Qualitative Research & Evaluation Methods

3

EDITION

Michael Quinn Patton

(e) informal discussions with program staff,

(f) a review of student projects and other documents, and

(g) 23 records from the files of each student (including employer evaluations of students, student products, test scores, and staff progress evaluations of students).

A set of guide questions was prepared for analyzing and reviewing each source (Fehrenbacher, Owens, and Haehnn 1976: 7-8). Information from all of these sources was integrated to produce a highly readable narrative that could be used by decision makers and funders to better understand what it was like to be in the program (Owens, Haehnn, and Fehrenbacher 1987). The evaluation staff of the Northwest Regional Educational Laboratory took great pains to carefully validate the information in the case studies. Different sources of information were used to cross-validate findings, patterns, and conclusions. Two evaluators reviewed the material in each case study to independently make judgments and interpretations about the content and meaning of the material in the case. In addition, an external evaluator reviewed the raw data to check for biases or unwarranted conclusions. Students were asked to read their own case studies and comment on the accuracy of fact and interpretation in the study. Finally, to guarantee the readability of the case studies, a newspaper journalist was employed to help organize and edit the final versions. Such a rigorous case study approach increases the confidence of readers that the cases are accurate and comprehensive. Both in its content and the process by which it was constructed, the Northwest Lab case study presented at the end of this chapter (Appen-

dix 8.2) exemplifies how an individual case study can be prepared and presented.

The same rigorous process would apply to case study data at the group or program level. For excellent examples of case studies in education, see Brizuela et al. (2000), Stake, Bresler, and Mabry (1991), Perrone (1985), and Alkin, Daillak, and White (1979); for family research see Sussman and Gilgun (1996); for international development see Salmen (1987) and Searle (1985); in government accountability see Kloman (1979); and for a detailed example of conducting and presenting an evaluation case study, see Hébert (1986).

How one compares and contrasts cases will depend on the purpose of the study and how cases were sampled. As discussed in Chapter 5, critical cases, extreme cases, typical cases, and heterogeneous cases serve different purposes. Other excellent resources for qualitative case analysis include Stake (1995), Merriam (1997), Yin (1994), Hamel (1993), and the U.S. General Accounting Office (1987). To pursue case studies as stories that build on and display the elements of good storytelling, see Glesne (1999).

Once case studies have been written, the analytic strategies described in the remainder of this chapter can be used to further analyze, compare, and interpret the cases to generate cross-case themes, patterns and findings.

⑤ Pattern, Theme, and Content Analysis

The ability to use thematic analysis appears to involve a number of underlying abilities, or competencies. One competency can be called *pattern recognition*. It is the ability to see patterns in seemingly random information. (Boyatzis 1998:7)

case

ɔply

ram

dies

ake,

)85),

for

gun

see

ern-

and

and

see

ises

and

1 in

/pi-

dif-

ces

ake

mel

Of-

ries

; of

the

iin-

an-

; to

ind

ars

, or

led

iat-

on.

No precise or agreed-on terms describe varieties and processes of qualitative analysis. Content analysis, for example, sometimes refers to searching text for recurring words or themes. For example, a speech by a politician might be analyzed to see what phrases or concepts predominate, or speeches of two politicians might be compared to see how many times and in what contexts they used a phrase such as "global economy" or "family values." Content analysis usually refers to analyzing text (interview transcripts, diaries, or documents) rather than observation-based field notes. **More generally, however, content analysis is used to refer to any qualitative data reduction and sense-making effort that takes a volume of qualitative material and attempts to identify core consistencies and meanings.** Case studies, for example, can be content analyzed.

The core meanings found through content analysis are often called patterns or themes. Alternatively, the process of searching for patterns or themes may be distinguished, respectively, as pattern analysis or theme analysis. I'm asked frequently about the difference between a pattern and a theme. There's no hard-and-fast distinction. The term *pattern* usually refers to a descriptive finding, for example, "Almost all participants reported feeling fear when they rappelled down the cliff," while a theme takes a more categorical or topical form: *Fear.* Putting these terms together, a report on a wilderness education study might state:

> The *content analysis* revealed a *pattern* of participants reporting being afraid when rappelling down cliffs and running river rapids; many also initially experienced the group process of sharing personal feelings as evoking some fear. Those patterns make "Dealing with fear" a major *theme* of the wilderness education program experience.

Inductive and Deductive Qualitative Analyses

Francis Bacon is known for his emphasis on *induction, the use of direct observation to confirm ideas and the linking together of observed facts to form theories or explanations of how natural phenomenon work.* Bacon correctly never told us how to get ideas or how to accomplish the linkage of empirical facts. Those activities remain essentially humanistic—you think hard. (Bernard 2000:12)

Bacon (1561-1626) is recognized as one of the founders of scientific thinking, but he also has been awarded "the dubious honor of being the first martyr of empiricism" (Bernard 2000:12). Still pondering the universe at age 65, he got an idea one day while driving his carriage in the snow in a farming area north of London. It occurred to him that cold might delay the biological process of putrefaction, so he stopped, purchased a hen from a farmer, killed it on the spot, and stuffed it with snow. His idea worked. The snow did delay the rotting process, but he caught bronchitis and died a month later. As I noted in Chapter 6, fieldwork can be risky. Engaging in analysis, on the other hand, is seldom life threatening, though you do risk being disputed and sometimes ridiculed by those who arrive at contrary conclusions.

Inductive analysis involves *discovering* patterns, themes, and categories in one's data. Findings emerge out of the data, through the analyst's interactions with the data, in contrast to *deductive analysis* where the data are analyzed according to an existing framework. Qualitative analysis is typically inductive in the early stages, especially when developing a codebook for content analysis or figuring out possible categories, patterns, and themes. This is often called "open coding" (Strauss and Corbin 1998:223) to emphasize the importance of being open to the

data. "Grounded theory" (Glaser and Strauss 1967) emphasizes becoming immersed in the data—being *grounded*—so that embedded meanings and relationships can emerge. The French would say of such an immersion process: *Je m'enracine.* "I root myself." The analyst becomes implanted in the data. The resulting analysis grows out of that groundedness.

Once patterns, themes, and/or categories have been established through inductive analysis, the final, confirmatory stage of qualitative analysis may be deductive in testing and affirming the authenticity and appropriateness of the inductive content analysis, including carefully examining deviate cases or data that don't fit the categories developed. Generating theoretical propositions or formal hypotheses after inductively identifying categories is considered deductive analysis by grounded theorists Strauss and Corbin (1998): "Anytime that a researcher derives hypotheses from data, because it involves interpretation, we consider that to be a deductive process" (p. 22). Grounded theorizing, then, involves both inductive and deductive processes: "At the heart of theorizing lies the interplay of making inductions (deriving concepts, their properties, and dimensions from data) and deductions (hypothesizing about the relationships between concepts)" (Strauss and Corbin 1998:22).

Analytic induction, in contrast to grounded theory, begins with an analyst's deduced propositions or theory-derived hypotheses and "is a procedure for verifying theories and propositions based on qualitative data" (Taylor and Bogdan 1984:127). Sometimes, as with analytic induction, qualitative analysis is first deductive or quasi-deductive and then inductive as when, for example, the analyst begins by examining the data in terms of theory-derived sensitizing concepts or applying a theoretical frame-

work developed by someone else (e.g., testing Piaget's developmental theory on case studies of children). After or alongside this deductive phase of analysis, the researcher strives to look at the data afresh for undiscovered patterns and emergent understandings (inductive analysis). I'll discuss both grounded theory and analytic deduction at greater length later in this chapter.

Because, as identified and discussed in Chapter 2, inductive analysis is one of the primary characteristics of qualitative inquiry, we'll focus on strategies for thinking and working inductively. There are two distinct ways of analyzing qualitative data inductively. First, the analyst can identify, define, and elucidate the categories developed and articulated by the people studied to focus analysis. Second, the analyst may also become aware of categories or patterns for which the people studied did not have labels or terms, and the analyst develops terms to describe these inductively generated categories. Each of these approaches is described below.

Indigenous Concepts and Practices

A good place to begin inductive analysis is to inventory and define key phrases, terms, and practices that are special to the people in the setting studied. What are the indigenous categories that the people interviewed have created to make sense of their world? What are practices they engage in that can be understood only within their worldview? Anthropologists call this *emic* analysis and distinguish it from *etic* analysis, which refers to labels imposed by the researcher. (For more on this distinction and its origins, see Chapter 6, which discusses emic and etic perspectives in fieldwork.) "Identifying the categories and terms used

by informants themselves is also called *in vivo* coding" (Bernard and Ryan 1998:608).

Consider the practice among traditional Dani women of amputating a finger joint when a relative dies. The Dani people live in the lush Baliem Valley of Irian Java, Indonesia's most remote province on the western half of New Guinea. The joint is removed to honor and placate ancestral ghosts. Missionaries have fought the practice as sinful and the government has banned it as barbaric, but many traditional women still practice it.

> Some women in Dani villages have only four stubs and a thumb on each hand. In tribute to her dead mother and brothers, Soroba, 38, has had the tops of six of her fingers amputated. "The first time was the worst," she said. "The pain was so bad, I thought I would die. But it's worth it to honor my family." (Sims 2001:6)

Analyzing such an indigenous practice begins by understanding it from the perspective of its practitioners, within the indigenous context, in the words of the local people, in their language, within their worldview.

> According to this view, cultural behavior should always be studied and categorized in terms of the inside view—the actors' definition—of human events. That is, the units of conceptualization in anthropological theories should be "discovered" by analyzing the cognitive processes of the people studied rather than "imposed" from cross-cultural (hence, ethnocentric) classifications of behavior. (Pelto and Pelto 1978:54)

Anthropologists, working cross-culturally, have long emphasized the importance of preserving and reporting the indigenous categories of people studied. Franz Boas (1943) was a major influence in this direction: "If it is our serious purpose to under-

stand the thoughts of a people, the whole analysis of experience must be based on their concepts, not ours" (p. 314).

In an intervention program, certain terms may emerge or be created by participants to capture some essence of the program. In the wilderness education program I evaluated, the idea of "detoxification" became a powerful way for participants to share meaning about what being in the wilderness together meant (Patton 1999a:49-52). In the Caribbean Extension Project evaluation, the term *liming* had special meaning to the participants. Not really translatable, it essentially means passing time, hanging out, doing nothing, shooting the breeze—but doing so agreeably, without guilt, stress, or a sense that one ought to be doing something more productive with one's time. *Liming* has positive, desirable connotations because of its social, group meaning—people just enjoying being together with nothing that has to be accomplished. Given that uniquely Caribbean term, what does it mean when participants describe what happened in a training session or instructional field trip as primarily "liming"? How much liming could acceptably be built into training for participant satisfaction and still get something done? How much programmatic liming was acceptable? These became key formative evaluation issues.

In evaluating a leadership training program, we gathered extensive data on what participants and staff meant by the term *leadership*. Pretraining and posttraining exercises involved participants in writing a paragraph on leadership; the writing was part of the program curriculum, not designed for evaluation, but the results provided useful qualitative evaluation data. There were small group discussions on leadership. The training included lectures and group discussions on leadership, which we observed. We participated in and took notes on informal

discussions about leadership. Because the very idea of "leadership" was central to the program, it was essential to capture variations in what participants meant when they talked about leadership. The results showed that ongoing confusion about what leadership meant was one of the problematic issues in the program. Leadership was an indigenous concept in that staff and participants throughout the training experience used it extensively, but it was also a *sensitizing concept* because we knew going into the fieldwork that it would be an important notion to study.

Sensitizing Concepts

In contrast to purely indigenous concepts, sensitizing concepts refer to categories that the analyst brings to the data. Experienced observers often use sensitizing concepts to orient fieldwork, an approach discussed in Chapter 6. These sensitizing concepts have their origins in social science theory, the research literature, or evaluation issues identified at the beginning of a study. Sensitizing concepts give the analyst "a general sense of reference" and provide "directions along which to look" (Blumer 1969: 148). Using sensitizing concepts involves examining how the concept is manifest and given meaning in a particular setting or among a particular group of people.

Conroy (1987) used the sensitizing concept "victimization" to study police officers. Innocent citizens are frequently thought of as the victims of police brutality or indifference. Conroy turned the idea of victim around and looked at what it would mean to study police officers as victims of the experiences of law enforcement. He found the sensitizing concept of victimization helpful in understanding the isolation, lack of interpersonal affect, cynicism, repressed anger, and sadness observed among police officers.

He used the idea of victimization to tie together the following quotes from police officers:

As a police officer and as an individual I think I have lost the ability to feel and to empathize with people. I had a little girl that was run over by a bus and her mother was there and she had her little book bag. It was really sad at the time but I remember feeling absolutely nothing. It was like a mannequin on the street instead of some little girl. I really wanted to be able to cry about it and I really wanted to have some feelings about it, but I couldn't. It's a little frightening for me to be so callous and I have been unable to relax.

I am paying a price by always being on edge and by being alone. I have become isolated from old friends. We are different. I feel separate from people, different, out of step. It becomes easier to just be with other police officers because they have the same basic understanding of my environment, we speak the same language. The terminology is crude. When I started I didn't want to get into any words like "scumbags" and "scrotes," but it so aptly describes these people.

I have become isolated from who I was because I have seen many things I wish I had not seen. It's frustrating to see things that other people don't see, won't see, can't see. I wish sometimes, I didn't see the things. I need to be assertive, but don't like it. I have to put on my police mask to do that. But now it is getting harder and harder to take that mask off. I take my work home with me. I don't want my work to invade my personal life but I'm finding I need to be alone more and more. I need time to recharge my batteries. I don't like to be alone, but must. (Conroy 1987:52)

Two additional points are worth making about these quotations. First, by presenting

113

the actual data on which the analysis is based, the readers are able to make their own determination of whether the concept "victimization" helps in making sense of the data. By presenting respondents in their own words and reporting the actual data that was the basis of his interpretation, Conroy invites readers to make their own analysis and interpretation. The analyst's constructs should not dominate the analysis, but rather should facilitate the reader's understanding of the world under study.

Second, these three quotations illustrate the power of qualitative data. The point of analysis is not simply to find a concept or label to neatly tie together the data. What is important is understanding the people studied. Concepts are never a substitute for direct experience with the descriptive data. **What people actually say and the descriptions of events observed remain the essence of qualitative inquiry.** The analytical process is meant to organize and elucidate telling the story of the data. Indeed, the skilled analyst is able to get out of the way of the data to let the data tell their own story. The analyst uses concepts to help make sense of and present the data, but not to the point of straining or forcing the analysis. The reader can usually tell when the analyst is more interested in proving the applicability and validity of a concept than in letting the data reveal the perspectives of the people interviewed and the intricacies of the world studied.

Having suggested how singular concepts can bring focus to inductive analysis, the next level of analysis, constructing typologies, moves us into a somewhat more complex analytical strategy.

Indigenous Typologies

Typologies are classification systems made up of categories that divide some aspect of the world into parts along a continuum. They differ from *taxonomies*, which completely classify a phenomenon through mutually exclusive and exhaustive categories, like the biological system for classifying species. Typologies, in contrast, are built on ideal-types or illustrative endpoints rather than a complete and discrete set of categories. Well-known and widely used sociological typologies include Redfield's folk-urban continuum (*gemeinschaft/gesellschaft*) and Von Wiese's and Becker's sacred-secular continuum (for details, see Vidich and Lyman 2000:52). Sociologists classically distinguish ascribed from achieved characteristics. Psychologists distinguish degrees of mental illness (neuroses to psychoses). Political scientists classify governmental systems along a democratic-authoritarian continuum. Economists distinguish laissez-faire from centrally planned economic systems. Systems analysts distinguish open from closed systems. In all of these cases, however, the distinctions involves matters of degree and interpretation rather than absolute distinctions. All of these examples have emerged from social science theory and represent theory-based typologies constructed by analysts. We'll examine that approach in greater depth in a moment. First, however, let's look at identifying indigenous typologies as a form of qualitative analysis.

Illuminating indigenous typologies requires an analysis of the continua and distinctions used by people in a setting to break up the complexity of reality into distinguishable parts. The language of a group of people reveals what is important to them in that they name something to separate and distinguish it from other things with other names. Once these labels have been identified from an analysis of what people have said during fieldwork, the next step is to identify the attributes or characteristics that distinguish one thing from another. In describing this

114

kind of analysis, Charles Frake (1962) used the example of a hamburger. Hamburgers can vary a great deal in how they are cooked (rare to well-done) or what is added to them (pickles, mustard, ketchup, lettuce), and they are still called hamburgers. However, when a piece of cheese is added to the meat, it becomes a cheeseburger. The task for the analyst is to discover what it is that separates "hamburger" from "cheeseburger," that is, to discern and report "how people construe their world of experience from the way they talk about it" (Frake 1962:74).

An analysis example of this kind comes from a formative evaluation aimed at reducing the dropout rate among high school students. In observations and interviews at the targeted high school, it became important to understand the ways in which teachers categorized students. With regard to problems of truancy, absenteeism, tardiness, and skipping class, the teachers had come to label students as either "chronics" or "borderlines." One teacher described the chronics as "the ones who are out of school all the time and everything you do to get them in doesn't work." Another teacher said, "You can always pick them out, the chronics. They're usually the same kids." The borderlines, on the other hand, "skip a few classes, waiting for a response, and when it comes they shape up. They're not so different from your typical junior high student, but when they see the chronics getting away with it, they get more brazen in their actions." Another teacher said, "Borderlines are gone a lot but not constantly like the chronics."

Not all teachers used precisely the same criteria to distinguish chronics from borderlines, but all teachers used these labels in talking about students. To understand the program activities directed at reducing high school dropouts and the differential impact of the program on students, it became important to observe differences in how bor-

derlines and chronics were treated. Many teachers, for example, refused even to attempt to deal with chronics. They considered it a waste of their time. Students, it turned out, knew what labels were applied to them and how to manipulate these labels to get more or less attention from teachers. Students who wanted to be left alone called themselves "chronics" and reinforced their "chronic" image with teachers. Students who wanted to graduate, even if only barely and with minimal school attendance, cultivated an image as "borderline."

Another example of an indigenous typology emerged in the wilderness education program I evaluated. During the second year of the project, one subgroup's members started calling themselves the "turtles." They contrasted themselves to the "truckers." On the surface, these labels were aimed at distinguishing different styles of hiking and backpacking, one slow and one fast. Beneath the surface, however, the terms came to represent different approaches to the wilderness and different styles of experience in relation to the wilderness and the program.

Groups, cultures, organizations, and families develop their own language systems to emphasize distinctions they consider important. Every program gives rise to a special vocabulary that staff and participants use to differentiate types of activities, kinds of participants, styles of participation, and variously valued outcomes. These indigenous typologies provide clues to analysts that the phenomena to which the labels refer are important to the people in the setting and that to fully understand the setting it is necessary to understand those terms and their implications.

Analyst-Constructed Typologies

Once indigenous concepts, typologies, and themes have been surfaced, the analyst

115

BEYOND NAMING PROBLEMS: HOLISTIC AND BROADLY GAUGED ANALYSES

Excerpts of Reflections of Philosopher Elizabeth Minnich

United States readers of critiques of popular culture such as Robert Bly's *Bowling Alone* tend to respond with enthusiasm to easily grasped analyses of what is wrong with us, from whatever left/middle/right stance they come—viz., to pick just a few, whoever wrote differently the popularity of 50's analyses of the organization man and the ugly American, or Baldwin's *The Fire Next Time*, Friedan's "the problem that has no name" in *The Feminine Mystique*, Harrington's *The Other America*, Bellah et al.'s *Habits of the Heart*, Bloom's *The Closing of the American Mind*, William Bennett's and Cornel West's politically opposite diagnoses of a moral crisis that is besetting the nation. Such analyses give us the relief of names to attach to widespread concerns; they catch on like a new kind of pill for a real social ill that, whether the catchily named pill works or not, gives us some sense that at least someone knows about our pain.

It is because they come into the vicinity of where we are hurting that we respond so strongly, poke my wound, even to help me heal, and I will react. But this "poking" is also not as healing as it could be insofar as it remains too narrow in ways that constrain and may misdirect the holistic help we want. Like many analysts before them, what they have done is to focus on where a problem becomes readily evident. But as analyses of wife-beating that focus on the victims tend to lead to proposals (often formulated by entirely other people than the analysts) that also focus on the women, excluding from the picture the male perpetrators and the systems that empower them, this won't do. We need other analyses and broader gauge ones. (Minnich 1999:8,11)

moves to a different task of induction—looking for patterns, categories, and themes for which the analyst can construct a typology to further elucidate findings. Such constructions must be done with considerable care to avoid creating things that are not really in the data. The advice of biological theorist John Maynard Smith (2000) is informative in this regard: Seek models of the world that make sense and whose consequences can be worked out, for "to replace a world you do not understand by a model of a world you do not understand is no advance" (p. 46).

Constructing ideal-types or alternative paradigms is one simple form of presenting qualitative comparisons. Exhibit 1.3 in

Chapter 1 presented my ideal-typical comparison of coming-of-age paradigms that contrasts tribal initiation themes with contemporary coming-of-age themes (Patton 1999a). A series of patterns is distilled into contrasting themes that create alternative ideal-types. The notion of "ideal-types" makes it explicit that the analyst has constructed and interpreted something that supersedes purely descriptive analysis.

In creating analyst-constructed typologies through inductive analysis, you take on the task of identifying and making explicit patterns that appear to exist but remain unperceived by the people studied. The danger is that analyst-constructed typologies impose a world of meaning on the participants

that better reflects the observer's world than the world under study. One way of testing analyst-constructed typologies is to present them to people whose world is being analyzed to find out if the constructions make sense to them.

> The best and most stringent test of observer constructions is their recognizability to the participants themselves. When participants themselves say, "yes, that is there, I'd simply never noticed it before," the observer can be reasonably confident that he has tapped into extant patterns of participation. (Lofland 1971:34)

Exhibit 8.6, using the problem of classifying people's ancestry, shows what can happen when indigenous and official constructions conflict, a matter of some consequence to those affected.

A good example of an analyst-generated typology comes from an evaluation of the National Museum of Natural History, Smithsonian Institution, done by Robert L. Wolf and Barbara L. Tymitz (1978). This has become a classic in the museum studies field. They conducted a naturalistic inquiry of viewers' reactions to the "Ice Age Mammals and Emergence of Man" exhibit. From their observations, they identified four different kinds of visitors to the exhibit. These descriptions are progressive in that each new category identifies a person more serious about the exhibit hall.

■ *The Commuter:* This is the person who merely uses the hall as a vehicle to get from the entry point to the exit point. . . .

■ *The Nomad:* This is a casual visitor, a person who is wandering through the hall, apparently open to becoming interested in something. The Nomad is not really sure why he or she is in the hall and

not really sure that s/he is going to find anything interesting in this particular exhibit hall. Occasionally the Nomad stops, but it does not appear that the nomadic visitor finds any one thing in the hall more interesting than any other thing.

■ *The Cafeteria Type:* This is the interested visitor who wants to get interested in something, and so the entire museum and the hall itself are treated as a cafeteria. Thus, the person walks along, hoping to find something of interest, hoping to "put something on his or her tray" and stopping from time to time in the hall. While it appears that there is something in the hall that spontaneously sparks the person's interest, we perceive this visitor has a predilection to becoming interested, and the exhibit provides the many things from which to choose.

■ *The V.I.P.—Very Interested Person:* This visitor comes into the hall with some prior interest in the content area. This person may not have come specifically to the hall, but once there, the hall serves to remind the V.I.P.'s that they were, in fact, interested in something in that hall beforehand. The V.I.P. goes through the hall much more carefully, much slower, much more critically— that is, he or she moves from point to point, stops, examines aspects of the hall with a greater degree of scrutiny and care. (Wolf and Tymitz 1978: 10-11)

This typology of types of visitors became important in the full evaluation because it permitted analysis of different kinds of museum experiences. Moreover, the evaluators recommended that when conducting interviews to get museum visitors' reactions to exhibits, the interview results should be dif-

117

find
ır ex-
mad
ɔ no-
ı the
ther

sted
1 in
ɔum
ete-
ɪop-
ɪing
ɪnd
ɪall.
ing
the
tor
ɪer-
ɪny

his
ɪne
his
·to
to
ct,
ɪe-
all
er,
he
n-
e-
ɪd

ɪe
it
ɪ-
ɪs
r-
o
ɪ-

EXHIBIT 8.6 Qualitative Analysis of Ancestry at the U.S. Census

To count different kinds of people—the job of the Census Bureau—you need categories to count them in. The long form of the 2000 census, given to 1 in 6 households, asked an open-ended, fill-in-the-blank question about "ancestry." Analysts then coded the responses into categories, 1 of 604 categories, up from 467 in 1980. The government doesn't ask about religion, so if people respond that they are Jewish, they don't get their ancestry counted. However, those who write in that they are Amish or Mennonite do get counted because those are considered cultural categories.

Ethnic minorities that cross national boundaries, such as French and Spanish Basques, and groups affected by geopolitical change, such as Czechs and Slovaks or groups within the former Yugoslavia, are counted in distinct categories. The Census Bureau, following advice from the U.S. State Department, differentiates Taiwanese Americans from Chinese Americans, a matter of political sensitivity.

Can Assyrians and Chaldeans be lumped together? When the Census Bureau announced that it would combine the two in the same ancestry code, an Assyrian group sued over the issue, but lost the lawsuit. Assyrian Americans trace their roots to a biblical-era empire covering much of what is now Iraq and believe that Chaldeans are a separate religious subgroup. A fieldworker for the Census Bureau did fieldwork on the issue.

"I went into places where there were young people playing games, went into restaurants, and places where older people gathered," says Ms. McKenney.... She paid a visit to Assyrian neighborhoods in Chicago, where a large concentration of Assyrian Americans live. At a local community center and later that day at the Assyrian restaurant next door, community leaders presented their case for keeping the ancestry code the same. Over the same period, she visited Detroit to look into the Chaldean matter....

"I found that many of the people, especially the younger people, viewed it as an ethnic group, not a religion," says Ms. McKenney. She and Mr. Reed (Census Bureau ancestry research expert) concurred that enough differences existed that the Chaldeans could potentially qualify as a separate ancestry group.

In a conference call between interested parties, a compromise was struck. Assyrians and Chaldeans would remain under a single ancestry code, but the name would no longer be Assyrian, it would be Assyrian/Chaldean/Syriac—Syriac being the name of the Aramaic dialect that Assyrians and Chaldeans speak. "There was a meeting of the minds between all the representatives, and basically it was a unified decision to say that we're going to go under the same name," says the Chaldean Federation's Mr. Yono. (Kulish 2001:1)

---

ferentially valued depending on the type of person being interviewed—commuter, nomad, cafeteria type, or VIP.

A different typology was developed to distinguish how visitors learn in a museum:

"Museum Encounters of the First, Second, and Third Kind," a take-off on the popular science fiction movie *Close Encounters of the Third Kind*, which referred to direct human contact with visitors from outer space.

- *Museum Encounters of the First Kind:* This encounter occurs in halls that use display cases as the primary approach to specimen presentation. Essentially, the visitor is a passive observer to the "objects of interest." Interaction is visual and may occur only at the awareness level. The visitor is probably not provoked to think or consider ideas beyond the visual display.

- *Museum Encounters of the Second Kind:* This encounter occurs in halls that employ a variety of approaches to engage the visitor's attention and/or learning. The visitor has several choices to become active in his/her participation. . . . The visitor is likely to perceive, question, compare, hypothesize, etc.

- *Museum Encounters of the Third Kind:* This encounter occurs in halls that invite high levels of visitor participation. Such an encounter invites the visitor to observe phenomena in process, to create, to question the experts, to contribute, etc. Interaction is personalized and within the control of the visitor. (Wolf and Tymitz 1978:39)

Here's a sample of a quite different classification scheme, this one developed from fieldwork by sociologist Rob Rosenthal (1994) as "a map of the terrain" of the homeless.

- *Skidders:* Most often women, typically in their 30s, grew up middle or upper class but "skidded" into homelessness as divorced or separated parents.

- *Street people:* Mostly men, often veterans, rarely married; highly visible net and know how to use the resources of the street.

- *Wingnuts:* People with severe mental problems, occasionally due to longterm alcoholism, a visible subgroup.

- *Transitory workers:* People with job skills and a history of full-time work who travel from town to town, staying months or years in a place, and then heading off to greener pastures.

Categories of how homeless people spend their time:

- Hanging out
- Getting by
- Getting ahead

As these examples illustrate, the first purpose of typologies is to distinguish aspects of an observed pattern or phenomenon *descriptively*. Once identified and distinguished, these types can later be used to make interpretations and they can be related to other observations to draw conclusions, but the first purpose is description based on an inductive analysis of the patterns that appear in the data.

⑤ The Intellectual and Mechanical Work of Analysis

Coding Data, Finding Patterns, Labeling Themes, and Developing Category Systems

*C*lassification is Ariadne's clue through the labyrinth of nature.

—George Sand, *Nouvelles Lettres d'un Voyageur,* 1869

119

Thus far, I've provided lots of examples of the fruit of qualitative inquiry: patterns, themes, categories, and typologies. Let's back up now to consider how you recognize patterns in qualitative data and turn those patterns into meaningful categories and themes. This chapter could have started with this section, but I think it's helpful to understand what kinds of findings can be generated from qualitative analysis before delving very deeply into the mechanics, especially because the mechanics vary greatly and are undertaken differently by analysts in different disciplines and working from divergent frameworks. That said, some guidance can be offered.

Raw field notes and verbatim transcripts constitute the undigested complexity of reality. Simplifying and making sense out of that complexity constitutes the challenge of content analysis. **Developing some manageable classification or coding scheme is the first step of analysis.** Without classification there is chaos and confusion. Content analysis, then, involves identifying, coding, categorizing, classifying, and labeling the primary patterns in the data. This essentially means analyzing the core *content* of interviews and observations to determine what's significant. In explaining the process, I'll describe it as done traditionally, which is without software, to highlight the thinking and mechanics involved. Software programs provide different tools and formats for coding, but the principles of the analytical process are the same whether doing it manually or with the assistance of a computer program.

I begin by reading through all of my field notes or interviews and making comments in the margins or even attaching pieces of paper or Post-it notes that contain my notions about what I can do with the different parts of the data. This constitutes the first cut at organizing the data into topics and files. Coming up with topics is like constructing an index for a book or labels for a file system: You look at what is there and give it a name, a label. The copy on which these topics and labels are written becomes the indexed copy of the field notes or interviews. Exhibit 8.7 shows a sampling of codes from the field note margins of the evaluation of the wilderness education program I described in the chapter on observation.

The shorthand codes are written directly on the relevant data passages, either in the margins or with an attached tab on the relevant page. Many passages will illustrate more than one theme or pattern. The first reading through the data is aimed at developing the coding categories or classification system. Then a new reading is done to actually start the formal coding in a systematic way. Several readings of the data may be necessary before field notes or interviews can be completely indexed and coded. Some people find it helpful to use colored highlighting pens—color coding different idea or concepts. Using self-adhesive colored dots or Post-it notes offers another option. Some use a color printer to print out transcripts in different colors to make it easy to track the source of a quote when cutting and pasting different quotes into a theme.

> If sensing a pattern or "occurrence" can be called *seeing,* then the encoding of it can be called *seeing as.* That is, you first make the observation that something important or notable is occurring, and then you classify or describe it. . . . [T]he *seeing as* provides us with a link between a new or emergent pattern and any and all patterns that we have observed and considered previously. It also provides a link to any and all patterns that others have observed and considered previously through reading. (Boyatzis 1998:4)

EXHIBIT 8.7	First-Cut Coding Examples: Sample Codes From the Field Note Margins

Code: Ps Re Prog (meaning: participants' reactions to the program)
Code: Ps Re Ps (participants' reactions to other participants)
Code: Ob PP (observations of participants' interactions)
Code: Ob SS (observations of staff's interactions)
Code: Ob SP (observations of staff/participant interactions)
Code: Phil (statements about program philosophy)
Code: Prc (examples of program processes)
Code: P/outs (effects of program on participants/outcomes)
Code: S-G (subgroup formations)
Code: GPrc (group process)
Code: C! (conflicts)
Code: C-PP (conflicts among participants)
Code: C-SP (conflicts between staff and participants)
Code: C-SS (conflicts among staff)

NOTE: P = participants, S = staff. These codes are from the field note margins of the evaluation of the wilderness education program described in the chapter on observation. The shorthand codes (abbreviations) are written in the margins directly on the relevant data passages or quotations. The full labels in parentheses are the designations for separate files that contain all similarly coded passages.

Where more than one person is working on the analysis, it is helpful to have each person (or small teams for large projects) develop the coding scheme independently, then compare and discuss similarities and differences. Important insights can emerge from the different ways in which two people look at the same set of data, a form of analytical triangulation.

Often an elaborate classification system emerges during coding, particularly in large projects where a formal scheme must be developed that can be used by several trained coders. In the study of evaluation use that is the basis for *Utilization-Focused Evaluation* (Patton 1997a), graduate students in the evaluation program at the University of Minnesota conducted lengthy interviews with 60 project officers, evaluators, and federal decision makers. We developed a comprehensive classification system that would provide easy access to the data by any of the student or faculty researchers. Had only one investigator been intending to use the data, such an elaborate classification scheme would not have been necessary. However, to provide access to several students for different purposes, every paragraph in every interview was coded using a systematic and comprehensive coding scheme made up of 15 general categories with subcategories. Portions of the codebook used to code the utilization of evaluation data appear in Appendix 8.1 at the end of this chapter as an example of one kind of qualitative analysis codebook. This codebook was developed from four sources: (a) the standardized open-ended questions used in interviewing, (b) review of the utilization literature for ideas to be examined and hypotheses to be reviewed, (c) our initial inventory review of the interviews in which two of us read all the

121

data and added categories for coding, and (d) a few additional categories added during coding when passages didn't fit well in the available categories.

Every interview was coded twice by two independent coders. Each individual code, including redundancies, was entered into our qualitative analysis database so that we could retrieve all passages (data) on any subject included in the classification scheme, with brief descriptions of the content of those passages. The analyst could then go directly to the full passages and complete interviews from which passages were extracted to keep quotations in context. In addition, the computer analysis permitted easy cross-classification and cross-comparison of passages for more complex analyses across interviews.

Some such elaborate coding system is routine for very rigorous analysis of a large amount of data. Complex coding systems with multiple coders categorizing every paragraph in every interview constitute a labor-intensive form of coding, one that would not be used for small-scale formative evaluation or action research projects. However, where data are going to be used by several people, or where data are going to be used over a long period of time, including additions to the data set over time, such a comprehensive and computerized system can be well worth the time and effort required.

Kibel (1999) developed a very sophisticated and comprehensive system for coding stories of successful outcomes attainment that he called "results mapping." His system permitted converting individualized stories into standardized categories that permitted aggregation, comparison, and even quantification. However, it required intensive training to use and proved too cumbersome and demanding for most human services and educational programs. As this was being

written, he had gone back to the drawing board and was working on a more usable coding framework to capture and code the stories of program participants in a standardized framework, an approach to be called "journey mapping."

Classifying and coding qualitative data produce a framework for organizing and describing what has been collected during fieldwork. (For published examples of coding schemes, see Bernard 1998:325-28, 387-89, 491-92, 624; Bernard 2000:447-50; Boyatzis 1998; Strauss and Corbin 1998; Miles and Huberman 1994.) This descriptive phase of analysis builds a foundation for the interpretative phase when meanings are extracted from the data, comparisons are made, creative frameworks for interpretation are constructed, conclusions are drawn, significance is determined, and, in some cases, theory is generated.

Convergence and Divergence in Coding and Classifying

In developing codes and categories, a qualitative analyst must first deal with the challenge of *convergence* (Guba 1978)—figuring out what things fit together. Begin by looking for *recurring regularities* in the data. These regularities reveal patterns that can be sorted into categories. Categories should then be judged by two criteria: *internal homogeneity* and *external heterogeneity*. The first criterion concerns the extent to which the data that belong in a certain category hold together or "dovetail" in a meaningful way. The second criterion concerns the extent to which differences among categories are bold and clear. "The existence of a large number of unassignable or overlapping data items is good evidence of some basic fault in the category system" (Guba 1978:53). The analyst then works back and forth between the data

and the classification system to verify the meaningfulness and accuracy of the categories and the placement of data in categories. If several different possible classification systems emerge or are developed, some priorities must be established to determine which are more important and illuminative. Prioritizing is done according to the utility, salience, credibility, uniqueness, heuristic value, and feasibility of the classification schemes. Finally, the category system or set of categories is tested for completeness.

1. The set should have internal and external plausibility, a property that might be termed "integratability." Viewed internally, the individual categories should appear to be consistent; viewed externally, the set of categories should seem to comprise a whole picture. . . .

2. The set should be reasonably inclusive of the data and information that do exist. This feature is partly tested by the absence of unassignable cases, but can be further tested by reference to the problem that the inquirer is investigating or by the mandate given the evaluator by his client/sponsor. If the set of categories did not appear to be sufficient, on logical grounds, to cover the facets of the problem or mandate, the set is probably incomplete.

3. The set should be reproducible by another competent judge. . . . The second observer ought to be able to verify that (a) the categories make sense in view of the data which are available, and (b) the data have been appropriately arranged in the category system. . . . The category system auditor may be called upon to attest that the category system "fits" the data and that the data have been properly "fitted into" it.

4. The set should be credible to the persons who provided the information which the set is presumed to assimilate. . . . Who is in a better position to judge whether the categories appropriately reflect their issues and concerns than the people themselves? (Guba 1978:56-57)

After analyzing for convergence, the mirror analytical strategy involves examining *divergence.* By this Guba means the analyst must "flesh out" the patterns or categories. This is done by processes of extension (building on items of information already known), bridging (making connections among different items), and surfacing (proposing new information that ought to fit and then verifying its existence). The analyst brings closure to the process when sources of information have been exhausted, when sets of categories have been saturated so that new sources lead to redundancy, when clear regularities have emerged that feel integrated, and when the analysis begins to "overextend" beyond the boundaries of the issues and concerns guiding the analysis. Divergence also includes careful and thoughtful examination of data that doesn't seem to fit including *deviant cases* that don't fit the dominant identified patterns.

This sequence, convergence then divergence, should not be followed mechanically, linearly, or rigidly. The processes of qualitative analysis involve both technical and creative dimensions. As noted early in this chapter, no abstract processes of analysis, no matter how eloquently named and finely described, can substitute for the skill, knowledge, experience, creativity, diligence, and work of the qualitative analyst. "The task of converting field notes and observations about issues and concerns into systematic categories is a difficult one. No infallible procedure exists for performing it" (Guba 1978:53).

Determining Substantive Significance

In lieu of statistical significance, qualitative findings are judged by their *substantive significance*. The analyst makes an argument for substantive significance in presenting findings and conclusions, but readers and users of the analysis will make their own value judgments about significance. In determining substantive significance, the analyst addresses these kinds of questions:

- How solid, coherent, and consistent is the evidence in support of the findings? (Triangulation, for example, can be used in determining the strength of evidence in support of a finding.)
- To what extent and in what ways do the findings increase and deepen understanding of the phenomenon studied (*Verstehen*)?
- To what extent are the findings consistent with other knowledge? (A finding supported by and supportive of other work has confirmatory significance. A finding that breaks new ground has discovery or innovative significance.)
- To what extent are the findings useful for some intended purpose (e.g., contributing to theory, informing policy, summative or formative evaluation, or problem solving in action research)?

The qualitative analyst's effort at uncovering patterns, themes, and categories includes using both creative and critical faculties in making carefully considered judgments about what is really significant and meaningful in the data. Since qualitative analysts do not have statistical tests to tell them when an observation or pattern is significant, they must rely first on their own intelligence, experience, and judgment; second,
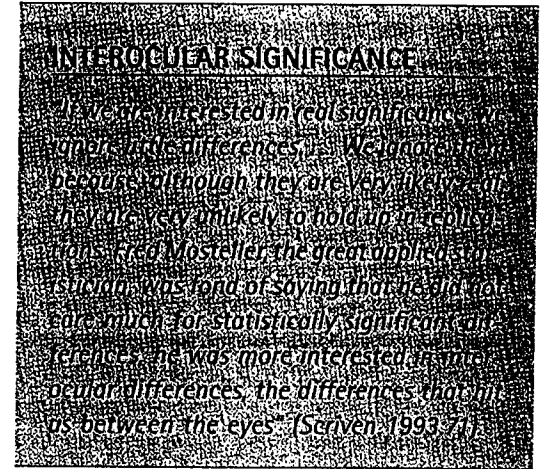
they should take seriously the responses of those who were studied or participated in the inquiry; and third, the researcher or evaluator should consider the responses and reactions of those who read and review the results. Where all three—analyst, those studied, and reviewers—agree, one has *consensual validation* of the substantive significance of the findings. Where disagreements emerge, which is the more usual case, you get a more interesting life and the joys of debate.

Determining substantive significance can involve the making of the qualitative analyst's equivalent of Type I and Type II errors from statistics: The analyst may decide that something is not significant when in fact it is, or, conversely, the analyst may attribute significance to something that is meaningless. A story illustrates this problem of making judgments about what is really significant.

Halcolm was approached by a woman who handed him something. Without hesitation, Halcolm returned the object to the woman. The many young disciples who followed Halcolm to learn his wisdom began arguing among themselves about the special meaning of this interchange. A variety of interpretations were offered.

124

When Halcolm heard of the argument among his young followers, he called them together and asked each one to report on the significance of what they had observed. They offered a variety of interpretations. When they had finished he said, "The real purpose of the exchange was to enable me to show you that you are not yet sufficiently masters of observation to know when you have witnessed a meaningless interaction."

⑤ Logical Analysis

While working inductively, the analyst is looking for emergent patterns in the data. These patterns, as noted in preceding sections, can be represented as dimensions, categories, classification schemes, themes, and categories. Once some dimensions have been constructed, using either participant-generated constructions or analyst-generated constructions, it is sometimes useful to cross-classify different dimensions to generate new insights about how the data can be organized and to look for patterns that may not have been immediately obvious in the initial, inductive analysis. Creating cross-classification matrices is an exercise in logic.

The logical process involves creating potential categories by crossing one dimension or typology with another, and then working back and forth between the data and one's logical constructions, filling in the resulting matrix. This logical system will create a new typology, all parts of which may or may not actually be represented in the data. Thus, the analyst moves back and forth between the logical construction and the actual data in a search for meaningful patterns.
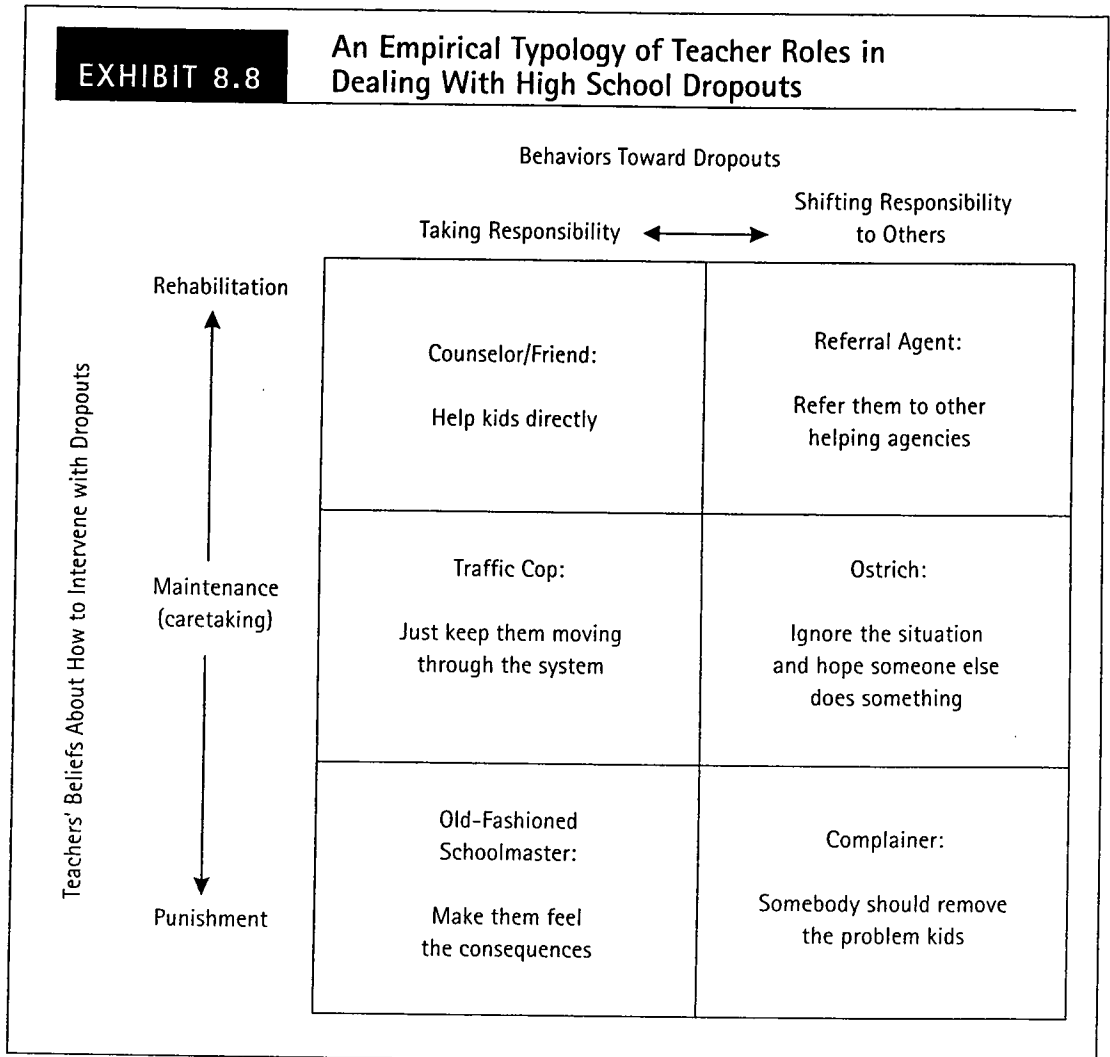
In the high school dropout program described earlier, the focus of the program was reducing absenteeism, skipping of classes, and tardiness. An external team of consul-

tant/change agents worked with teachers in the school to help them develop approaches to the dropout problem. Observations of the program and interviews with the teachers gave rise to two dimensions. The first dimension distinguished *teachers' beliefs about what kind of programmatic intervention was effective* with dropouts, that is, whether they primarily favored maintenance (caretaking or warehousing of kids to just keep the schools running), rehabilitation efforts (helping kids with their problems), or punishment (no longer letting them get away with the infractions they had been committing in the past). *Teachers' behaviors toward dropouts* could be conceptualized along a continuum from taking direct responsibility for doing something about the problem, at one end, to shifting responsibility to others at the opposite end. Exhibit 8.8 shows what happens when these two dimensions are crossed. Six cells are created, each of which represents a different kind of teacher role in response to the program.

The evaluator analyst working with these data had been struggling in the inductive analysis to find the patterns that would express the different kinds of teacher roles manifested in the program. He had tried several constructions, but none of them quite seemed to work. The labels he came up with were not true to the data. When he described to me the other dimensions he had generated, I suggested that he cross them, as shown in Exhibit 8.8. When he did, he said that "the whole thing immediately fell into place." Working back and forth between the matrix and the data, he generated a full descriptive analysis of diverse and conflicting teacher roles.

The description of teacher roles served several purposes. First, it gave teachers a mirror image of their own behaviors and attitudes. It could thus be used to help teach-

| EXHIBIT 8.8 | An Empirical Typology of Teacher Roles in Dealing With High School Dropouts |

Behaviors Toward Dropouts

Taking Responsibility ◄——► Shifting Responsibility to Others

Teachers' Beliefs About How to Intervene with Dropouts

Rehabilitation ↑ — Maintenance (caretaking) — ↓ Punishment

	Taking Responsibility	Shifting Responsibility to Others
Rehabilitation	**Counselor/Friend:** Help kids directly	**Referral Agent:** Refer them to other helping agencies
Maintenance (caretaking)	**Traffic Cop:** Just keep them moving through the system	**Ostrich:** Ignore the situation and hope someone else does something
Punishment	**Old-Fashioned Schoolmaster:** Make them feel the consequences	**Complainer:** Somebody should remove the problem kids

ers make more explicit their own understanding of roles. Second, it could be used by the external team of consultants to more carefully gear their programmatic efforts toward different kinds of teachers who were acting out the different roles. The matrix makes it clear that an omnibus strategy for helping teachers establish a program that would reduce dropouts would not work in this school; teachers manifesting different roles would need to be approached and worked with in different ways. Third, the

description of teacher roles provided insights into the nature of the dropout problem. Having identified the various roles, the evaluator analyst had a responsibility to report on the distribution of roles in this school and the observed consequences of that distribution.

One must be careful about this kind of logical analysis. It is easy for a matrix to begin to manipulate the data as the analyst is tempted to force data into categories created by the cross-classification to fill out the ma-

trix and make it work. Logical analysis to generate new sensitizing concepts must be tested out and confirmed by the actual data. **Such logically derived sensitizing concepts provide conceptual hypotheses to test.** Levin-Rozalis (2000), following American philosopher Charles Sanders Pierce of the pragmatic school of thought, suggests labeling the logical generation and discovery of hypotheses and findings *abduction* to distinguish such logical analysis from data-based inductive analysis and theory-derived deductive analysis.

Denzin (1978b) has explained abduction in qualitative analysis as a combination of inductive and deductive thinking with logical underpinnings:

> Naturalists inspect and organize behavior specimens in ways which they hope will permit them to progressively reveal and better understand the underlying problematic features of the social world under study. They seek to ask the question or set of questions which will make that world or social organization understandable. They do not approach that world with a rigid set of preconceived hypotheses. They are initially directed toward an interest in the routine and taken-for-granted features of that world. They ask how it is that the persons in question know about producing orderly patterns of interaction and meaning. . . . They do not use a full-fledged deductive hypothetical scheme in thinking and developing propositions. Nor are they fully inductive, letting the so-called "facts" speak for themselves. Facts do not speak for themselves. They must be interpreted. Previously developed deductive models seldom conform with empirical data that are gathered. The method of abduction combines the deductive and inductive models of proposition development and theory construction. It can be defined as *working from consequence back to cause or antecedent.* The observer records the

occurrence of a particular event, and then works back in time in an effort to reconstruct the events (causes) that produced the event (consequence) in question. (pp. 109-10)

Famous fictional detective Sherlock Holmes relied on abduction more than deduction or induction, at least according to a review by William Sanders (1976) of Holmes's analytical thinking in *The Sociologist as Detective.* We've already suggested that the qualitative analyst is part scientist and part artist. Why not add the qualitative analyst as detective? The empty cell of a logically derived matrix (the cell created by crossing two dimensions for which no name or label immediately occurs) creates an intersection of a possible consequence and antecedent that begs for abductive exploration and explanation. Each such intersection of consequence and antecedent sensitizes the analyst to the possibility of a category of activity or behavior that either has been overlooked in the data or is logically a possibility in the setting but has not been manifested. The latter cases are important to note because their importance derives from the fact that they did not occur. The next section will look in detail at a process/outcomes matrix ripe with abductive possibilities.

Nick Smith (1980) used a matrix to draw important distinctions among different kinds of evaluation use by asking if "techniques of effective evaluation utilization differ with regard to audience or entity studied." His matrix crossed a programs/policies dimension (what can be studied?) with a program managers/policymakers distinction (who is to be aided?) to show different kinds of utilization in each case. Exhibit 8.9 shows a matrix for mapping stakeholders' stakes in a program or policy. This matrix can be used to guide data collection as well as analysis. Later this chapter presents a process/outcomes matrix for crossing

MODUS OPERANDI ANALYSIS

Modus operandi (MO) was conceptualized by evaluation theorist Michael Scriven (1976) as a way of inferring causality when experimental designs are impractical or inappropriate. The MO approach, drawing from forensic science, makes the inquirer a detective. Detectives compare clues discovered at a crime scene to known patterns of possible suspects. Those suspects whose MO (method of operating) does not fit the crime scene pattern are eliminated from further investigation.

Translated to research and evaluation, the inquirer/detective observes some pattern and makes a list of possible causes. Evidence from the inquiry is compared to the list of suspects (possible causes). Those possible causes that do not fit the pattern of evidence can be eliminated from further consideration. Following the autopsy-like logic of Occam's razor, as each possible cause is compared to the evidence, that cause supported by the preponderance of evidence and offering the simplest interpretation among competing possibilities is preferred and considered most likely.

program processes with program outcomes as a qualitative analysis framework.

To study how schools used planning and evaluation processes, Campbell (1983) developed a 500-cell matrix (Exhibit 8.10) that begins (but just begins) to reach the outer limits of what one can do in three-dimensional space. Campbell used this matrix to guide data collection and analysis in studying how the mandated, statewide educational planning, evaluation, and reporting system in Minnesota was used. She examined 5 levels of use (high school, . . . commu-

nity, district), 10 components of the statewide project (planning, goal-setting, . . . student involvement), and 10 factors affecting utilization (personal factor, political factors, . . .). Exhibit 8.10 again illustrates matrix thinking for both data organization and analytical/conceptual purposes.

Miles and Huberman (1994) have provided a rich source of ideas and illustrations of how to use matrices in qualitative analysis. They include examples of a time-ordered matrix, role-ordered matrix, role-by-time matrix, role-by-group matrix, conceptually clustered matrix, site dynamics matrix, and predictor-outcome matrix, among others. Their *Sourcebook* provides a variety of ideas for analytical approaches to qualitative data including a variety of concept mapping and visual display techniques.

Other ways of graphing and mapping findings include concept mapping and cognitive mapping. For a detailed discussion of concept mapping as a way of visually displaying data to facilitate analytic clarity and depicting relationships in a network or system, see Trochim (1989). For an example of cognitive mapping as a way of displaying qualitative results showing the "structure and content of decision schemas" among senior managers, see Clarke and Mackaness (2001).

🔳 A Process/Outcomes Matrix

The linkage between processes and outcomes constitutes such a fundamental issue in many program evaluations that it provides a particularly good focus for illustrating qualitative matrix analysis. As discussed in Chapter 4, qualitative methods can be particularly appropriate for evaluation where program processes, impacts, or both are largely unspecified or difficult to measure. This can be the case because the outcomes are meant to be individualized;

EXHIBIT 8.9	Mapping Stakeholders' Stakes		
	Estimate of Various Stakeholders' Inclination Toward the Program		
How high are the stakes for various primary stakeholders?	Favorable	Neutral or Unknown	Antagonistic
High			
Moderate			
Low			

SOURCE: Patton (1997a:344).
NOTE: Construct illustrative case studies for each cell based on fieldwork.

sometimes the program is simply uncertain what its outcomes will be; and, in many programs, neither processes nor impacts have been carefully articulated. Under such conditions, one purpose of the evaluation may be to illuminate program processes, program impacts, and the linkages between the two. This task can be facilitated by constructing a process/outcomes matrix to organize the data.

Exhibit 8.11 (p. 474) shows how such a matrix can be constructed. Major program processes or identified implementation components are listed along the left side. Types or levels of outcomes are listed across the top. The category systems for program processes and outcomes are developed from the data in the same way that other typologies are constructed (see previous sections). The cross-classification of any process with any outcome produces a cell in the matrix; for example, the first cell in Exhibit 8.11 is created by the intersection of process 1 with outcome a. The information that goes in cell 1a (or any other cell in the matrix) describes
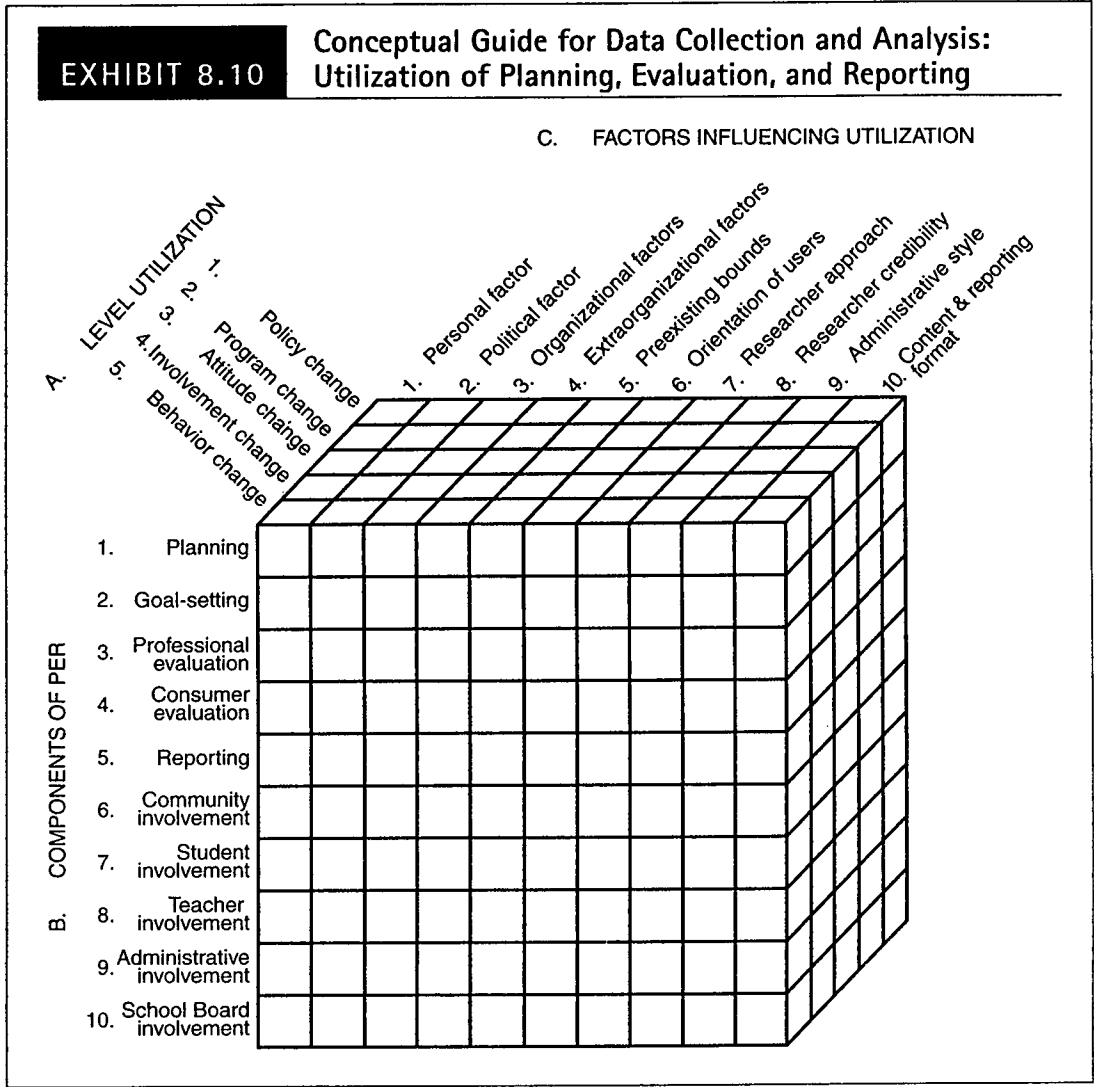
linkages, patterns, themes, experiences, content, or actual activities that help us understand the relationships between processes and outcomes. Such relationships may have been identified by participants themselves during interviews or discovered by the evaluator in analyzing the data. In either case, the process/outcomes matrix becomes a way of organizing, thinking about, and presenting the qualitative connections between program implementation dimensions and program impacts.

An example will help make the notion of the process/outcomes matrix more concrete. Suppose we have been evaluating a juvenile justice program that places delinquent youth in foster homes. We have visited several foster homes, observed what the home environments are like, and interviewed the juveniles, the foster home parents, and the probation officers. A *regularly recurring process theme* concerns the importance of "letting kids learn to make their own decisions." A *regularly recurring outcomes theme* involves "keeping the kids

129

im

,con-
p us
ı pro-
ıships
ıpants
vered
In ei-
ix be-
ıbout,
ctions
imen-

ion of
con-
g a ju-
delin-
have
what
inter-
ə par-
ularly
npor-
their
; out-
kids

	Conceptual Guide for Data Collection and Analysis:
EXHIBIT 8.10	**Utilization of Planning, Evaluation, and Reporting**

C.   FACTORS INFLUENCING UTILIZATION

A. LEVEL UTILIZATION
1.
2.
3. *Policy change*
4. *Program change*
5. *Attitude change*
Involvement change
Behavior change

1.   Personal factor
2.   Political factor
3.   Organizational factors
4.   Extraorganizational factors
5.   Preexisting bounds
6.   Orientation of users
7.   Researcher approach
8.   Researcher credibility
9.   Administrative style
10. Content & reporting format

B.   COMPONENTS OF PER

1.   Planning
2.   Goal-setting
3.   Professional evaluation
4.   Consumer evaluation
5.   Reporting
6.   Community involvement
7.   Student involvement
8.   Teacher involvement
9.   Administrative involvement
10. School Board involvement

straight" (reduced recidivism). By crossing the program process ("kids making their own decisions") with the program outcome ("keeping kids straight"), we create a data analysis question: What actual decisions do juveniles make that are supposed to lead to reduced recidivism? We then carefully review our field notes and interview quotations looking for data that help us understand how people in the program have answered this question based on their actual behaviors and practices. By describing what

decisions juveniles actually make in the program, the decision makers to whom our findings are reported can make their own judgments about the strength or weakness of the linkage between this program process and the desired outcome. Moreover, once the process/outcomes descriptive analysis of linkages has been completed, the evaluator is at liberty to offer interpretations and judgments about the nature and quality of this process/outcomes connection.

130

| EXHIBIT 8.11 | Matrix of Linkages Between Program Processes and Impacts |

Types or Levels of Program Outcomes

a   b   c   d

Program
Processes
or
Implementation
Components

1

2   LINKAGES EXPRESSED AS THEMES,
PATTERNS, QUOTATIONS, PROGRAM

3   CONTENT OR ACTUAL ACTIVITIES

4

SOURCE: Campbell (1983).

An Analysis Example: Recognizing Processes, Outcomes, and Linkages in Qualitative Data

Because of the centrality of the sensitizing concepts "program process" and "program outcome" in evaluation research, it may be helpful to provide a more detailed descrip-tion of how these concepts can be used in qualitative analysis. How does one recog-nize a program process? Learning to identify and label program processes is a critical evaluation skill. This sensitizing notion of "process" is a way of talking about the com-mon action that cuts across program activi-ties, observed interactions, and program content. The example I shall use involves

131

data from the wilderness education program I evaluated and discussed throughout the observations chapter (Chapter 6). That program, titled the Southwest Field Training Project, used the wilderness as a training arena for professional educators in the philosophy and methods of experiential education by engaging those educators in their own experiential learning process. Participants went from their normal urban environments into the wilderness for 10 days at a time, spending at least one day and night completely alone in some wilderness spot "on solo." At times, while backpacking, the group was asked to walk silently so as not to be distracted from the wilderness sounds and images by conversation. In group discussions, participants were asked to talk about what they had observed about the wilderness and how they felt about being in the wilderness. Participants were also asked to write about the wilderness environment in journals. *What do these different activities have in common, and how can that commonality be expressed?*

We begin with several different ways of abstracting and labeling the underlying process:

- Experiencing the wilderness
- Learning about the wilderness
- Appreciating the wilderness
- Immersion in the environment
- Developing awareness of the environment
- Becoming conscious of the wilderness
- Developing sensitivity to the environment

Any of these phrases, each of which consists of some verb form (experiencing, learning, developing, and so on) and some noun form (wilderness, environment), captures some

nuance of the process. The qualitative analyst works back and forth between the data (field notes and interviews) and his or her conception of what it is that needs to be expressed to find the most fitting language to describe the process. What language do people in the program use to describe what those activities and experiences have in common? What language comes closest to capturing the essence of this particular process? What level of generality or specificity will be most useful in separating out this particular set of things from other things? How do program participants and staff react to the different terms that could be used to describe the process?

It's not unusual during analysis to go through several different phrases before finally settling on exact language that will go into a final report. In the Southwest Field Training Project, we began with the concept label "Experiencing the wilderness." However, after several revisions, we finally described the process as "developing sensitivity to the environment" because this broader label permitted us to include discussions and activities that were aimed at helping participants understand how they were affected by and acted in their normal institutional environments. "Experiencing the wilderness" became a specific subprocess that was part of the more global process of "developing sensitivity to the environment." Program participants and staff played a major role in determining the final phrasing and description of this process.

Below are other processes identified as important in the implementation of the program:

- Encountering and managing stress
- Sharing in group settings
- Examining professional activities, needs, and commitments

- Assuming responsibility for articulating personal needs

- Exchanging professional ideas and resources

- Formally monitoring experiences, processes, changes, and impacts

As you struggle with finding the right language to communicate themes, patterns, and processes, keep in mind that there is no absolutely "right" way of stating what emerges from the analysis. There are only more and less useful ways of expressing what the data reveal.

Identifying and conceptualizing program outcomes and impacts can involve induction, deduction, and/or logical analysis. *Inductively*, the evaluator analyst looks for changes in participants, expressions of change, program ideology about outcomes and impacts, and ways that people in the program make distinctions between "those who are getting *it*" and "those who aren't getting *it*" (where *it* is the desired outcome). In highly individualized programs, the statements about change that emerge from program participants and staff may be global. Such outcomes as "personal growth," increased "awareness," and "insight into self" are difficult to operationalize and standardize. That is precisely the reason qualitative methods are particularly appropriate for capturing and evaluating such outcomes. The task for the evaluator analyst, then, is to describe what actually happens to people in the program and what they say about what happens to them. Appendix 8.3 at the end of this chapter presents portions of the report describing the effects on participants of their experiences in the wilderness education program. The data come from in-depth, open-ended interviews. This report excerpt shows how descriptive data (direct quotations) are used to support and explain inductive thematic analysis.

Deductively, the evaluator analyst may draw from outcomes identified in similar programs or from goal statements found in program proposals, brochures, and planning documents that were used to guide data collection.

Logically (or abductively), constructing a process/outcomes matrix can suggest additional possibilities. That is, where data on both program processes and participant outcomes have been sorted, analysis can be deepened by organizing the data through a logical scheme that links program processes to participant outcomes. Such a logically derived scheme was used to organize the data in the Southwest Field Training Project. First, a classification scheme that described different types of outcomes was conceptualized:

(a) changes in skills,

(b) changes in attitudes,

(c) changes in feelings,

(d) changes in behaviors, and

(e) changes in knowledge.

These general themes provided the reader of the report with examples of and insights into the kinds of changes that were occurring and how those changes that were perceived by participants to be related to specific program processes. I emphasize that the process/outcomes matrix is merely an organizing tool; the data from participants themselves and from field observations provide the actual linkages between processes and outcomes.

What was the relationship between the program process of "developing sensitivity to the environment" and these individual-level outcomes? Space permits only a few examples from the data.

133

Skills: "Are you kidding? I learned how to survive without the comforts of civilization. I learned how to read the terrain ahead and pace myself. I learned how to carry a heavy load. I learned how to stay dry when it's raining. I learned how to tie a knot so that it doesn't come apart when pressure is applied. You think those are metaphors for skills I need in my work? You're damn right they are."

Attitudes: "I think it's important to pay attention to the space you're in. I don't want to just keep going through my life oblivious to what's around me and how it affects me and how I affect it."

Feelings: "Being out here, especially on solo, has given me confidence. I know I can handle a lot of things I didn't think I could handle."

Behaviors: "I use my senses in a different way out here. In the city you get so you don't pay much attention to the noise and the sounds. But listening out here I've also begun to listen more back there. I touch more things too, just to experience the different textures."

Knowledge: "I know about how this place was formed, its history, the rock formations, the effects of the fires on the vegetation, where the river comes from and where it goes."

A different way of thinking about organizing data around outcomes was to think of different levels of impact: effects at the individual level, effects on the group, and effects on the institutions from which participants came into the program. The staff hoped to have impacts at all of these levels. Thus, it also was possible to organize the data by looking at what themes emerged when program processes were crossed with levels of impact. How did "developing sensitivity to the environment" affect individuals? How did the process of "developing sensitivity to the environment" affect the group? What was the effect of "developing sensitivity to the environment" on the institutions to which participants returned after their wilderness experiences? The process/outcomes matrix thus becomes a way of asking questions of the data, an additional source of focus in looking for themes and patterns in hundreds of pages of field notes and interview transcriptions.

🖺 Interpreting Findings

S imply observing and interviewing do not ensure that the research is qualitative; the qualitative researcher must also interpret the beliefs and behaviors of participants.

—Valerie J. Janesick (2000:387)

Interpreting for Meaning

Qualitative interpretation begins with elucidating meanings. The analyst examines a story, a case study, a set of interviews, or a collection of field notes and asks, What does this mean? What does this tell me about the nature of the phenomenon of interest? In asking these questions, the analyst works back and forth between the data or story (the evidence) and his or her own perspective and understandings to make sense of the ev-