engage it. See, for example, the list of hard questions that McQuillan and Muncey (1990) suggest a school should ask a qualitative researcher before it agrees to proceed:

1. What is the focus of the research? What are the guiding questions? Why and for whom is the research being done?
2. How will data be collected?
3. Who will be asked to participate in this research?
4. What role(s) will school personnel be asked to play in this research?
5. How will participants' confidentiality be protected?
6. Will research participants assist in data analysis?
7. What feedback will the school receive, what form will it take, and at what stages of the research process will it be provided?

The list concludes with an additional heading—"You Can Say 'No' "—and encourages initial and regular negotiation about conditions the school considers unsatisfactory.

It does seem clear that dialogue and ongoing renegotiation are needed (Erickson, 1986; May, 1987; L. M. Smith, 1991) for voluntary, informed decisions about the researcher-researched relationship. Wax (1982) suggests, in fact, that reciprocity is far more important than informed consent.

Even this fairly enlightened list is silent on such matters as potential benefits to school and researcher, risks and costs, who owns the data, and control of publications.

Implications for analysis. Weak consent usually leads to poorer data: Respondents will try to protect themselves in a mistrusted relationship, or one formed with the researcher by superiors only.

Ambiguity about later stages of analysis also can be damaging to study quality, and to the interests of people in the case. If you plan to use "member checks" to verify or deepen conclusions, that expectation and specific procedures need to be clear and reclarified as the study proceeds. See McTaggart's (1991) account: The researcher thought he had developed a clear understanding, codified in written "principles of procedure" about matters including who could veto or edit a report, criteria of fairness, relevance and accuracy, and how reports would be released. But a principal concerned about the report ignored the principles and, by threatening her dismissal, forced a teacher to withdraw data.

Benefits, costs, and reciprocity. What will each party to the study gain from having taken part? What do they have to invest in time, energy, or money? Is the balance equitable?

Researchers are often paid; they usually enjoy their work and learn from it; they may get a dissertation out of it; their papers, articles, and books not only contribute to "science" but also can bring them recognition, royalties, new funding, and career advancement.

Study participants have a somewhat different set of benefits: They get to be listened to; they may gain insight or learning; they may improve their personal practice; a program or policy they are involved with may be strengthened; they may get help in taking effective action on some recurring problem. (In a complex case, these benefits often are distributed unequally: Elites, managers, and key respondents may gain more than "workers" [e.g., students, teachers] or those with peripheral contact with the study.)

But study participants rarely share in publication, and they usually don't become famous. They don't get paid for their research contributions. McLaren (1991) calls this situation "ethnographic vampirism."

The question of costs and who bears them is important. The researcher's time is repaid—usually not fully—in cash or a class grade or dissertation approval. Research participants normally must take time from or beyond whatever else they are doing, and are usually not recompensed. The local organization may well have added costs (e.g., for teacher substitutes).

Implications for analysis. Study participants' concern about the inequity of benefits and costs serves to jeopardize access and thin out data. (We are always moved and amazed when people do keep talking to us thoughtfully, inviting us into their lives day after day, when the benefits to them seem so slim, so intangible, and often so delayed.)

Researchers traffic in *understanding.* Most study participants are preoccupied with *action*—how to work and live better. It can be argued that if you approach your analytic work with a deeper sense of its action implications, your understanding will be deeper—and the benefits to participants more equitable.

Harm and risk. What might this study do to hurt the people involved? How likely is it that such harm will occur? McCall and Simmons (1969) believe that real or feared harm will always occur to someone in a qualitative study. They state the question bluntly: "To whom shall harm be done in this study, and in what magnitude?" (p. 276).

Harm to participants can come in many varieties: from blows to self-esteem or "looking bad" to others, to threats to one's interests, position, or advancement in the organization, to loss of funding for a program, on up to being sued or arrested. The information from a qualitative study is *never* value-free, and it may have negative consequences (as an extreme example, take Garfinkel's [1981] comment about the morality of the "descriptive" statement, "Anne Frank is in the attic," if it were made in the presence of a Nazi search party).

Harm cuts both ways. We like the story told by a *New York Times* reporter who asked a drug dealer if he really felt comfortable about talking frankly. The dealer said cheerfully, "Sure. If I don't like what you write, I'll kill you." As researchers we have occasionally been threatened with litigation, and with promises to intervene with our funding agency when a draft report was seen as threatening a key interest.

As Sieber (1992) points out, it's important to think of varying vulnerability to harm. More-vulnerable persons (and institutions) include those who are "visible, famous and have deep pockets," lack resources or autonomy, are stigmatized, are weakened or institutionalized, cannot speak for themselves, are involved in illegal acts, or are closely associated with those studied.

Setting risk levels for potential harm is very difficult—perhaps impossible—in qualitative studies. It's wise to assume that the chances of some type of harm are better than even, and to consider, in advance, ways of reducing that likelihood. (For example, Erickson [1986] advises distinguishing between qualitative information in a report that is "news" or "not news" to people in the case, *and* that is likely to be regarded positively or negatively; it makes sense to think carefully about the probable impact of "bad news.")

Implications for analysis. As with inequitable benefits and costs, if harm is expected, access and data quality may suffer. The prospect of immediately impending harm—which well may occur when reports are made to local participants, sponsors, funding agencies—can lead to pressure on you to revise or delete conclusions, or to self-censor them in advance.

Honesty and trust. What's my relationship with the people I am studying? Am I telling the truth? Do we trust each other?

Most qualitative researchers are unlikely to lie, cheat, or steal in the course of their work. But broken promises are not unknown. And some researchers have reported deceiving respondents about the true nature of the inquiry (as in the style of "investigative social research" reported by Douglas [1976] and in some "participant observation" studies [e.g., Humphreys' (1970) study of homosexuals, where he posed as a "Watch Queen" lookout outside public toilets]). The trend toward "investigative evaluation" (N. L. Smith, 1992), carried out for accountability or the detection of wrongdoing, may also involve deception.

More typically, dishonesty is more subtle. The field-worker projects a "fake persona" (the friendly listener, the would-be insider) in order to gain knowledge or access, or to keep access (Deyhle et al., 1992). Or, as Punch (1986) remarks, "Subjects are conning you until you gain their trust, and then once you have their confidence you begin conning them" (pp. 72-73).

At some levels, as Van Maanen (1979) notes, whenever the field-worker works to "penetrate fronts," symbolic violence is being done: "People are, to a degree, coaxed, persuaded, pushed, pressured, and sometimes almost blackmailed into providing information to the researcher that they might otherwise prefer to shield" (p. 545).

Van Maanen says there are always individually drawn moral limits to this violence: A researcher may decide not to push on a delicate matter, or to leave an embarrassing scene. Nevertheless, the question of just how coercive and unauthentic relationships with respondents are cannot be ignored, or be defined away by the pious stipulation that "my relationship is fully honest."

Implications for analysis. If people feel betrayed by you when they read a report, it becomes almost impossible for them to accept it as a reasonable interpretation of what happened, because of their natural defensiveness when "the truth hurts," as it may well, and their feelings of anger at having been misled.

There is also a longer-term ecological consequence. When Punch (1986) speaks of "ripping and running" ethnography, or when more traditional researchers speak of "poisoning the well," they are concerned about access by future researchers, not only to the case at hand but also to others in its region, type, or network. Our deceptiveness and broken promises—especially if benefits and costs have been inequitable or harm has occurred—will make any continuation of inquiry problematic. We will have wronged not only our respondents but also our colleagues.

Privacy, confidentiality, and anonymity. In what ways will the study intrude, come closer to people than they want? How will information be guarded? How identifiable are the individuals and organizations studied?

Sieber (1992) makes these helpful distinctions among three terms, which often are confused in research practice:

Privacy: control over others' access to oneself and associated information; preservation of boundaries against giving protected information or receiving unwanted information.

Confidentiality: agreements with a person or organization about what will be done (and may not be done) with their data; may include legal constraints.

Anonymity: lack of identifiers, information that would indicate which individuals or organizations provided which data.

As Sieber notes, issues of privacy are often subtle and misunderstood by the researcher, surfacing only when there are unexpected reluctances, or an outpouring of information beyond what the person meant to say, or a confidence overheard by others. Privacy is, in part, about secrecy; see also Mitchell's (1991) discussion of secrets as serving important functions for individuals and groups (mainly protecting status or access to resources). It's easy for a zealous researcher to treat secrets casually, or to seduce respondents into telling them.

Confidentiality and anonymity are usually promised—sometimes very superficially—in initial agreements with respondents. For example, unless the researcher explains very clearly what a fed-back case will look like, people may not realize that they will *not* be anonymous at all to other people within the setting who read the case. Or, as an indignant principal once told one of us, "There are 111 high schools in this city, and only one of them has a female principal. That's me." (In the report, we moved the school to another city and changed the principal's gender. Was violence done to truth? Yes, though the basic lessons of the case were preserved. Her concerns about anonymity were important because the report pointed to negative aspects of the school, and she did not want it to jeopardize the school's genuine improvement efforts.)[1]

Implications for analysis. When you realize that privacy has been threatened, new analytic moves (e.g., triangulation with other sources) may be needed to protect data quality. If privacy has, in fact, been breached, questions of the report's impact when it is fed back to respondents become central; can the information involved be connected with an identifiable person?

Using member checks to verify or extend interpretations and conclusions helps with anonymity problems, particularly if you begin with the most vulnerable respondents before moving more broadly; they usually (though not always) can spot information that would identify them and thus threaten their interests.

The basic problem of identifiability when the case is a complex site must be considered before and during report preparation. Local people nearly always can tell (or will assume) who is being depicted. You may need to err on the side of protecting anonymity, if it has been promised, and to rely on dialogue and negotiation for corrective action before the report is finalized.

Shulman (1990) found that, in a collaborative research project, local participants'—actually, case writers'—sense of empowerment (and decreased need for "protection") during the project made them want to be identified, rather than to remain anonymous. But it took 6 months of negotiation with the school district (and funding agency), and legal advice, before publication (with minor changes to "foul language" and negative descriptions of students and district conditions) was agreed to.

Explicit confidentiality agreements about where raw data and analyses will be stored, and who will have access to them, probably enhance data quality by increasing trust. However, courts have not normally treated social science researchers' data as a legally privileged communication, as in priest-confessor or doctor-patient relationships (Douvanis & Brown, 1992). If a case is brought to litigation, your field notes, tapes, and other data could become "unconfidential."

Intervention and advocacy. What do I do when I see harmful, illegal, or wrongful behavior on the part of others during a study? Should I speak for anyone's interests besides my own? If so, whose interests do I advocate?

Take, for example, the vivid instance of police brutality witnessed by Van Maanen (1983a) and written up in his field notes. Although he did not assist the police, he did not protest or intervene either; he had what Fetterman

(1984, 1989) calls "guilty knowledge," which he kept to himself, in the service of continued access. He thus protected the police by default, and avoided supporting the victim. Later, in a suit brought by police officers against a newspaper, he opted to "protect the best interests of the patrolmen I knew in Union City" by refusing to release his field notes.

Ethical choices are not always so dramatic, but they are still there when we see indifferent teachers, medical malpractice, abused children, or evidence of misspent funds—and do or don't report them. Ethical choices are also present when we study people who are breaking the law as we study them: drug dealers and users, corrupt agency heads, traffic scofflaws, prostitutes and their pimps and customers, violent gangs, embezzlers. . . . It is important to know more about deviance, but are we somehow condoning it by our neutral, noncondemning presence? Does the "immediate, morally unquestionable and compelling good end of keeping one's promises to one's subjects" (Klockars, 1979) outweigh the harm to specific victims and to society in general? With a formal investigative mandate, our allegiances are clear, but without it, they easily become clouded.

Implications for analysis. A situation in which you have "dirty hands"—cannot evade doing wrong to one party or another—is personally anguishing. And no matter how it is resolved, it skews your understanding. If you decide to withhold "guilty knowledge" in favor of continued access, then not only your public reports but also your conceptualization and explanatory theories may become lopsided. For example, it seems likely that Van Maanen's (1983a) views of the incident he saw may have underplayed the role of institutionalized racism.

But had Van Maanen opted for the victim and gone public on the incident, he might well have forfeited not only his future access (and that of colleagues) but also his understanding of the marginality of police work, its deeply frustrating aspects, and its attractions for authoritarian personalities.

Research integrity and quality. Is my study being conducted carefully, thoughtfully, and correctly in terms of some reasonable set of standards?

This is more than a technical issue. If we provide a set of conclusions based on sloppy (or even fraudulent) work and claim their validity, then we are being dishonest with our sponsors, colleagues, supervisors, respondents, and anyone else who reads and trusts our reports.

Outright scientific fraud is rare, but it does occur when stakes are high (funding, primacy of discovery, career advancement). It would be naive to expect that a qualitative study could never be faked; as Guba and Lincoln (1981) note, "An unethical case writer could so select from among available data that virtually anything he wished could be created" (p. 378).

Far more common, as noted in Adler's (1991) report of a conference on integrity in behavioral science research, is "poor science": sloppy data recording; insufficient, selective, or misleading reporting of findings; unwillingness to share or retain data; undisclosed conflicts of interest; and inappropriate citation.

Implications for analysis. In Chapter 10 we acknowledge the difficulties of assessing the quality of qualitative studies and suggest a series of questions to be asked about the goodness of any particular study: its confirmability, dependability, credibility, and potential transferability to other settings.

The practical implication here is: If you have not attended to the issue of goodness criteria in your study, you are on shaky intellectual ground. It is not just that we must somehow "please" our critical colleague audiences; the deeper issue is avoiding self-delusion. After that we can turn to the task of being honest with our readers about how we did the study, and what worried us about its quality. Without such methodological frankness, we run the risk of reporting "knowledge that ain't so."

As for fraud, the truth eventually will out, though it may take as long as the decades before Cyril Burt's famous results on the intelligence similarity of separated twins were detected as faked. We should concentrate on not lying to *ourselves.*

Ownership of data and conclusions. Who owns my field notes and analyses: I, my organization, my funders? And once my reports are written, who controls their diffusion?

The issue of confidentiality requires being clear about who can have access to data. Most researchers consider their data and analyses as belonging to them and, absent some form of litigation, consider themselves to be responsible for safeguarding its confidentiality, preserving anonymity, and making data available to others for auditing, reanalyses, secondary analyses, and replications.

But the issues here are not simple. As Akeroyd (1991) points out in her thoughtful discussion, the spreading use of computerized databases, while it makes retrieval and analysis vastly easier, raises acute problems of identifiability, misuse of information, harm to respondents, and, indeed, data security. Most European nations now have laws strictly regulating the use of personal data covering privacy, confidentiality, identifiability, and data protection, and it is possible that the United States may follow suit. (In any case, note that the Freedom of Information Act in the United States now permits access to fieldwork data from federally funded projects.) Akeroyd suggests a worst-scenario in which qualitative research may become impossible; at the least, she says, personal data laws "impose constraints and procedures to which social researchers will have to adapt."

The control of report diffusion is also a crucial issue. In many case reports (e.g., McCutcheon, 1990; McTaggart, 1991; Shulman, 1990), local respondents or their organizations have threatened, sometimes successfully, to block publication because of some aspect of the findings. Funding agencies, too, may exert this right; at the 1993 meeting of the American Educational Research Association, there were at least two instances of a government agency's refusing to allow a public report of completed research: In one it appeared that the agency saw findings release as politically unwise; in the other the agency's "internal review had not been completed."

Implications for analysis. In Chapter 10 we encourage the broader use of audits, often done informally by colleagues on a less than full-scale basis. That use improves the goodness of conclusions. But there are potential risks to respondents, and you need to think carefully about how others can access your database, and how confidentiality and nonidentifiability can be maintained.

In our experience, nothing seems to hit a researcher harder than the realization that some person, group, or organization may block the dissemination of months or years of research effort. Freedom of scholarly inquiry is a strong value, as are the values of career advancement, recognition, and funding. Researchers can easily become obsessed about this (one colleague of ours wrote a 60-page paper on how a school district suppressed and ignored his evaluation findings, and still felt unvindicated). You need to be clear about the political context of your work, and to guard against agreeing too easily to others' veto efforts, or altering important substantive aspects as a way of assuring publication or continued funding.

Use and misuse of results. Do I have an obligation to help my findings be used appropriately? What if they are used harmfully or wrongly?

Researchers placed farther along the continuum that starts with description and moves toward evaluation, intervention, and critique (Soltis, 1990) can expect to encounter increasing ambiguity: Do I have the right to change others? Is that paternalistic? Am I sure my findings are correct? How far beyond my findings am I prepared to go? Could my findings or actions based on them lead to harm? Is consciousness-raising without action support enough? Who am I trying to help, specifically?

Once our interventions go beyond reporting to taking local action—even if it is done collaboratively—we enter a larger ethical arena. In a discussion of the ethics of organizational development "change agents," Miles (1979a) points to such issues as congruence between the "client's" and the change agent's values; exaggerated competence claims; prescriptions unconnected to diagnoses; clarification of whose interests are to be served in the local setting; the balance of individual and organizational priorities; the choice of intervention means; use of coercion, persuasion, and manipulation; and accountability. Most qualitative researchers have not pondered such questions, to put it mildly.

Even though, as Everhart put it (quoted by May, 1987), "all fieldwork is politically and potentially reformist," being *directly* reformist by encouraging and supporting the use of qualitative findings is not only technically ambitious (see Chapter 10, section C, and Chapter 12's discussion of using reports) but also ethically demanding.

Examples of misuse of study findings are not hard to find; Judd et al. (1991) point to several types and cite examples. Findings may be misinterpreted (deliberately or not) and used to support wrong policies (e.g., black-white test-score differences as a justification for segregation). Or the findings, correctly understood, can be used for a purpose the researcher deplores (e.g., a new drug found to reduce people's resistance to being attacked might be used by military forces). Or long-range outcomes for different parties may be inequitable (e.g., counterinsurgency research could lead to the suppression of revolt against an unjust government).

Knowledge is not exactly power. Rather, the use of our knowledge raises our awareness of—and enmeshes us in—the power relationships in the world we have been studying.

Implications for analysis. On any given project, it probably pays to be as clear as possible from the start just how committed you are to supporting the use of your findings (and/or to warding off misuse from various quarters). Such clarity not only encourages strong technical attention to utilization effort (as in Chapter 12) but also focuses your attention on the ethical issues. You are less likely to get ethically blindsided if the project drifts—or suddenly shifts—toward a "helping" direction.

A clear commitment to use of your findings also serves to delimit and constrain the conceptual issues being considered, to focus the data collection more than in a straight "descriptive" study, and to narrow the analytic strategies followed.

Conflicts, Dilemmas, and Trade-offs

Most of us would like to believe, as Erickson (1986) says, that "ethical responsibility and scientific adequacy go hand in hand." But the typical research experience is full of dilemmas. You often face a choice between two goods, where choosing one means, to some degree, forgoing the other.

For example, there is often a conflict between the demands of *validity versus avoiding harm.* One of us once studied a school where the new principal, a bit shaky on the job, received regular consultation from an experienced ex-principal—who was called by some staff members "the principal's baby-sitter." During feedback of a report draft, several respondents said, "You don't need to put that in," acknowledging tacitly that it would damage the principal's self-esteem. Right.

But when the phrase was excised, readers did not know that the staff believed their principal needed a "baby-sitter" *and* was a decent man who did not deserve to be hurt—nor that they were, perhaps, asking the researcher to collude with them in keeping their feelings from the principal.

We've alluded to the dilemma of *anonymity versus visibility* (Shulman, 1990); participants involved in action research, collaborative research, and perhaps in efforts intended to be "consciousness-raising" or "empowering" may well want to be credited for their work, be recognized as inquirers who do not need, any more than the "researcher" does, the protection of anonymity.

The long-term study of syphilis patients begun in 1932 at Tuskegee Institute (Jones, 1982) exemplifies the dilemma of *scientific understanding versus individual rights,* which in this case was resolved in a particularly inhumane and destructive way. The scientific aim was understanding the long-term course of syphilis, and it was given priority. So individuals in the study were left without treatment, and were not informed of the availability of penicillin (in 1943) as an effective treatment. The study was not terminated until 1972, by which time most participants had died. At a less tragic level, this dilemma appears in studies where we do not intervene in harmful situations for fear of jeopardizing our access (and thus perhaps our future scientific understanding).

Another dilemma appears when use of results is considered: *detached inquiry versus help.* In any given situation, we can define our role as that of "understanding" what is going on and providing "help" only vaguely, in the long-term future when the fieldwork is over and the report appears. That stance defines out actual assistance to respondents in solving problems they face. If, on the other hand, we choose to make helping primary, we risk becoming co-opted, "going native," and forfeiting our intellectual independence.

Cutting things another way, we can see the dilemma of *help-giving versus confidentiality.* It's ordinarily quite difficult to assist (or get assistance for) a respondent you have discovered to be in need (let's say, a new teacher who is floundering unassisted) without breaching the agreement that no one else will have access to what the person told you.

Freedom of inquiry versus political advantage represents another dilemma. Most researchers opt for the first horn, and are distressed when their findings are blocked. Yet suppose you were strongly committed to the improvement of early childhood education at a point when legislators were about to reauthorize an important national program. And let's suppose that your data (like some in 1993) were fairly pessimistic about that program's effects. Might you want to temper the "whole truth" a bit by saying that the findings were only partial or methodologically flawed?

This list of dilemmas is illustrative; it certainly is not exhaustive. And more often than not, multiple dilemmas will be in play. Many of the accounts we've read of ethical dilemmas in qualitative studies describe "compromises," "trade-offs," "balances," and "unhappy choices." If ethical issues were clear-cut and totally approachable on a deontological basis, life would be simple. But we have to keep thinking and talking about ethical dilemmas, weighing how we and our respondents are choosing one side or the other. As Punch (1986) says, "Morals in research are too important to be left to moralists" (p. 73).

Advice

Awareness. Consider what your general ethical position looks like: What theory or frame do you seem to be applying to day-to-day choices in your research? Reasoning inductively from past situations in which you felt uncertain about the right thing to do can be helpful. So can looking at situations in which you felt okay, but others did not. Another avenue is discussing cases like those provided by Mathison et al. (1992) and comparing your ideas with others'.

Anticipation. Most of the specific issues raised above can benefit from advance thinking during the early stages of project design. Do not delude yourself that they can be pre-resolved, but taking thought—perhaps running through the issues as a sort of checklist—can help you avoid problems later.

Preliminary agreements. We've discussed the dimensions of early "contracting" with case participants (see Figure 3.5) and won't repeat them here. Our main advice is to attend to such agreements explicitly during entry and access, and to commit them to paper (including a brief handout brochure for use with people who were not close participants in the agreements).

Documentation and reflection. Our previous advice on study documentation should be expanded here. It's easy to become preoccupied with the demands of data collection and analysis, and to miss latent, potentially painful ethical issues until it is too late. Some routinized structure, such as an occasional "ethical issues" memo or study team meeting, helps surface the sorts of mild worries that often prove to be a distant early warning.

Third parties. Because ethical issues often tend to be masked by our taken-for-granted assumptions, beliefs, and values, engaging a trusted third party can be very helpful in bringing them to attention. Such a person can raise unheeded issues, suggest alternate viewpoints, help surface tacit assumptions, be an advocate for respondents, or serve as a mediator between respondents and researchers when there are unresolved problems (e.g., in the account by King et al., 1993). See also the helpful advice of Bermant and Warwick (1978).

Regular checking and renegotiation. The evolution of any qualitative study normally involves some twists and turns that no one fully anticipated, even with the best of good will. Initial agreements and working procedures almost always need updates. In our experience, agreement revision is quite typical as issues of feedback, member checks, and reporting come into focus. It's useful from the start to create the expectation that agreements may need renegotiation, and that "recheck" meetings can be called at any point by either researchers or respondents.

Summary Comments

We've examined general ethical theories, and a series of specific ethical issues, ranging from early ones (the project's worthiness, your own competence, informed consent, anticipated benefits and costs) to those occurring as a project develops (harm and risk, your relationship with respondents, privacy/confidentiality/anonymity, and intervention) and those prominent late (research quality, data ownership, and use of results). All have clear implications for analysis and the quality of conclusions. Dealing with ethical issues effectively involves heightened awareness, negotiation, and making trade-offs among ethical dilemmas, rather than the application of rules.

In the next chapter we consider study products in the form of reports, and how they're produced.

Note

1. Issues of privacy, confidentiality, and anonymity take a completely new turn when visual media (still pictures, video) are used in data collection and reporting. Individuals and groups still need to have control over how they are depicted, and can resist privacy invasions (e.g., via vetoing specific pictures or sequences), but the possibility of anonymity is almost surely gone, even with facial-blur special effects. The best discussion of this subject we have seen is by Gold (1989), who also notes that visual material, when fed back, has potential for harm beyond that of text because it is seen as "definitive," especially if no associated explanation is provided. Gold has many useful suggestions for dealing with such issues.

12

Producing Reports

The General Problem

Qualitative research is not just an exercise for our private enjoyment. Some form of reporting the results to others will be in play. But what form? It's clear that the conventional formats long familiar to quantitative researchers, something like this, are too schematic and constraining:

Statement of the problem
Conceptual framework
Research questions
Methodology
Data analysis
Conclusions
Discussion

A qualitative study *could* follow that format, but it looks Procrustean, forced. Normally we'd have other expectations for a qualitative report. For example, we might expect a close description of the history, the context, and the major actors, and we might want to see some of those details in the very statement of the problem. We might look for a more "circular" linkage between research questions, methods, data collection, and interim analyses, as each analysis opened up new leads. And many qualitative analysts would not start with a conceptual framework, but rather aim to end up with one. Qualitative data have special strengths, as we note in Chapter 1: local groundedness, holism, temporal extension, access to causality, emphasis on meanings. Reporting formats must respect them.

Zeller (1991) suggests that qualitative studies don't report out "data"; they report "scenes"—that is, accounts of researchers' engagements over time with informants in their surroundings. Often these reports may not be compiled into factually accurate accounts, but rather serve as a corpus from which the researcher actively selects, transforms, and interprets the material at hand—sometimes without leaving traces of the successive steps taken along the way (Chapter 10, section D).

Zeller asks further whether qualitative field studies are different from the nonfiction novel, or from slice-of-life case studies from the New Journalism. Do qualitative studies have a distinctive reporting format? As we saw earlier, even for data segments within a study, a set of field observations can be rendered differently—in a poetic form (Chapter 5, section B) or as a series of vignettes (Chapter

4, section G) or as scenes or stories. A blurring of the frontiers seems to occur between social scientific reporting and "figurative" or "rhetorical" renditions of aesthetic material.

The reporting of qualitative data may be one of the most fertile fields going; there are no fixed formats, and the ways data are being analyzed and interpreted are getting more and more various. As qualitative data analysts, we have few shared canons of how our studies should be reported. Should we have normative agreement on this? Probably not now—or, some would say, ever. Yet it's important to consider the choices you face in designing and writing reports.

The challenge is to combine theoretical elegance and credibility appropriately with the many ways social events can be described; to find intersections between the propositional thinking of most conventional studies and more figurative thinking. Just as good analysis nearly always involves a blend of variable-oriented, categorizing, "paradigmatic" moves, and case-oriented, contextualizing, narrative ones, so does good reporting.

Overview. As in our section on qualitative study "goodness" (Chapter 10, section C), we do not offer a fixed set of ideas about reports and reporting, but rather identify a series of *choices* you face in producing study reports. They include choices about the report's audiences and the hoped-for effects on them, the "voice" or genre of the report, its writing style, its structure and format, and what will be done to support use of the report. Such choices apply, of course, to interim, as well as final, reports.

Our stance is that these choices should be made clearly and deliberately in preliminary form—for interim reports, quite early, and for final reports, somewhat before the midpoint of a study. Otherwise much effort can be wasted. Wolcott (1990b) has an even stronger position: *"You cannot begin writing early enough....* Would that mean you might write a first draft before ever venturing into the field ... ? Correct" (p. 20). The point here is not advocacy of data-free writing, but of early and continuing writing as a way to make your ideas—more and more informed by data—explicit.

Other researchers may be more leisurely, and more indeterminate. But the issue is that *not* facing choices about writing is a choice—a de facto action that may lead you into unhappy, unforeseen circumstances when the crunch of producing final reports begins to make itself felt.

As we emphasize throughout the book, data analysis includes selecting, condensing, and transforming *data; displaying* these data in an organized way; and drawing and verifying *conclusions* from the condensed, displayed data. Any interim or final report will deal, more or less explicitly, with this flow of analytical processes and their resulting products.

As we said earlier, reporting is not separate from thinking, from analysis. Rather, it *is* analysis. We can adapt the old saying into a new form: "How do I know what I think until I see what I've written?" Or, as Atkinson (1991) says, "The analytic induction of categories, themes and relationships; the explication of meaning; and the understanding of action may all proceed via the writing itself. . . . The 'writing up' of the qualitative study is not merely a major and lengthy task; it is *intrinsic* to the 'analysis,' the 'theory' and the 'findings' " (p. 164).

Here, then, is a set of report-related issues you face from the start, whether or not you decide to confront them directly.

Audiences and Effects

The reader and the writer. Reports are supposed to be written for specific audiences, to achieve specific effects. But the familiar label "audience" obscures some important, primal issues. As Agar (1986) notes, "Ethnographer, intended audience, and group [studied] all represent traditions that limit, but do not fully determine, the possible ethnographic reports that can emerge. Ethnography is neither subjective nor objective. It is interpretive, mediating two worlds through a third" (p. 16). Or, as Noblit (1989) puts it, ethnographies are interpretations of the interpretations people (respondents) make of their situations.

So you cannot exactly manage a fully rational specification of audiences, nor of the effects you want to achieve with them. Rather, in Erickson's (1986) terms, the reader is a co-analyst, experiencing the original setting vicariously, looking at the evidence, weighing the writer's interpretations and perspective—and noting how they have changed along the way. Harris (1990) has a similar view: The writer starts with an "interior text" that moves into a draft "generative text" and then to a public text (the actual report). The reader, he suggests, starts with the text, then works back to a personal generative text, and then to interior text.

Reasonable clarity about audiences helps this process. Richardson (1990) notes in her helpful book on reaching diverse audiences (academic, trade book, mass circulation): "All writing is *encoded,* serving the rhetorical function of locating it in a particular genre, helping the reader to know what to expect, and thereby, in actuality, helping the writer, who can then draw on the habits of thought, glosses, and specialized knowledge of the reader" (p. 32).

And Schatzman and Strauss (1973) speak of "audience conjuring":

> Since one can hardly write or say anything without there being some real or imagined audience to receive it, any description necessarily will vary according to the audience to which it is directed. Audiences "tell" what substance to include, what to emphasize, and the level and complexity of abstractions needed to convey essential facts and ideas. (p. 118)

Types of readers. Generally speaking, you need to make choices of reader types from a list like this:

- Local respondents: the people who provided data
- Program operators: in evaluation studies, people running and/or deciding about the program being looked at
- Practitioners: people engaged in the same sort of work as those studied, but in different settings
- Other researchers:
  - colleagues in your own setting
  - members of dissertation committees
  - colleagues in your academic "field"
- Policymakers: governing boards, legislators, agency officials
- General readers: purchasers of trade books
- Mass readers: purchasers of magazines and newspapers

Types of effects. What are you hoping will occur when a particular type of reader engages with your report? Although that cannot be really predicted or controlled, getting your *intentions* clear can make a profound difference in what the report will look like, and how you produce it. Any qualitative research report may be written from certain general stances:

Aesthetic:
- entertain, amuse, arouse feeling
- enable vicarious experiencing

Scientific:
- heighten insight, illuminate, deepen understanding
- add to existing information on a topic
- expand or revise existing concepts, theory, explanations
- convince the reader of the report's worth, truth, and value
- advance the methodological craft of research

Moral:
- clarify and sharpen moral issues
- emancipate, raise consciousness, free the reader from unrealized oppression

Activist:
- enable improved decisions, provide guidance for action
- show connections between findings and local problems
- empower the reader, increase the sense of control
- mobilize specific action
- support the reader in future use of the findings

What an array of intentions! The critical question is *which* effects you are intending for *which* types of reader. If your report is a dissertation and your audience is your committee (and only secondarily other researchers), then the effects of theoretical and methodological advancement—or perhaps, mostly, convincing the readers of the report's credibility—are likely to be central.

On the other hand, if your audiences are policymakers and trade book readers, your intended effects may be different—for example, the illumination, moral clarification, and mobilization that Kozol (1991) doubtless intended in his riveting depiction of urban schools in *Savage Inequalities*.

Advice. Make a matrix for your study of reader types by intended effect types. You may find that more than one report is needed, as did Richardson (1990), who wrote in-progress and final academic papers, a trade book, magazine articles, and an answer column—all based on her study of single women's liaisons with married men.

Voices, Genres, and Stances

How do we speak to the reader? Many choices are to be made, and no standard lexicon describes the overall tone, mode, and orientation of the report. Yet when we read a research report, we can almost immediately detect a pervasive stance, a flavor, a tone that defines the relationship between writer and reader.

Van Maanen's (1988) eloquent *Tales of the Field* distinguishes among several possible "voices," with many examples from his own and others' work:

- *Realist:* a direct, matter-of-fact portrait, with methods left mostly undescribed. Many details. The field-worker is invisible and "interpretively omnipotent."
- *Confessional:* written from field-worker viewpoint, with personalized authority. Methods are "unmasked" through frank, fallible description.
- *Impressionist:* personalized, atheoretical accounts, often storylike, aiming to link reality and the field-worker, and to enable reader reliving of the experience.

These "voices" have profound consequences for what is included in a report and what can be learned from it. For example, Van Maanen notes that a "realist" voice tends to rule out alternative interpretations and/or to pretend that the interpretations come straight from the respondents. "Confessional" tales may overfocus on the field-worker as doing "vanity ethnography," blurring "what happened," and may lead to a paralysis of method. An "impressionist" account may provide a portrait that is far more coherent than disorderly reality. (These are only sketchy summaries: We commend *Tales of the Field* to the reader seeking an articulate, experienced researcher's reflections on writing.)[1]

"Tales" in Van Maanen's terms are sometimes labeled differently by others. Richardson (1990) speaks of "narrative stances," such as "omniscient," "exploratory," and the "limited sociologist." Geertz (1980) speaks of general "genres," like those seeking explanations through "laws-and-instances," as versus "cases-and-interpretations," and discusses the blurring he sees between more specific genres: "philosophical inquiries looking like literary criticism . . . scientific discussions looking like belles lettres *morceaux* . . . histories that consist of equations and tables . . . documentaries that read like true confessions . . . parables posing as ethnographies" (p. 165). Geertz's article makes it clear that stock "genres" can no longer be pulled off the shelf and used automatically.

Advice. You have *choices* about the voice, genre, and stance of your report, and they need to be made wittingly, if you can. Sometimes these issues come clear only after you have looked at some first-draft material. In the revised outline for this book, for example, we originally intended a certain "voice": that of a "good, knowledgeable, fallible, concrete mentor; trusted guide through familiar, sometimes surprising, territory." A good deal of the first-draft material, however, was full of references to "the researcher," "the analyst," and had to be recast to speak more directly to the reader, someone known as "you."

Style

Should qualitative reports be "well written"? It can't hurt. The classic criteria of "clarity, force, and ease" apply to all writing. Becker (1986) and Wolcott (1990b) are particularly useful for qualitative researchers. Strunk and White (1979) and Zinsser (1976) are classic guides to good writing in general. We all need to eschew obfuscation, er, uh . . . be *clear* about what we have to say.

But wait a minute. There may be some problems. Can a qualitative report be very well written and yet be "wrong"? Certainly, at least in the qualified sense of "wrong" we discuss in Chapter 10. Can a report hide its conceptual emptiness by being interesting, articulate, and vivid? Quite possibly. And as Clough (1992) shows, it's not too hard to deconstruct the "plain" style sometimes advocated as good writing and argue that it "engulfs the reader in a nostalgia for the absence of conflict and difference." Nevertheless, we commit ourselves: Goodness of style is good.

Matters of style are connected with choice of voice. Using Latinate instead of Anglo-Saxon language, passive instead of active verbs, "objective" stances rather than honestly personal ones, and indirect locutions instead of straightforward talk have a long and dishonorable history in traditional research reporting. They serve only to wrap the writer in the mantle of "science," while mystifying and alienating the reader.

Advice. Like any writer, you need feedback and revision of successive drafts from trusted colleagues with strong editorial skills, and/or from intended recipients of the report. If you still don't get it, try on-paper consultants like Becker or Zinsser.[2]

Formats and Structures

The analytic mix. We've said it before and we'll say it again: Good qualitative research—and thus good reporting—requires an interactive mix of two basic views of the world. The views are construed slightly differently by different writers, but they generally line up like this:

Variable-oriented	Case-oriented	(Ragin, 1987)
Categorizing	Contextualizing	(Maxwell & Miller, 1992)
Analytical	Synthetic	(Werner & Schoepfle, 1987b)
Etic	Emic	(Headland, Pike, & Harris, 1990)
Variance theory	Process theory	(Mohr, 1982)

The views on the left side can often be represented most easily (no surprise) in matrix form; the views on the right typically benefit from a network display.

Vitz (1990) echoes these two views: He sees conventional data analysis, involving *propositional* thinking, as the fruit of abstract reasoning, leading to formal, theoretical interpretations. *Figural* genres, such as narratives, entail more concrete, holistic reasoning; they are "stories" retaining the temporal configurations of the original events.

Recent studies on modes of thinking and representation (Fuhrman & Wyer, 1988; Howard, 1991) have made the strong case that we retrieve information from memory in different forms; two prominent ones are the *paradigmatic* and the *narrative.* We represent much of what we think or remember either through propositions or through stories. Both modes are useful when we report our findings. They will be better attended to, remembered, and integrated into the reader's previous conceptual map.

A good case history, by definition, must trace the flow of events over time. Careful description of the settings, people, and events is one of the main contributions of qualitative research. But such descriptions also have an *analytic, interpretive* purpose: to illuminate the constant, influential, determining factors shaping the course of events. After all, our main findings entail themes and constructs derived from the interactions between settings and people.

In our own work, we've aimed to keep the chronology of events conceptually tied to variables emerging during interim analyses. For example, an emerging theme of "distrust of school officials by parents" may be derived from many "scenes," episodes, public interactions, and private conversations—each with clear antecedents and consequences. The sense and coherence of these happenings, seen over time, is the grounding of the emerging theme.

Stories without variables do not tell us enough about the meaning and larger import of what we are seeing. Variables without stories are ultimately abstract and unconvincing—which may explain certain scrupulous rules for reporting quantitative studies, as well as the familiar comment, "I couldn't really understand the numbers until I looked at the open-end data."

Is there an optimum balance of stories and variables? Lofland (1974) cites a reviewer's rule of thumb: 50-70% should be events, anecdotes, episodes, and 30-40% should be conceptual. Those figures seem arbitrary, but we would surely be unhappy with 95% story and 5% conceptual, and the converse would be worse. Essentially, of course, as Atkinson (1991) puts it, "The analytic frame and the qualitative data it comments on should be 'interpenetrated' if the text is to be judged satisfactory" (p. 169).

Presentation modes. In this book we stress two basic modes of presentation: *text* (with varying degrees of detail and organization), associated with organized *displays* in either matrix or network form. We think these modes are an enormous help in drawing coherent meaning from data and in confirming and deepening conclusions. By extension they communicate clearly and well to readers.

But there are other possibilities, too. Could a qualitative study report be a lyrical poem? (Yes, as in Chapter 5, section B.) Could it be rendered as a drama or a comedy? As a musical composition? Technically, the answer is probably "yes,"—but applying the sorts of goodness criteria outlined in Chapter 10 becomes more and more difficult. Could the qualitative "fictions" that we and our respondents generate (Geertz, 1980; Van Maanen, 1979) be rendered in a novelistic form, blurring the distinction between real and fictional conversation, as Oscar Lewis (1961) does in *Children of Sanchez?* Such accounts are typically "larger than life"; like all good fiction, they make individuals compelling to the reader because they describe situations that are more extreme or "archetypal" than most instances of daily life. But this is shaky ground. The reader should look at Clifford's (1988) illuminating, well-documented account of the life of Edyth Astrid Ferris, an educational researcher in the 1920s. At the end of her account, Clifford reveals that Ferris never existed; she is only a historical composite. You feel both infuriated and appalled at how easily the writer manipulated you. Coherence, plausibility, and compellingness are not enough to ensure the goodness of a qualitative report. As in TV, "docudramas" are certainly coherent—but different from most people's versions of what happened.

That comment reminds us that reports can go beyond words. We've restricted our analytic and presentation modes in this book, but many qualitative researchers use presentation modes of drawings, still photos, and videotapes. They, too, have issues that transcend compellingness (the best discussions of these modes we have seen are by Harper, 1989, and Ball & Smith, 1992).

Even within text, qualitative researchers characteristically employ a range of literary devices. The description of the study's context is often rendered in quasi-pictorial form ("Yellow Falls lies at the end of a fertile valley, with small farms in its hollow and a creek meandering through the flood plain"). Important events or interactions may be reported in the form of vignettes ("This happened near the entrance to the outpatient clinic, when the cardiologist, two residents, and several patients were within earshot."). Many codes—especially pattern codes—are captured in the form of metaphors ("dwindling efforts," "interactive glue"), where they can synthesize large blocks of data in a single trope.

In the wake of postmodernism, it's even been claimed that *all* modes of reporting on social science research are essentially rhetorical (Howard, 1991; Noblit, 1989; Richardson, 1990). Metaphor, irony, metonymy—even comedy, satire, tragedy, farce, and allegory—are seen as workable tools. Probability estimates, graphs, and theorems can be called instances of a classic trope (synecdoche, where the part stands for the whole), meant to reinforce a "management metaphor": The researcher is in charge, has "manipulated" the key variables, has "managed" the data set (Richardson, 1990). But qualitative research reports are both something more and something less than sheer "literature." Our tropes are about something that we have realistic reasons for believing in; we are real authors, not fictional narrators (Whaley, 1991). Warrants are behind our claims.

Alternative structures, outlines. How are qualitative reports organized? Not surprisingly, there are no standard setups, except at a very general level. Each researcher must craft a report structure that fits the intellectual and local context of the particular study.

Here are some general and specific illustrations. Werner and Schoepfle (1987b) limit their advice: Combine analysis and synthesis and choose between a "climax first" and "climax last" structure. Not much help. Bogdan and Biklen (1992) suggest a three-part report: introduction (including background and research methods), the "core" (arguing a thesis, presenting a theme, illuminating some topics), and the conclusion. Lofland and Lofland (1984) propose these sections: general questions; overview of data; review of prior work; "main part of the report," with a balance and interpenetration of description and analysis, plus elaboration; and conclusions/implications. They relegate actual data collection and analysis processes to "a footnote, early section, or appendix."

R. Ginsberg (personal communication, 1993) suggests a structure starting with "what" (will be done in the report), followed by "why" (rationale, conceptual framework, etc.), "how" (the methods), "analysis" (the findings, often in several chapters), and "meaning" (conclusions, connections to theory).

Judd et al. (1991), not limiting themselves to qualitative studies, suggest an "hourglass" outline, beginning with a broad introduction, becoming more specific with the problem definition and literature review, and most specific in a methods and results section, and finally opening out more broadly to conclusions, discussion, and implications.

For qualitative evaluations, Patton (1980) includes sections on purpose/context, methodology, presentation of data, validation/verification of findings, conclusions, and recommendations. But the 1990 edition of Patton's text abjures any such outline, settling for advice on "focus," balancing description and interpretation, and inclusion of an executive summary/abstract.

Lofland (1974) helps a bit more specifically with an empirical look at editors' evaluations of more than 200 articles and reports. He sorts reports into several "styles": "generically framed" ones, including "moral" (diatribe), "protocol" (data speak for themselves), "then they do this" (sustained narratives), and "vacillating." He also mentions would-be "novel frames," "elaborated frames," and "eventful frames" focusing on concepts and events, and "interpenetrated frames" that link them closely.

It seems clear that a study's questions, context, and audiences drive the design of reports more than any general canon could. So let's look at a few specific outlines.

Figure 4.13 (p. 84) shows the outline of a "prestructured" case focusing on reform in urban high schools. Note that it says nothing about the general study's research questions because it is one of several subreports; still the headings and associated displays are driven by these questions. When the case reports (70-100 pages), verified through local site feedback, were compressed into a book for practitioners and researchers (Louis & Miles, 1990), the case report structure was streamlined considerably into 20-25-page case summaries, like this:

A picture of the school in 1985
The context
The improvement program
The story of implementation
Preliminary results
Epilogue (2 years later)
Why did this happen? A reflective review

The book's early portions were mostly variable-oriented; they presented summaries of the problem and past literature, and gave a brief cross-case overview. Then came the cases as extended narratives. The latter portion of the book was organized around cross-case examination of key variables, such as "vision building," adding vignettes (mini-stories) to clarify the variables. The book concluded with action implications. The methods used were described in appendices.

By contrast we can look at Lieblich's (1993) headings for a narrative report of the life of a young woman who had immigrated from Russia to Israel:

Part I: Life in Transition
- The Present Situation
- Natasha's Past in Moldavia
- The Beginning of Transition: Natasha's Years as a Student in Moldavia
- Departure and Arrival

Part II: What Was Lost?
- Parental Authority
- Friends and Social Network
- Loss of Status
- Loss of Clarity of Norms
- Loss of Self-Confidence
- Loss of a Sense of Cultural Belonging
- Clarity of Career Path

Part III: Changing and Not Changing
- First Signs of Changing
- Establishing New Friendships
- A Boyfriend
- New Career Plans
- New Gender Expectations

Can We Speak About a Newly Acquired Identity?

Lieblich describes her methods briefly at the start (two extended open-end interviews 6 months apart) and then turns to a combination of present and retrospective data to bring the reader into Natasha's story. Part II is variable-oriented, reviewing what Natasha experienced as losing during the transition; they are cut-across themes illustrated with excerpts and vignettes. Part III returns to the current situation, looking for signs of change in terms of both events and themes. The last section summarizes conclusions.

As a third exhibit, here's the report outline for an action-oriented evaluation study (Saxl, Springer, Sauer, Sclan, & Miles, 1990). The audience is managers, funders, and operators of a program for helping schools adopt better methods of reading and writing instruction. The program was just completing its fourth year, and its stakeholders needed information that would help them decide about the program's future.

Executive Summary
- I. Introduction
  - Program Overview
  - Evaluation Plan
  - Methods
- II. Documentation Summary of Services Delivered
- III. Survey Findings
- IV. Case Studies
  - A. City Intermediate School
    - Introduction
    - The School Context
    - Goals
    - Services Delivered
    - Staff Developer Support at the School
    - Implementation
    - Accomplishments
    - Problems Encountered by the Team, and Strategies Used to Cope With Them
    - The Basic Skills Improvement Plan: Quality and Level of Implementation
    - Classroom Implementation
    - Impact and Capacity Building
    - Integrative Summary
  - B. Roundtree Junior High School (parallel structure)
  - C. Union Prep High School (parallel structure)
  - D. Cross-Site Summary
- V. Recommendations

In this study, quantitative data from teachers, program staff, and students across 18 schools were organized according to key variables, such as "staff developer support" and "implementation." The extensive case study material showed through narrative how the same variables played out over time.

The final recommendations were anchored around a series of 10 key findings, such as "The role of the staff developer is still vital in project success. Yet staff developer presence will probably be thinner during the fifth and possible subsequent years." Each finding had a general recommendation, such as "Consider how the functions performed by the staff developer can be built in to the local setting, and try some prototype approaches during the coming year." Then the report presented a series of four to eight possible specific actions exemplifying the recommendation ("Staff developers 'contract' with their schools as to how their functions can be assumed locally during the coming year."

Advice. If we look across these examples (and consider the dozens of other qualitative studies we might encounter in a year), do we see any minimum guidelines for report structure? Grossly, yes.

1. The report should tell us what the study was about or came to be about.
2. It should communicate a clear sense of the social and historical context of the setting(s) where data were collected.
3. It should provide us with what Erickson (1986) calls the "natural history of the inquiry," so we see clearly what was done, by whom, and how. More deeply than in a sheer "methods" account, we should see how key concepts emerged over time; which variables appeared and disappeared; which codes led into important insights.[3]
4. A good report should provide basic data, preferably in focused form (vignettes, organized narrative, photographs, or our data displays) so that the reader can, in parallel with the researcher, draw warranted conclusions. (Conclusions without data are a sort of oxymoron.)
5. Finally researchers should articulate their conclusions, and describe their broader meaning in the worlds of ideas and action they affect. This guideline brings us full circle to the "goodness" questions explored in Chapter 10.[4]

Using Reports

Regardless of how eloquent, well-reasoned, and coherent a report is, we must make choices about how it will be disseminated, by whom, to whom, and with what anticipated uses. Here we return to audiences and effects, but with a bit more juice; by now it is clear that your work may not be over when the page proofs have been corrected.

This work requires differentiating audiences and effects somewhat more finely than we suggested earlier, as shown in Table 12.1. For example, you need to think not just of general audiences such as "practitioners," but also of subaudiences such as innovators and gatekeepers. The effects, too, are more differentiated, and have to be seen as levels of diffusion and use—ranging from simple awareness of the existence of your message to understanding, an "adoption" decision, implementation, and integration into regular routines.

Here are a few examples of dissemination and utilization efforts. King et al. (1993) describe what happened when Wasley's qualitative "snapshots" of radical school

Table 12.1
Matrix for Objectives of Dissemination Strategy Planning

	Size of Audience			
Level of Diffusion and Use	Innovators: those few with special interest, skill, motivation	Key persons: gatekeepers, leaders of opinion	Majority of potential audience	Everyone who might be affected
Awareness				
Reception of basic message				
Understanding				
Acceptance				
Adoption decision				
Utilization/ Implementation				
Integration				
Routinization				

change efforts were fed back for discussion to school faculties, represented in the paper by King and Louth. There were strongly negative reactions, and although both school people and researcher were committed to using the results for improvement, there was little evidence of increased capacity to do so, after the first waves of affect had cooled down. Continuing effort seemed to be required before real use of the findings could proceed.

Miles, Saxl, and Lieberman (1988) studied the skills possessed by effective school "change agents." Although they reported the results in the form of papers and journal articles, they were committed from the start to developing a set of training materials built around the research findings. Having done this, they found an "intermediary" professional organization to publish and distribute the materials (Saxl, Miles, & Lieberman, 1990) and host a series of "training of trainers" workshops for state and regional chapters, designed and led by the authors.

Huberman and Gather-Thurler (1991) examined what happened as staffs of scientific research projects on vocational education in Switzerland carried out their mandate to disseminate the findings of their studies. But before dissemination started, Huberman and two consultants (Miles and R. Havelock) led a week-long training program for the researchers, helping them identify the basic "message" they wished to disseminate, the audiences they wanted to reach, the level of "use" aimed for (using Table 12.1), the dissemination media they would use, the intermediary organizations and networks they would work through, and how they would know whether their strategies were being effective.

During the training program, researchers planning their dissemination strategies used the matrix to define their objectives by starring cells they regarded as a minimum set of objectives, and adding check marks for cells they hoped they could accomplish beyond the minimum.

Glaser (1978) once reflected: "What a man in the know does not want is to be told what he already knows. What he wants is to be told how to handle what he knows with some increase in control and understanding of his area of action" (p. 13).

The message of the several thousand available studies of knowledge use (see Havelock et al., 1979), however, is that meeting the needs of Glaser's "man" goes far, far beyond being "told." Supportive assistance, usually in an interactive setting, is almost always essential. The knowledge from qualitative studies is *not* "self-implementing," once we move beyond the objectives of entertainment and illumination. Normally, the user needs direct, sustained interpersonal contact with a source of help—if not the original researcher, then an intermediary, as in the Saxl, Springer, et al. (1990) example.

Take, for example, the book you are reading at this moment. For the ideas in it to be well used in your regular work, you will need direct, interactive experience: in a graduate class on qualitative research methods; in an ongoing research project where you and colleagues try out and adapt the methods; in a design session where a consultant is helping you put together a new project; in a workshop on qualitative data analysis methods. Few people's practice will be changed by reading alone.

Advice. It's clear that once you choose to act in the dissemination/utilization area, rather than leaving such issues to the actions of others, you will be investing a good deal of energy. It's a new mode of work, a new project. Detailed interactive strategies need to be designed and carried out, usually over a period of months, sometimes years. Fortunately the wheel need not be completely reinvented; very helpful discussions of dissemination and use strategies appear in Glaser, Abelson, and Garrison (1983), Havelock (1973), and Rothman (1980).

Summary Comments

We've identified a series of choices that need to be made about study reports, preferably before the midpoint of the work. They include specifying audiences and desired effects on them, clarifying the "voice" of the report, working on issues of style, designing formats using a range of presentation modes that combine "variable" and "story" views, and attention to strategies supporting the use of reports.

The next and final chapter concludes our journey.

Notes

1. Van Maanen (1988) also lists several other "genres" and types of tales, including:

Critical: focus on the meaning of local events as embedded in larger social, political, or economic structures.

Formal: systematized use of inductive and inferential logic to develop theory (e.g., semiotics, conversational analysis, ethnomethodology).

Literary: theatrical, emotion-arousing accounts shaped by the author; little reference to past "theory" (e.g., Tom Wolfe, James McPhee, Norman Mailer, Truman Capote as "New Journalists").

Jointly told: texts produced by both respondents and field-worker; "polyvocal" accounts.

Van Maanen also mentions "tales" told during the course of applied work: "hired-hand" tales reporting on an issue at the request of others; "clinical" tales focusing on an improvement program headed by a "client"; and "action" tales where the field-worker is a direct agent of change.

2. See also Zeller's (1987) ideas on "craftsmanship," which, as extended by Lincoln and Guba (1990), may include many features:

power and elegance (level of discourse, precision, evocative value of metaphor)

creativity (revealing new understandings, meanings, questions)

a quality of openness (to problematics, to reinterpretations)

independence (operating outside the "given" frames; evidence of the intellectual "wrestling" during the study)

demonstration of an emotional and intellectual commitment to craftsmanship

"courage" (going beyond "safe" limits; putting your ego on the line)

egalitarianism (especially with respect to informants and respondents)

3. Punch (1986) is quite helpful on the issue of "coming clean" with study documentation. The reader ought to know, he suggests, such things as:

The problems encountered getting into and getting out of the field

The micro-politics of the site and the role played by your study

In detail, the conflicts, ambiguities, and more sordid understandings gained during fieldwork

4. How can we judge the "goodness" of the report itself? The best discussions we have seen on this are by Lincoln and Guba (1990) and Zeller (1987). They suggest criteria grouped under several headings:

Resonance: fit between the report and the researcher's paradigm

Rhetorical (persuasive): structural unity, coherence, corroboration, simplicity, clarity, use of respondent language

Overall organization: flow of action, narrator role

Craftsmanship: (see Note 2)

Empowerment: consciousness-raising, increased sense of control, "actionability"

Applicability: vicarious linking to own life, illumination, transfer, reconsideration of own frame

13

Concluding Remarks

Seeing Qualitative Analysis Whole

Chapter by chapter, we have looked at a sequence of issues in qualitative data analysis—from initial study design and conceptualization to data collection, coding, innumerable types of displays, and on to conclusion drawing and reporting. Studying these features close up, it's easy to lose sight of the overall ecosystem they form.

Figure 13.1, adapted from training workshops we have led, pulls together all of the aspects of the approach to qualitative data analysis we have explored in the book. The aim is to provide an integrated overview. As with causal networks, it helps to walk through the figure, piece by piece. The time dimension, for once, goes vertically.

Starting at the top, we see the early mutual influence of conceptual framework and research questions. Both lead to plans for sampling (within and across cases) and for instrumentation. Once the sampling plan is clear, access to cases starts and data collection begins.

During this time it's a good idea to build a well-defined data management system, which will be used—with evolutionary changes—throughout the project. We also advocate building a documentation scheme for regular tracking of project data analysis activities.

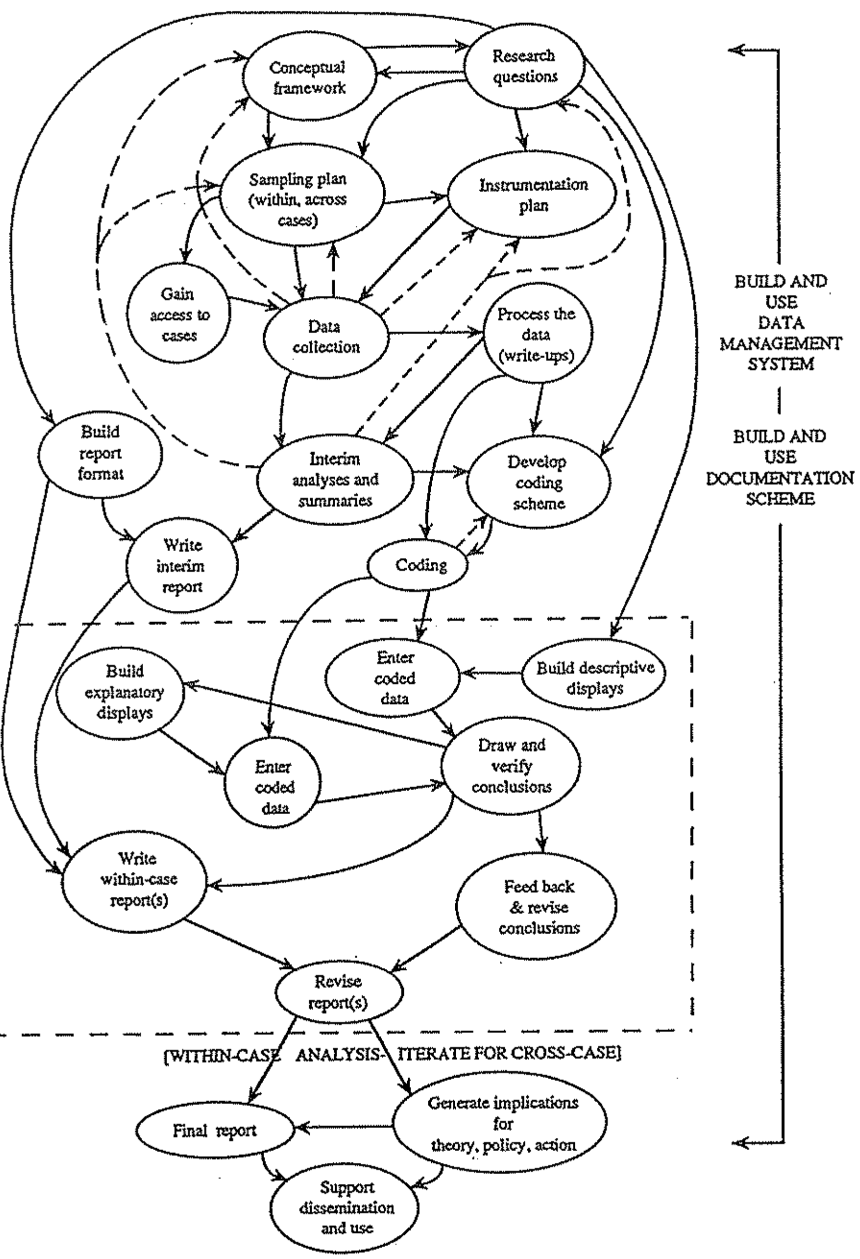
The first experiences of data collection nearly always have back effects (dashed lines), which induce reconsideration not only of sampling and instrumentation but also of the conceptual framework itself. Initial processing of the data often leads to interim summaries of various kinds, which also have effects back to the sampling and instrumentation approaches being taken, as well as, more conceptually, to the framework and the research questions.

The initial coding scheme is usually influenced by the research questions, but it develops and iterates steadily as further coding is carried out. Interim analyses also make a clear contribution here. If interim reports are contemplated, a format is designed at this point and the interim summaries fed into it.

Assuming a body of coded data (not necessarily complete or full), the next move is into within-case analysis. The general strategy is to build, first, *descriptive* display formats designed to answer the research questions. Coded data are entered and, from them, conclusions of a descriptive sort are drawn and verified. With those conclusions in hand, *explanatory* displays can be built, coded data entered, and explanatory conclusions drawn.

Both descriptive and explanatory conclusions can be fed back for verification and revision. By now, a format for within-site reporting, based as always on the research questions, as well as on the interim results, has been prepared. Within-case reports are written, and then revised on the basis of respondent feedback.

Figure 13.1
Overview of Qualitative Data Analysis Processes



If more than one case is included, this entire cycle of within-case analysis is repeated to draw cross-case conclusions. These conclusions and the within-case results go into the final report, which will include implications for theory, policy, and action. Depending on the study's audiences and intentions, some degree of support for dissemination and use of the findings will be provided.

Any particular project will always loop among the nodes of Figure 13.1 in idiosyncratic ways. The issue is to be aware of where you are in the flow right now, and where you want to be by X time.

Reflections

Writing this book has been a long, sometimes arduous, and always fascinating journey. We have encountered many new and thoughtful treatments of the core issues of qualitative data analysis. The experience has reconfirmed the old saw that trying to teach something deepens your understanding of it. The process of clarifying, reformulating, and synthesizing the ideas in this second edition has extended and enriched our thinking about qualitative data analysis methods. We are grateful to the many colleagues who helped us advance.

Throughout the book, we have aimed to stay practical, close to the reader's elbow, talking aloud as we went, offering both variations and advice. Some methodological texts tend toward the abstract, with brief examples that always seem to work out clearly, even effortlessly. Yet when you actually come to grips with collecting and analyzing real-life data, things seldom work out that way. Research-in-use is almost always more intractable, disjointed, and perverse than research-in-theory, and we have tried to take careful account of that fact. In short, *doing* qualitative analysis is the way you get better at it—and we believe that holds true not just for novices and new entrants to qualitative work, but for wise old dogs as well.

We believe that methodological quagmires, mazes, and dead ends are not necessarily a product of researcher incapacity, but of qualitative data themselves. Like the phenomena they mirror, they are usually complex, ambiguous, and sometimes downright contradictory. Doing qualitative analysis means living for as long as possible with that complexity and ambiguity, coming to terms with it, and passing on your conclusions to the reader in a form that clarifies and deepens understanding. It's not surprising that the mechanics of analysis seem formidable or elusive even to experienced analysts—and that researchers have often shied away from making them fully explicit.

It's right to say that qualitative data analysis is a craft—one that carries its own disciplines. There are *many* ways of getting analyses "right"—precise, trustworthy, compelling, credible—and they cannot be wholly predicted in advance.

But we are not alone in experiencing such indeterminacy. The same holds true for experimental and correlational researchers. They use a variety of designs and instruments to study the same problems. They lose subjects from their samples. Their instrumentation behaves strangely. The data from some cases are equivocal, sometimes unfathomable. They make intuitive leaps when looking at data outputs. Quantitative researchers get better at their work by learning how to contend with these normal problems, not by seeking for an ideal design and a fully spelled-out conceptualization. They are not Platonists, but analytic pragmatists, just like qualitative researchers.

We've found that making the steps of analysis explicit makes them less formidable and uncertain, and more manageable. You don't need prolonged socialization or arcane technologies. The core requisites for qualitative analysis seem to be a little creativity, systematic doggedness, some good conceptual sensibilities, and cognitive flexibility—the capacity to rapidly undo your way of construing or transforming the data and to try another, more promising tack. (Those, and a little help from your friends. . . .) None of these qualities is contingent on a battery of advanced "methods courses." In revising this book, we have continued our hope from the first edition: to create a self-help book for researchers, rather than coining a new, only remotely accessible set of methods.

We also don't think that good qualitative analysis necessarily calls for formal prerequisites (e.g., long ethnographic experience, knowledge of scientific logic, deep background in epistemology). Those are all useful tools and can empower the analyses you are interested in doing. But working without them, at least initially, is not fatal—and can help you see what else you need to know to do good analysis.

Qualitative researchers come in many varieties and flavors. The views we have expressed will almost certainly be seen as narrow-minded, even arrogant, by some of our colleagues. We don't want to be seen as pretending to be ideology-free, or that our biases are noblest, or that we are just being plain-folks pragmatic and practical without understanding the issues. But these intentions may not always work out.

To those who believe that analysis is an intuitive, nearly incommunicable act, we have insisted that analyses can be workably replicated, and that to be taken seriously you should be fully explicit about what is being done each step of the way.

To those who believe that serious explanation must involve converting words into numbers and manipulating the numbers according to conventional statistical canons, we have said that better, more powerful methods of data analysis are available that illuminate the web of local causality.

To those enamored of long narrative accounts as the sole route to understanding, we have counterposed the idea of focused, organized displays that permit systematic analyses and enhance confidence in findings.

To those who believe that qualitative data analysis requires years of training and apprenticeship, we have offered an expanded set of working methods that can be easily learned, tried on for size, and developed further for particular projects—all in the service of a stronger methodology.

To those colleagues who are pleased with what we have done in this second edition, we express our return pleasure, along with our hope for skeptical testing and revision of these methods, so that we all can advance our craft—or sullen art.

Advice

Throughout the book, we've provided detailed advice on a method-by-method basis. Here we offer some generalized last words of encouragement to our colleagues, of whatever persuasion.

Think display. Given a research question or a puzzling issue in a qualitative database, consider what forms of display—matrices and networks—are most likely to bring together relevant, transformed data in a way that will permit good conclusion drawing—and strengthening of those conclusions.

Be open to invention. The wide range of useful displays we and many others have created reinforces our belief that the universe of useful displays is very large and, like other universes, constantly expanding.

Expect iteration. The mode of analysis we've advocated throughout the book involves shuttling among data reduction, display, and preliminary and verified conclusions. New data enter the picture, new display forms evolve, conclusions get bent and revised. All of these will have back effects on each other, effects that are crucial to the evolving analysis.

Seek formalization, and distrust it. We have steadily emphasized a structured approach to drawing meaning from qualitative data. Becoming more systematic, whatever your epistemological position, strikes us as a priority for those who wish to advance analysis methodology. Of course increased formalization carries its own risks: narrowness, overconfidence, obsessiveness, blindness to the emergent—and the risk of orthodoxy. The field of qualitative data analysis did not, in 1984, need a narrow set of canons to strike fear into the hearts of graduate students and inspire endless casuistry and disputation—and it does not need them now. Patience, trying to do things better, and sharing are more like it.

Entertain mixed models. We have sought to make a virtue of avoiding polarization, polemics, and life at the extremes. Quantitative and qualitative inquiry can support and inform each other. Narratives and variable-driven analyses need to interpenetrate and inform each other. Realists, idealists, and critical theorists can do better by incorporating other ideas than by remaining pure. Think of it as hybrid vigor.

Stay self-aware. Our own experience showed us vividly how useful it is to maintain a part of your attention on the processes involved in analysis—from the selection of research questions through coding, the creation of displays, data entry, conclusion drawing, and verification. Only through such sustained awareness can regular self-correction occur—not just during specific analysis episodes, but over time, as the methods themselves iterate and develop. We have suggested supports for self-awareness in the form of documentation logs and—perhaps most essentially—"critical friends" who can supportively counter your taken-for-granted approaches and suggest alternatives.

Share methodological learnings. The methodological sections of most reports of qualitative studies are still thin. Articles focusing directly on analysis issues and approaches are all too rare. We believe that anyone who wants to advance the craft of qualitative analysis owes it to colleagues to communicate what they have learned to others. We advise stronger methodological emphasis in articles and books drawn from qualitative data, and encourage reports of training methods in courses and workshops that have successfully expanded analysis skills. We continue to urge much parallel and divergent effort in order to develop, gradually, a stronger, clearer consensus on how to draw valid conclusions from qualitative data.

In sum, we hope that more and more qualitative researchers will tell each other, concretely and specifically, just how they went about it, and what they learned. Perhaps we can all be as vivid and rich in describing our own work as we are in describing the inner and outer lives of the people we are studying. We owe them, and ourselves, at least that much, and probably a great deal more.

Appendix

CHOOSING COMPUTER PROGRAMS FOR QUALITATIVE DATA ANALYSIS

Matthew B. Miles
Eben A. Weitzman

1. Overview and Suggestions for Use

This Appendix begins (section 2) with a review of software types and functions, explaining what programs can do for the qualitative researcher. Then (section 3) comes advice on how to choose software on the basis of your computer sophistication, what your data are like, and the analyses you are planning to do.

In sections 4 and 5, we provide a matrix showing the general characteristics of each of 22 programs, followed by a list of the software developers and distributors involved with the programs; they can be contacted directly.

An overview like this cannot provide all of the information you need for deciding which software to use. For detailed reviews of these programs, see Weitzman and Miles (1994) and Tesch (1990). We also provide, in section 6, a brief annotated bibliography of recent work in this domain, with emphasis on specific programs.

Qualitative analysis software is a fast-moving field. Expect the information here to decay as programs are refined and new programs appear. But the suggestions on choosing software will still be useful to you. In any case, we believe that new programs will inevitably emerge that cover more aspects of qualitative data analysis in improved ways.

2. Software Types and Functions

General Types

We begin with some gross distinctions. Most programs are actually a blend or combination of these general types, so it does not pay to be purist here. You have to look at the specific functions of a program. But as a starter, we can name the following general types (see also Richards & Richards, 1994, for more detail). We begin with three types of "generic" programs, which were not necessarily developed with the needs of qualitative researchers in mind, then turn to three others that are qualitative-specific.

Word processors. These tools are basically designed for the production and revision of text, and thus are helpful for taking, transcribing, writing up, or editing field notes, for memoing, for preparing files for coding and analysis, and for writing report text. We do not list specific word processors here.

Word retrievers. Programs such as Metamorph, Sonar Professional, The Text Collector, and WordCruncher specialize in finding all of the instances of words, phrases, and combinations of these you are interested in locating, in one or several files. Some have content-analytic capabilities as well: counting, displaying words in their context, and creating word lists and concordances (organized lists of all words and phrases in their contexts).

Text base managers. These programs organize text more systematically for search and retrieval. Examples are askSam, FolioVIEWS, MAX, Orbis, and ZyINDEX. Some deal with highly structured text organized into "records" (specific cases) and "fields" (numerical or text information appearing for each case). Others easily manage "free-form" text. Some can include quantitative information. Essentially, they search for and retrieve various combinations of words, phrases, coded segments, memos, or other material (some are able to refer you to externally stored documents, pictures, or audio- or videotapes).

Code-and-retrieve programs. Software of this sort—often developed specifically by qualitative researchers—helps you divide text into segments or chunks, attach codes to the chunks, and find and display all instances of coded chunks (or combinations of coded chunks). Examples are ATLAS/ti, HyperQual, Kwalitan, MAX, NUDIST, QUALPRO, and The Ethnograph.

Theory builders. These programs, too, are often researcher-developed. They usually include code-and-retrieve capabilities, but also allow you to make connections between codes (categories of information); to develop higher-order classifications and categories; to formulate propositions or assertions, implying a conceptual structure that fits the data; and/or to test such propositions to determine whether they apply. They're often organized around a system of rules, or are based on formal logic. Examples are AQUAD, ATLAS/ti, HyperRESEARCH, NUDIST, and QCA.

Conceptual network builders. Software of this sort also helps you build and test theory, but you work with systematically-built graphic networks. You can see your variables as nodes, linked with other nodes by specified relationships (e.g., "belongs to," "leads to," "is a kind of"). The networks are not just casually hand-drawn, but are real "semantic networks" that develop from your data and your concepts (usually higher level codes) and the relationships you see among them. Examples are ATLAS/ti, MECA, and SemNet. Programs with strong network-drawing capabilities include Inspiration and MetaDesign.

Specific Functions: What to Look For

Let's turn to the specific things programs can do for you. Any given program incorporates a set of specific functions for doing qualitative analysis work, and may overlap several of the categories above. We sketch these briefly; they are indicated generally in the matrix in section 4. To make sensible choices, you will need to contact someone who uses the program, the developer or distributor, or read the specific descriptions in Weitzman and Miles (1994).

Coding. In this domain, programs can help you segment or "chunk" your data (e.g., into separate words, phrases, lines, sentences, paragraphs, free-form segments) and then attach codes or key words to the chunks. Some have one-step (on-screen) coding; others require you to work with hard copy printout.

Codes usually can be attached in several ways: one or several to a chunk, or on nesting or overlapping chunks. Many programs let you do hierarchical or multilevel coding (e.g., "rain" has successively higher-level codes of "precipitation," "weather," and "climate"), and some have features that make renaming and copying of codes easy (e.g., if you add "snow," you can tell the program to include it under the higher-level codes, and if you decide that lower-level codes of "powder snow" and "corn snow" need to be connected with a new code, "skiing conditions," you can easily revise and update your originally assigned codes).

Many programs also attach "source tags," so you can see later where a retrieved chunk came from (which interview, when, with whom, etc.). Some programs will show you your complete code list in list, hierarchical tree, or other network form.

Memoing/annotation. Qualitative analysis is aided when you can reflect on your data, mark it up, write memos about it, and so on. Some programs let you make marginal or side notes, and some let you write extended memos about the meaning of data or codes. Several let you apply codes to these "second-level" inputs. By implication: Some programs do not permit any of these functions.

Data linking. As you work on data, you are constantly seeing connections between different parts of the qualitative database. Some programs are good at making and recording links *within* parts, such as the field note text, the codes, annotations, memos, and "off-line" data (e.g., audiotapes or documents that are not actually in your database but that can be referred to when you need them). They also may make links *across* these parts of the database (e.g., linking codes to memos). Programs with strong linking functions, perhaps in a "hypertext" mode, let you "browse" or "navigate" easily through the database.

Search and retrieval. Most programs can look for material you want and show it to you, either alone or together with other instances of what you asked for. One issue is how fast this search takes place. Another is whether the program will find segments by code and/or can search for "strings" of characters (usually words or phrases occurring naturally in the text). Some programs will find not only the word or phrase you asked for, but all synonyms for it. Some have "wild-card" searching ("decisi*" will get you "decisive," "decision," "decision-making," etc.).

An important issue is how search requests are made. Many programs support "Boolean" requests (power AND conflict, power OR conflict, power AND NOT conflict, etc.). These requests can be strung together using algebraic parentheses to formulate very precise search requests. Some programs can look for items that are near each other in the text or precede or follow each other. Some can look just within a certain portion of the database (e.g., only female respondents, only in second-grade classrooms, classes taught by women).

More functions: When a chunk is retrieved (sometimes called a search "hit"), how do you see it? Is it highlighted? How much of the surrounding context do you see? Can you tell which person, time, or part of the database this came from? Can you go directly to the hit's location in the file? Some programs will send hits to another file so that you can see them all together.

An important function to look for here is whether the program makes a log or record of searches and retrievals as you go. That function enables you to build up a database that includes the ones that did something useful for you, along with annotations and memos made along the way.

Conceptual/theory development. As noted above, some programs help you clarify your ideas via rule-based or logical approaches, and some through building semantic networks. For network building, look for whether nodes simply are labeled, or can have text attached—and whether links are of a single, unspecified type or of multiple types (e.g., "is part of," "leads to," "follows," "is a kind of," "belongs to"), or of types you yourself specify. Some programs permit you to develop and test "if-then" propositions/hypotheses. Others allow "configurational" analysis across several cases to find case-specific patterns of predictors associated with an outcome, as described by Ragin (1993).

Data display. Most programs show you results on-screen, or will send them to a printer or to another file. Some work on an alternate or split screen, or in windows. Some programs can output text or numerical data to help you enter them in a matrix display; others will display networks. If the program does frequency counts, it may be able to send results directly to a statistical program.

Graphics editing. Some programs let you create and edit networks composed of "nodes" connected by "links." Nodes can be variables or codes or memos, or even have large amounts of text attached to them. The links can be one- or two-way and may also be labeled with titles, as noted above. Look for editing flexibility (e.g., when you move nodes, the links should move along with them) and a range of graphic styles.

Other Considerations

We've just summarized the main functions that programs can carry out. As you examine a particular program, some considerations also cut across all of these functions.

One is *flexibility.* Does the program do only what it is built for, or can you "work around" its features, doing other things you want to do without too much hassle? Must the program be used as is, off the shelf, or can you customize it, adapt it, write "macros," create new routines within it?

The second—a very fundamental one—is *user friendliness.* How hard is the program to learn? And, once you have learned the program, how easy is it to use? Will the time and effort required to learn and use the program be justified in terms of what you are getting? And how good are the supports for learning: the manual or other documentation, the tutorials, the help screens, and the technical support on the other end of the phone line? For both of these issues, ask other users, and see the descriptions in Weitzman and Miles (1994).

3. How to Choose Software: Key Questions

"What's the best program?" There is no answer in the abstract. Choosing the right software, for you, depends on your own level of work with computers, on the particular project you have in mind, and on the type of analysis you expect to do. A careful look at the issues below will help you know what to look for in more detail, in relation to specific programs.

What Kind of Computer User Are You?

If you are new to computers, your best bet is probably to choose a word processing program with advice from friends and to use it, learning your computer's operating system (MS-DOS or Mac) and getting comfortable with the idea of creating text, moving around in it, and revising it. That practice would bring you to what we call Level 1.

Or you may be acquainted with several programs, use your operating system easily, and be able to see your way clearly to exploring and learning new programs (Level 2).

Or you may have an active interest in the ins and outs of how programs work (Level 3) and feel easy with customization, writing macros, and so on. (We do not deal here with the "hacker," a Level 4 person who lives and breathes computing.)

If you are Level 1, give plenty of attention to the user friendliness aspects of the programs you are considering. Find a friend who is at Level 2, preferably one who is also a qualitative researcher, to help you interpret unclear features of the programs you are considering. Always connect the features to your project and analysis plans. Your friend can also support your choice of a program and initial tryout of it.

If your computer use is at Level 2, you should have little difficulty in using the detailed information in Weitzman and Miles (1994) to make some preliminary choices. Be sure it fits your project and its analysis needs, especially if you choose an ambitious or demanding program. Try to find friends who are already using the program you are interested in. Clarify any questions you have with a Level 2 or 3 friend—who is also a qualitative researcher. You may want to consult a Level 3 friend for assistance during installation and early use.

If you are a Level 3 computer user, you'll have no difficulty in assessing detailed program features. Do not let particularly interesting features seduce you away from the questions, What is my project like? and What sort of analysis am I planning? Stay with those questions.

In general, both during the choice of programs and looking forward to early use of programs, we strongly emphasize the importance of friends. They help you with the pros and cons. When you begin program use, trying to do it all alone, or from the manual and tutorial alone, is often difficult. Learning anything new is an interactive process. Friends supply support, tips, new ideas, and tricks of the trade quickly and easily. Friends don't make you feel dumb. Friends learn from each other.

Finally a nonminor question is whether you are an MS-DOS or Mac user. Although more than a few programs are available for either, it is often one or the other. People used to the intuitive, point-and-click, icon-grabbing, flexible Mac style get impatient with typing out commands. MS-DOS users like being able to issue a typed, precise command from anywhere in the program or its menus.

What Kind of Database and Project Is It?

The second general question is about your project and contemplated database. As you look at detailed features, you need to play them against these issues.

Data sources: Single versus multiple. You may be collecting data from many sources (say, your case is a student, and you talk with several teachers, the parents, friends, and the student herself). Look for programs that are good at making links, such as those that have hypertext capability and that attach "source tags" so you know where information is coming from.

Single versus multiple cases. If you have multiple cases, you may want to sort them according to different patterns or configurations, or to work with only some of the cases, or do cross-case comparisons. Look for software that will select portions of the database, and/or do configurational or other types of cross-case analysis.

Fixed records versus revised. Will you be working with data that are fixed (e.g., official documents, the plays of Shakespeare, survey responses) or data that will be revised (with added codes, commentary, memos, corrections, etc.)? Some programs make database revision easy; others are quite rigid: Revising can use up a lot of time and energy. Some will not let you revise data at all without starting over.

Structured versus open. Are your data strictly organized (e.g., responses to a standard questionnaire or interview) or free-form (running field notes, participant observation, etc.)? Highly organized data usually can be managed more easily, quickly, and powerfully in programs set up to accommodate them, with well-defined "fields" and "records."

Uniform versus diverse entries. Your data may all come from interviews. Or you may have information on the same person or organization of many sorts (documents, observations, questionnaires, pictures, audiotapes, videotapes). Some programs handle diverse data types easily; others are narrow and stern in their requirements. Look for good linking features—hypertext again.

Size of database. A program's database capacity may be expressed in terms of numbers of cases, files, size of files, and/or total kilobytes (K). (Roughly, consider that a single-spaced page of printed text is about 2K.) Estimate your total size and double it. Most programs are more than ample, but you should check.

What Kind of Analysis Is Anticipated?

Choice of software also depends on how you expect to go about analysis, generally speaking. This does not mean a detailed analysis plan, but a general sense of the style and approach you are expecting.

Exploratory versus confirmatory. Are you mainly planning to poke around in your data to see what they are like, or do you have some specific hypotheses in mind that you would like to check out? If the former, look for features of fast search and retrieval, easy coding and revision, along with good graphic display. If the latter, programs with strong theory-building features are better bets. Programs largely limited to word retrieving can only test limited hypotheses.

Coding scheme firm at start versus evolving. Does your study have a fairly well-defined a priori scheme for codes (categories, key words) that you will apply to your data? Or will such a scheme evolve as you go, in the style of the "constant comparative" method (Strauss & Corbin, 1990)? If the latter, look for on-screen and/or automated coding (rather than programs that require you to code on hard copy) for easy revision of codes. Hypertext link-making capabilities are helpful here, too.

Multiple versus single coding. Some programs let you assign several different codes to the same segment of text, including higher-order codes, and may let you overlap or nest coded chunks. A few are stern: one chunk, one code.

Iterative versus one pass. Do you have the time and the wish to keep walking through your data several times, taking different and revised cuts? Or will you limit yourself to one pass, for intellectual or resource reasons? If the former, look for programs that are flexible, invite repeated runs, and can make a log of your work as you go. (This step is related to the question of whether your records are fixed or revisable during analysis.)

Fineness of analysis. Will your analysis focus on specific words? Or lines of text? Or free-form segments? Or sentences? Paragraphs? Pages? Whole files? Look to see what the program permits (or requires or forbids) you to do. How flexible is it? Can you look at varying sizes of chunks in your data?

Interest in context of data. When the program pulls out chunks of text for you to look at, how much surrounding information do you want to have? Do you need only the word or phrase itself? Do you want the preceding and following lines/sentences/paragraphs? Do you want to be able to see the entire file? Do you want the information to be marked with a "source tag" that tells you where it came from (e.g., Interview 2 with Brenda Keeler, page 29, line 12)? Programs vary widely on this.

Intentions for displays. Analysis goes much better when you can see organized, compressed information in one place. Some programs produce displays in list form. Some can help you produce matrix displays; others can give you networks or hierarchical diagrams.

Qualitative only, or numbers included. If your data and/or your analyses include the possibility of number-crunching, check whether the program will count things, and/or whether it can send information to other programs specifically designed for quantitative analysis.

Final comment: A significant issue is whether you are choosing software just for this project, or for the next few years. Your word processor does not care what you are writing about, so most people pick one and stick with it until something better comes along *and* they feel motivated to learn it. But qualitative analysis programs tend to be good for certain types of analyses. Switching costs time and money. Think whether you should choose the best program for *this* project, or the program that best covers the kinds of projects you are considering over the next few years.

4. Program Characteristics

(See Table A.1, p. 316.)

5. Program Developers and Distributors

AQUAD: Günter Huber, University of Tübingen, Department of Pedagogical Psychology, Munzgasse 22-30, D-72070 Tübingen, Germany. Phone: 49-7071-292113. Fax: 49-7071-294954. E-mail: 100115.230@compuserve.com. *

askSam: P.O. Box 1428, 119 S. Washington St., Perry, FL 32347. Phone: (800) 800-1997; (904) 584-6590. Fax: (904) 584-7481. Tech support: (904) 584-6590.

ATLAS/ti: Thomas Muhr, Trautenaustr. 12, D10717 Berlin, Germany. Phone and Fax: 49- 30-861 1415. E-mail: muhr@cs.tu-berlin.de. *

FolioVIEWS: Folio Corporation, 2155 N. Freedom Blvd., Suite 150, Provo, UT 84604. Phone (800) 543-6546.

HyperQual: Raymond V. Padilla, 3327 N. Dakota, Chandler, AZ 85224. Phone (602) 892-9173. *

HyperRESEARCH: Researchware, Inc., 20 Soren St., Randolph, MA 02368-1945. Phone: (617) 961-3909.

Inspiration: Inspiration Software, Inc., 2920 S.W. Dolph Ct., Suite 3, Portland, OR 97219. Phone: (503) 245-9011.

Kwalitan: Vincent Peters, Department of Research Methodology, Social Sciences Faculty, University of Nijmegen, Th. van Acquinostraat 4, 6525 GD Nijmegen, The Netherlands. Phone 31-80-612038. Fax: 31-80-612351. E-mail: U211384@HNYKUN11.

MAX: Udo Kuckartz, Free University of Berlin, Institute for Social and Adult Education, Arnimallee 12, D-14195 Berlin, Germany. Phone: 49-30-838 5539. Fax: 49-30-838 5889. *

MECA: Kathleen Carley, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA 15568. Phone: (412) 268-3225. E-mail: Kathleen.Carley%CENTRO.SOAR.CS.CMU.EDU@Carnegie.Bitnet

Table A.1
Program Characteristics

	Version	Operating System	Coding	Search & Retrieval	Database Mgmt	Memoing	Data linking	Matrix building	Network display	Theory-building	User friendliness
AQUAD	[3.2]	D	■	⊖	□	■		⊖		■	✓✓
askSam	[5.1]	DW		●	■	□	■				✓✓
ATLAS/ti	[1.0e]	D	■	⊖	□	■	■	○	● [a]	■	✓✓✓
Folio Views	[3.0]	DW	■	●	■	■	■	○			✓✓✓
HyperQual[b]	[4.3]	Mc	■	○	■	■	■	○			✓✓
HyperRESEARCH	[1.5]	McW	■	⊖	□			○		■	✓✓✓
Inspiration	[4.0]	Mc		○		■	■		●		✓✓✓
Kwalitan	[3.1]	D	■	⊖	□	■	□			□	✓✓✓
MAX[b]	[3.x]	D	■	○	■		■				✓✓
MECA	[1.0]	McDU					■		⊖	■	✓✓
Meta Design	[4.0]	McW		○		□	■		●		✓✓
Metamorph[c]	[3.5]	McDUM		● [d]	□	□	■				✓
	[4.0]	McWU	□	● [d]	□	■	■				✓✓
NUDIST	[2.3]	McWUM	■	●	□		□	⊖		■	✓
	[3.0]	Mc	■	●	□	■	□	●	⊖	■	✓✓
Orbis[b,e]	[1.0]	DW	□	●	■	□	□	●			✓✓
QCA	[3.0]	D		○		□		⊖		■	✓
QUALPRO	[4.0]	D	■	⊖	□	□					✓✓✓
SemNet	[1.0.2]	Mc		⊖		■	■		●	■	✓✓
Sonar Professional	[8.4]	McW		●	□	■	□			□	✓✓
The Ethnograph	[4.0]	D	■	⊖	□	■		○			✓✓
The Text Collector	[1.7]	D		●	□						✓✓✓
WordCruncher	[4.5]	D		⊖	□		■				✓
	[Beta]	W		●	□	■	■				✓✓
ZyIndex	[5.0]	DW		●	□	■	■				✓

■ = Designed for this purpose, as we see it (May be more or less effective)
□ = Not really designed for it, but can do at least in a limited way
blank = Can't do this

● = Strong
⊖ = OK
○ = Weak
blank = absent

✓ = Not too friendly
✓✓ = Fairly friendly
✓✓✓ = Very friendly

Operating System: Mc = Macintosh; D = DOS; U = Unix; M = Main Frame; W = Windows

[a] Weaker than Inspiration and MetaDesign on graphics, but can create networks from data.
[b] Reviewed from documentation, so ratings are somewhat speculative.
[c] Based on version 3.5 and a demo of version 4.0, so ratings for 4.0 are somewhat speculative.
[d] For raw text retrieval, in a class by itself.
[e] Available as add-on module for XyWrite 4.0 or NotaBene 4.0 word processors. Ratings include word processor features.

MetaDesign: Meta Software Corporation, 125 Cambridge Park Dr., Cambridge, MA 02140. Phone: (617) 576-6920. Fax: (617) 661-2008.

Metamorph: Thunderstone Expansion Programs International, Inc., 11115 Edgewater Dr., Cleveland, OH 44102. Phone: (216) 631-8544. Fax: (216) 281-0828.

NUDIST: Tom and Lyn Richards, Qualitative Solutions and Research Pty Ltd., 2 Research Drive, La Trobe University, Melbourne, Vic. 3083, Australia. Phone: 61-3-479-1311. Fax: 61-3-479-4441. E-mail: nudist@latcs1.lat.oz.au.

Orbis: XYQuest, The Technology Group, Inc., 36 S. Charles St., Baltimore, MD 21201. Phone: (410) 576-2040. Fax: (410) 576-1968.

QCA: Kriss Drass and Charles Ragin, Center for Urban Affairs and Policy Research, Northwestern University, Evanston, IL 60208. Phone: (708) 491-8712. E-mail: kadrass@nevada.edu; cragin@nwu.edu.

QUALPRO: *Impulse* Development Company, 3491-11 Thomasville Rd., Suite 202, Tallahassee FL 32308, or Bernard Blackman, 2504 Debden Ct., Tallahassee, FL 32308-3035. Phone: (904) 668-9865. Fax: (904) 668-9866. *

SemNet: Dr. Joseph Faletti, SemNet Research Group, 1043 University Ave., San Diego, CA 92103. Phone: (619) 594-4453.

Sonar Professional: Virginia Systems, Inc., 5509 West Bay Ct., Midlothian, VA 23112. Phone: (804) 739-3200.

The Ethnograph: Qualis Research Associates, P.O. Box 2070, Amherst, MA 01004. Phone: (413) 256-8835. E-mail: Qualis@mcimail.com. *

The Text Collector: O'Neill Software, P.O. Box 26111, San Francisco, CA 94126. Phone: (415) 398-2255.

WordCruncher: Johnston & Co., 314 E. Carlyle Ave., Alpine, UT 84004. Phone: (801) 756-1111.

ZyINDEX: ZyLAB Corporation, 100 Lexington Dr., Buffalo Grove, IL 60089. Phone: (800) 544-6339; (708) 459-8000. Fax: (708) 459-8054.

*These programs are also distributed and supported by Qualitative Research Management, 73425 Hilltop Rd., Desert Hot Springs, CA 92240, Phone: (619) 329-7026. In addition, QRM provides consulting and training services for qualitative data analysis.

6. References

Fielding, N. G., & Lee, R. M. (1991). *Using computers in qualitative research.* London: Sage.

An edited collection focusing on underlying issues. Includes the developers of The Ethnograph and NUDIST.

Richards, T., & Richards, L. (1994). Using computers in qualitative analysis. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research.* Thousand Oaks, CA: Sage.

A thoughtful overview of types of programs and their underlying "architecture." Emphasis on analysis issues.

Tesch, R. (1990). *Qualitative research: Analysis types and software tools.* New York: Falmer.

Classifies types of qualitative research and describes six programs in detail (ETHNO, TAP, QUALPRO, The Ethnograph, TEXTBASE ALPHA, and HyperQual).

Tesch, R. (Ed.). (1991). Computers and qualitative data [Special issues, Parts 1 and 2]. *Qualitative Sociology, 14*(3 & 4).

An edited collection of leading-edge papers. Includes developers of HyperQual, ATLAS/ti, HyperRESEARCH, NUDIST, and AQUAD.

Weitzman, E., & Miles, M. B. (1994). *Computer programs for qualitative data analysis.* Thousand Oaks, CA: Sage.

Detailed reviews of 22 programs (those in this Appendix).

References

Abbott, A. (1990). A primer on sequence methods. *Organization Science, 1*(4), 375-392.

Abbott, A. (1992a). From causes to events: Notes on narrative positivism. *Sociological Methods and Research, 20*(4), 428-455.

Abbott, A. (1992b). What do cases do? Some notes on activity in sociological analysis. In C. Ragin & H. Becker (Eds.), *What is a case? Exploring the foundations of social inquiry* (pp. 53-82). New York: Cambridge University Press.

Abramson, P. R. (1992). *A case for case studies: An immigrant's journal.* Newbury Park, CA: Sage.

Adams, R., & Preiss, J. (Eds.). (1960). *Human organization research.* Homewood, IL: Dorsey.

Adler, T. (1991, December). Outright fraud rare, but not poor science. *APA Monitor,* p. 11.

Agar, M. H. (1980). *The professional stranger: An informal introduction to ethnography.* New York: Academic Press.

Agar, M. H. (1986). *Speaking of ethnography* (Qualitative Research Methods Series, Vol. 2). Beverly Hills, CA: Sage.

Akeroyd, A. V. (1991). Personal information and qualitative research data: Some practical and ethical problems arising from data protection legislation. In N. G. Fielding & R. M. Lee (Eds.), *Using computers in qualitative research* (pp. 88-106). London: Sage.

Akmajian, A. (1990). *Linguistics: An introduction to language and communication.* Cambridge: MIT Press.

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis* (Quantitative Applications in the Social Sciences Series, No. 44). Beverly Hills, CA: Sage.

Allison, P. D. (1984). *Event history analysis* (Quantitative Applications in the Social Sciences Series, No. 46). Beverly Hills, CA: Sage.

American Educational Research Association. (1992). Ethical standards of the American Educational Research Association. *Educational Researcher, 21*(7), 23-26.

American Psychological Association (APA). (1992). Ethical principles of psychologists and code of conduct. *American Psychologist, 47*(12), 1597-1628.

American Sociological Association. (1989). *Code of ethics.* Washington, DC: Author.

Antaki, C. (Ed.). (1988). *Analysing everyday explanation: A casebook of methods.* London: Sage.

Argyris, C., Putnam, R., & Smith, D. M. (1985). *Action science: Concepts, methods, and skills for research and intervention.* San Francisco: Jossey-Bass.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership, and men* (pp. 177-190). Pittsburgh: Carnegie Press.

Asher, H. B. (1983). *Causal modeling* (2nd ed.). Beverly Hills, CA: Sage.

Atkinson, P. (1991). Supervising the text. *Qualitative Studies in Education, 4*(2), 161-174.

Atkinson, P. (1992). *Understanding ethnographic texts* (Qualitative Research Methods Series, Vol. 25). Newbury Park, CA: Sage.

Axelrod, R. (Ed.). (1976). *Structure of decision: The cognitive maps of political elites.* Princeton, NJ: Princeton University Press.

Bailey, K. D. (1982). *Methods of social research* (2nd ed.). New York: Free Press.

Ball, M. S., & Smith, G. W. H. (1992). *Analyzing visual data* (Qualitative Research Methods Series, Vol. 24). Newbury Park, CA: Sage.

Barker, R. G., & Wright, H. F. (1971). *Midwest and its children: The psychological ecology of an American town.* Hamden, CT: Archon.

Barone, T. (1990). Using the narrative text as an occasion for conspiracy. In E. W. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 305-326). New York: Teachers College Press.

Bartlett, L. (1990). *The dialectic between theory and method in critical interpretive research.* St. Lucia, Queensland, Australia: University of Queensland, Department of Education.

Becker, H. S. (1958). Problems of inference and proof in participant observation. *American Sociological Review, 23,* 652-660.

Becker, H. S. (1970). *Sociological work.* Chicago: Aldine.

Becker, H. S. (1978). Do photographs tell the truth? *After Image, 5,* 9-13.

Becker, H. S. (1986). *Writing for social scientists: How to finish your thesis, book, or article.* Chicago: University of Chicago Press.

Becker, H. S. (1990). Generalizing from case studies. In E. W. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 233-242). New York: Teachers College Press.

Becker, H. S., Geer, B., Hughes, E. C., & Strauss, A. L. (1961). *Boys in white.* Chicago: University of Chicago Press.

Becker, H. S., Gordon, A. C., & LeBailly, R. K. (1984). Field work with the computer: Criteria for assessing systems [Entire issues]. *Qualitative Sociology, 7*(1-2).

Berelson, B. (1971). *Content analysis in communication research.* New York: Hafner.

Berg, B. L. (1989). *Qualitative research methods for the social sciences.* Boston: Allyn & Bacon.

Bergman, R. G. (1989). Irreproducibility in the scientific literature: How often do scientists tell the whole truth and nothing but the truth? *Ethical and Policy Perspectives on the Professions, 8*(2), 2-3.

Bermant, G., & Warwick, D. P. (1978). The ethics of social intervention: Power, freedom, and accountability. In G. Bermant, H. C. Kelman, & D. P. Warwick (Eds.), *The ethics of social intervention* (pp. 377-418). Washington: Hemisphere.

Bernard, H. R. (1988). *Research methods in cultural anthropology.* Newbury Park, CA: Sage.

Bernard, H. R. (1991). About text management and computers. *Cultural Anthropology Methods Newsletter, 3*(1), 1-4, 7, 12.

Berry, W. D. (1984). *Nonrecursive causal models* (Quantitative Applications in the Social Sciences Series, No. 37). Beverly Hills, CA: Sage.

Bhaskar, R. (1978). *A realist theory of science.* Leeds: Leeds Books.

Bhaskar, R. (1989). *Reclaiming reality: A critical introduction to contemporary philosophy.* London: Verso.

Blalock, H. M. (1964). Causal inferences in nonexperimental research. Chapel Hill: University of North Carolina Press.

Blalock, H. M. (1971). *Causal models in the social sciences.* Chicago: Aldine.

Bliss, J., Monk, M., & Ogborn, J. (1983). *Qualitative data analysis for educational research: A guide to uses of systemic networks.* London: Croom Helm.

Blumer, H. (1962). Society as symbolic interaction. In A. M. Rose (Ed.), *Human behavior and social processes* (pp. 179-192). Boston: Houghton Mifflin.

Blumer, H. (1969). *Symbolic interactionism: Perspective and method.* Englewood Cliffs, NJ: Prentice-Hall.

Bogdan, R., & Biklen, S. K. (1992). *Qualitative research for education: An introduction to theory and methods* (2nd ed.). Boston: Allyn & Bacon.

Bogdan, R., & Easterday, L. (n. d.). *What's in good field notes?* Syracuse, NY: Syracuse University, Center on Human Policy.

Bogdan, R., & Taylor, S. J. (1975). *Introduction to qualitative research methods.* New York: John Wiley.

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models.* Newbury Park, CA: Sage.

Borman, K. M. (1985, April). *What is qualitative analysis, and why it doesn't work.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Borman, K. M. (1991). *The first "real" job: A study of young workers.* Albany, NY: State University of New York Press.

Brandt, R. M. (1981). *Studying behavior in natural settings.* Lanham, MD: University Press of America.

Brewer, J., & Hunter, A. (1989). *Multimethod research: A synthesis of styles* (Sage Library of Social Research, Vol. 175). Newbury Park, CA: Sage.

Bronfenbrenner, U. (1976). The experimental ecology of education. *Teachers College Record, 78*(2), 157-178.

Brown, M. J. M. (1991). *Validity and the problem of reality: An issue of trust.* Paper presented at the American Educational Research Association Annual Meeting, Chicago.

Bruner, E. M. (1984). Introduction: The opening up of anthropology. In S. Plattner, & E. M. Bruner (Eds.), *Text, play, and story: The construction and reconstruction of self and society* (pp. 1-16). Washington: American Ethnological Society.

Bruyn, S. (1966). *Human perspective in sociology.* Englewood Cliffs, NJ: Prentice-Hall.

Bryman, A. (1988). *Quantity and quality in social research.* London: Unwin Hyman.

Bulmer, M. (1979). Concepts in the analysis of qualitative data. *Sociological Review, 27*(4), 651-677.

Burgess, R. G. (Ed.). (1989). *The ethics of educational research.* New York: Falmer.

Campbell, D. T. (1966). Pattern matching as an essential in distal knowing. In K. Hammond (Ed.), *The psychology of Egon Brunswik* (pp. 81-106). New York: Holt, Rinehart.

Campbell, D. T. (1974, August-September). *Quantitative knowing in action research.* Kurt Lewin Award Address, Society for the Psychological Study of Social Issues. Presented at the Annual Meeting of the American Psychological Association, New Orleans.

Campbell, D. T. (1975). Degrees of freedom and the case study. *Comparative Political Studies, 8,* 178-193.

Campbell, D. T. (1979). "Degrees of freedom" and the case study. In T. D. Cook & C. S. Reichardt (Eds.), *Qualitative and quantitative methods in evaluation research* (pp. 49-67). Beverly Hills, CA: Sage.

Campbell, D. T. (1986). Relabeling internal and external validity for applied social scientists. In W. M. K. Trochim (Ed.), *Advances in quasi-experimental design and analysis* (pp. 67-77) (New Directions for Program Evaluation, No. 31). San Francisco: Jossey-Bass.

Campbell, D. T., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Caracelli, V. J., & Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 15*(2), 195-208.

Carley, K. (1990). Content analysis. In R. E. Asher et al. (Eds.), *The encyclopedia of language and linguistics.* Elmsford, NY: Pergamon.

Carley, K. (1993). Coding choices for textual analysis: A comparison of content analysis and map analysis. *Sociological Methodology, 23,* 75-126.

Carney, T. F. (1990). *Collaborative inquiry methodology.* Windsor, Ontario, Canada: University of Windsor, Division for Instructional Development.

Carspecken, P. F., & Apple, M. (1992). Critical qualitative research: Theory, methodology, and practice. In M. LeCompte, W. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education* (pp. 507-554). New York: Academic Press.

Cassell, J., & Jacobs, S. E. (1987). *Handbook on ethical issues in anthropology.* Washington, DC: American Anthropological Association.

Casti, J. L. (1992). *Searching for certainty: What scientists can know about the future.* New York: Morrow.

Champagne, A. B., Klopfer, L. E., Desena, A. T., & Squires, D. A. (1981). Structural representation of students' knowledge before and after science instruction. *Journal of Research in Science Teaching, 18*(2), 97-111.

Chenail, R. J. (1991). *Medical discourse and systematic frames of comprehension* (Advances in Discourse Processes, Vol. XLII). Norwood, NJ: Ablex.

Chesler, M. (1987). *Professionals' views of the "dangers" of self-help groups* (CRSO Paper 345). Ann Arbor, MI: Center for Research on Social Organization.

Christensen-Szalanski, J. J. J., & Beach, L. R. (1984). The citation bias: Fad and fashion in the judgment and decision literature. *American Psychologist, 39,* 75-78.

Clark, C. (1990). What you can learn from applesauce: A case of qualitative inquiry in use. In E. W. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 327-338). New York: Teachers College Press.

Clauset, K. H., Jr. (1989, March). *The dynamics of project implementation.* Paper presented at the Annual Meeting of the Comparative and International Education Society, Harvard Graduate School of Education, Cambridge, MA.

Clauset, K. H., Jr., & Gaynor, A. K. (1992). Educational policy analysis: An analysis of effective and ineffective schools using the system dynamics method. In R. L. Levine & H. E. Fitzgerald (Eds.), *Analysis of dynamic psychological systems: Vol. 2. Methods and applications* (pp. 303-323). New York: Plenum.

Cleveland, W. S. (1985). *The elements of graphing data.* Belmont, CA: Wadsworth.

Clifford, G. J. (1988). The historical recovery of Edyth Astrid Ferris. *Educational Researcher, 17,* 4-7.

Clough, P. T. (1992). *The end(s) of ethnography: From realism to social criticism.* Newbury Park, CA: Sage.

Connelly, F. M., & Clandinin, D. J. (1990). Stories of experience and narrative inquiry. *Educational Researcher, 19*(4), 2-14.

Constas, M. A. (1992). Qualitative analysis as a public event: The documentation of category development procedures. *American Educational Research Journal, 29*(2), 253-266.

Corbett, D., & Wilson, B. L. (1991). *Testing, reform, and rebellion.* Norwood, NJ: Ablex.

Counelis, J. S. (1991). Review of B. L. Berg, *Qualitative research methods for the social sciences. Qualitative Studies in Education, 4*(3), 267-269.

Crandall, D. P., & Associates. (1983). *People, policies, and practices: Examining the chain of school improvement* (Vols. I-X). Andover, MA: Network.

Crane, J. G., & Angrosino, M. V. (1974). *Field projects in anthropology: A student handbook.* Morristown, NJ: General Learning.

Cressey, D. R. (1953). *Other people's money: A study in the social psychology of embezzlement.* New York: Free Press.

Cusick, P. (1985a). Review of *Reading, writing, and resistance. Anthropology and Education Quarterly, 16,* 69-72.

Cusick, P. (1985b). Comment on the Everhart/Cusick reviews. *Anthropology and Education Quarterly, 16,* 246-247.

Dalin, P., Ayano, T., Biazen, A., Jahan, M., Miles, M. B., & Rojas, C. (1992). *How schools improve: International report.* Oslo, Norway: IMTEC.

Davis, F. (1959). The cabdriver and his fare: Facets of a fleeting relationship. *American Journal of Sociology, 65,* 158-165.

Davis, J. A. (1985). *The logic of causal order* (Quantitative Applications in the Social Sciences Series, Vol. 55). Beverly Hills, CA: Sage.

Dawes, R. (1971). A case study of graduate admissions: Applications of three principles of human decision-making. *American Psychologist, 26,* 180-188.

Dawson, J. A. (1979, April). *Validity in qualitative inquiry.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Dawson, J. A. (1982, March). *Qualitative research findings: What do we do to improve and estimate their validity?* Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Dean, J., Eichorn, R., & Dean, L. (1967). Observation and interviewing. In J. T. Doby (Ed.), *An introduction to social research* (pp. 284-286). New York: Meredith.

Degener, D. (Ed.). (1983). Improving school improvement: New study shows that most schools improve under SIP [Entire issue]. *Research and Educational Practice in the Far West,* February.

Denzin, N. K. (1978). *Sociological methods: A source book* (2nd ed.). New York: McGraw-Hill.

Denzin, N. K. (1983). Interpretive interactionism. In G. Morgan (Ed.), *Beyond method: Strategies for social research* (pp. 129-146). Beverly Hills, CA: Sage.

Denzin, N. K. (1989a). *Interpretive biography* (Qualitative Research Methods Series, Vol. 17). Newbury Park, CA: Sage.

Denzin, N. K. (1989b). *Interpretive interactionism* (Applied Social Research Methods Series, Vol. 16). Newbury Park, CA: Sage.

Denzin, N., & Lincoln, Y. S. (Eds.). (1994). *Handbook of qualitative research.* Thousand Oaks, CA: Sage.

Deyhle, D. L., Hess, G. A., Jr., & LeCompte, M. D. (1992). Approaching ethical issues for qualitative researchers in education. In M. D. LeCompte, W. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education* (pp. 597-642). New York: Academic Press.

Dilthey, W. (1911/1977). *Descriptive psychology and historical understanding.* Translated by R. M. Zaner & K. L. Heiges. The Hague, Netherlands: Nijhoff.

Donmoyer, R. (1990). Generalizability and the single case study. In E. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 175-200). New York: Teachers College Press.

Douglas, J. (1976). *Investigative social research.* Beverly Hills, CA: Sage.

Douvanis, C. J., & Brown, J. A. (1992, March). *Confidential and privileged communications: Legal and ethical concepts in research.* Paper presented at the Annual Conference of the Eastern Educational Research Association, Hilton Head, SC.

Draper, S. W. (1988). What's going on in everyday explanation? In C. Antaki (Ed.), *Analyzing everyday explanation: A casebook of methods* (pp. 15-31). Newbury Park, CA: Sage.

Dreitzel, H. P. (1970). Introduction: Patterns of communicative behavior. In H. Dreitzel (Ed.), *Recent sociology* (No. 2, pp. vii-xxii). London: Macmillan.

Duncker, K. (1945). On problem-solving [Entire issue]. *Psychological Monographs, 58*(5).

Eco, U. (1982). *Le nom de la rose.* Paris: Ed. Grasset.

Eden, C., Jones, S., & Sims, D. (1983). *Messing about in problems: An informal structured approach to their identification and management.* Elmsford, NY: Pergamon.

Edwards, W. (1968). Conservatism in human information processing. In K. B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17-52). New York: John Wiley.

Eisenhardt, K. M. (1989a). Building theories from case study research. *Academy of Management Review, 14*(4), 532-550.

Eisenhardt, K. M. (1989b). Making fast strategic decisions in high-velocity environments. *Academy of Management Journal, 32*(3), 543-576.

Eisenhart, M. A., & Howe, K. R. (1992). Validity in educational research. In M. D. LeCompte, W. L. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education* (pp. 643-680). New York: Academic Press.

Eisner, E. W. (1991). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice.* New York: Macmillan.

Ely, M., Anzul, M., Friedman, T., Garner, D., & Steinmetz, A. M. (1991). *Doing qualitative research: Circles within circles.* London: Falmer.

Erickson, F. (1977). Some approaches to inquiry in school-community ethnography. *Anthropology and Education Quarterly, 8*(2), 58-69.

Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119-161). New York: Macmillan.

Erickson, F., & Wilson, J. (1982). *Sights and sounds of life in schools: A resource guide to film and videotape for research and education.* East Lansing: Michigan State University, College of Education, Institute for Research on Teaching.

Everhart, R. (1985a). Review of *The egalitarian ideal and the American high school. Anthropology and Education Quarterly, 16,* 73-77.

Everhart, R. (1985b). Comment on the Everhart/Cusick reviews. *Anthropology and Education Quarterly, 16,* 247-248.

Faulconer, J. E., & Williams, R. N. (1985). Temporality in human action: An alternative to positivism and historicism. *American Psychologist, 40*(11), 1179-1188.

Faust, D. (1982). A needed component in prescriptions for science: Empirical knowledge of human cognitive limitations. *Knowledge: Creation, Diffusion, Utilization, 3,* 555-570.

Festinger, L. (1957). *A theory of cognitive dissonance.* Evanston, IL: Row, Peterson.

Fetterman, D. M. (Ed.). (1984). *Ethnography in educational evaluation.* Beverly Hills, CA: Sage.

Fetterman, D. M. (1989). *Ethnography step by step* (Applied Social Research Methods Series, Vol. 17). Newbury Park, CA: Sage.

Fielding, N. G., & Fielding, J. L. (1986). *Linking data: The articulation of qualitative and quantitative methods in social research* (Qualitative Research Methods Series, Vol. 4). Beverly Hills, CA: Sage.

Fielding, N. G., & Lee, R. M. (1991). *Using computers in qualitative research.* London: Sage.

Firestone, W. A. (1987). Meaning in method: The rhetoric of quantitative and qualitative research. *Educational Researcher, 16*(7), 16-21.

Firestone, W. A. (1990). Accommodation: Toward a paradigm-praxis dialectic. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 105-124). Newbury Park, CA: Sage.

Firestone, W. A. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher, 22*(4), 16-23.

Firestone, W. A., & Herriott, R. E. (1983). The formalization of qualitative research: An adaptation of "soft" science to the policy world. *Evaluation Review, 7,* 437-466.

Fischer, C., & Wertz, F. (1975). Empirical phenomenological analyses of being criminally victimized. In A. Giorgi (Ed.), *Phenomenology and psychological research* (pp. 135-158). Pittsburgh: Duquesne University Press.

Flinders, D. J. (1992). In search of ethical guidance: Constructing a basis for dialogue. *Qualitative Studies in Education, 5*(2), 101-116.

Fornell, C. (Ed.). (1982). *A second generation of multivariate analysis.* New York: Praeger.

Forrester, J. W. (1973). *Principles of systems.* Cambridge, MA: Wright-Allen.

Forrester, J. W. (1990). System dynamics as a foundation for pre-college education. In D. F. Anderson, G. P. Richardson, & J. D. Sterman (Eds.), *System Dynamics '90, 1* (pp. 367-380). Chestnut Hill, MA: System Dynamics Society.

Fredericks, M., & Miller, S. (1988). Some notes on confirming hypotheses in qualitative research: An application. *Social Epistemology, 2*(4), 345-352.

Freedland, K. F., & Carney, R. M. (1992). Data management and accountability in behavioral and biomedical research. *American Psychologist, 47*(5), 640-645.

Freeman, D. (1983). *Margaret Mead and Samoa: The making and unmaking of an anthropological myth.* Cambridge, MA: Harvard University Press.

Freeman, J. W., Klein, G., Riedl, T., & Musa, J. (1991, April). *A knowledge elicitation technique for educational development: The critical decision method.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Frick, T. W. (1990). Analysis of patterns in time: A method of recording and quantifying temporal relations in education. *American Educational Research Journal, 27*(1), 180-204.

Fuhrman, R., & Wyer, R. (1988). Event memory. *Journal of Personality and Social Psychology, 54*(3), 365-384.

Fullan, M. (1982). *The meaning of educational change.* New York: Teachers College Press.

Fullan, M., & Steigelbauer, S. (1991). *The new meaning of educational change.* New York: Teachers College Press.

Furman, G. C. (1990, April). *Abduction or induction? The semantics of data analysis in qualitative research.* Paper presented at the Annual Meeting of the American Educational Research Association, Boston.

Garfinkel, A. (1981). *Forms of explanation: Rethinking the questions in social theory.* New Haven, CT: Yale University Press.

Garfinkel, H. (1967). *Studies in ethnomethodology.* Englewood Cliffs, NJ: Prentice-Hall.

Gaynor, A. K., & Clauset, K. H., Jr. (1985, July). *Implementing and institutionalizing school improvement programs: A theoretical reformulation of the work of Huberman and Miles.* Paper presented at the International Conference of the Systems Dynamics Society, Keystone, CO.

Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In C. Geertz, *The interpretation of cultures* (pp. 3-30). New York: Basic Books.

Geertz, C. (1980). Blurred genres: The refiguration of social thought. *American Scholar, 49,* 165-179.

Geertz, C. (1983). *Local knowledge: Further essays in interpretive anthropology.* New York: Basic Books.

Gertner, D., & Grudin, J. (1985). The evolution of mental metaphors in psychology: A 90-year retrospective. *American Psychologist, 40*(2), 181-192.

Gherardi, S., & Turner, B. A. (1987). Real men don't collect soft data. Quaderno 13, Dipartimento di Politica Sociale, Università di Trento.

Gibbons, J. D. (1992a). *Nonparametric measures of association* (Quantitative Applications in the Social Sciences Series, No. 91). Newbury Park, CA: Sage.

Gibbons, J. D. (1992b). *Nonparametric statistics* (Quantitative Applications in the Social Sciences Series, No. 90). Newbury Park, CA: Sage.

Gilgun, J. F. (1992). Hypothesis generation in social work research. *Journal of Social Service Research, 15,* 113-135.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life.* New York: Free Press.

Ginsberg, R. (1990, April). *Ethnographic countertransference: Beneath bias and subjectivity in qualitative research.* Paper presented at the Annual Meeting of the American Educational Research Association, Boston.

Giorgi, A. (1975). Phenomenological method. In A. Giorgi (Ed.), *Phenomenology and psychological research* (pp. 3-17). Pittsburgh: Duquesne University Press.

Giorgi, A. (1986). Theoretical justification for the use of descriptions. In P. Ashworth, A. Giorgi, & A. de Konig (Eds.), *Qualitative research in psychology* (pp. 3-22). Pittsburgh: Duquesne University Press.

Gladwin, C. H. (1989). *Ethnographic decision tree modeling* (Qualitative Research Methods Series, Vol. 19). Newbury Park, CA: Sage.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Chicago: Aldine.

Glaser, B. G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory.* Mill Valley, CA: Sociology Press.

Glaser, B. G. (1992). *Emergence vs. forcing: Basics of grounded theory analysis.* Mill Valley, CA: Sociology Press.

Glaser, B. G., & Strauss, A. L. (1970). Discovery of substantive theory. In W. Filstead (Ed.), *Qualitative methodology* (pp. 288-297). Chicago: Rand McNally.

Glaser, E. M., Abelson, H. H., & Garrison, K. N. (1983). *Putting knowledge to use.* San Francisco: Jossey-Bass.

Glesne, C., & Peshkin, A. (1992). *Becoming a qualitative researcher: An introduction.* New York: Longman.

Goetz, J. P., & LeCompte, M. D. (1984). *Ethnography and qualitative design in educational research.* New York: Academic Press.

Goffman, E. (1959). *The presentation of self in everyday life.* Garden City, NY: Doubleday.

Gold, S. J. (1989). Ethical issues in visual fieldwork. In G. Blank et al. (Eds.), *New technology in sociology: Practical applications in research and work* (pp. 99-109). New Brunswick, NJ: Transaction Books.

Goldberg, L. (1970). Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin, 73*(4), 422-432.

Goleman, D. (1992, May 12). Jurors hear evidence and turn it into stories. *New York Times,* pp. C1, C11.

Gouldner, A. (1958). Cosmopolitans and locals: Toward an analysis of latent social roles—II. *Administrative Science Quarterly, 2,* 444-480.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11*(2), 255-274.

Guba, E. G. (1978). *Toward a methodology of naturalistic inquiry in educational evaluation.* Los Angeles: University of California, Graduate School of Education, Center for the Study of Evaluation.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology Journal, 29,* 75-92.

Guba, E. G. (1990). Carrying on the dialog. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 368-378). Newbury Park, CA: Sage.

Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the effectiveness of evaluation results through responsive and naturalistic approaches.* San Francisco: Jossey-Bass.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation.* Newbury Park, CA: Sage.

Gummeson, E. (1991). *Qualitative methods in management research.* Newbury Park, CA: Sage.

Hage, J. (1972). *Techniques and problems of theory construction in sociology.* New York: John Wiley.

Halpern, E. S. (1983, April). *Auditing naturalistic inquiries: Some preliminary applications. Part 1: Development of the process. Part 2: Case study application.* Paper presented at the Annual Meeting of the American Educational Research Association, Montreal.

Hammersley, M. (1992). Some reflections on ethnography and validity. *Qualitative Studies in Education, 5*(3), 195-204.

Hammond, P. E. (Ed.). (1964). *Sociologists at work.* New York: Basic Books.

Hanson, N. (1958). *Patterns of discovery.* Cambridge, UK: Cambridge University Press.

Harper, D. (1987). *Working knowledge: Skill and community in a small shop.* Chicago: University of Chicago Press.

Harper, D. (1989). Visual sociology: Expanding sociological vision. In G. Blank et al. (Eds.), *New technology in sociology: Practical applications in research and work* (pp. 81-97). New Brunswick, NJ: Transaction Books.

Harré, R., & Madden, E. H. (1975). *Causal powers.* Oxford, UK: Basil Blackwell.

Harré, R., & Secord, P. F. (1973). *The explanation of social behavior.* Totowa, NJ: Littlefield, Adams.

Harris, J. (1990). *Expressive discourse.* Dallas: Southern Methodist University Press.

Harrod, R. (1956). *Foundations of inductive logic.* New York: Macmillan.

Hartwig, F., & Dearing, B. E. (1979). *Exploratory data analysis.* Beverly Hills, CA: Sage.

Havelock, R. G. (1973). *The change agent's guide to innovation.* Englewood Cliffs, NJ: Educational Technology.

Havelock, R. G., Cox, P., Huberman, A. M., & Levinson, N. (1983). *School-university collaboration supporting school improvement* (Vol. 4). Washington, DC: American University, Knowledge Transfer Institute.

Havelock, R. G., Guskin, A., Frohman, M., Havelock, M., Hill, M., & Huber, J. (1979). *Planning for innovation through dissemination and utilization of knowledge.* Ann Arbor: University of Michigan, Institute of Social Research, Center for Research on Utilization of Scientific Knowledge.

Headland, T. N., Pike, K. L., & Harris, M. (1990). *Emics and etics: The insider/outsider debate.* (Frontiers of Anthropology Series, Vol. 7). Newbury Park, CA: Sage.

Heider, F. (1944). Social perception and phenomenal causality. *Psychological Review, 51,* 358-373.

Heritage, J. (1988). Explanations as accounts: A conversation analytic perspective. In C. Antaki (Ed.), *Analyzing everyday explanations: A casebook of methods* (pp. 127-144). Newbury Park, CA: Sage.

Herriott, R. E., & Firestone, W. A. (1983). Multisite qualitative policy research: Optimizing description and generalizability. *Educational Researcher, 12*(2), 14-19.

Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine, 58,* 295-300.

Hodson, R. (1991). The active worker: Compliance and autonomy at the workplace. *Journal of Contemporary Ethnography, 20*(1), 47-78.

Holland, B. (1992, October 27). Pavarotti lip-syncs, and the echoes are far-reaching. *New York Times,* p. C13.

Holland, D., & Eisenhart, M. A. (1990). *Educated in romance: Women, achievement, and college culture.* Chicago: University of Chicago Press.

Holsti, O. R. (1968). Content analysis. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology: Vol. 2. Research methods* (2nd ed., pp. 596-692). Reading, MA: Addison-Wesley.

Holsti, O. R. (1969). *Content analysis for the social sciences and the humanities.* Reading, MA: Addison-Wesley.

Homans, G. C. (1949). The strategy of industrial sociology. *American Journal of Sociology, 54*(4), 330-337.

Hook, S. (1953). Dialectics in science and history. In H. Feigl & M. Brodbeck (Eds.), *Readings in the philosophy of science* (pp. 701-713). New York: Appleton.

Horowitz, I. (Ed.). (1974). *The rise and fall of Project Camelot: Studies in the relationship between social science and practical politics.* Cambridge: MIT Press.

House, E. R. (1990a). An ethics of qualitative field studies. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 158-164). Newbury Park, CA: Sage.

House, E. R. (1990b). Methodology and justice. In K. A. Sirotnik (Ed.), *Evaluation and social justice: Issues in public education* (pp. 23-26) (New Directions for Program Evaluation, No. 45). San Francisco: Jossey-Bass.

House, E. R. (1991). Realism in research. *Educational Researcher, 20*(6), 2-9.

Howard, G. S. (1991). Culture tales: A narrative approach to thinking, cross-cultural psychology, and psychotherapy. *American Psychologist, 46*(3), 187-197.

Howe, K. R. (1985). Two dogmas of educational research. *Educational Researcher, 14*(8), 10-18.

Howe, K. R. (1988). Against the quantitative-qualitative incompatibility thesis or dogmas die hard. *Educational Researcher, 17*(8), 10-16.

Howe, K. R., & Dougherty, K. (1992, April). *Ethics, IRBs, and the changing face of educational research.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Howe, K. R., & Eisenhart, M. (1990). Standards for qualitative (and quantitative) research: A prolegomenon. *Educational Researcher, 19*(4), 2-9.

Huber, G., & Marcelo Garcia, C. (1991). Computer assistance for testing hypotheses about qualitative data: The software package AQUAD 3.0. *Qualitative Sociology, 14*(4), 325-347.

Huberman, A. M. (1978). *Evaluation of three objectives of an experimental primary school: Summary of outcomes.* Bern: Swiss Scientific Research Council.

Huberman, A. M. (1980). *A further replication study of autonomous and dependent behavior in unstructured learning environments.* Geneva: University of Geneva, Faculty of Psychology and Education.

Huberman, A. M. (1981a). *School-university collaboration supporting school improvement: Vol. 1. The Midwestern state case.* Washington, DC: American University, Knowledge Transfer Institute.

Huberman, A. M. (1981b). Splendeurs, miseres et promesses de la recherche qualitative. *Education et Recherche, 3*(3), 233-249.

Huberman, A. M. (1983). Recipes for busy kitchens: A situational analysis of everyday knowledge use in schools. *Knowledge: Creation, Diffusion, Utilization, 4*(4), 478-510.

Huberman, A. M. (1986). *Engagements in educational change throughout the teaching career.* Geneva: University of Geneva, Faculty of Psychology and Education.

Huberman, A. M. (1989). The professional life cycle of teachers. *Teachers College Record, 91*(1), 31-57.

Huberman, A. M. (1993). *The lives of teachers.* London: Cassell.

Huberman, A. M., & Gather-Thurler, M. (1991). *La mise en pratique de la recherche.* Bern: P. Lang.

Huberman, A. M., & Miles, M. B. (1983a). Drawing valid meaning from qualitative data: Some techniques of data reduction and display. *Quality and Quantity, 17,* 281-339.

Huberman, A. M., & Miles, M. B. (1983b). *Innovation up close: A field study in 12 school settings.* Andover, MA: Network.

Huberman, A. M., & Miles, M. B. (1984). *Innovation up close: How school improvement works.* New York: Plenum.

Huberman, A. M., & Miles, M. B. (1985). Assessing local causality in qualitative research. In D. N. Berg & K. K. Smith (Eds.), *Exploring clinical methods for social research* (pp. 351-382). Beverly Hills, CA: Sage.

Huberman, A. M., & Miles, M. B. (1989). Some procedures for causal analysis of multiple-case data. *Qualitative Studies in Education, 2*(1), 55-68.

Huck, S. W., & Sandler, H. M. (1979). *Rival hypotheses: "Minute mysteries" for the critical thinker.* New York: Harper & Row.

Humphreys, L. (1970). *Tearoom trade: Impersonal sex in public places.* Chicago: Aldine.

Jacob, E. (1987). Qualitative research traditions: A review. *Review of Educational Research, 57*(1), 1-50.

Jaffrey, M. (1975). *An invitation to Indian cooking.* New York: Random House.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data.* Beverly Hills, CA: Sage.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly, 24,* 602-611.

Johnson, J. C. (1990). *Selecting ethnographic informants* (Qualitative Research Methods Series, Vol. 22). Newbury Park, CA: Sage.

Johnson, M. (Ed.). (1981). *Philosophical perspectives on metaphor.* Minneapolis: University of Minnesota Press.

Jones, J. (1982). *Bad blood.* New York: Free Press.

Josselson, R., & Lieblich, A. (Eds.). (1993). *The narrative study of lives.* Newbury Park, CA: Sage.

Judd, C., & Kenny, D. (1981). *Estimating the effects of social interventions.* Cambridge, UK: Cambridge University Press.

Judd, C. M., Smith, E. R., & Kidder, L. H. (1991). *Research methods in social relations* (6th ed.). New York: Harcourt Brace Jovanovich.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430-454.

Kaplan, A. (1964). *The conduct of inquiry.* Scranton, PA: Chandler.

Katz, J. (1983). A theory of qualitative methodology: The social system of analytic fieldwork. In R. M. Emerson (Ed.), *Contemporary field research* (pp. 127-148). Boston: Little, Brown.

Kell, D. G. (1990). *Multimethod approach to analyzing effects of computers on classroom teaching and learning.* Andover, MA: Author.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15, pp. 192-238). Lincoln: University of Nebraska Press.

Kelley, H. H. (1971). Causal schemata and the attribution process. In E. E. Jones et al. (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 151-174). Morristown, NJ: General Learning Press.

Kennedy, M. (1979). Generalizing from single case studies. *Evaluation Quarterly, 3,* 661-678.

Khattri, N., & Miles, M. B. (1993). *Mapping restructuring: Final technical report.* New York: Center for Policy Research.

Kidder, T. (1985). *House.* Boston: Houghton Mifflin.

Kim, J. (1981). Causes as explanations: A critique. *Theory and Decision, 13,* 293-309.

King, S., Louth, C., & Wasley, P. (1993, April). *Roses, retrievers, and research: Collaborative inquiry to foster better schools.* Address presented at the American Educational Research Association Annual Meeting, Atlanta.

Kirk, J., & Miller, M. L. (1986). *Reliability and validity in qualitative research* (Qualitative Research Methods Series, Vol. 1). Beverly Hills, CA: Sage.

Klockars, C. B. (1979). Dirty hands and deviant subjects. In C. B. Klockars & F. W. O'Connor (Eds.), *Deviance and decency: The ethics of research with human subjects* (pp. 261-282). Beverly Hills, CA: Sage.

Korzybski, A. H. (1933). *Science and sanity.* Clinton, CT: Colonial.

Kozol, J. (1991). *Savage inequalities: Children in America's schools.* New York: Crown.

Krathwohl, D. (1991). *Methods of educational and social science research: An integrated approach.* New York: Longman.

Krippendorff, K. (1980a). Clustering. In P. R. Monge & J. N. Capella (Eds.), *Multivariate techniques in human communication research* (pp. 259-308). New York: Academic Press.

Krippendorff, K. (1980b). *Content analysis: An introduction to its methodology.* Beverly Hills, CA: Sage.

Kruglanski, A. (1989). *Lay epistemics and human knowledge.* New York: Plenum.

Kurzman, C. (1991). Convincing sociologists: Values and interests in the sociology of knowledge. In M. Burawoy et al. (Eds.), *Ethnography unbound* (pp. 250-270). Berkeley: University of California Press.

Kuzel, A. J. (1992). Sampling in qualitative inquiry. In B. F. Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (pp. 31-44) (Research Methods for Primary Care Series, Vol. 3). Newbury Park, CA: Sage.

Kvale, S. (1988). The 1000-page question. *Phenomenology and Pedagogy, 6*(2), 90-106.

Kvale, S. (Ed.). (1989a). *Issues of validity in qualitative research.* Lund, Sweden: Studentlitteratur.

Kvale, S. (1989b). To validate is to question. In S. Kvale (Ed.), *Issues of validity in qualitative research* (pp. 73-92). Lund, Sweden: Studentlitteratur.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Lazarsfeld, P. F., & Barton, A. H. (1972). Some principles of classification in social research. In P. F. Lazarsfeld, *Qualitative analysis: Historical and critical essays* (pp. 225-240). Boston: Allyn & Bacon.

Lazarsfeld, P. F., Pasanella, A. K., & Rosenberg, M. (1972). *Continuities in the language of social research.* New York: Free Press.

Leary, W. E. (1988, November 15). Novel methods unlock witnesses' memories. *New York Times,* pp. C1, C15.

LeCompte, M. D. (1975). Institutional constraints on teacher styles and the development of student work norms (Doctoral dissertation, University of Chicago). *Dissertation Abstracts International, 36,* 43A.

LeCompte, M. D., & Goetz, J. P. (1982). Problems of reliability and validity in ethnographic research. *Review of Educational Research, 52*(1), 31-60.

LeCompte, M. D., & Goetz, J. P. (1983, April). *Playing with ideas: Analysis of qualitative data.* Paper presented at the Annual Meeting of the American Educational Research Association, Montreal.

LeCompte, M. D., Millroy, W. L., & Preissle, J. (1992). *The handbook of qualitative research in education.* New York: Academic Press.

LeCompte, M. D., & Preissle, J., with Tesch, R. (1993). *Ethnography and qualitative design in educational research* (2nd ed.). New York: Academic Press.

Lee, A. A. (1991). Integrating positivist and interpretive approaches to organizational research. *Organization Science, 2*(4), 342-365.

Lee, D., Kessling, W., & Melaragno, R. (Eds.). (1981). *Parents and federal education programs: Vol. 7. Methodologies employed in the study of parental involvement.* Santa Monica, CA: System Development Corp.

Leinhardt, G. (1987). Development of an expert explanation: An analysis of a series of subtraction lessons. *Cognition and Instruction, 4*(4), 225-282.

Leithwood, K., Jantzi, D., & Dart, B. (1991, February). *How the school improvement strategies of transformational leaders foster teacher development.* Paper presented at the Sixth Annual Conference of the Department of Educational Administration and Centre for Leadership Development, OISE, Toronto.

Levine, H. G. (1985). Principles of data storage and retrieval for use in qualitative evaluations. *Educational Evaluation and Policy Analysis, 7*(2), 169-186.

Lewis, O. (1961). *Children of Sanchez: Autobiography of a Mexican family.* New York: Random House.

Lieblich, A. (1993). Looking at change: Natasha, 21: New immigrant from Russia to Israel. In R. Josselson & A. Lieblich (Eds.), *The narrative study of lives* (pp. 92-129). Newbury Park, CA: Sage.

Liebow, E. (1967). *Tally's corner: A study of Negro streetcorner men.* Boston: Little, Brown.

Lincoln, Y. S. (1990). The making of a constructivist. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 67-87). Newbury Park, CA: Sage.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.

Lincoln, Y. S., & Guba, E. G. (1990). Judging the quality of case study reports. *Qualitative Studies in Education, 3*(1), 53-59.

Lindee, M. S., Speaker, S. L., & Thackray, A. (1992). Conference report: Writing history while it happens. *Knowledge: Creation, Diffusion, Utilization, 13*(4), 479-486.

Lindesmith, A. (1947). *Opiate addiction.* Bloomington, IN: Principia.

Lindesmith, A. (1968). *Addiction and opiates.* Chicago: Aldine.

Lofland, J. (1971). *Analyzing social settings: A guide to qualitative observation and analysis.* Belmont, CA: Wadsworth.

Lofland, J. (1974). Styles of reporting qualitative field research. *American Sociologist, 9,* 101-111.

Lofland, J., & Lofland, L. H. (1984). *Analyzing social settings: A guide to qualitative observation and analysis* (2nd ed.). Belmont, CA: Wadsworth.

Lortie, D. C. (1975). *School teacher: A sociological study.* Chicago: University of Chicago Press.

Louis, K. S. (1982). Multisite/multimethod studies. *American Behavioral Scientist, 26*(1), 6-22.

Louis, K. S., & Miles, M. B. (1990). *Improving the urban high school: What works and why.* New York: Teachers College Press.

Madey, D. L. (1978). *Some benefits and costs of integrating qualitative and quantitative methods in program evaluation.* Durham, NC: NTS Research Corp.

Mailer, N. (1959). *Advertisements for myself.* New York: Signet.

Manicas, P. T., & Secord, P. F. (1982). Implications for psychology of the new philosophy of science. *American Psychologist, 38,* 390-413.

Manning, P. K. (1977). *Police work: The social organization of policing.* Cambridge: MIT Press.

Manning, P. K. (1982). Analytic induction. In R. B. Smith & P. K. Manning (Eds.), *Qualitative methods: A handbook of social science methods* (Vol. 2, pp. 273-302). Cambridge, MA: Ballinger.

Manning, P. K. (1987). *Semiotics and fieldwork.* Newbury Park, CA: Sage.

Mark, M. M., & Shotland, R. L. (Eds.). (1987). *Multiple methods in program evaluation* (New Directions for Program Evaluation, No. 35). San Francisco: Jossey-Bass.

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology, 35*(2), 63-78.

Marshall, C. (1990). Goodness criteria: Are they objective or judgment calls? In E. G. Guba (Ed.), *The paradigm dialog* (pp. 188-201). Newbury Park, CA: Sage.

Marshall, C., & Rossman, G. B. (1989). *Designing qualitative research.* Newbury Park, CA: Sage.

Martin, J. (1990). Deconstructing organizational taboos: The suppression of gender conflict in organizations. *Organization Science, 1*(4), 339-359.

Martin, P. Y., & Turner, B. A. (1986). Grounded theory and organizational research. *Journal of Applied Behavioral Science, 22*(2), 141-157.

Mathison, S. (1988). Why triangulate? *Educational Researcher, 17*(2), 13-17.

Mathison, S., Ross, E. W., & Cornett, J. W. (1993). *A casebook for teaching about ethical issues in qualitative research.* Washington, DC: American Educational Research Association, Qualitative Research SIG.

Maxwell, J. A. (1984, November). *Using ethnography to identify causes.* Paper presented at the Annual Meeting of the American Anthropological Association, Denver.

Maxwell, J. A. (1992a). *The logic of qualitative research.* Unpublished manuscript. Cambridge, MA: Harvard University, Graduate School of Education.

Maxwell, J. A. (1992b, December). *A synthesis of similarity/continuity distinctions.* Poster session presented at the Annual Meeting of the American Anthropological Association, San Francisco.

Maxwell, J. A. (1992c). Understanding and validity in qualitative research. *Harvard Educational Review, 62*(3), 279-300.

Maxwell, J. A. (n. d.). *Using qualitative research for causal explanation.* Unpublished manuscript. Cambridge, MA: Harvard University, Graduate School of Education.

Maxwell, J. A., Bashook, P. G., & Sandlow, L. J. (1986). Combining ethnographic and experimental methods in educational evaluation: A case study. In D. M. Fetterman & M. A. Pitman (Eds.), *Educational evaluation: Ethnography in theory, practice, and politics* (pp. 121-144). Beverly Hills, CA: Sage.

Maxwell, J. A., & Miller, B. A. (1992). *Categorization and contextualization as components of qualitative data analysis.* Unpublished manuscript. Cambridge, MA: Harvard University, Graduate School of Education.

May, W. T. (1987, April). *On the potential to be an unethical researcher of children.* Paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC.

McCall, G. J., & Simmons, J. L. (Eds.). (1969). *Issues in participant observation.* Reading, MA: Addison-Wesley.

McCutcheon, G. (1990, April). *Conflict about conflict: Between a rock and a hard place.* Paper presented at the Annual Meeting of the American Educational Research Association, Boston.

McEwan, H., & Bull, B. (1991). The pedagogic nature of subject matter knowledge. *American Educational Research Journal, 28*(2), 316-334.

McGrath, J. E., & Brinberg, D. (1983). External validity and the research process: A comment on the Calder-Lynch dialogue. *Journal of Consumer Research, 10*(1), 115-124.

McKeown, M. G., & Beck, I. L. (1990). The assessment and characterization of young learners' knowledge of a topic in history. *American Educational Research Journal, 27*(4), 688-726.

McLaren, P. (1991). Field relations and the discourse of the other. In W. B. Shaffir & R. A. Stebbins (Eds.), *Experiencing fieldwork* (pp. 149-163). Newbury Park, CA: Sage.

McPhee, R. D. (1990). Alternate approaches to integrating longitudinal case studies. *Organization Science, 1*(4), 393-405.

McQuillan, P., & Muncey, D. (1990). *Protecting the interests of your school while promoting quality research: Some issues to consider when allowing research to be conducted in your school* (Working Paper #2, School Ethnography Project). Providence, RI: Brown University.

McTaggart, R. (1991). When democratic evaluation doesn't seem democratic. *Evaluation Practice, 12*(1), 9-21.

Mead, M. (1928). *Coming of age in Samoa.* Magnolia, MA: Peter Smith.

Meehl, P. (1954). *Clinical versus statistical prediction.* Minneapolis: University of Minnesota Press.

Meehl, P. (1965). Clinical versus statistical prediction. *Journal of Experimental Research in Personality, 63*(1), 81-97.

Melnick, C. R., & Beaudry, J. S. (1990, April). *A qualitative research perspective: Theory, practice, essence.* Paper presented at the Annual Meeting of the American Educational Research Association, Boston.

Merriam, S. B. (1988). *Case study research in education: A qualitative approach.* San Francisco: Jossey-Bass.

Merryfield, M. M. (1990). *Constructing scenes and dialogues to display findings in case study reporting.* Columbus: Ohio State University, College of Education.

Miles, M. B. (1979a). Ethical issues in OD intervention: Should we care? *OD Practitioner, 11*(3), 1-10.

Miles, M. B. (1979b). Qualitative data as an attractive nuisance: The problem of analysis. *Administrative Science Quarterly, 24,* 590-601.

Miles, M. B. (1980). Innovation from the ground up: Dilemmas of planning and implementing new schools. *New York University Education Quarterly, 11*(2), 2-9.

Miles, M. B. (1982). A mini-cross-site analysis. *American Behavioral Scientist, 26*(1), 121-132.

Miles, M. B. (1986, April). *Improving the urban high school: Some preliminary news from five cases.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Miles, M. B. (1990). New methods for qualitative data collection and analysis: Vignettes and pre-structured cases. *Qualitative Studies in Education, 3*(1), 37-51.

Miles, M. B. (1992, April). *Forty years of change in schools: Some personal reflections.* Invited address, Division A, American Educational Research Association Annual Meeting, San Francisco.

Miles, M. B., Calder, P. H., Hornstein, H. A., Callahan, D. M., & Schiavo, R. S. (1966, August). *Data feedback and organizational change in a school system.* Paper presented at the American Sociological Association Meeting, Miami Beach.

Miles, M. B., Farrar, E., & Neufeld, B. (1983). *Review of effective schools programs: Vol. 2. The extent of adoption of effective schools programs.* Cambridge, MA: Huron Institute.

Miles, M. B., & Huberman, A. M. (1980). *The realities of school improvement programs: Analysis of qualitative data.* Proposal to National Institute of Education (funded as NIE Grant G-81-0018).

Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis: A sourcebook of new methods.* Beverly Hills, CA: Sage.

Miles, M. B., & Huberman, A. M. (1990). Animadversions and reflections on the uses of qualitative inquiry. In E. W. Eisner & A. Peshkin (Eds.), Qualitative inquiry in education: The continuing debate (pp. 339-357). New York: Teachers College Press.

Miles, M. B., & Rosenblum, S. (1987). *Chester High School: Working against the odds.* New York: Center for Policy Research, Project on Improving the Urban High School.

Miles, M. B., Saxl, E. R., James, J. A., & Lieberman, A. (1986). *New York City Teacher Centers Consortium evaluation report, 1985-1986.* New York: Center for Policy Research.

Miles, M. B., Saxl, E. R., & Lieberman, A. (1988). What skills do educational "change agents" need? An empirical view. *Curriculum Inquiry, 18*(2), 157-193.

Miles, M. B., Saxl, E. R., & Robinson, J. J. (1987). *New York City Teacher Centers Consortium evaluation report, 1986-1987.* New York: Center for Policy Research.

Miles, M. B., Sullivan, E. W., Gold, B. A., Taylor, B. L., Sieber, S. D., & Wilder, D. E. (1978). *Designing and starting innovative schools: A field study of social architecture in education* (Final report, NIE Grant G-74-0051. ED 170 828-834). New York: Center for Policy Research.

Miller, A. (1986). *Imagery in scientific thought.* Cambridge: MIT Press.

Miller, S. I. (1982). Quality and quantity: Another view of analytic induction as a research technique. *Quality and Quantity, 16,* 281-295.

Miller, S. I. (1983). Some comments on the logic of triangulation. *International Journal of Experimental Research in Education, 20*(2), 200-212.

Miller, S. I. (n. d.). *Qualitative research methods: A philosophical and practical inquiry* [prospectus for a monograph]. Chicago: Loyola University, School of Education.

Miller, S. I., & Fredericks, M. (1988). Uses of metaphor: A qualitative case study. *Qualitative Studies in Education, 1*(3), 263-272.

Miller, S. I., & Fredericks, M. (1991). Some notes on the nature of methodological indeterminacy. *Synthese, 88,* 359-378.

Mills, C. W. (1959). On intellectual craftsmanship. In C. W. Mills, *The sociological imagination* (pp. 195-226). New York: Oxford University Press.

Mirvis, P. H., & Seashore, S. E. (1982). Creating ethical relationships in organizational research. In J. E. Sieber (Ed.), *The ethics of social research: Vol. 1. Surveys and experiments* (pp. 79-104). New York: Springer-Verlag.

Mishler, E. G. (1979). Meaning in context: Is there any other kind? *Harvard Educational Review, 49*(1), 1-19.

Mishler, E. G. (1986). *Research interviewing: Context and narrative.* Cambridge, MA: Harvard University Press.

Mishler, E. G. (1990). Validation in inquiry-guided research: The role of exemplars in narrative studies. *Harvard Educational Review, 60*(4), 415-441.

Mitchell, R. G., Jr. (1991). Secrecy and disclosure in fieldwork. In W. B. Shaffir & R. A. Stebbins (Eds.), *Experiencing fieldwork* (pp. 97-108). Newbury Park, CA: Sage.

Mohr, L. B. (1982). *Explaining organizational behavior.* San Francisco: Jossey-Bass.

Morgan, G. (1980). Paradigms, metaphors, and puzzle solving in organizational theory. *Administrative Science Quarterly, 25*(4), 605-622.

Morgan, G. (1983). More on metaphor: Why we cannot control tropes in administrative science. *Administrative Science Quarterly, 28,* 601-607.

Morine-Dershimer, G. (1991, April). *Tracing conceptual change in preservice teachers.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Morse, J. M. (Ed.). (1989). *Qualitative nursing research: A contemporary dialogue.* Newbury Park, CA: Sage.

Morse, J. M., & Bottorff, J. L. (1992). The emotional experience of breast expression. In J. M. Morse (Ed.), *Qualitative health research* (pp. 319-332). Newbury Park, CA: Sage.

Mulhauser, F. (1975). Ethnography and policy-making: The case of education. *Human Organization, 34,* 311-315.

Nash, N., & Culbertson, J. (Eds.). (1977). *Linking processes in educational improvement: Concepts and applications.* Columbus, OH: University Council for Educational Administration.

Neff, B. E. (1987). *The impact of prescriptive planning models on preservice English teachers' thought and on the classroom environments they create.* Unpublished doctoral dissertation. Virginia Polytechnic Institute and State University, Blacksburg.

The Network, Inc. (1979). *Conceptual framework: A study of dissemination efforts supporting school improvement.* Andover, MA: Author.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Noblit, G. W. (1988, February). *A sense of interpretation.* Paper presented at the Ethnography in Education Research Forum, Philadelphia.

Noblit, G. W. (1989, April). *Ethnography as literature: The literary devices of qualitative research.* Paper presented at the Annual Meeting of the Southern Sociological Society, Norfolk, VA.

Noblit, G. W., & Hare, R. D. (1983, April). *Meta-ethnography: Issues in the synthesis and replication of qualitative research.* Paper presented at the Annual Meeting of the American Educational Research Association, Montreal.

Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies* (Qualitative Research Methods Series, Vol. 11). Newbury Park, CA: Sage.

Novak, J. D., & Gowan, D. B. (1984). *Learning how to learn.* Cambridge, UK: Cambridge University Press.

Ogbu, J. (1981). School ethnography: A multilevel approach. *Anthropology and Education Quarterly, 12*(1), 1-24.

Oja, S., & Smulyan, L. (1989). *Collaborative action research: A developmental approach* (Social Research and Educational Studies Series, 7). London: Falmer.

Ortony, A. (Ed.). (1979). *Metaphor and thought.* Cambridge, UK: Cambridge University Press.

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Counseling Psychology, 29*(3), 261-265.

Owen, J., & Hartley, R. (1987). Federal intervention in the mid-1980s: Charting the influence of the Participation and Equity Program at the institutional level. *Australian Educational Researcher, 15*(2), 71-85.

Patton, M. Q. (1980). *Qualitative evaluation and research methods.* Beverly Hills, CA: Sage.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.

Pearsol, J. A. (1985, April). *Controlling qualitative data: Understanding teachers' value perspectives on a sex equity education project.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Pecheux, M. (1982). *Language, semantics, and ideology: Stating the obvious.* New York: Macmillan.

Pelto, P. J., & Pelto, G. H. (1975). Intra-cultural diversity: Some theoretical issues. *American Ethnologist, 2,* 1-18.

Pelto, P. J., & Pelto, G. H. (1978). *Anthropological research: The structure of inquiry* (2nd ed.). Cambridge, UK: Cambridge University Press.

Peshkin, A. (1993). The goodness of qualitative research. *Educational Researcher, 22*(2), 23-29.

Phillips, D. C. (1990). Postpositivistic science: Myths and realities. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 31-45). Newbury Park, CA: Sage.

Pitman, M. A., & Maxwell, J. A. (1992). Qualitative approaches to evaluation: Models and methods. In M. D. LeCompte, W. L. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education* (pp. 729-770). New York: Academic Press.

Platt, J. R. (1964). Strong inference. *Science, 146,* 347-353.

Poggie, J. J., Jr. (1972). Toward control in key informant data. *Human Organization, 31,* 23-30.

Polkinghorne, D. E. (1988). *Narrative knowing and the human sciences.* Albany: State University of New York Press.

Popkewitz, T. S. (1990). Whose future? Whose past? Notes on critical theory and methodology. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 46-66). Newbury Park, CA: Sage.

Popper, K. (1968). *The logic of scientific discovery.* New York: Harper & Row.

Potter, J., & Wetherell, M. (1987). *Discourse and social psychology: Beyond attitudes and behavior.* Newbury Park, CA: Sage.

Preissle, J. (1991, January). *The choreography of design: A personal view of what design means in qualitative research.* Paper presented at Qualitative Research Conference, University of Georgia, Athens.

Punch, M. (1986). *The politics and ethics of fieldwork: Muddy boots and grubby hands* (Qualitative Research Methods Series, No. 3). Beverly Hills, CA: Sage.

Putnam, H. (1987). *Many faces of realism.* LaSalle, IL: Open Court.

Ragin, C. C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies.* Berkeley: University of California Press.

Ragin, C. C. (1993). Introduction to qualitative comparative analysis. In T. Janoski & A. Hicks (Eds.), *The comparative political economy of the welfare state* (pp. 299-319). New York: Cambridge University Press.

Ragin, C. C., & Becker, H. S. (1989). How the microcomputer is changing our analytic habits. In G. Blank et al. (Eds.), *New technology in sociology: Practical applications in research and work* (pp. 47-55). New Brunswick, NJ: Transaction Books.

Read, S. J., Druian, P. R., & Miller, L. C. (1989). The role of causal sequence in the meaning of actions. *British Journal of Social Psychology, 28,* 341-351.

Reichardt, C. S., & Cook, T. D. (1979). Beyond qualitative versus quantitative methods. In T. D. Cook & C. S. Reichardt (Eds.), *Quantitative methods in evaluation research* (pp. 7-32). Beverly Hills, CA: Sage.

Rein, M., & Schon, D. (1977). Problem setting in policy research. In C. Weiss (Ed.), *Using social policy research in public policy-making* (pp. 235-251). Lexington, MA: D. C. Heath.

Reinharz, S. (1993). Empty explanations for empty wombs: An illustration of a secondary analysis of qualitative data. In M. Schratz (Ed.), *Qualitative voices in educational research* (pp. 157-178). London: Falmer.

Reynolds, A., & Wineburg, S. S. (1990). *Qualitative queries about qualitative quandaries: Confessions of qualitative researchers in education.* Unpublished manuscript, Stanford University, CERAS, Stanford, CA.

Richards, T., & Richards, L. (1994). Using computers in qualitative analysis. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research.* Thousand Oaks, CA: Sage.

Richardson, L. (1990). *Writing strategies: Reaching diverse audiences* (Qualitative Research Methods Series, Vol. 21). Newbury Park, CA: Sage.

Richardson, L. (1992). The consequences of poetic representation: Writing the other, rewriting the self. In C. Ellis & M. G. Flaherty (Eds.), *Investigating subjectivity: Research on lived experience* (pp. 125-140). Newbury Park, CA: Sage.

Robson, C. (1993). *Real world research: A resource for social scientists and practitioners-researchers.* Oxford, UK, and Cambridge, MA: Blackwell.

Rosenberg, S. (1982). *Handbook of applied psycholinguistics.* Hillsdale, NJ: Lawrence Erlbaum.

Rosenthal, R. (1976). *Experimenter effects in behavioral research.* New York: Irvington.

Ross, L., & Lepper, M. R. (1980). The perseverance of beliefs: Empirical and normative considerations. In R. A. Shweder (Ed.), *Fallible judgment in behavioral research* (pp. 17-36). San Francisco: Jossey-Bass.

Rossman, G. B., & Wilson, B. L. (1984, April). Numbers and words: Combining quantitative and qualitative methods in a single large-scale evaluation study. *Evaluation Review, 9*(5), 627-643.

Rossman, G. B., & Wilson, B. L. (1991). Numbers and words revisited: Being "shamelessly eclectic." *Evaluation Review, 9*(5), 627-643.

Roth, S., & Cohen, L. J. (1986). Approach, avoidance, and coping with stress. *American Psychologist, 41,* 813-819.

Rothman, J. (1980). *Using research in organizations: A guide to successful application.* Beverly Hills, CA: Sage.

Runkel, P. J. (1990). *Casting nets and testing specimens: Two grand methods of psychology.* New York: Praeger.

Sadler, D. R. (1981). Intuitive data processing as a potential source of bias in naturalistic evaluations. *Educational Evaluation and Policy Analysis, 3,* 25-31.

Salomon, G. (1991). Transcending the qualitative-quantitative debate: The analytic and systemic approaches to educational research. *Educational Researcher, 20*(6), 10-18.

Sanjek, R. (Ed.). (1990). *Fieldnotes: The makings of anthropology.* Ithaca, NY: Cornell University Press.

Saxl, E. R., Miles, M. B., & Lieberman, A. (1985). *New York City Teacher Centers Consortium evaluation report, 1984-1985.* New York: Center for Policy Research.

Saxl, E. R., Miles, M. B., & Lieberman, A. (1990). *Assisting change in education: A training program for school improvement facilitators.* Washington, DC: Association for Supervision and Curriculum Development.

Saxl, E. R., Springer, C. M., Sauer, E., Sclan, E., & Miles, M. B. (1990). *Project Basics: Linking staff development and school improvement. The Project Basics Research and Evaluation Report, 1988-89.* New York: Center for Policy Research.

Schatzman, L., & Strauss, A. L. (1973). *Field research.* Englewood Cliffs, NJ: Prentice-Hall.

Scheffler, I. (1967). *Science and subjectivity.* New York: Bobbs-Merrill.

Schensul, J. J., & Schensul, S. L. (1992). Collaborative research: Methods of inquiry for social change. In M. D. LeCompte, W. L. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education* (pp. 161-200). New York: Academic Press.

Schillemans, L. et al. (n. d.). *Treating victims of incest.* Antwerp, Belgium: Flemish Institute for General Practice, and University of Antwerp, Department of Family Medicine.

Schofield, J. W. (1990). Increasing the generalizability of qualitative research. In E. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 201-232). New York: Teachers College Press.

Schuman, H. (1982). Artifacts are in the mind of the beholder. *American Sociologist, 17,* 21-28.

Schusky, E. (1983). *Manual for kinship analysis* (2nd ed.). Lanham, MD: University Press of America.

Schwandt, T. A. (1989). The politics of verifying trustworthiness in evaluation auditing. *Evaluation Practice, 10*(4), 33-40.

Schwandt, T. A. (1990, April). *On judging trustworthiness in interpretivist methodologies.* Paper presented at the Annual Meeting of the American Educational Research Association, Boston.

Schwandt, T. A. (1990). Paths to inquiry in the social disciplines: Scientific, constructivist, and critical theory methodologies. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 258-276). Newbury Park, CA: Sage.

Schwandt, T. A., & Halpern, E. S. (1988). *Linking auditing and metaevaluation: Enhancing quality in applied research* (Applied Social Research Methods Series, Vol. 11). Newbury Park, CA: Sage.

Schwartz, N., & Schwartz, C. (1955). Problems in participant observation. *American Journal of Sociology, 60,* 343-354.

Scriven, M. (1966). Causes, connections, and conditions in history. In W. H. Dray (Ed.), *Philosophical analysis and history* (pp. 240-254). New York: Harper & Row.

Scriven, M. (1974). Maximizing the power of causal investigations: The modus operandi method. In W. J. Popham (Ed.), *Evaluation in education: Current applications* (pp. 68-84). Berkeley, CA: McCutchan.

Seidman, I. E. (1991). *Interviewing as qualitative research: A guide for researchers in education and the social sciences.* New York: Teachers College Press.

Sherman, R., & Titus, W. (1982). Covariation information and cognitive processing: Effects of causal implications on memory. *Journal of Personality and Social Psychology, 42,* 989-1000.

Shulman, J. (1990). Now you see them, now you don't: Anonymity versus visibility in case studies of teachers. *Educational Researcher, 19*(6), 11-15.

Sieber, J. E. (Ed.). (1982). *Ethics of social research: Vol. 1. Surveys and experiments. Vol. 2. Field work, regulation, and publication.* New York: Springer-Verlag.

Sieber, J. E. (1992). *Planning ethically responsible research: A guide for students and internal review boards* (Applied Social Research Methods Series, Vol. 31). Newbury Park, CA: Sage.

Sieber, S. D. (1973). The integration of fieldwork and survey methods. *American Journal of Sociology, 78*(6), 1335-1359.

Sieber, S. D. (1976). *A synopsis and critique of guidelines for qualitative analysis contained in selected textbooks.* New York: Center for Policy Research, Project on Social Architecture in Education.

Sieber, S. D. (1981). *Fatal remedies: The ironies of social intervention.* New York: Plenum.

Silverstein, A. (1988). An Aristotelian resolution of the ideographic versus nomothetic tension. *American Psychologist, 43*(6), 425-430.

Smith, A. G., & Louis, K. S. (Eds.). (1982). Multimethod policy research: Issues and applications [Entire issue]. *American Behavioral Scientist, 26*(1).

Smith, A. G., & Robbins, A. E. (1982). Structured ethnography. *American Behavioral Scientist, 26,* 45-61.

Smith, A. G., & Robbins, A. E. (1984). Multimethod policy research: A case study of structure and flexibility. In D. M. Fetterman (Ed.), *Ethnography in educational evaluation* (pp. 115-132). Beverly Hills, CA: Sage.

Smith, J. K. (1983). Quantitative vs. qualitative research: An attempt to clarify the issue. *Educational Researcher, 12*(3), 6-13.

Smith, J. K. (1990). Alternative paradigms and the problem of criteria. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 167-187). Newbury Park, CA: Sage.

Smith, J. K., & Heshusius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate among educational inquirers. *Educational Researcher, 15,* 4-12.

Smith, L. M. (1990). Ethics, field studies, and the paradigm crisis. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 139-157). Newbury Park, CA: Sage.

Smith, L. M. (1991). Ethics in qualitative field research: An individual perspective. In E. W. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 258-276). New York: Teachers College Press.

Smith, L. M. (1992a). *Doing ethnographic biography: A reflective practitioner at work during a spring in Cambridge.* St. Louis: Washington University.

Smith, L. M. (1992b). Ethnography. In M. C. Alkin (Ed.), *Encyclopedia of educational research, Volume 2* (6th ed., pp. 458-462). New York: Macmillan.

Smith, L. M., & Keith, P. (1971). *The anatomy of educational innovation.* New York: John Wiley.

Smith, N. L. (1987). Toward the justification of claims in evaluation research. *Evaluation and Program Planning, 10,* 209-314.

Smith, N. L. (1992). Varieties of investigative evaluation [Entire issue]. *New Directions for Program Evaluation, 56,* Winter.

Smith, R. B. (1982). Enumerative induction: Quantification in symbolic interaction. In R. B. Smith & P. K. Manning (Eds.), *Qualitative methods: A handbook of social science methods* (Vol. 2, pp. 303-318). Cambridge, MA: Ballinger.

Smith, R. B., & Manning, P. K. (1982). *Qualitative methods: A handbook of social science methods* (Vol. 2). Cambridge, MA: Ballinger.

Soltis, J. F. (1990). The ethics of qualitative research. In E. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 247-257). New York: Teachers College Press.

Spradley, J. (1979). *The ethnographic interview.* New York: Holt, Rinehart & Winston.

Stake, R. (1976). *Evaluating educational programs: The need and the response.* Washington, DC: OECD Publications Center.

Stake, R., & Easley, J. (Eds.). (1978). *Case studies in science education.* Urbana, IL: Center for Instructional Research and Curriculum Evaluation.

Star, S. L., & Strauss, A. (1985). *Sampling in qualitative research.* San Francisco: Tremont Research Institute.

Stearns, M. S., Greene, D., David, J. L., & Associates. (1980). *Local implementation of PL 94-142: First year report of a longitudinal study* (SRI Project 7124). Menlo Park, CA: SRI International.

Stephenson, W. (1953). *The study of behavior.* Chicago: University of Chicago Press.

Stewart, D. W., & Shamdasani, P. N. (1990). *Focus groups: Theory and practice* (Applied Social Research Methods Series, Vol. 20). Newbury Park, CA: Sage.

Stiegelbauer, S., Goldstein, M., & Huling, L. L. (1982). Through the eye of the beholder: On the use of qualitative methods in data analysis (R & D Report 3137). Austin: University of Texas, R & D Center for Teacher Education.

Strauss, A. L. (1987). *Qualitative analysis for social scientists.* Cambridge, UK: Cambridge University Press.

Strauss, A. L. (1988). Teaching qualitative research methods courses: A conversation with Anselm Strauss. *International Journal of Qualitative Studies in Education, 1*(1), 91-99.

Strauss, A. L., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage.

Stringfield, S., & Teddlie, C. (1990). School improvement efforts: Qualitative and quantitative data from four naturally occurring experiments in Phases III and IV of the Louisiana School Effectiveness Study. *School Effectiveness and School Improvement, 1*(2), 139-161.

Stringfield, S., & Teddlie, C. (1991). Observers as predictors of schools' multiyear outlier status on achievement tests. *Elementary School Journal, 91*(4), 357-376.

Strunk, W., Jr., & White, E. B. (1979). *The elements of style* (3rd ed.). New York: Macmillan.

Taft, R. (1955). The ability to judge people. *Psychological Bulletin, 52*(1), 1-23.

Taylor, M., MacLean, H., Pallister, R., & White, N. (1988). *Minimizing chronic illness effects* (Research proposal, National Health Research and Development Program, Health and Welfare Canada). Montreal and Hamilton: Concordia University and McMaster University.

Teich, A. H., & LaFollette, M. (Eds.). (1992). Understanding scientific misconduct: What do we know? [Special section]. *Knowledge: Creation, Diffusion, Utilization, 14*(2), 149-211.

Templin, P. A. (1982). Still photography in evaluation. In N. L. Smith (Ed.), *Communication strategies in evaluation* (pp. 121-175). Beverly Hills, CA: Sage.

Tesch, R. (1989). Computer software and qualitative analysis: A reassessment. In G. Blank et al. (Eds.), *New technology in sociology: Practical applications in research and work* (pp. 141-154). New Brunswick, NJ: Transaction Books.

Tesch, R. (1990). *Qualitative research: Analysis types and software tools.* New York: Falmer.

Tesch, R. (Ed.). (1991). Computers and qualitative data II [Special issues, Parts 1 & 2]. *Qualitative Sociology, 14*(3 & 4), Fall, Winter.

Thomas, J. (1993). *Doing critical ethnography* (Qualitative Research Methods Series, No. 26). Newbury Park, CA: Sage.

Trochim, W. M. K. (1989a). Concept mapping: Soft science or hard art? *Evaluation and Program Planning, 12,* 87-110.

Trochim, W. M. K. (1989b). *The concept system.* Ithaca, NY: Cornell University, Department of Human Services Studies.

Trochim, W. M. K. (1989c). An introduction to concept mapping for planning and evaluation. *Evaluation and Program Planning, 12,* 1-6.

Tuchman, B. (1981). *Practicing history.* New York: Knopf.

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.

Tufte, E. R.(1986). Designing statistical presentations. *Social Science, 71*(1), 75-80.

Tufte, E. R. (1990). *Envisioning information.* Cheshire, CT: Graphics Press.

Tuma, N. (1982). Nonparametric and partially parametric approaches to event history analysis. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 1-60). San Francisco: Jossey-Bass.

Turner, B. A. (1981). Some practical aspects of qualitative data analysis: One way of organizing the cognitive processes associated with the generation of grounded theory. *Quality and Quantity, 15*(3), 225-247.

Turner, B. A. (1988). Connoisseurship in the study of organizational culture. In A. Bryman (Ed.), *Doing research in organizations* (pp. 100-122). London: Routledge.

Turner, S. P. (1980). *Sociological explanation as translation.* New York: Cambridge University Press.

Tversky, A., & Kahneman, D. (1971). The belief in the law of small numbers. *Psychological Bulletin, 76*(2), 105-110.

van der Vegt, R., & Knip, H. (1990). Implementing mandated change: The school as change contractor. *Curriculum Inquiry, 20*(2), 183-203.

Van Dijk, T. A. (Ed.). (1985). *Handbook of discourse analysis* (Vol. 4). New York: Academic Press.

Van Maanen, J. (1979). The fact of fiction in organizational ethnography. *Administrative Science Quarterly, 24,* 539-611.

Van Maanen, J. (1983a). The moral fix: On the ethics of fieldwork. In R. M. Emerson (Ed.), *Contemporary field research* (pp. 269-287). Prospect Heights, IL: Waveland.

Van Maanen, J. (Ed.). (1983b). *Qualitative methodology.* Beverly Hills, CA: Sage.

Van Maanen, J. (1988). *Tales of the field: On writing ethnography.* Chicago: University of Chicago Press.

van Manen, M. (1977). Linking ways of knowing with ways of being practical. *Curriculum Inquiry, 6*(3), 205-228.

Van Parijs, P. (1981). *Evolutionary explanation in the social sciences: An emerging paradigm.* Totowa, NJ: Rowman & Littlefield.

Vitz, P. C. (1990). The use of stories in moral development: New psychological reasons for an old education method. *American Psychologist, 45*(6), 709-720.

Wagner, J. (Ed.). (1979). *Images of information.* Beverly Hills, CA: Sage.

Wainer, H. (1982). *How to display data badly* (Program Statistics Research Technical Report No. 82-33). Princeton, NJ: Educational Testing Service.

Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher, 21*(1), 14-23.

Walker, R. (1981). The conduct of educational case studies: Ethics, theory, and procedures. In W. B. Dockrell & D. Hamilton (Eds.), *Rethinking educational research.* London: Hodder & Stoughton.

Wallis, W. A., & Roberts, H. V. (1956). *Statistics: A new approach.* New York: Free Press.

Warner, W. (1991, February). *Improving interpretive validity of camera-based qualitative research.* Paper presented at the Qualitative Health Research Conference, Edmonton, Alberta.

Watkins, K. E. (1991, April). *Validity in action research.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Wax, M. (1982). Research reciprocity rather than informed consent. In J. E. Sieber (Ed.), *Ethics of social research: Vol. 2. Fieldwork, regulation, and publication* (pp. 33-48). New York: Springer-Verlag.

Wax, R. (1971). *Doing fieldwork: Warnings and advice.* Chicago: University of Chicago Press.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1965). *Unobtrusive measures.* Chicago: Rand McNally.

Webb, R. B., & Glesne, C. (1992). Teaching qualitative research. In M. D. LeCompte, W. L. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education* (pp. 771-859). New York: Academic Press.

Weber, R. P. (1990). *Basic content analysis* (2nd ed.) (Quantitative Applications in the Social Sciences Series, Vol. 49). Newbury Park, CA: Sage.

Weick, K. (1979). *The social psychology of organizing.* Reading, MA: Addison-Wesley.

Weick, K. (1989). Theory construction as disciplined imagination. *Academy of Management Review, 14*(4), 516-531.

Weinstein, E. A., & Tamur, J. M. (1978). Meanings, purposes, and structural resources in social interaction. In J. G. Manis & B. N. Meltzer (Eds.), *Symbolic interaction* (3rd ed., pp. 138-140). Boston: Allyn & Bacon.

Weitzman, E. A., & Miles, M. B. (1993). *Computer-aided qualitative data analysis: A review of selected software.* New York: Center for Policy Research.

Weitzman, E., & Miles, M. B. (1994). *Computer programs for qualitative data analysis.* Thousand Oaks, CA: Sage.

Weller, S. C., & Romney, A. K. (1988). *Systematic data collection* (Qualitative Research Methods Series, No. 10). Newbury Park, CA: Sage.

Werner, O. (1992). How to record activities. *Cultural Anthropology Methods Newsletter, 4*(2), 1-3.

Werner, O., & Schoepfle, G. M. (1987a). *Systematic fieldwork: Vol. 1. Foundations of ethnography and interviewing.* Newbury Park, CA: Sage.

Werner, O., & Schoepfle, G. M. (1987b). *Systematic fieldwork: Vol. 2. Ethnographic analysis and data management.* Newbury Park, CA: Sage.

Whaley, T. R. (1991, April). *Narrative, fiction, and interpretation in the social sciences.* Paper presented at the American Educational Research Association Annual Meeting, Chicago.

Whyte, W. F. (1943). *Street corner society.* Chicago: University of Chicago Press.

Whyte, W. F. (1984). *Learning from the field: A guide from experience.* Beverly Hills, CA: Sage.

Whyte, W. F. (Ed.). (1991). *Participatory action research.* Newbury Park, CA: Sage.

Wicker, A. W. (1985). Getting out of our conceptual ruts: Strategies for expanding conceptual frameworks. *American Psychologist, 40*(10), 1094-1103.

Williams, D. D. (1986). Naturalistic evaluation: Potential conflicts between evaluation standards and criteria for conducting naturalistic inquiry. *Educational Evaluation and Policy Analysis, 8*(1), 87-99.

Wilson, J. V., & Bolland, J. M. (1992, January). *Modeling anomalies: A case for qualitative research.* Paper presented at the Conference on Qualitative Research in Education, Athens, GA.

Wolcott, H. F. (1973). *The man in the principal's office: An ethnography.* New York: Holt, Rinehart & Winston.

Wolcott, H. F. (1982). Differing styles of on-site research, or, "If it isn't ethnography, what is it?" *The Review Journal of Philosophy and Social Science, 7*(1 & 2), 154-169.

Wolcott, H. F. (1990a). On seeking—and rejecting—validity in qualitative research. In E. W. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 121-152). New York: Teachers College Press.

Wolcott, H. F. (1990b). *Writing up qualitative research* (Qualitative Research Methods Series, Vol. 20). Newbury Park, CA: Sage.

Wolcott, H. F. (1992). Posturing in qualitative inquiry. In M. D. LeCompte, W. L. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education* (pp. 3-52). New York: Academic Press.

Wolcott, H. F. (1994). *Transforming qualitative data: Description, analysis, interpretation.* Thousand Oaks, CA: Sage.

Wolfe, R. (1992). Data management. In M. C. Alkin (Ed.), *Encyclopedia of educational research, Volume 1* (6th ed., pp. 293-299). New York: Macmillan.

Yin, R. K. (1984). *Case study research: Design and methods* (Applied Social Research Methods Series, Vol. 5). Beverly Hills, CA: Sage.

Yin, R. K. (1991). *Applications of case study research.* Washington, DC: Cosmos Corp.

Yngve, V. H. (1986). *Linguistics as a science.* Bloomington: Indiana University Press.

Zelditch, M. (1962). Some methodological problems of field studies. *American Journal of Sociology, 67,* 566-576.

Zeller, N. (1987). *A rhetoric for naturalistic inquiry.* Unpublished doctoral dissertation, Indiana University, Bloomington.

Zeller, N. (1991, April). *A new use for new journalism: Humanizing the case report.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Zinsser, W. (1976). *On writing well.* New York: Harper & Row.

Author Index

Subject Index

About the Authors

Matthew B. Miles, a social psychologist, has had a career-long interest in educational reform and strategies for achieving it (Miles, 1992). His first full-scale venture into qualitative research was a 4-year study of the creation of six new, innovative public schools (Miles, 1980; Miles et al., 1978). His "attractive nuisance" paper (Miles, 1979b) was at the same time a rueful reflection and a sort of manifesto for work to follow.

Michael Huberman's long-term interests have been in scientific epistemology—how scientific theories are developed and validated—and in adult cognition and knowledge use. His first extensive qualitative project was a 4-year study of an experimental elementary school seeking to implement Piagetian theories in a classroom setting (Huberman, 1978, 1980). His reflection/manifesto paper (Huberman, 1981b) spoke of the "splendors and miseries" of qualitative research.

Our Joint Work

When we had the chance to collaborate on a major field study, we leaped at it. The project was a study of the dissemination of educational innovations carried out during 1979-1983. Many of the examples in this text come from that study. (More detailed information appears in the first edition.)

The field study. The study was nested in a larger study of school improvement (Crandall & Associates, 1983), covering 145 school buildings and nearly 4,000 people throughout the United States involved in the implementation of educational innovations.

Joined by two colleagues, Beverly Loy Taylor and Jo Anne Goldberg, we repeatedly visited a stratified sample of 12 field sites across the country throughout the school year 1979-1980, with follow-up contacts the next year to verify the main findings.

The volume of data collected included 440 interviews, 85 observations, some 259 documents, and 2,713 pages of transcribed field notes. Interview and observation notes were dictated and transcribed. We developed a common set of data displays and, for each of the 12 field sites, used them to draw conclusions, resulting in a case report ranging from 70 to 140 pages, with a common format. Our subsequent cross-case analysis (Huberman & Miles, 1983a, 1984) was built from the appropriate sections of the 12 case reports.

The documentation study. In the course of the cross-case analysis, we began another study, sponsored by the National Institute of Education (Miles & Huberman, 1980). The task was to document the procedures we used for analyses, from the initial coding of site-level notes to the more explanatory cross-case analyses.

Each analysis fed into a detailed self-documentation form (see Chapter 10, section D); this exercise provided many of the illustrations and rules of thumb in the present book.

Further work. The analytic methods reported in the first edition have been debugged, refined, and extended as we have worked with them—and learned of many interesting procedures used by colleagues.

In a second multiple-case field study (Havelock et al., 1983; Huberman, 1981a), we replicated many of the techniques for collecting, reducing, and analyzing qualitative data from the school improvement study, finding that they "traveled" quite well.

Additional studies have included a 2-year extensive interview study of 17 school "change agents" and their clients and managers (Miles et al., 1988); a 4-year study combining data from a national survey of urban high schools with five case studies of successful reform (Louis & Miles, 1990); a 4-year study of educational reform in Bangladesh, Ethiopia, and Colombia (Dalin et al., 1992); a study of the life cycles of teachers (Huberman, 1986, 1989, 1993); and a 4-year study of the use of scientific knowledge in vocational education (Huberman & Gather-Thurler, 1991).

We learned a great deal through teaching from the first edition in seminars in the United States, Canada, Western Europe, and Australia. The generic issues of qualitative data analysis come clearer as they are confronted not only across researchers, but across cultures.