

Matthew B. Miles ■ A. Michael Huberman ■ Johnny Saldaña
Arizona State University

Qualitative Data Analysis

A Methods Sourcebook

Edition 3



Los Angeles | London | New Delhi
Singapore | Washington DC

Brief Table of Contents

List of Displays	xiii
Preface to the Third Edition by Johnny Saldaña	xvii
Acknowledgments From the Second Edition by Matthew B. Miles and A. Michael Huberman	xxi
About the Authors	xxiii

Part One – The Substantive Start 1

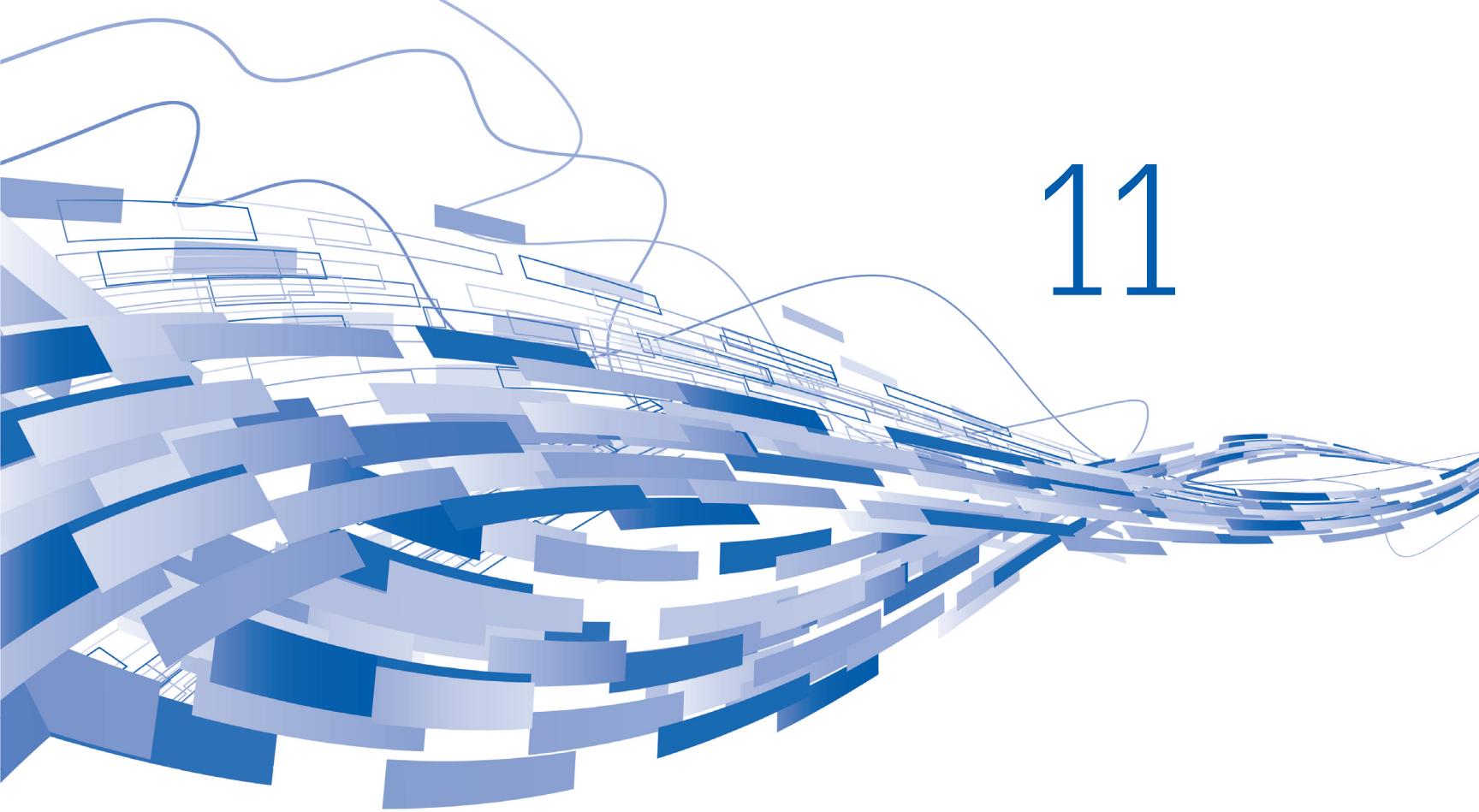
Chapter 1 - Introduction	3
Chapter 2 - Research Design and Management	17
Chapter 3 - Ethical Issues in Analysis	55
Chapter 4 - Fundamentals of Qualitative Data Analysis	69

Part Two – Displaying the Data 105

Chapter 5 - Designing Matrix and Network Displays	107
Chapter 6 - Methods of Exploring	121
Chapter 7 - Methods of Describing	161
Chapter 8 - Methods of Ordering	193
Chapter 9 - Methods of Explaining	221
Chapter 10 - Methods of Predicting	255

Part Three – Making Good Sense 273

Chapter 11 - Drawing and Verifying Conclusions	275
Chapter 12 - Writing About Qualitative Research	323
Chapter 13 - Closure	339
Appendix – An Annotated Bibliography of Qualitative Research Resources	345
References	363
Author Index	369
Subject Index	373



Drawing and Verifying Conclusions

Chapter Summary

This chapter offers additional analytic tactics for generating meaning from data and for testing or confirming findings. Standards for assessing the quality of conclusions are proposed, along with methods for documenting a researcher's analytic processes.

Contents

Introduction

Tactics for Generating Meaning

1. Noting Patterns, Themes
2. Seeing Plausibility
3. Clustering
4. Making Metaphors
5. Counting
6. Making Contrasts/Comparisons
7. Partitioning Variables

8. Subsuming Particulars Into the General
9. Factoring
10. Noting the Relations Between Variables
11. Finding Intervening Variables
12. Building a Logical Chain of Evidence
13. Making Conceptual/Theoretical Coherence

Tactics for Testing or Confirming Findings

1. Checking for Representativeness
2. Checking for Researcher Effects
3. Triangulating
4. Weighting the Evidence
5. Checking the Meaning of Outliers
6. Using Extreme Cases
7. Following Up Surprises
8. Looking for Negative Evidence
9. Making If-Then Tests
10. Ruling Out Spurious Relations
11. Replicating a Finding
12. Checking Out Rival Explanations
13. Getting Feedback From Participants

Standards for the Quality of Conclusions

- Objectivity/Confirmability
- Reliability/Dependability/Auditability
- Internal Validity/Credibility/Authenticity
- External Validity/Transferability/Fittingness
- Utilization/Application/Action Orientation

Analytic Documentation

The Problem

Illustration

Closure and Transition

Introduction

This is a substantial 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*—that is, ways of drawing and verifying conclusions that you use during the process. These tactics have been named in **boldface** in the preceding chapters. Now, we discuss each in detail. Following that, we turn to the question of the goodness—the *quality*—of the conclusions you reach. The chapter concludes with a look at the documentation of qualitative analyses.

The creativity scholar Sir Ken Robinson is attributed with offering this cautionary advice about making a convincing argument: “Without data, you’re just another person with an opinion.” We add to that the following: “Without verification, you’re just another researcher with a hunch.”

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, sometimes referring back to previous sections (where the tactic was noted in **boldface**). If we can muster advice, we present that too. But the real test of these tactics comes in the using.

People are meaning finders and meaning makers; they can usually 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 trustworthy and “right.” The following section discusses tactics for testing or confirming meanings, minimizing bias, and ensuring to your best ability the quality of the 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 analog of a familiar quantitative technique; **noting the 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 initial displays, you often note recurring patterns, themes, or “gestalts” that pull together many separate pieces of data. Something jumps out at you and suddenly makes sense.

Some examples of patterns from the school improvement study were

- 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, and
- the use of “administrative latitude”—the 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 bounded context.

Pattern finding can be very productive when the number of cases and/or the data overload is severe. McCammon et al. (2012), for example, analyzed responses from 234 returned open-ended surveys. The patterns constructed from the mixed-methods data resulted in 30 major categories, which were then synthesized into three major themes, then into one key assertion.

The human mind finds patterns almost intuitively; it needs no how-to advice. But patterns don't just happen; we construct them from our observations of reoccurring phenomena. The important thing is to be able to (a) see added evidence of the same pattern and (b) remain open to disconfirming evidence when it appears. 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 can represent useful knowledge.

2. Seeing Plausibility

It often happens during analysis that a conclusion is plausible, makes good sense, and “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 (see the films *π*, *A Beautiful Mind*, and *The Number 23*). So plausibility can easily become the refuge of analysts who are too ready to jump to conclusions.

During the documentation of our own analysis efforts, we often found ourselves giving the “plausibility” basis for the 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: McCammon et al. (2012) analyzed their survey data by *gender* to observe whether men and women responded differently to their high school arts experiences. Saldaña hypothesized and “felt sure” that major differences would appear, based on his earlier research with audience response trends. But after data analysis, no *major* gender differences were found. There were nuances of experience that women remembered and men did not, but the projected differences were just not present in the data. Saldaña even thought, “This can’t be right,” and he subjected the qualitative data to reanalysis tactics—**counting** and finding new ways of **making contrasts/comparisons** with the data—but nothing new was discovered. (The most revealing differences actually occurred between other stratifications, such as age cohort.)

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 a *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 will discuss the tactic of **following up surprises** later.)

Most of the conclusions drawn during analysis are substantive, based on the content. But the analyst is constantly drawing *procedural* conclusions along the way as well: to transpose two rows in a matrix, to add or discard a column, to collapse the data into a summary table, and to change a decision rule for data entry. It's important to log and report the procedural decisions made (the final matrix, the operative decision rules, etc.).

3. Clustering

In daily life, we're constantly clumping things into classes, categories, and bins: Things that do not move around but grow are called "plants," things that move around and have babies are called "animals," and things that move around, have four wheels, have an engine run by fossil fuels or electricity, and carry people are called "automobiles." Most categories require other categories to define them: "wheel," "engine," "babies."

The next step often is to figure out what to call clusters. For example, the process of transforming hundreds of pages of interview transcripts and field notes into a one-page matrix or network display might be called "condensing" the data. But to others, it might be called "distilling," "synthesizing," "abstracting," "transforming," and even the abhorrent "reducing" the data. Even though these are different words and processes, they *kind of* mean the same thing. We have clustered them and now need to assign a word or phrase that represents all of them. Shall one of the six words be chosen as their representative? Or shall we compose a completely new umbrella label for the six of them? (This is where a thesaurus comes in handy. But, ultimately, it's your call as the analyst.)

Just as we are pattern-finding people, we are also cluster-making people. We categorize and organize the vast array of things in our lives to feel a sense of comfort and security in their ordering. Roles, rules, relationships, routines, and rituals are clusters of patterned actions—what some generically label as our habits (Duhigg, 2012). Clustering is a tactic that can be applied at many levels to qualitative data: at the level of events or acts, of individual participants, of processes, of settings/locale, of sites or cases as wholes, of time periods, and so on. In all instances, we're trying to understand a phenomenon better by *grouping* and then *conceptualizing* objects that have similar patterns or characteristics.

But, occasionally, clusters are not always mutually exclusive and may overlap. For example, a student most often is in *either* an undergraduate *or* a graduate program. That's easy enough. But what's the undergraduate student's major? Well, it's complicated. He's a double major in social work and religious studies, *and* he's concurrently enrolled in the university's Honors College. Yes, he's male. And gay. And Hispanic. And a first-generation college student. And so on. Life is not neatly bounded and discrete as it once might have seemed. As the analyst works at clustering, the 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, and floor polishers) into sensible clusters. Clustering is our best attempt to categorize what seems to belong together.

Here are some additional examples.

Focusing on *participants*, we asked high school students to name the cliques they observed at their school. They identified clusters such as “preppies,” “jocks,” “goths,” “skaters,” “band kids,” “stoners,” “math geeks,” “cheerleaders,” and so on.

At the level of *processes*, we clustered the activities involved in coping with the problems of later implementation of an innovation: “reaching up,” “improving,” “debugging,” “refining,” “integrating,” “adapting,” and “extending.” Processes are best identified in the form of gerunds (“-ing” words) to connote action.

At the level of both *processes* and *phases*, 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* where site-specific *actions* occur. For example, in schools, we might sort places where people interact into the following clusters:

- Formal instructional (classroom, gym)
- Informal instructional (library, media center)
- Formal adult work (meeting room, office)
- Informal adult association (faculty lounge, corridor)
- Mixed (cafeteria, playground)

Clustering also can be seen as a process of moving to higher levels of abstraction (see **subsuming particulars into the general** later in this chapter).

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, participants, processes, settings, and so on—into those categories. When less complex things are being sorted (events, participants, etc.), the clustering tactic typically relies on aggregation and comparison (What things are like each other/unlike each other?) and is closely interwoven with the creation and use of attribute codes.

Clustering techniques do not have to be completely self-invented. There is a long tradition of content-analytic, coding, and categorizing techniques dealing directly with issues of unitizing and clustering qualitative data that can be very helpful (see Richards, 2009; Saldaña, 2013; Schreier, 2012).

4. Making Metaphors

The notion that research should focus only on matter-of-fact, literal 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, poverty, misery). Metaphors, seen as one major type of trope or literary device, involve

comparing two things via their similarities and ignoring their differences. 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.

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 looked 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 “a parent’s separation anxiety,” would. Our metaphor lets us see new theoretical possibilities (e.g., maybe if socialization for independence is weak, the child will regress).

What else is true of metaphors? They are *data-condensing 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, the treatment of those perceived as “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 ask, “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. Metaphoric thinking effectively unites reason and imagination. Lakoff and Johnson’s (1980) *Metaphors We Live By* is essential reading on this topic.

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. So a few words of advice for metaphor makers:

- 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.
- Don't look for overarching metaphors too early in the study. It distracts you from fieldwork, and it leads to hasty judgments that clamp down on the meaning of what you are studying. You start to look around less, and you project the metaphor on things that are, at best, remotely related to it.
- Being cognitively playful helps generate metaphors. Ask yourself, "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. Instead of the social or the personal, go to biological, mechanical, or spatial domains to find useful metaphors.
- Interaction helps. Groups stimulate their members' thinking by increasing the inputs, bringing in ideas from a new angle, and creating a contagiously playful thinking environment.
- 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.

There are other literary devices available to qualitative inquirers, such as 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). Gibbs (2007) describes how romance, tragedy, and other literary genres can be used effectively for narrative inquiry.

5. Counting

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*, which *matter more* than others, which *go together*, and so on. When we say that something is "important," "significant," or "recurrent," we have come to that estimate, in part, by making counts, comparisons, and weights.

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*.

There are three good reasons to resort to numbers: (1) to see rapidly what you have in a large batch of data, (2) to verify a hunch or hypothesis, and (3) 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 participants why they were using the new school practices we were studying. We got a mass of answers from several participants at each of the 12 field sites. It seemed that many people were saying that 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 Display 11.1.

It turns out that 62% (35 of the 56 respondents) mentioned administrative 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.

Verifying a Hypothesis

McCammon et al.’s (2012) mixed-methods survey examined, in part, whether any gender differences occurred between adult men’s and women’s responses about their high school arts experiences. Saldaña hypothesized that there would be. On a 4.00 “strongly

Display 11.1

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

Source: Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage Publications.

“agree” scale, the female respondents’ mean was 3.85 and the male respondents’ mean was 3.93 to the prompt “I have good memories from my [high school] speech and/or drama participation.” But when we “crunched the numbers” through a *t* test, there was no statistically significant difference on a $p < .05$ standard between these two mean ratings. And the *qualitative* content analysis of their narrative responses to this prompt also confirmed no major differences between men’s and women’s experiences. So counting in the form of a statistical operation disconfirmed the hypothesis and made for a more trustworthy analysis.

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, was potentially controversial, and might have been a result of our expectation. So 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 earlier 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. Doing qualitative analysis with the occasional aid of numbers is a good way of testing for possible bias and seeing how robust our insights are.

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, variables, cases as a whole—that are known to differ in some other important respect. This is the classic “method of differences” heuristic. A few examples from our work described in earlier chapters are as follows:

- *Contrast tables* show how exemplary cases vary in extreme ranges of their change, from “high” to “low” to “negative” degrees of user change (see Display 6.21).
- *Growth gradients* display time on one axis and degree on another axis, enabling a comparison of changes across time (see Display 8.3).
- *Predictor-outcome-consequence 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 the *magnitude* of change, showing which predictors were present in “high-change” sites but not in “small-/no-change” sites (see Display 10.6).

We advise you to make sure that the comparisons made are the right ones and that they make sense. Take a moment before you display a comparison, and think, “How big must a difference be before it makes a difference?” You do not always need a statistical significance test to fall back on. The *practical* significance is what you need to assess.

7. Partitioning Variables

There are times when *differentiation* is more important than integration. It may occur to you at some point in your analysis: “Whoa—that variable is not really one variable but two, or maybe even three.”

Partitioning or subdividing 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 general variable of “preparedness” to carry out an innovation can be partitioned into subvariables or components, ranging from the state of the user (“commitment,” “understanding,” “skills”) to the availability of materials and the actions taken by administrators (“time allocation,” “in-service workshops”).

When you are designing matrix formats, variable partitioning is also useful; more differentiation lets you see differences that might otherwise be blurred or buried. Rather than a single column of *Outcomes*, it may help to partition that into two columns: (1) *Short-Term Effects* and (2) *Long-Term Consequences*.

When is variable partitioning a good tactic? The first answer: Divide variables in the 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, and poor mapping of events and processes. When you divide a variable, it should be in the service of finding coherent, integrated descriptions and explanations.

8. Subsuming Particulars Into the General

Clustering involves clumping together things that “go together” by using single or multiple dimensions. 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 similar to the analytic processes for pattern coding and focused coding (see Chapter 4). Classic grounded theory calls this the “constant comparative method.” You are trying to categorize a particular action, event, participant, state, and so on, into a more abstractly defined class. That class may have been predefined, or it may have emerged as a result of coding or analytic memoing.

For example, in the school improvement study, we noted specific statements made by teachers and administrators, such as the following:

- “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.

Subsuming particulars into more general classes is a conceptual and theoretical activity 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 abstraction, however, gets you nowhere. Suppose you observed a teacher writing her name on the dry-erase board on the first day of school. That specific action can be subsumed into a larger class of “written communication,” then into a larger class of “information transmission,” and finally into a still larger class of “human action.” That is a sort of taxonomic classification without useful meaning, however. 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.

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 hypothetical variables. These second-order variables (factors) may have some “communality.” So what is the qualitative researcher’s version of factoring? Making patterns of patterns, or categories of categories.

Most of the tactics we’ve discussed are designed to do two things: (1) condense the bulk of the data and (2) find patterns in them. Noting patterns/themes, clustering, and making metaphors 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 create a Pattern Code (see Chapter 4), you are proposing that several disparate but related pieces of data have something in common. What they do or are is the factor, and the process by which you generate it is factoring. In other words, we’re tightening up the data even further by making a smaller number of patterns from a larger number of patterns, or a smaller number of categories from a larger number of categories. Time for an illustration.

Hager, Maier, O’Hara, Ott, and Saldaña (2000; see also Saldaña, 2013) examined a state department of education’s new arts standards document and how it was received by high school theatre teachers. (Long story short, the majority of teachers rejected it because it was poorly written and developed with virtually none of their input.) The focus group interview transcript data generated 52 different Versus Codes, which were then clustered into comparable groups to form eight descriptive categories:

1. Arts Standards Development
2. Curricula
3. Teacher Resistance
4. People
5. Political Ideologies
6. Testing and Graduation Requirements
7. Exclusion and Marginalization
8. Institutions

The next analytic step was to *factor* these categories—in other words, to categorize these eight categories into an even tighter and smaller number of categories. Given

below is how the analyst factored the eight into the resulting three categories, which were labeled with war imagery because tensions were high among the participants and the metaphor seemed apt for this study:

Category 1: Human and Institutional Conflicts—The “Fighters”

Subcategories:

People

Institutions

Political Ideologies

Category 2: Standards and Curriculum Conflicts—The “Stakes”

Subcategories:

Curricula

Arts Standards Development

Testing and Graduation Requirements

Category 3: Results of Conflicts—The “Collateral Damage”

Subcategories:

Exclusion and Marginalization

Teacher Resistance

These three new categories were eventually transformed into “versus” phrases (e.g., Your Way vs. Our Way) that served to thematically represent the data.

The consequential question for this analytic tactic is “Do these factors 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, paper, and cardboard.

10. Noting the Relations Between Variables

The idea of the interrelationship between variables has been discussed throughout the book thus far, but let’s examine it as an analytic tactic.

Networks are most easily depicted as sets of boxes and arrows; the boxes are the 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 relationship can we envision between variables A and B? A variable is something that varies. Thus, we might have the following:

1. A+, B+ (both are high, or both are low at the same time)
2. A+, B− (A is high, while B is low, or vice versa)

3. $A \uparrow, B \uparrow$ (A has increased, and B has increased)
4. $A \uparrow, B \downarrow$ (A has increased, and B has decreased)
5. $A \uparrow$, then $B \uparrow$ (A increased first, then B increased)
6. $A \uparrow$, then $B \uparrow$, then $A \uparrow$ (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 the inverse. With Relationship 3, we are noting that changes have occurred recently in A and in B in the same direction; Relationship 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 causation: 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 relation: A change in A leads to a subsequent change in B , 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 Morrison (2009) reminds us, causation can be direct or indirect, mediated through other variables.

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 or concepts*, not necessarily specific actions.

Even when we focus 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 chemicals in the beverage involved and the body’s ability to metabolize alcohol, the amount consumed, the time intervening, and so on.

How are relationships detected? We propose in this book that displays are an especially strategic 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.

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 (discussed later), 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.

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 an inconclusive interrelationship. 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. Your initial analysis may be showing that $A \rightarrow C$, but you may be thinking, “Yeah, but something’s missing. Where’s ‘B’? And what is ‘B’ anyway?”

To start the example, look at Display 11.2.

In the school improvement study, we observed that schools adopting innovations accompanied with large 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 (Display 11.2) could be understood much more realistically when several other variables entered the picture (see Display 11.3). Here, we see that “Size of funding” is part of a web of other variables.

Display 11.2

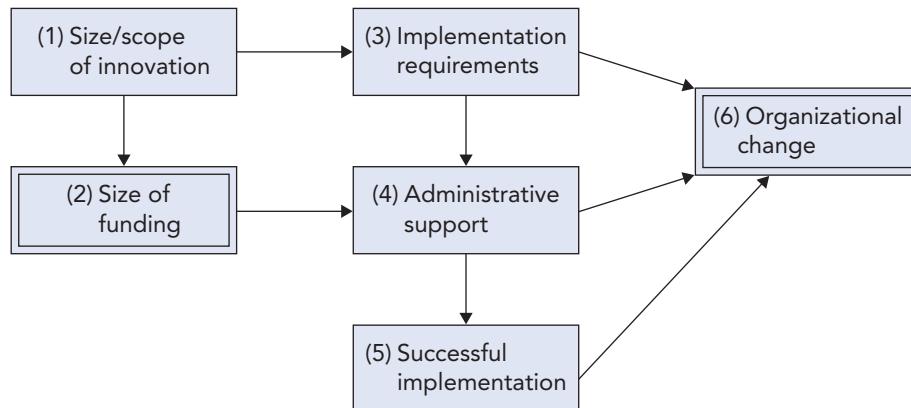
Two-Variable Relationship



Source: Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Display 11.3

Two-Variable Relationship With Intervening Variables



Source: Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Larger innovations (Box 1) carry more funds with them (Box 2). The funds increase the support administrators give (Box 4), but so do the heavier implementation requirements (Box 3) of larger innovations. Organizational change (Box 6) comes from at least three sources: (1) the direct requirements of the implementation itself (Box 3), (2) administrative support (Box 4), and (3) the degree to which implementation is successful (Box 5). As the network revealed, *Administrative support* is a very central intervening variable.

In this example, 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. Finding intervening variables is easiest with multiple examples of the two-variable relationship to look at, contrast, and compare.

12. Building a Logical Chain of Evidence

We've discussed **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 you can use? Let's start with an example.

In the study of interorganizational arrangements between schools and universities, we happened on 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.

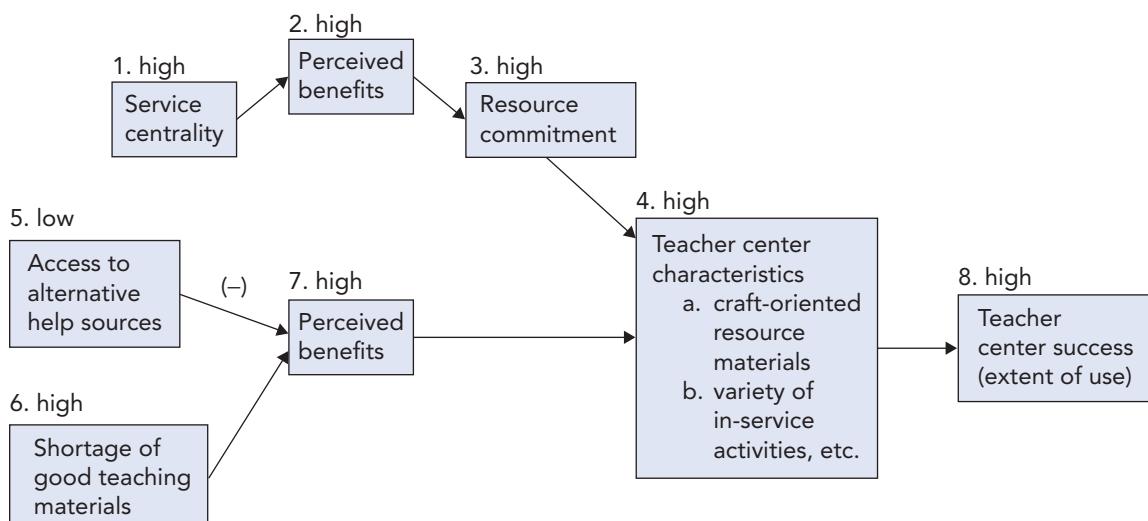
We developed a logical chain of factors that could be leading to success, as seen from the state college side and from the school side (see Display 11.4). The logical chain of evidence goes like this.

The state college might regard service and outreach activities as very central (1). Because of that, we would expect college staff to see the benefits (2) of a teacher

Display 11.4

Example of a Chain of Evidence Supporting an Observed Outcome

STATE COLLEGE



SCHOOLS

Note: (-) = inverse influence

Source: Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage Publications.

center as high (which 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* participants 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 in order 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 causal network methods? This approach is more tactically and 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, . . . ”

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 Display 11.4, the link between 3 and 4 is not so obvious. The college might have committed resources and might 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.

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.

13. Making Conceptual/Theoretical Coherence

When you’re trying to determine what someone’s actions “mean,” the mental exercise involves connecting a discrete fact with other discrete facts, then grouping these into 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 dealing with just 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 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 participants 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 there really is. The metaphors, in gerund form, might be "group thinking," "self-deluding," or "wish fulfilling." We also could sketch out a logical flowchart like the one shown in Display 11.4 to get a fix on how this happens at the several sites. But we are still within the confines of our study. The analyst now has to ask, "Do any broader constructs put these facts together the way I am putting them together?" In principle, gerunds should help. The first one (group thinking) points toward group behavior, the next one (self-deluding) toward individual cognitive process, and the third one (wish fulfilling) toward motivational dynamics.

We picked up an appropriate and powerful construct from classic 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 analog, 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 are (a) establishing the discrete findings, (b) relating the findings to each other, (c) naming the pattern, and (d) identifying a corresponding construct. We are making conceptual "additions" to the observed data to make them applicable more broadly.

It is perfectly legitimate, and sometimes necessary, to work from the top down—from a conceptual framework or theory 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.

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* or *trustworthiness*. 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 participants 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. In this section, we review some general guidelines for judging the "goodness" of qualitative research—that is, what researchers can do at the tactical, operating level to test and confirm the findings.

First, let's take a general view of the problem: Many, if not 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, condensing the information, analyzing it, interpreting it, and 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”). 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 there are few guidelines for explaining to their colleagues what they did, and how. (And, in all fairness, the page limit for journal articles also forces researchers to briefly sketch their methodology paragraphs.)

We need to be mindful 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 research methods textbooks; the archetypical ones include the following:

- *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—that is, sloppy research
- *Elite bias*: Overweighting data from articulate, well-informed, usually high-status participants and underrepresenting data from less articulate, lower status ones
- *Personal bias*: The researcher’s personal agenda, personal demons, or personal “axes to grind,” which skew the ability to represent and present fieldwork and data analysis in a trustworthy manner
- *Going native*: Losing your perspective or your “bracketing” ability, being co-opted into the perceptions and explanations of local participants

We draw on some of the research methods literature as we discuss tactics for testing and 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?

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 concluding 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 actions it is about—the participants who supplied the original data. The tactic of **getting feedback from participants** (13) concludes our list.

1. Checking for Representativeness

When we develop 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 is it?

People typically make a generalization and then illustrate it (“For example, my friend . . .”), but they 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.

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:

The first pitfall, **sampling nonrepresentative participants**, 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 workload, lack of cooperativeness, or both. Anthropologists have often warned of field-workers’ tendency to rely too much on articulate, insightful, attractive, and intellectually responsive participants; such people often turn out to be in the local elite.

The second pitfall, **generalizing from nonrepresentative events or activities**, 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 that 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.

The third pitfall, **drawing inferences from nonrepresentative processes**, is looking for underlying processes explaining what you’ve seen and heard. But if the samples of events and activities are faulty, the explanation cannot be generalized beyond them. If an emerging account makes good logical sense (to you) and fits well with other, independently derived analyses, you lock onto it and begin to make a stronger case for it. The problem is that you might have *plausibility* but not *confirmability*.

The real problem with selective sampling and hasty 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 reinforced the former approach (critical friends, colleagues, other field-workers, research team members, auditors) throughout the methods profiles, so let's offer some additional safeguards.

Automatically assume that you are selectively sampling and drawing inferences from a nonrepresentative sample of cases, be they people, events, or processes, in the first place. You are guilty until you prove yourself innocent by extending the “universe” of your study. How?

1. Increase the number of cases.
2. Look purposively for contrasting cases (negative, extreme, countervailing).
3. Order the cases in various ways in a matrix to see who or what may be missing.
4. Randomly sample people and phenomena within the site(s) you're studying.

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.

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 “researchees.” 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 (an appropriate term in this instance) the natural characteristics of the setting with the artificial effects of the researcher–researchee relationship. Unconfounding them is like moving through a hall of mirrors.

So we have two possible sources of bias here:

1. The effects of the researcher on the case
2. The effects of the case on the researcher

Field study researchers are often less worried about Bias 1 because they typically spend enough time on-site to become part of the local landscape. But that, of course, increases the hazard of Bias 2: being co-opted, going native, swallowing the agreed-on or taken-for-granted version of local events.

Although we discuss these biases as they occur during site visits, they influence the analysis deeply, both during and after data collection. The researcher who has gone native remains native during analysis. The researcher who has influenced the site in un-understood ways suffers unawares from that influence during analysis.

Bias 1 occurs when the researcher threatens or disrupts ongoing social and institutional relationships by doing nothing more than simply being there. 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, participants 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, participants will often craft their responses to appear amenable to the researcher and to protect their self-interests. For some analysts, local participants' 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 mislead and must shift into a more investigative mode.

Field research can, at bottom, be considered as an act of betrayal, no matter how well-intentioned or well integrated the researcher is. You make the private public and leave the locals to take the consequences.

But that is not the only way Bias 1 can occur. In some instances, Biases 1 and 2 can team up to create “artifactual” effects as a result of the complicity between the researcher and local actors. This is Rosenthal’s (1976) famous “experimenter” effect.

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 cancelled 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 university researchers 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 1 can take still other forms. For example, local participants can implicitly or explicitly boycott the researcher, who is seen variously as a spy, a voyeur, or a pest. Or the researcher can inhibit the 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 2: 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.

A. Avoiding Biases Stemming From Researcher Effects on the Site

- Stay on-site as long as possible; spend some time simply hanging around, fitting into the landscape, taking a lower profile.
- Use unobtrusive measures where you can, such as reading the site's publicly accessible documents.

- Make sure your intentions are clear for participants: why you are there, what you are studying, how you will collect information, and what you will do with it.
- Consider co-opting a participant—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, participant'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 participants; include lower status participants 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 your site visits.
- 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 a participant 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 just on talk, just on observation, or just on documents to make sense of the setting.
- If you sense you are being misled, try to understand and focus on why a participant 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 participant you are trying to co-opt.

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 1 and 2. However, we take that with a grain of salt. Long exposure can just push up Bias 2 and make Bias 1 harder to see.

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.

3. Triangulating

Much has been written about triangulation as a near-obligatory method of confirming findings. Stripped to its basics, triangulation is supposed to support a finding by showing that at least three independent measures of it agree with it or, at least, do not contradict it. But it's not as simple as that. For one thing, if two measures agree and one does not, you are stuck with a deeper question: Which do you believe? (See the film *Minority Report*.)

Triangulation is similar to the modus operandi approach used by detectives, mechanics, and primary care physicians. 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 (2001) 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*. To this, we add *data type* (e.g., qualitative texts, audio/video recordings, quantitative data). How to choose which? The aim is to pick triangulation sources that have different foci and different strengths, so that they can complement each other. In some senses, we are *always* faced with triangulating data, whether we attend to it or not.

As an example, Saldaña's studies on child audience responses to theatre employed multiple measures: (a) live audience observations and field notes of the events, (b) audio and video recordings of audience responses to the performance, (c) audio-recorded focus group interviews with children after the performance, (d) drawings or written assessments by children of the productions they viewed, (e) play script dramaturgical analyses, (f) interviews with selected teachers and adult production company members, and (g) relevant theories from the professional research literature in areas such as child development, psychology, education, media, and performance studies.

What can we expect from triangulation? We may get *corroboration* from three different sources, which enhances the trustworthiness of our analysis. But sometimes we may get *inconsistent* or even directly *conflicting* findings. At best, this can push us to more closely examine the integrity of the data collection methods and even the data themselves. Inconsistent and conflicting findings force us to explain why they exist. It may be due to an undetected variability of something that we need to consider or unanticipated anomalies and outliers in the phenomena we're studying that merit

closer examination. Sometimes the inconsistent and conflicting findings suggest that the researcher made a mistake somewhere along the way and needs to figure out what may have gone wrong. But sometimes the conflicting findings are a blessing because the different data collection methods used gather different facets of data, and their combined effects *build* on each other to compose a more three-dimensional perspective of the phenomenon.

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 along. 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.

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 and reports.

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 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 to back it up.

Basically, there is a very large range of reasons why certain data are stronger (valid) or weaker (less valid) than others. First, data from some participants are “better.” The participant may be articulate and reflective and may enjoy talking about events and processes. Or the participant may be knowledgeable and 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 district budget matters 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. Seeing with your own eyes what happened in a classroom is more reliable than overhearing faculty lounge gossip about the incident. Interviewing a participant during later rather than earlier stages of fieldwork may get you richer responses since you’ve had time to build rapport and trust.

Third, data quality may be stronger because of a field-worker’s validation efforts. These may consist of several varieties (the bolded ones are discussed elsewhere in this chapter):

- **Checking for researcher effects** and biases
- **Checking for representativeness**
- **Getting feedback from participants**

- **Triangulating**
- 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. Douglas offers a few strategies, which we have adapted:

- Check against “hard facts.”
- Check against alternative accounts.
- Look for “the trapdoor”—what is going on beyond the obvious.
- Share your own personal story to open up the respondent.
- Share what you genuinely think is going on, and see how the respondent reacts.

Field-workers who rely mainly on trust may quail at such interventions or dismiss them as too intrusive. We have found that it pays to be skeptical, to expect to be lied to sometimes, to look for respondent self-deception, and to gently and tactfully question respondents from time to time on such matters.

Two added suggestions. First, we've found it useful to keep a running log of data quality issues (often in the form of analytic memos or OCs—observer's comments—in field notes), together with recurrent efforts to improve data quality in subsequent site visits. Second, when approaching the 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.

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 it, 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 participants 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 participants 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 became clearer that the innovation was like a “miracle cure” only if it was technically well carried out. Furthermore, these outliers told us that there were more people like them around than the advocates had admitted. We realized then that we had oversampled the contented users and, in a sense, had been “sucked in” to the taken-for-granted version of events among the local teachers. 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 consist of 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 [using extreme cases](#) next). 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, or settings. And, on following up, stay open to the eventuality that the exceptional cases can turn out to be prototypical ones.

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 modification.

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.

In Saldaña’s (1995) exit interviews with sixth graders about their 7 years of theatre-viewing experiences since kindergarten, one interview question we posed to the young participants was “Is theatre necessary?” All except one child said “Yes”; it was the young boy who said “No” that was the extreme case in this instance. His explanation that

“we don’t have to have” theatre was not unexpected, as he was one of the lowest rated participants in the study over the past 7 years. But his “no” became part of the data record, and we now had to explain why, after so many years of free and mostly high-quality theatre experiences, he alone out of 30 children felt that theatre was unnecessary. Long story short, his “no” forced us to closely examine the varying *value* children attribute to the arts, and thus the landscapes of meanings and payoffs they offered to their young participants. That single “no” helped crystallize the conclusions of the study.

This is a tactic of “holding everything else constant”—looking for the most extreme case, where there should have been consensus but there 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.

The second sort of extreme case 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** (discussed earlier). 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.

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, such as “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 in the attention, and maybe reflect on their kindness, and perhaps on the sneakiness of your significant other, who colluded with the plot to delude you—for 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, and taken-for-granted assumptions? And where can I go in my data to help me rebuild my theory?

Something that surprised Saldaña on his first day of fieldwork in an inner-city school in the mid-1990s Southwest was the combed back and “slicked” hairstyles of most of the Hispanic boys—a grooming nuance he realized had been within the culture since his own childhood 40 years ago. Also surprising was the realization that many young girls

did not wear shorts or pants but wore dresses to school—unlike the clothing observed at middle-class, predominantly White elementary schools. These small fashion details made him rethink the larger cultural dynamics he was now surrounded by and how selected traditions and values still played a role in these Hispanic children's lives.

Following up surprises has three aspects. You (1) reflect on the surprise to surface your violated theory, (2) consider how to revise it, and (3) look for evidence to support your revision. You may also work from (2) to (3), hunting in a sniffing mode, to find new aspects of the case that possibly could lead to a new theory.

8. Looking for Negative Evidence

The tactic **looking for negative evidence** is easy to describe but, given people's pattern-making proclivities, not naturally instinctive. When a preliminary conclusion is at hand, the tactic is to ask, "Do any data oppose this conclusion, or are any inconsistent with this conclusion?" This is a more extreme version of looking for **outliers** (discussed earlier) and for **rival explanations** (discussed later). You are actively seeking disconfirmation of what you think is true.

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 few data to convince us of their solidity, and we are not eager to encounter the many brutal facts that could doom our frameworks. A good case study of such an effort is 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.

Miller (n.d.) advises that "disconfirming instances should be handled with care." Discarding your original hypothesis too quickly and 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.

Commission a friendly but curmudgeonly skeptic to take a good look at your 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, proceed to the formulation of an alternative conclusion that deals with the evidence.

Finally, note what might be called the "delusional error." The absence of negative evidence can never be decisive as a confirmatory tactic, as in the following exchange:

Question: Why do you have that blue ribbon on your little finger every day?

Answer: It's to keep elephants from following me.

Question: But there are no elephants here.

Answer: See? It's working.

9. Making If-Then Tests

If-then tests are the workhorse of qualitative data analysis. They are more focused than a generalized working hypothesis, which supports a general analytic direction.

A formal statement is “If X, then Y.” It is a statement of expected relationship. Assuming X to be true (an important condition), we look to see whether Y is true. If Y is true, then we have a building block for understanding. We are a long way from a “law” about the relation of X and Y, 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.

Saldaña’s (1999) study of adolescent social class and social consciousness was an opportunistic study that investigated why two different groups of teenagers, each from two very different social and economic backgrounds, responded quite differently to the same social issues-oriented workshop content. The lower socioeconomic group embraced the seriousness of the social issues addressed at a forum workshop and participated with respectful enthusiasm. But the upper socioeconomic group transformed the same content at their forum workshop into a parody and satire of the social issues we tried to address.

After reflection on these experiences, Saldaña put forth the following general proposition: *If* adolescents are from an upper social class background, *then* their social consciousness may not be as heightened as those from a lower social class. Of course, a reflective practitioner might also develop the following proposition: *If* workshop content is not relevant or meaningful for its participants, *then* they will reject the provided experiences. One can always think of exceptions or **negative evidence** that will force a revision of these propositions, but for this particular event, this is what was observed.

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. Therefore, if-then statements are a way to formalize *propositions* for testing. The method of generating *predictions* involves linking together a large number of “ifs” to a single major “then.” If-then statements are just a step away from constructing theories.

10. Ruling Out Spurious Relations

“Spurious” means something that’s falsely attributed; a spurious relation means that you’re connecting things together incorrectly. Suppose you’ve been able, through assorted tactics, to establish that variable A is indeed related to B. Whenever we see A, we see B, and vice versa. Before breathing a sigh of relief and proceeding to the next conclusion, it pays to consider the picture you are drawing:

$$A \rightarrow B$$

may in fact be more accurately portrayed as



where some third factor is in play, causing both A and B to occur.

This is an old logic problem, and we can draw a nice example from a classic Wallis and Roberts (1956) 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 less (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 (1956) suggest that a third variable may be influencing both A (transportation) and B (mortality)—the *seriousness of the initial attack*. All of the patients were seen in a certain hospital: Willard Parker, a noted center for the treatment of contagious diseases. Polio patients who lived farther away were probably brought to Willard Parker only if their condition was serious; milder cases would be treated nearer their own homes. Thus, the picture develops as shown in Display 11.5.

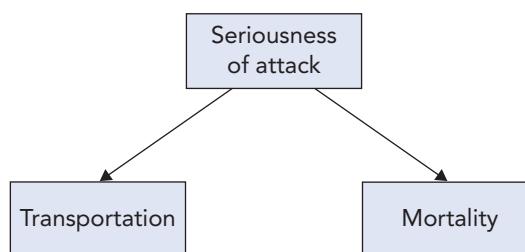
This interpretation can be checked out through the sort of display found in Display 11.6, with Ns and mortality rates entered in the cells.

As Wallis and Roberts (1956) 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.

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

Display 11.5

Possible Explanation of Spurious Relationship



Source: Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Display 11.6

Display for Testing Explanations in Display 11.5

	Local patients	Distant patients
Mild attack		
Severe attack		

Source: Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage Publications.

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 that they are causally associated, wait a bit and consider whether some third variable might be underlying/influencing/causing them both. Use a knowledgeable but detached colleague to help in the search. Then, consider new displays that will give you a clean look at 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**, but there we were concerned with understanding a tepid relationship better. Here, the issue is *undoing* a relationship you think looks plausible and strong.

11. Replicating a Finding

As we showed earlier in the section 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 data collection instrument measuring the same thing.

Still, the fact that usually one person is doing all of this measuring with homemade instruments is a ground 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 slippery 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 do you protect against this?

One line of attack is to think in terms of study replication. If I can observe and reproduce the finding in a new context or in another part of my database, it is most likely a dependable one. If someone else can reproduce it, better still.

This tactic can be used in several ways. At the most basic level, you are replicating as you collect new information from new participants, 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. 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 better, replicate in a brand new case, though we acknowledge that may be an extreme and ephemeral goal. Some brief words of advice:

- 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. Be prepared in the course of fieldwork to do the kind of corroborative testing described here and in the section on **triangulation**.

- 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.
- Doing replication at the very end of fieldwork, during the 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.

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 database and methods, how easy or difficult would it be *for your 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 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.

But we think that the search for rival explanations is often more thorough in qualitative research than 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 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.

The trick is to hold onto 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?

One of the most intriguing yet puzzling findings in McCammon et al.’s (2012) survey of adults from various generations reflecting on their high school experiences was an outcome from data analysis of the most recent age cohort. When participants were asked about the challenges they faced as an adolescent, those who graduated during 2000–2010 described a general sense of disconnect and “fear” more than older age-groups. Saldaña hypothesized this as the result of living in an uncertain post-9/11 world and affirmation of the then current emotional angst in America’s youth projected by Goleman (1995, 2007). But Saldaña was a researcher in his 50s who was trying to understand the responses of a generation that was in their 20s, so he turned to younger colleagues—his research assistants Matt Omasta and Angie Hines—for their perspectives. And they did indeed offer other factors to consider for hypothesis development, such as the generational search for “safe spaces” that school arts programs provided and the socio-psychological effects on the most recent generation by the rapidly developing social media and entertainment technology.

Foreclosing too early on alternative explanations is a harbinger of bias. 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—thus further increasing your confidence in your erroneous thesis.

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.

It also helps to fasten onto 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, ask yourself what kind of alternative case these bits of information could build, and then check them out further.

A useful subtactic is as follows: During the 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 the fieldwork. “Next bests” have more pulling power than fanciful alternatives.

13. Getting Feedback From Participants

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. In that sense, local participants can act as judges, evaluating the major findings of a study through what has been colloquially labeled “member checking.” Some of the display methods profiles encouraged feedback from case participants: commenting on a short summary of findings, evaluating the accuracy of a causal network, verifying researcher predictions.

Feedback may happen during data collection, too. When a finding begins to take shape, the researcher may check it out with new participants and/or with key participants. The delicate issue here, of course, is that of introducing bias (see **researcher effects**). Feeding things back in the course of a study may change participants’ actions or perspectives. (In some genres, such as action research, critical ethnography, or community-based collaborative research, however, changing participants’ actions and perspectives for the better *is* the goal.)

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 the basis for shared reflection. Warner (1991) not only had children run camcorders of adult-child interaction and added their comments, but he also played back the video tapes for yet more comment by the children. The more emic the study, the more useful early feedback is likely to be.

There are also good reasons for conducting feedback after the final analysis. For one thing, you know more. You also know *better* what you do know—you are less tentative,

have more supporting evidence, and can illustrate it. In addition, you can get feedback at a higher level of inference: on main factors, on causal relationships, and 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.

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:

- 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. There will be too many competing claims on your time.
- Think carefully about displays. As with analysis, matrices and networks work much better than text alone to help participants access the information. They will find it easier to get an overview, to see how the pieces fit together.
- Providing information at more macro-analytical levels of inference (e.g., main factors and relationships, causal determinants) has to be done very carefully by working up from particulars. If this is not done, participants 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.” Format your data, and feed them back in a participant-friendly way. Beyond this is the possibility of mutual enlightenment.
- 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.)
- Don't expect that participants 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.

Data feedback is an occasion to learn more about the *case*, not only about the feedback.

Standards for the Quality of Conclusions

We've reviewed 26 tactics for drawing and 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, trustworthy, 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.

In this section, we explore some practical standards that could help us judge the quality of conclusions. The battles in this domain have been extensive, and they continue. Many interpretivist researchers suggest that it is not really possible to establish standards or 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, and personally interpretive nature of any qualitative study.

But the problem of quality, of trustworthiness, of authenticity of the findings will not go away. The fact is that some accounts read better than others. Although we may acknowledge that “getting it all right” may be an unrealistic goal, we should, as Wolcott (1990) 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. These are matters of merit, rigor, integrity, ethics, and accountability.

We cannot enter here into a discussion of how goodness criteria flow from epistemological positions. Rather, we remain broadly in the critical realist tradition and discuss five main, somewhat overlapping, issues: (1) the **objectivity/confirmability** of qualitative work, (2) **reliability/dependability/auditability**, (3) **internal validity/credibility/authenticity**, (4) **external validity/transferability/fittingness**, and (5) **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 (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 that can be applied to qualitative work—your own or that of others. These are not rules to be stiffly applied but guidelines to consider when you reflect on the question “How good is this research report?”

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. This domain is sometimes labeled “external reliability,” borrowed from classic quantitative terminology.

Some useful points to consider about this issue for a qualitative study are as follows:

1. The study’s general methods and procedures are described explicitly and in detail. We feel that we have a complete picture, including “backstage” information.
2. We can follow the actual sequence of how data were collected, processed, condensed/transformed, and displayed for specific conclusion drawing.

3. The conclusions are explicitly linked with exhibits of condensed/displayed data.
4. There is a record of the study's methods and procedures, detailed enough to be audited by an outsider (Lincoln & Guba, 1985).
5. The researcher has been explicit and as self-aware as possible about personal assumptions, values and biases, and affective states—and how they may have come into play during the study.
6. Competing hypotheses or rival conclusions have been considered. The plausibility of rival conclusions has been examined.
7. The study's data are retained and available for reanalysis by others (as allowed by institutional review board regulations and any researcher-participant agreements).

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 are addressing issues of quality and integrity: Have things been done with reasonable care?

Some useful points to consider about this issue for a qualitative study are as follows:

1. The research questions are clear, and the features of the study design are congruent with them.
2. The researcher's role and status within the site have been explicitly described.
3. The findings show meaningful parallelism across data sources (participants, contexts, times).
4. Basic paradigms and analytic constructs are clearly specified. (Reliability depends, in part, on its connectedness to theory.)
5. Data were collected across the full range of appropriate settings, times, respondents, and so on, as suggested by the research questions.
6. If multiple field-workers are involved, they have comparable data collection protocols.
7. When appropriate, intercoder agreement checks were made with adequate results.
8. Data quality checks have been made (e.g., for bias, deceit).
9. Multiple observers' accounts converge—in instances, settings, or times—when they might be expected to.
10. Forms of peer or colleague review are 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 is a contested term among selected qualitative researchers. Some feel that this traditional quantitative construct (with its components of face, content, predictive, etc., validity) has no place in qualitative inquiry. Alternative terms such as *verisimilitude* and *a persuasively written account* are preferred. But other qualitative methodologists continue to use the term purposefully because it suggests a more rigorous stance toward our work.

Wolcott (1990) went so far as to reject validity in our field and proposed that we should come to a deep *understanding* instead. Maxwell's (1992) 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* (the concepts, and their relationships, used to explain actions and meanings), and *evaluative* (judgments of the worth or value of actions and meanings).

Saldaña calls this domain the “That’s right!” factor. When an oral presentation is made of one’s research and people in the audience are nodding their heads affirmatively, saying “M-hm” in agreement and (even better) exclaiming out loud, “That’s right!” to the speaker after an assertion has been made, you can feel confident that some sense of resonance has been achieved between the research and the audience. Call this domain what you will; it’s the write-up itself that matters in the end.

Some useful points to consider about this issue for a qualitative study are as follows:

1. Descriptions are context-rich, meaningful, and “thick” (Geertz, 1973).
2. The account rings true, makes sense, seems convincing or plausible, and enables a vicarious presence for the reader.
3. Triangulation among complementary methods and data sources produced generally converging conclusions. If not, the procedures for reconciling the differences and their results are explained.
4. The data presented are well linked to the categories of prior or emerging theory. The measures reflect the constructs at work.
5. Findings are clear, coherent, and systematically related—that is, unified (Charmaz, 2006; Eisner, 1991).
6. Confirmation procedures for assertions, propositions, hypotheses, conclusions, and so on, are described.
7. Any areas of uncertainty have been identified.
8. Negative evidence was sought (and, if applicable, found and accounted for in the analysis and write-up).
9. Rival explanations have been actively considered.
10. When possible, findings have been replicated in other parts of the database than the one they arose from.
11. The conclusions were considered to be accurate by the original participants. If not, a coherent explanation is offered.
12. If predictions were made in the study, their accuracy is reported.

External Validity/Transferability/Fittingness

We need to know whether the conclusions of a study—a case study, in particular—have any larger import. Are they transferable to other contexts? Do they fit? How far can they be generalized?

Grounded theorists attest that the methodology develops concepts and abstractions at a level that supports their transferability to other populations and contexts (Glaser, 2005). Some methodologists purport that any transfer of a study's findings to other contexts is the responsibility of the reader, not the researcher (Erickson, 1986). The generalizability of the case study has been a contentious issue, ranging from the researcher's analytic ability to find levels of universality in the case (Spradley, 1979, 1980) to frank admission that complex and site-specific contexts problematize the ability to construct theory and, thus, generalization (Clarke, 2005).

The generalizing process is far from mechanical, as Noblit and Hare (1988) note in their discussion of meta-ethnography: It is more like synthesizing two or more studies of similar phenomena. It is careful interpretation, not just “adding up.” In this case, we again conclude that it’s the write-up itself that matters in the end. How *persuasive* a case can the researcher make that the findings of $N = 1$ have meaning and resonance to other individuals, sites, and times?

Some useful points to consider about this issue for a qualitative study are as follows:

1. The characteristics of the original sample of persons, settings, processes, and so on, are sufficiently fully described to permit adequate comparisons with other samples.
2. The report specifies any limits on sample selection and critically examines its ability to generalize to other settings and contexts.
3. The sampling is theoretically diverse enough to encourage broader applicability when relevant.
4. The findings include enough “thick description” for readers to assess the potential transferability and appropriateness for their own settings.
5. A range of readers report that the findings are consistent with their own experiences.
6. The findings are congruent with, connected to, or confirmatory of prior theory.
7. The processes and outcomes described in the conclusions are applicable in comparable settings.
8. Any theories and their transferability are explicitly stated.
9. The report suggests settings where the findings could fruitfully be tested further.
10. When possible, the findings have been replicated in other studies to assess their robustness.

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.

Evaluation and policy studies in particular are supposed to lead to more positive and constructive actions; whether or not they do, real people's lives are being affected, and large amounts of money are being spent (or misspent). Action research, critical ethnography, and other community-based research projects are designed to clarify and rectify particular local problems through participatory engagement. At the very least, they heighten awareness among participants of selected social issues that affect them directly.

These research genres also raise questions of ethics—Who benefits from a qualitative study, and who may be harmed?—and of “evaluative validity” (Maxwell, 1992): judgments made about the worth, legitimacy, or goodness of actions or meanings.

Some useful points to consider about this issue for a qualitative study are as follows:

1. Value-based or ethical concerns and dilemmas are raised explicitly in the report.
2. The findings are intellectually and physically accessible to potential users.
3. The findings stimulate and provide intellectual “payoff” for a reader, and possibly ideas for his or her own related research project.
4. The level of usable knowledge offered is worthwhile, ranging 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 or specific action images.
5. The actions taken actually help solve a local problem.
6. Users of the findings have experienced a sense of empowerment or increased control over their lives (Lincoln & Guba, 1985).
7. Users of the findings have learned or developed new capacities.

Analytic 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*: well-organized electronic and hard copy files that help keep track of what was done along the way, suggesting ways of improving next steps and documenting all logistical matters related to the study.

The second audience is the *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 and journal peer reviewers make such judgments carefully, even obsessively. And other readers such as local participants, policymakers, managers, and the general public often raise questions such 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, who want to carry out a metasynthesis of several studies, or who want to replicate the findings to strengthen or modify them.

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 mixed-methods studies, it's also expected that you report relevant statistical data in standardized display formats.

There is, in other words, a traditional set of conventions for documenting empirical research and for reporting it, and a corresponding set of methods for verifying the report. But qualitative researchers don't have very clear alternatives to fall back on. We've borrowed a few conventions from the culture of quantitative-oriented journals. But the hybridity of the field has created multiple, open-ended reporting structures for progressive genres such as poetic inquiry, ethnodrama, narrative inquiry, and others. And selected schools of thought reject the very notion of "conclusions," considering them abhorrent abuses of researcher power to force fixed meanings on an uncertain social world. (The irony, of course, is that these same schools of thought seem to authoritatively assert their stances as fixed meanings.)

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 very little about how the 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 seemingly 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 always be verified because researchers don't always report clearly on their methodology, and they don't report clearly on their methodology because there are not always required 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 do we do?

Methods sections are much more typical in dissertations, theses, and journal reports of traditional qualitative studies. They tend to follow a traditional outline and sequence:

- Central and related research questions
- Cursory literature review
- Conceptual or theoretical framework
- Methodology
 - Participants
 - Data collection methods
 - Data analysis procedures

We need to know a few additional things, if possible:

- Which kinds of qualitative designs do researchers actually use?
- How are sampling decisions actually made?

- What does the instrumentation look like, and how does the researcher know whether it is measuring accurately what it was meant to measure?
- How does fieldwork actually proceed?
- How are the data aggregated, condensed, partitioned, displayed, analyzed, and interpreted?
- Most fundamentally, how do researchers get from hundreds if not thousands of pages of field notes, interview transcripts, and documents to a final report?

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.

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.

Lincoln and Guba (1985) 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 first and basic audience for good documentation is the *self*. Even if no audit is ever intended, devices such as the researcher’s journal, diary, and analytic memos strengthen the study *as it goes*. Second, 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. There are also confidentiality issues to consider, with outsiders looking at personal data; institutional review board regulations may prohibit this unless approval has been made in advance to the regulatory body.

Illustration

We (Miles and Huberman) developed a documentation form focusing on analytic operations. (*Note:* It does not deal with issues of sampling, instrumentation, fieldwork, and so on.) 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 that the form requires some time and care.

We show in Display 11.7 a streamlined version of more elaborate efforts. We include it here for essentially heuristic, illustrative reasons, and 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

Display 11.7

Qualitative Analysis Documentation Form

1. Research Issue being explored: _____
FORM NO. _____ Analyst _____ Date _____

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.

- (a) Indicate whether single or multi-site. May include: field notes; write-ups; summaries; documents; figures; matrices; tables; tape recordings; photos; film/video; others (specify).

(b) PROVIDE ALL RELEVANT DISPLAYS; 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): _____

(c) Explicit list of actual rules used for "readying" data (clustering, sorting, scaling, etc.); may also apply to drawing/confirming conclusions.

Source: Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage Publications.

(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 the 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 that 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 *displays* (materials used or developed in the course of analysis), so that they can be referred to easily. For a reviewer, they should be appended.

Make running notes for the *Procedural Steps* section as you go along; much gets lost in recollection. Then, it's easy to fill in the section with a few summary phrases, such as the following:

- Reviewed preliminary matrix format; entered data from two sites.
- 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.

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 along is very useful; it reduces error and can aid self-correction. Here are some examples:

- Theme was coded as present for a participant if mentioned repeatedly or with strong affect during the interview.
- Item was entered in display if mentioned by more than one participant at sites with more than three participants; at sites with fewer users, item was entered if mentioned by one participant, with no contradiction from others.
- Cutting point between “high” and “moderate” cases was made when half or fewer of the indicators were absent.

Making decision rules explicit is not only critical for data-readying operations but also important for conclusion drawing and confirmation. 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 (Miles and Huberman) developed an “in-house” code list (see Display 11.8). Scanning the list serves as a kind of prompt for the analyst, both helping 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 *Readyng 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 earlier sections. Because there is no standardized language about analysis operations and there are a variety of partial dialects from different disciplines,

Display 11.8

Code List for Analysis Operations

<u>Readyng data for analysis</u>		TEMP	determining temporal order/temporal relationships
TAB	tabulating coded segments	INF	making inferences
MAT	filling in matrices	INF-COMP	by computations
CLAS	classifying, categorizing	INF-DED	by deduction
RANK	ranking/weighting data	INF-IND	by induction (e.g., determining antecedents, covariates, consequences)
SMM	summarizing phrases, generating key words		
SUB	subsuming data under higher level variable		
SCAL	scaling, summing indices	REPR	checking for representativeness
COMP	computing, tabulating	RES-EFF	checking for researcher effects
SPLT	splitting one variable into two	BIAS-CONTR	control for bias (specify)
PAR	partitioning	TRI	triangulation
AGG	aggregating	TRI-DATA	from different data sources
<u>Drawing conclusions</u>		TRI-METH	from different methods
PLAUS	seeing "plausibility" only	TRI-CONC	conceptually (different theories)
GES	seeing a Gestalt, pattern, theme	TRI-RES	from different researchers
MET	generating metaphors	WT	weighting of evidence
CLUS	clustering	OUT	user of outliers, exceptions
COUNT	counting/frequencies	EXTR-SIT	extreme situation verification
CEN	establishing central tendencies	EXTR-BIAS	extreme bias verification
CONT	systematic contrasts/comparisons	SURP	following up surprises
FAC	establishing factors	EMP	empirical evidence from elsewhere
REL	establishing relationships between variables/sets of variables	NONEG	absence of negative evidence
LOG	logical chain of evidence	IF-THEN	testing if-then relationships
INTV	establishing intervening/linking conditions	FALSE-REL	checking false relation due to third variable
COH	conceptual/theoretical coherence	REPL	replication
CAUSE	determining directional influence	RIV	test of rival explanation
		FB	corroboration from informant feedback

Source: Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage Publications.

we tried to find items that would have meaning to most researchers. We've aimed to keep our own 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 and how they actually work. Other researchers may have more success in this.

The next to last column of the Display 11.7 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 *Displays* 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.

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 the final conclusions. The value of feedback from a friend, colleague, reader, replicator, meta-evaluator, or auditor comes later as an add-on. A number of CAQDAS software programs automatically develop a dated log of analysis operations of certain sorts, such as code development, searches and queries, and analytic memo records.

Less complex documentation methods are not hard to find. Sometimes a detailed researcher journal or analytic memos may be all that are needed for your purposes. We offer these final recommendations:

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, or 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.
3. Documentation detail 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, make rough running notes; then order them, logging in the procedural steps, decision rules, and conclusions, giving enough detail to be clear; and then assemble the exhibits (tables, worksheets, text, etc.). After the analysis is complete, review the entries, cleaning them up where needed and adding analysis codes; then fill out the more reflective part.
5. 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. Avoid “laundering” or retrospective enlightenment. Do not let incomplete documentation forms pile up—that defeats your purposes. Do them as you go along.
6. Code lists 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.



Closure and Transition

Tremendous amounts of tactics, guidance, advice, and recommendations were covered in this chapter. We do not mean to overwhelm you with this vast array of factors to consider in your data analyses. But we do feel they are important guidelines that can strengthen the quality and integrity of your qualitative work. Use these methods on an “as needed” or “as relevant” basis.

In the next chapter, we will explore the write-ups researchers must make for sharing their data-analytic efforts and findings with audiences.