

# One

## An Introduction to Codes and Coding

Any researcher who wishes to become proficient at doing qualitative analysis must learn to code well and easily. The excellence of the research rests in large part on the excellence of the coding.

(Anselm L. Strauss, *Qualitative Analysis for Social Scientists*, 1987, p. 27)

### Purposes of the Manual

The three primary purposes of *The Coding Manual for Qualitative Researchers* are:

- to briefly discuss the functions of codes, coding, and analytic memo writing during the qualitative data collection and analytic processes
- to profile a selected yet diverse repertoire of coding methods generally applied in qualitative data analysis, and
- to provide readers sources, descriptions, examples, recommended applications, and exercises for coding and further analyzing qualitative data.

This manual does not address such matters as qualitative research design or how to conduct interviews or participant observation fieldwork. These topics are already masterfully discussed in other textbooks. *The Coding Manual for Qualitative Researchers* is intended as a reference to supplement those existing works. This manual focuses exclusively on codes and coding and how they play a role in the qualitative data analytic process. For newcomers to qualitative inquiry it presents a repertoire of coding methods in broad brushstrokes. Additional information and extended discussion of the methods can be found in most of the cited sources. Grounded theory (discussed in Chapter Two), for example, is elegantly profiled, streamlined, and re-envisioned in Kathy Charmaz's (2006) *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*; while Graham R. Gibbs' (2007) *Analyzing Qualitative Data* provides an elegant survey of basic analytic processes.

*The Coding Manual* does not maintain allegiance to any one specific research genre or methodology. Throughout this book you'll read a breadth of perspectives on codes and coding, sometimes purposely juxtaposed to illustrate and highlight the diverse opinions among scholars in the field. No one, including myself, can claim final authority on the "best" way to code qualitative data. In fact, there are a few instances where I take moderate liberty with adapting and even renaming prescribed coding methods for clarity or flexibility's sake. This is not intended to standardize terminology within the field, but simply to employ consistency throughout this particular manual. My perspective acknowledges and promotes the pragmatist paradigm (Patton, 2002), which chooses "the right tool for the right job" since all research questions, methodologies, conceptual frameworks, and fieldwork parameters are context-specific.

I also wrote this manual because I find it problematic (but not difficult) to teach coding in my own qualitative research methods course. I provide students with an array of readings about the process from multiple sources because I have yet to find that single satisfactory book (to me) that focuses exclusively on the topic. General introductory texts in qualitative inquiry are so numerous and well-written that it becomes difficult not to find the best one to use, but which one of such quality works to select as the primary textbook. This manual supplements introductory works in the subject because most limit their discussions about coding to the writer's prescribed, preferred, or signature methods. I wanted to provide in a single resource a selected collection of various coding methods developed by other researchers (and myself) that provides students and colleagues a handy reference for classroom exercises and assignments, and for their own independent research for thesis and dissertation fieldwork and future qualitative studies. But by no means is it an exhaustive resource. I deliberately exclude such discipline-specific methods as Behavior Coding (which notes problematic respondent and interviewer behaviors during survey interviews [Singleton & Straits, 2002, p. 65]) and such signature methods as the Davis Observation Code system (for medical interviews [Zoppi & Epstein, 2002, p. 375]). If you need additional information and explanation about the coding methods, check the References.

*The Coding Manual* is intended primarily as a reference work. It is not necessarily meant to be read cover-to-cover, but it certainly can be if you wish to acquaint yourself with all 29 coding methods profiles and their analytic possibilities. There are, in fact, several principles related to coding matters not discussed in the first two chapters that are unique to some of the profiles. If you choose to review all the contents, read selected sections at a time, not all of them in one sitting, otherwise it can overwhelm you. If you're scanning the manual to see which coding method(s) might be appropriate for your particular study,

read the profiles' Description and Applications sections to see if further reading of the profile is merited. It's doubtful you'll use every coding method included in this manual for your particular research endeavors throughout your career, but they are available here on an "as needed" basis for your unique projects. Like an academic curriculum, the sequential order of the profiles has been carefully considered. They don't necessarily progress in a linear manner from simple to complex, but are clustered generally from the fundamental to the intermediate to the advanced.

### **What is a Code?**

A code in qualitative inquiry is most often a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data. The data can consist of interview transcripts, participant observation field notes, journals, documents, literature, artifacts, photographs, video, websites, e-mail correspondence, and so on. The portion of data to be coded during First Cycle coding processes can range in magnitude from a single word to a full sentence to an entire page of text to a stream of moving images. In Second Cycle coding processes, the portions coded can be the exact same units, longer passages of text, and even a reconfiguration of the codes themselves developed thus far. Just as a title represents and captures a book or film or poem's primary content and essence, so does a code represent and capture a datum's primary content and essence.

### ***Coding examples***

An example of a coded datum, as it is presented in this manual, looks like this when taken from a set of field notes about an inner city neighborhood. The one-word capitalized code in the right column is called a Descriptive Code, which summarizes the primary topic of the excerpt:

<sup>1</sup> I notice that the grand majority of homes have chain link fences in front of them. There are many dogs (mostly German shepherds) with signs on fences that say "Beware of the Dog."

<sup>1</sup> SECURITY

Here is an example of several codes applied to data from an interview transcript in which a high school senior describes his favorite teacher. The codes are based on what outcomes the student receives from his mentor. Note that one of the codes is taken directly from what the participant himself says and is placed in quotation marks – this is called an In Vivo Code:

- <sup>1</sup> He cares about me. He has never told me but he does. <sup>2</sup> He's always been there for me, even when my parents were not. He's one of the few things that I hold as a constant in my life. So it's nice. <sup>3</sup> I really feel comfortable around him.
- <sup>1</sup> SENSE OF SELF-WORTH  
<sup>2</sup> STABILITY  
<sup>3</sup> "COMFORTABLE"

Did you agree with the codes? Did other words or phrases run through your mind as you read the data? It's all right if your choices differed from mine. Coding is not a precise science; it's primarily an interpretive act.

Also be aware that a code can sometimes *summarize* or *condense* data, not simply *reduce* it. The introductory examples above were kept purposely simple and direct. But depending on the researcher's academic discipline, ontological and epistemological orientations, theoretical and conceptual frameworks, and even the choice of coding method itself, some codes can attribute more evocative meanings to data. In the excerpt below, a mother describes her teenage son's troubled school years. The codes emerge from the perspective of middle- and junior high school years as a difficult period for most youth. They are not specific types of codes; they are "first impression" phrases derived from an open-ended process called Initial Coding:

- <sup>1</sup> My son, Barry, went through a really tough time about, probably started the end of fifth grade and went into sixth grade. <sup>1</sup> MIDDLE-SCHOOL HELL
- <sup>2</sup> When he was growing up young in school he was a people-pleaser and his teachers loved him to death. <sup>2</sup> TEACHER'S PET
- <sup>3</sup> Two boys in particular that he chose to try to emulate, wouldn't, were not very good for him. <sup>3</sup> BAD INFLUENCES
- <sup>4</sup> They were very critical of him, they put him down all the time, and he kind of just took that and really kind of internalized it, I think, for a long time. <sup>4</sup> TWEEN ANGST
- <sup>5</sup> In that time period, in the fifth grade, early sixth grade, they really just kind of shunned him all together, and so his network as he knew it was gone. <sup>5</sup> THE LOST BOY

Note that when we reflect on a passage of data to decipher its core meaning, we are *decoding*; when we determine its appropriate code and label it, we are *encoding*. For ease of reference throughout this manual, *coding* will be the sole term used. Simply understand that coding is the transitional process between data collection and more extensive data analysis.

**Coding for patterns**

In the examples presented thus far, each unit of data was assigned its own unique code. This is due primarily to the short length of the excerpts. In larger and complete data sets, you will find that several to many of the same codes will be used repeatedly throughout. This is both natural and deliberate – natural because there are mostly repetitive patterns of action and consistencies in human affairs, and deliberate because one of the coder's primary goals is to find these repetitive patterns of action and consistencies in human affairs as documented in the data. In the example below, note how the same Process Code (a word or phrase which captures action) is used twice during this small unit of elementary school classroom activity:

- |  |   |
|--|---|
| <p><sup>1</sup> Mrs. Jackson rises from her desk and announces, "OK, you guys, let's get lined up for lunch. Row One." Five children seated in the first row of desks rise and walk to the classroom door. Some of the seated children talk to each other.</p> | <p><sup>1</sup> LINING UP FOR LUNCH</p> |
| <p><sup>2</sup> Mrs. Jackson looks at them and says, "No talking, save it for the cafeteria.</p>   | <p><sup>2</sup> MANAGING BEHAVIOR</p>   |
| <p><sup>3</sup> Row Two." Five children seated in the second row of desks rise and walk to the children already standing in line.</p>  | <p><sup>3</sup> LINING UP FOR LUNCH</p> |

Another way the above passage could be coded is to acknowledge that MANAGING BEHAVIOR is not a separate action or an interruption of the routine that disrupts the flow of LINING UP FOR LUNCH, but to interpret that MANAGING BEHAVIOR is an embedded or interconnected part of the larger social scheme that composes LINING UP FOR LUNCH. The coding might appear thusly, using a method called Simultaneous Coding (which applies two or more codes within a single datum):

- |  |  |
|--|--|
| <p><sup>1</sup> Mrs. Jackson rises from her desk and announces, "OK, you guys, let's get lined up for lunch. Row One." Five children seated in the first row of desks rise and walk to the classroom door. Some of the seated children talk to each other.</p> | <p><sup>1</sup> LINING UP FOR LUNCH</p>                  |
| <p><sup>1a</sup> Mrs. Jackson looks at them and says, "No talking, save it for the cafeteria.</p>  | <p style="margin-left: 20px;"><sup>1a</sup> MANAGING</p> |
| <p><sup>1</sup> Row Two." Five children seated in the second row of desks rise and walk to the children already standing in line.</p>  | <p style="margin-left: 20px;">BEHAVIOR</p>               |

Take note of some important caveats when it comes to understanding patterns and regularity: idiosyncrasy *is* a pattern (Saldana, 2003, pp. 118–22) and

there can be patterned variation in data (Agar, 1996, p. 10). Sometimes we code and categorize data by what participants talk about. They may all share with you their personal perceptions of school experiences, for example, but their individual value, attitude, and belief systems about education may vary greatly from being bored and disengaged to being enthusiastic and intrinsically motivated. When you search for patterns in coded data to categorize them, understand that sometimes you may group things together not just because they are exactly alike or very much alike, but because they might also have something in common – even if, paradoxically, that commonality consists of differences.

For example, each one of us may have a strong opinion about who should be leading our country. The fact that we each have an individual opinion about that issue is what we have in common. As for *whom* we each believe should be leading the country, that's where the differences and variations occur. Acknowledge that a confounding property of category construction in qualitative inquiry is that data within them cannot always be precisely and discretely bounded; they are within "fuzzy" boundaries at best (Tesch, 1990, pp. 135–8). That's why a method called Simultaneous Coding is an option we have, when needed. Finally, Hatch (2002) offers that you think of patterns not just as stable regularities but as varying forms. A pattern can be characterized by:

- similarity (things happen the same way)
- difference (they happen in predictably different ways)
- frequency (they happen often or seldom)
- sequence (they happen in a certain order)
- correspondence (they happen in relation to other activities or events)
- causation (one appears to cause another) (p. 155)

### **Coding filters**

The act of coding requires that you wear your researcher's analytic lens. But how you perceive and interpret what is happening in the data depends on what type of filter covers that lens. For example, consider the following statement from an older male: "There's just no place in this country for illegal immigrants. Round them up and send those criminals back to where they came from." One researcher, a grounded theorist using In Vivo Coding to keep the data rooted in the participant's own language, might code the datum this way:

<sup>1</sup> There's just no place in this country for illegal immigrants. Round them up and send those criminals back to where they came from.

<sup>1</sup> "NO PLACE"

A second researcher, an urban ethnographer employing Descriptive Coding to document and categorize the breadth of opinions stated by multiple participants, might code the same datum this way:

<sup>1</sup> There's just no place in this country for illegal immigrants. Round them up and send those criminals back to where they came from.      <sup>1</sup> IMMIGRATION ISSUES

And a third researcher, a critical race theorist employing Values Coding to capture and label subjective perspectives, may code the exact same datum this way:

<sup>1</sup> There's just no place in this country for illegal immigrants. Round them up and send those criminals back to where they came from.      <sup>1</sup> XENOPHOBIA

The collection of coding methods in this manual is a repertoire of possible filters to consider and apply to your approaches to qualitative inquiry. But even before that, your level of personal involvement as a participant observer – as a peripheral, active, or complete member during fieldwork – filters how you perceive, document, and thus code your data (Adler & Adler, 1987). So do the types of questions you ask and the types of responses you receive during interviews (Kvale, 1996; Rubin & Rubin, 1995), the detail and structuring of your field notes (Emerson, Fretz, & Shaw, 1995), the gender and race/ethnicity of your participants – and yourself (Behar & Gordon, 1995; Stanfield & Dennis, 1993), and whether you collect data from adults or children (Greene & Hogan, 2005; Zwiers & Morrissette, 1999).

Merriam (1998) states, “our analysis and interpretation – our study’s findings – will reflect the constructs, concepts, language, models, and theories that structured the study in the first place” (p. 48). And it is not only your approach to qualitative inquiry (e.g., case study, ethnographic, phenomenological) and ontological, epistemological, and methodological issues that influence and affect your coding decisions (Creswell, 2007; Mason, 2002). Sipe and Ghiso (2004), in their revealing narrative about coding dilemmas for a children’s literacy study, note that “All coding is a judgment call” since we bring “our subjectivities, our personalities, our predispositions, [and] our quirks” to the process (pp. 482–3).

### **Coding as a heuristic**

The majority of qualitative researchers will code their data both during and after collection as an analytic tactic, for coding *is* analysis (Miles & Huberman, 1994, p. 56). Differing perspectives, however, attest that “Coding and analysis are not synonymous, though coding is a crucial aspect of analysis” (Basit, 2003, p. 145).

Coding is a heuristic (from the Greek, meaning “to discover”) – an exploratory problem-solving technique without specific formulas to follow. Coding is only the initial step toward an even more rigorous and evocative analysis and interpretation for a report. Coding is not just labeling, it is *linking*: “It leads you from the data to the idea, and from the idea to all the data pertaining to that idea” (Richards & Morse, 2007, p. 137).

And, coding is a cyclical act. Rarely is the first cycle of coding data perfectly attempted. The second cycle (and possibly the third and fourth, and so on) of recoding further manages, filters, highlights, and focuses the salient features of the qualitative data record for generating categories, themes, and concepts, grasping meaning, and/or building theory. Coffey and Atkinson (1996) propose that “coding is usually a mixture of data [summation] and data complication ... breaking the data apart in analytically relevant ways in order to lead toward further questions about the data” (pp. 29–31).

Dey (1999) posits, though his original intent was to be critical, “With categories we impute meanings, with coding we compute them” (p. 95). To some, *code* is a dirty “four-letter word.” A few research methodologists perceive a code as mere shorthand or an abbreviation for the more important category yet to be discovered. Unfortunately, some use the terms *code* and *category* interchangeably and even in combination when they are, in fact, two separate components of data analysis. I advocate that qualitative codes are essence-capturing and essential elements of the research story that, when clustered together according to similarity and regularity – a pattern – they actively facilitate the development of categories and thus analysis of their connections. Ultimately, I like one of Charmaz’s (2006) metaphors for the process when she states that coding “generates the bones of your analysis. ... [I]ntegration will assemble those bones into a working skeleton” (p. 45).

### **Codifying and Categorizing**

To codify is to arrange things in a systematic order, to make something part of a system or classification, to categorize. When codes are applied and reapplied to qualitative data, you are codifying – a process that permits data to be “segregated, grouped, regrouped and relinked in order to consolidate meaning and explanation” (Grbich, 2007, p. 21). Bernard (2006) succinctly states that analysis “is the search for patterns in data and for ideas that help explain why those patterns are there in the first place” (p. 452). Coding is thus a method that enables you to organize and group similarly coded data into categories or “families” because they share some characteristic – the beginning of a pattern (see the examples in Pattern Coding and Focused Coding in Chapter Four). You use



classification reasoning plus your tacit and intuitive senses to determine which data “look alike” and “feel alike” when grouping them together (Lincoln & Guba, 1985, p. 347).

### ***From codes to categories***

For example, in Harry, Sturges, & Klingner’s (2005) ethnographic study of the overrepresentation of minorities in special education programs, data initially coded as classroom MATERIALS, COMPUTERS, and TEXTBOOKS were categorized under the major heading, **Resources**. As their study continued, another major category emerged labeled **Teacher Skills** with the subcategories **Instructional Skills** and **Management Skills**. The codes subsumed under these subcategories – part of the overall hierarchical “coding scheme” (Lewins & Silver, 2007) – were:

#### **Category: Teacher Skills**

##### **Subcategory 1: Instructional Skills**

Code: PEDAGOGICAL  
 Code: SOCIO-EMOTIONAL  
 Code: STYLE/PERSONAL EXPRESSION  
 Code: TECHNICAL

##### **Subcategory 2: Management Skills**

Code: BEHAVIORIST TECHNIQUES  
 Code: GROUP MANAGEMENT  
 Code: SOCIO-EMOTIONAL  
 Code: STYLE (overlaps with instructional style)  
 Code: UNWRITTEN CURRICULUM

As another example, in Basit’s (2003) study of the aspirations of teenage British Muslim girls, analysis of interview data with the girls, their parents, and their teachers brought forth 23 major categories that clustered under six major themes. One major theme was IDENTITY, and its related categories were **Ethnicity**, **Language**, and **Religion**. Under the theme CAREER ASPIRATIONS, the categories were **Career Choices**, **Unrealistic Aspirations**, and **Career Advice**.

Rubin & Rubin (1995) recommend that you refine the contents of each category (working within) from your data before you start comparing them with each other (working across) (pp. 241, 251). Maykut & Morehouse (1994) refine each category by developing a rule for inclusion in the form of a propositional statement, coupled with sample data. For example, if an emergent category in a case study is labeled **Physical Health**, its rule for inclusion as a propositional statement might read:

**Physical Health:** The participant shares matters related to physical health such as wellness, medication, pain, etc.: “I’m on 25 milligrams of amitriptyline each night”; “I hate going to the gym.”

Emergent categories might also evolve as conceptual processes rather than descriptive topics such as:

***Inequity:*** Participants perceive unfair treatment directed toward themselves and favoritism directed toward others: “I’ve been working here for over 25 years and some newcomers are making higher salaries than me.”

The categories’ propositional statements are then compared with each other to discern possible relationships to create an *outcome proposition* based on their combination.

### ***Recoding and recategorizing***

Rarely will anyone get coding right the first time. Qualitative inquiry demands meticulous attention to language and deep reflection on the emergent patterns and meanings of human experience. Recoding can occur with a more attuned perspective using First Cycle methods again, while Second Cycle methods describe those processes that might be employed during the second (and third and possibly fourth ... review of data.

As you code and recode, expect – or rather, strive for – your codes and categories to become more refined. Some of your First Cycle codes may be later subsumed by other codes, relabeled, or dropped all together. As you progress toward Second Cycle coding, there may be some rearrangement and reclassification of coded data into different and even new categories. Abbott (2004) cleverly likens the process to “decorating a room; you try it, step back, move a few things, step back again, try a serious reorganization, and so on” (p. 215).

For example, I observed and interviewed fourth and fifth grade children to learn the ways they hurt and oppress each other (Saldana, 2005b). This was preparatory fieldwork before an action research project that attempted to empower children with strategies, learned through improvised dramatic simulations and role-playing, for dealing with bullying in the school environment. I initially categorized their responses into **Physical** and **Verbal** forms of oppression. Some of the codes that fell under these categories were:

#### ***Category: Physical Oppression***

Code: PUSHING  
Code: FIGHTING  
Code: SCRATCHING

#### ***Category: Verbal Oppression***

Code: NAME-CALLING  
Code: THREATENING  
Code: LAUGHING AT

As coding continued, I observed that a few oppressions were a combination of both physical *and* verbal actions. For example, a child can EXCLUDE others

physically from a game, accompanied with a verbal statement such as, “You can’t play with us.” Hence, a third major category emerged: **Physical and Verbal Oppression**.

As the study continued, more data were collected through other methods, and gender differences in children’s perceptions and enactment of oppression became strikingly apparent. To participants, oppression was not about the body and voice; it was about “*force*” and “*feelings*.” The three initial categories were eventually reduced to two, and renamed based on what seemed to resonate with gender-based observations. The new categories and a few sample codes and rearranged subcodes included:

**Category: Oppression through Physical Force** (primarily but not exclusively by boys)

Code: FIGHTING  
 Subcode: SCRATCHING  
 Subcode: PUSHING  
 Subcode: PUNCHING

**Category: Oppression through Hurting Others’ Feelings** (primarily but not exclusively by girls)

Code: PUTTING DOWN  
 Subcode: NAME-CALLING  
 Subcode: TEASING  
 Subcode: TRASH TALKING

See the Domain and Taxonomic Coding profile in Chapter Three for an extended discussion of this case, and the Initial and Focused Coding examples in Chapters Three and Four respectively to see how a series of codes gets categorized.

### **From codes and categories to theory**

Some categories may contain clusters of coded data that merit further refinement into subcategories. And when the major categories are compared with each other and consolidated in various ways, you begin to transcend the “reality” of your data and progress toward the thematic, conceptual, and theoretical. As a very basic process, codifying usually follows an ideal and streamlined scheme illustrated in Figure 1.1. Keep in mind that the actual act of reaching theory is much more complex and messy than illustrated. Richards & Morse (2007) clarify that “categorizing is how we get ‘up’ from the diversity of data to the shapes of the data, the sorts of things represented. *Concepts* are how we get up to more general, higher-level, and more abstract constructs” (p. 157). Our ability to show how these themes and concepts systematically interrelate lead toward the development of theory (Corbin & Strauss, 2008, p. 55), though Layder (1998) contends that pre-established sociological theories can inform, if not drive, the initial coding process itself. The development of an original theory is not always a necessary outcome for qualitative

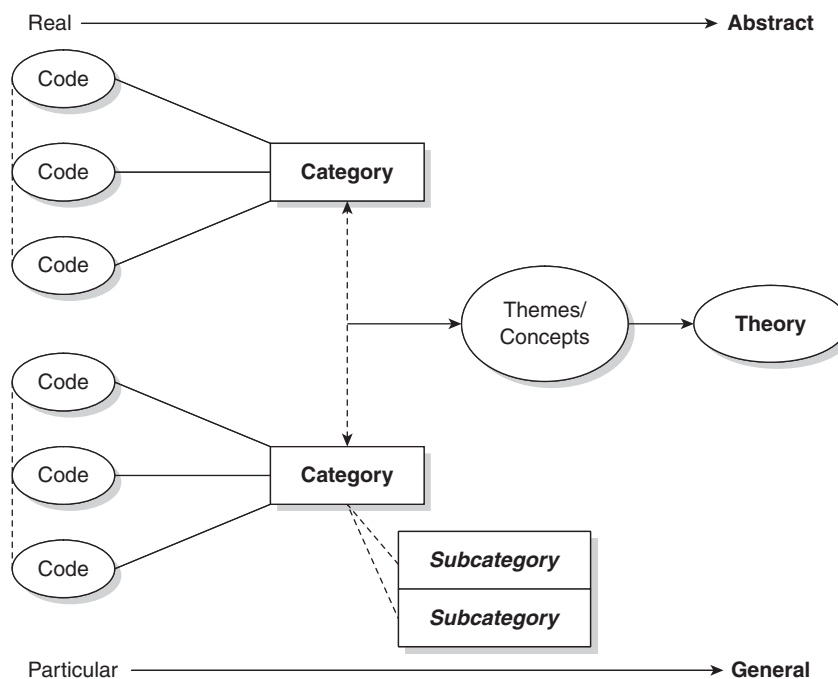


FIGURE 1.1 A streamlined codes-to-theory model for qualitative inquiry

inquiry, but acknowledge that preexisting theories drive the entire research enterprise, whether you are aware of them or not (Mason, 2002).

In the example above of children's forms of oppression, there were two major categories that emerged from the study: **Oppression through Physical Force**, and **Oppression through Hurting Others' Feelings**. So, what major themes or concepts can be developed from these categories? An obvious theme we noticed was that, in later childhood, *peer oppression is gendered*. One higher-level concept we constructed – an attempt to progress from the real to the abstract – was *child stigma*, based on the observation that children frequently label those who are perceived different in various ways “weird,” and thus resort to oppressive actions (Goffman, 1963). We could not, in confidence, formulate a formal theory from this study due to the limited amount of fieldwork time in the classrooms. But a key assertion (Erickson, 1986) we did develop and put forth, based on the contexts of this study, was:

To artist and activist Augusto Boal, adult participation in theatre for social change is “rehearsal for the revolution.” With ages 9–11 children, however, their participation in theatre for social change seems more like an “audition” for preadolescent social interaction. The key assertion of this study is: *Theatre for social change overtly reveals the interpersonal social systems and power hierarchies within an elementary school classroom microculture,*

*because the original dramatic simulations children create authentically reflect their statuses and stigmas.* It diagnostically shows which children are leaders, followers, resisters, and targets; who is influential and who is ignored; which children may continue to assert dominance in later grade levels; and which children may succumb to those with more authority in later grade levels. (adapted from Saldaña, 2005b, p. 131)

This key assertion, like a theory, attempts to progress from the particular to the general by inferring *transfer* – that what was observed in just six elementary school classrooms at one particular site may also be observed in comparable elementary school classrooms in other locations. This assertion also progresses from the particular to the general by *predicting patterns* of what may be observed and what may happen in similar present and future contexts.

### **The differences between codes and themes**

Several qualitative research texts recommend that you initially “code for themes.” That, to me, is misleading advice because it muddies the water. A theme is an *outcome* of coding, categorization, and analytic reflection, not something that is, in itself, coded (that is why there is no “Theme Coding” method in this manual, but there *are* references to thematic analysis and a section called Themeing the Data). A datum is initially and, when needed, secondarily coded to discern and label its content and meaning according to the needs of the inquiry. Rossman & Rallis (2003) explain the differences: “think of a category as a *word or phrase* describing some segment of your data that is *explicit*, whereas a theme is a *phrase or sentence* describing more *subtle and tacit* processes” (p. 282, emphasis added). As an example, SECURITY can be a code, but A FALSE SENSE OF SECURITY can be a theme.

Qualitative researchers are not algorithmic automatons. If we’re carefully reading and reviewing the data before and as we’re formally coding them, we can’t help but notice a theme or two (or a pattern, trend, or concept) here and there. Make a note of it in an analytic memo (see Chapter Two) when it happens, for it can sometimes guide your continued coding processes. A set of themes is a good thing to emerge from analysis, but at the beginning cycles there are other rich discoveries to be made with specific coding methods that explore such phenomena as participant process, emotions, and values.

### **What Gets Coded?**

Richards & Morse (2007) humorously advise for analytic work, “If it moves, code it” (p. 146). But what exactly *gets* coded in the data?

**Units of social organization**

Lofland, Snow, Anderson, & Lofland (2006) note that social life happens at four coordinates, “the intersection of one or more *actors* [participants] engaging in one or more *activities* (behaviors) at a particular *time* in a specific *place*” (p. 121, emphasis in original). The authors first outline major *units* of social organization into:

- 1 cultural practices (daily routines, occupational tasks, microcultural activity, etc.);
- 2 episodes (unanticipated or irregular activities such as divorce, championship games, natural disasters, etc.);
- 3 encounters (a temporary interaction between two or more individuals such as sales transactions, panhandling, etc.);
- 4 roles (student, mother, customer, etc.) and social types (bully, tight-ass, geek, etc.);
- 5 social and personal relationships (husband and wife, party-goers, etc.);
- 6 groups and cliques (gangs, congregations, families, jocks, etc.);
- 7 organizations (schools, fast-food restaurants, prisons, corporations, etc.);
- 8 settlements and habitats (villages, neighborhoods, etc.); and
- 9 subcultures and lifestyles (the homeless, skinheads, gay leather bears, etc.)

But you won't find in this manual any coding methods based on the major units outlined above such as “Encounter Coding,” “Organization Coding” or “Lifestyle Coding.” When the units above are combined with *aspects* listed below, they then become topics for study *and coding*. Lofland et al.'s aspects include:

- 1 cognitive aspects or meanings (e.g., ideologies, rules, self-concepts, identities);
- 2 emotional aspects or feelings (e.g., sympathy in health care, road rage, workplace satisfaction);
- 3 hierarchical aspects or inequalities (e.g., racial inequality, battered women, high school cliques)

Aside from examining the magnitude and frequency of social life outlined above, Lofland et al. also recommend examining how participant agency interacts and interplays with structures and processes, plus causes and consequences observed in the data (2006, pp. 144–67).

Aspects *in combination with* units lend themselves to such First Cycle coding methods (see Chapter Three) as Emotion Coding, Values Coding, and Versus Coding. Structures and processes can be discerned through Descriptive Coding, Process Coding, and Domain and Taxonomic Coding, while causes and consequences can be discerned through Pattern Coding or grounded theory's

Second Cycle coding methods (see Chapter Four). But note that some question qualitative research's ability to assert causality: "the understanding of human experience is a matter of chronologies more than of causes and effects" (Stake, 1995, p. 39).

### **Amounts of data to code**

One related issue with which qualitative research methodologists disagree is the amount of the data corpus – the total body of data – that should be coded. Some (e.g., Lofland et al., 2006; Strauss, 1987; cf. Wolcott, 1999) feel that every recorded fieldwork detail is worthy of consideration, for it is from the patterned minutiae of daily life that we might generate significant social insight. Others (e.g., Seidman, 2006), if not most, feel that only the most salient portions of the corpus merit examination, and that even up to one half of the total record can be summarized or deleted, leaving the primary half for intensive data analysis. The danger is that the portions deleted might contain the as yet unknown units of data that could pull everything together, or include the negative case that motivates a rethinking of a code, category, theme, concept, theory, or assertion. Postmodern perspectives on ethnographic texts consider all documentation and reports partial and incomplete anyway, so the argument for maintaining and coding a full or reduced data corpus seems moot. Amount notwithstanding, insure that you have not just sufficient qualitative but sufficient *quality* data with which to work that have been appropriately transcribed and formatted (see Poland, 2002).

I have learned from years of qualitative data analysis that, only with experience, I now feel more secure knowing and feeling what is important in the data record and what is not, and thus code what rises to the surface – "relevant text" as Auerbach & Silverstein (2003) label it. The beginning of my fieldwork career, however, was a major learning curve for me, and I coded anything and everything that was collected. I advise the same for novices to qualitative research. You, too, will eventually discover from experience what matters and what doesn't in the data corpus. (Of course, there will always be brief passages of minor or trivial consequence scattered throughout interviews and field notes. Code these N/A – not applicable.)

So, what *gets* coded? Slices of social life recorded in the data – participant activities, perceptions, and the tangible documents and artifacts produced by them. Your own reflective data in the form of analytic memos (discussed in Chapter Two) and observer's comments in field notes are also substantive material for coding. The process does not have to be approached as if it were some elusive mystery or detective story with deeply hidden clues and misleading red herrings scattered throughout. If "human actions are based upon, or infused by, social meanings: that is, by intentions, motives, beliefs, rules, and values"

(Hammersley & Atkinson, 1995, p. 7), then why not just code these actions and social meanings directly (assuming they are represented in your data and your inferential skills are working at an optimum)? The entire process *and products* of creating data about the data in the form of codes, analytic memos, and graphical summaries are “metadata activities” (MacQueen & Guest, 2008, p. 14).

### **The Mechanics of Coding**

As you prepare text-based qualitative data for manual (i.e., paper and pencil) coding and analyzing, lay out printed interview transcripts, field notes, and other researcher-generated materials in double-spaced format on the left half or left two-thirds of the page, keeping a wide right-hand margin for writing codes and notes. Rather than keeping your data running together as long unbroken passages, separate the text into short paragraph-length units with a line break in-between them whenever the topic or subtopic appears to change (as best as you can, because in real life “social interaction does not occur in neat, isolated units” [Glesne, 2006, p. 150]). Gee, Michaels, & O’Connor (1992) call these unit breaks and their rearrangement into poetic-like verses for discourse analysis “stanzas” of text, and emphasize that “formatting choices are a part of the analysis and may reveal or conceal aspects of meaning and intent” (p. 240). Unit divisions will also play a key role in formatting data for CAQDAS – Computer Assisted Qualitative Data Analysis Software – programs (discussed later).

#### **Pre-coding**

In addition to coding with words and short phrases, never overlook the opportunity to “pre-code” (Layder, 1998) by circling, highlighting, bolding, underlining, or coloring rich or significant participant quotes or passages that strike you – those “codable moments” worthy of attention (Boyatzis, 1998). Creswell (2007, pp. 168–9) recommends that such quotes found in data contained in a CAQDAS program file can be simultaneously coded *as* QUOTES with their other codes to enable later retrieval. Selected programs have areas dedicated to storing intriguing quotations for later access. These data can become key pieces of the evidentiary warrant to support your propositions, assertions, or theory (Booth, Colomb, & Williams, 2003; Erickson, 1986; Lofland et al., 2006), and serve as illustrative examples throughout your report. The codes or quotes may even be so provocative that they become part of the title, organizational framework, or through-line of the report. For example, in my study of theatre of the oppressed (i.e., theatre for social change) with elementary school children, I was puzzled why young



people continually employed combative tactics during improvisational dramatic simulations to resolve imbalanced power issues, when I was trying to teach them proactive peace-building efforts. A fourth-grade girl poignantly provided the answer when we discussed my concerns by explaining to me, “Sometimes, you can’t be nice to deal with oppression” (Saldaña, 2005b, p. 117). The quote was so powerful that it began my final research report as a datum that would both capture the reader’s interest and later explain the through-line of the study.

### ***Preliminary jottings***

Start coding *as* you collect and format your data, not after all fieldwork has been completed. When you’re writing up field notes, transcribing recorded interviews, or filing documents you gathered from the site, jot down any preliminary words or phrases for codes on the notes, transcripts, or documents themselves, or as an analytic memo or entry in a research journal for future reference. They don’t have to be accurate or final at this point, just ideas for analytic consideration while the study progresses. Be wary of relying on your memory for future writing. Get your thoughts, however fleeting, documented in some way.

Also make certain that these code jottings are distinct in some way from the body of data – bracketed, capitalized, italicized, bolded, etc. Liamputtong & Ezzy (2005, pp. 270–3) recommend formatting pages of data into three columns rather than two. The first and widest column contains the data themselves – interview transcripts, field notes, etc. The second column contains space for preliminary code notes and jottings, while the third column lists the final codes. The second column’s ruminations or first impressions may help provide a transitional link between the raw data and codes:

#### **COLUMN 1**

##### ***Raw Data***

<sup>1</sup> The closer I get to retirement age, the faster I want it to happen. I’m not even 55 yet and I would give anything to retire now. But there’s a mortgage to pay off and still a lot more to sock away in savings before I can even think of it. I keep playing the lottery, though, in hopes of winning those millions. No luck yet.

#### **COLUMN 2**

##### ***Preliminary Codes***

*“retirement age”*

*financial obligations*

*dreams of early retirement*

#### **COLUMN 3**

##### ***Final Code***

<sup>1</sup> RETIREMENT ANXIETY

Virtually all methodologists recommend initial and thorough readings of your data while writing analytic memos or jotting in the margins tentative ideas for codes, topics, and noticeable patterns or themes. Write your code words or phrases completely rather than abbreviating them to mnemonics or assigning them reference numbers. Avoid such truncations as “PROC-AN CD” or “122-A,” which just make the decoding processes of your brain work much harder than they need to during analysis.

### **Questions to consider as you code**

Auerbach & Silverstein (2003, p. 44) recommend that you keep a copy of your research concern, theoretical framework, central research question, goals of the study, and other major issues on one page in front of you to keep you focused and allay your anxieties because the page focuses your coding decisions. Emerson, Fretz, & Shaw (1995) advise a general list of questions to consider when coding field notes (in chronological order), regardless of research purpose:

- What are people doing? What are they trying to accomplish?
- How, exactly, do they do this? What specific means and/or strategies do they use?
- How do members talk about, characterize, and understand what is going on?
- What assumptions are they making?
- What do I see going on here? What did I learn from these notes?
- Why did I include them? (p. 146)

I would add to this list the question I ask myself during all cycles of coding and data analysis: “What strikes you?” Creswell (2007, p. 153) notes that a code can emerge from data that is not only expected but even surprising, unusual, or conceptually interesting.

### **Coding contrasting data**

If you are working with multiple participants in a study, it may help to code one participant’s data first, then progress to the second participant’s data. You might find that the second data set will influence and affect your recoding of the first participant’s data, and the consequent coding of the remaining participants’ data. The same may hold true for a coding system applied to an interview transcript first, then to a day’s field notes, then to a document. Bazeley (2007) recommends that the second document coded should contrast “in some important way with the first ... to maximize the potential for variety in concepts (or in their forms of expression) early in the process” (p. 61). Be aware

that, depending on the coding method(s) chosen, some codes may appear more frequently in selected types of data than others. Selected CAQDAS program functions can keep you abreast of the codes and their frequencies as analysis progresses.

### The Numbers of Codes

The actual number of codes, categories, themes and/or concepts you generate for each project will vary and depend on many contextual factors, yet one question students ask most is how often codes “should” get applied to qualitative data. The answer depends on the nature of your data, which particular coding method you select for analysis, and how detailed you want or need to be – in other words, more filters to consider.

#### **“Lumping” and “splitting” the data**

For example, the following data excerpt is from a speech by a second year, inner city, grades K–8 school teacher speaking to pre-service education majors enrolled in a university teaching methods course (Saldaña, 1997). She has just completed several poignant vignettes about some of her most difficult students. Notice that just one In Vivo Code is applied to capture and represent the essence of this entire 145-word excerpt – a broad brush-stroke representation called Holistic Coding:

<sup>1</sup> I'm not telling you this to depress you or scare you but it was a reality for me. I thought I was so ready for this population because I had taught other groups of kids. But this is such a unique situation, the inner city school. No, I should take that back: It's not as much of a unique situation *anymore*. There are more and more schools that are turning into inner city schools ... I really had to learn about the kids. I had to learn about the culture, I had to learn the language, I had to learn the gang signals, I had to learn what music was allowed, what t-shirts they could wear on certain days and not on other days. There was just a lot to learn that I had never even thought about.

<sup>1</sup> “A LOT TO LEARN”

The above has been colloquially called “lumper” coding. The opposite is someone who codes as a “splitter,” or, one who splits the data into smaller codable moments. Thus, more detailed In Vivo Coding of the exact same passage might appear thusly:

I'm not telling you this to depress you  
 or scare you but it was a <sup>1</sup> reality  
 for me. <sup>2</sup> I thought I was so ready  
 for this population because I had  
 taught other groups of kids. But this  
 is such a <sup>3</sup> unique situation, the  
 inner city school. No, I should take that  
 back: It's not as much of a unique  
 situation *anymore*. There are more and  
 more schools that are turning into <sup>4</sup> inner  
 city schools ... <sup>5</sup> I really had to learn  
 about the kids. I had to learn about  
<sup>6</sup> the culture, I had to learn the language,  
 I had to learn the gang signals, I had to  
 learn what music was allowed, what  
 t-shirts they could wear on certain days  
 and not on other days. There was just  
<sup>7</sup> a lot to learn that I had never even  
 thought about.

<sup>1</sup> "REALITY"  
<sup>2</sup> "I THOUGHT IWAS SO READY"  
<sup>3</sup> "UNIQUE SITUATION"  
<sup>4</sup> "INNER CITY SCHOOLS"  
<sup>5</sup> "I REALLY HAD TO LEARN"  
<sup>6</sup> "THE CULTURE"  
<sup>7</sup> "A LOT TO LEARN"

Now this 145-word excerpt is represented with seven codes rather than one. I state the numbers not to suggest that more is better or that less is more, but to highlight that lumping is an expedient coding method (with future detailed subcoding still possible), while splitting generates a more nuanced analysis from the start. Each approach has its advantages and disadvantages aside from the obvious factors of time and mental energy required. Lumping gets to the essence of categorizing a phenomenon while splitting encourages careful scrutiny of social action represented in the data. But lumping may lead to a superficial analysis if the coder does not employ conceptual words and phrases, while fine-grained splitting of data may overwhelm the analyst when it comes time to categorize the codes. During Second Cycle coding, you might collapse the original number of First Cycle codes into a smaller number as you reanalyze the data and find that larger segments of text are better suited to just one key code rather than several smaller ones. It is only from experience that you'll discover which approach works best for you, your particular study, and your particular research goals.

### ***The quantities of qualities***

Lichtman (2006) projects that most qualitative research studies in education will generate 80–100 codes that will be organized into 15–20 categories which eventually synthesize into five to seven major concepts (pp. 164–5). Creswell (2007) begins his analyses with a short-list of five to six Provisional Codes to begin the process of “lean coding.” This expands to no more than 25–30 categories that then combine into five to six major themes (p. 152). Other disciplines and varying approaches to qualitative inquiry may prescribe different sets of numbers

as general guidelines for analysis. The final number of major themes or concepts should be held to a minimum to keep the analysis coherent, but there is no standardized or magic number to achieve. Unlike Lichtman's five to seven central concepts and Creswell's five to six major themes, anthropologist Harry F. Wolcott (1994, p. 10) generally advises throughout his writings that three of anything major seems an elegant quantity for reporting qualitative work.

### ***The codebook or code list***

Since the number of codes can accumulate quite quickly and change as analysis progresses, keep a record of your emergent codes in a separate file as a codebook – a compilation of the codes, their content descriptions, and a brief data example for reference. CAQDAS programs, by default, will maintain a list of codes you have created for the project and provide space to define them. This can be reviewed periodically – both on the monitor screen and on hard copy – as coding progresses to assess its current contents and possible evolution. Maintaining this list provides an analytic opportunity to organize and reorganize the codes into major categories and subcategories. This management technique also provides a comparative list if you are working with multiple participants and sites. One school site's data, for example, may generate a list of codes significantly different from another school site.

Codebooks or CAQDAS code lists become especially critical as a set of coding standards when multiple team members work together on the same project's data. MacQueen et al. (2008) strongly recommend that each codebook entry should contain “the code, a brief definition, a full definition, guidelines for when to use the code, guidelines for when not to use the code, and examples” (p. 121). Also note that a codebook differs from an *index*, the latter being a coded composite of the data corpus, organized alphabetically, hierarchically, chronologically, categorically, etc. CAQDAS programs are superior for indexing functions with a qualitative data corpus.

### **Manual and CAQDAS Coding**

Some instructors of statistics and quantitative data analysis require that their students first learn how to “crunch the numbers” manually using only a pocket/hand calculator to provide them with cognitive understanding and ownership of the formulas and results. Once a statistical test has been administered this way, they can then use computers with software specifically designed to calculate numeric data.

Coding and qualitative data analysis have their equivalent trial. I am one of those instructors who require that my students first perform “manual” coding and qualitative data analysis using paper and pencil on hard copies of data

entered and formatted with basic word-processing software only. The reason is that each class assignment of data gathering is relatively small-scale and thus a manageable project to analyze in this manner. But if a student's dissertation project or my own independent research studies will require multiple participant interviews or extended fieldwork and extensive field note-taking, then CAQDAS becomes a vital and indispensable tool for the enterprise. Basit (2003) compared personal experiences between manual and electronic coding and concluded, "the choice will be dependent on the size of the project, the funds and time available, and the inclination and expertise of the researcher" (p. 143).

### ***Coding manually***

Trying to learn the basics of coding and qualitative data analysis simultaneously with the sometimes complex instructions and multiple functions of CAQDAS programs can be overwhelming for some, if not most. Your mental energies may be more focused on the software than the data. I recommend that for first-time or small-scale studies, code on hard-copy printouts first, not via a computer monitor (cf. Bazeley, 2007, p. 92). There is something about manipulating qualitative data on paper and writing codes in pencil that give you more control over and ownership of the work. Perhaps this advice stems from my admitted lack of technological expertise and old-fashioned ways of working that have become part of my "modus" operandi. Nevertheless, there is something to be said for a large area of desk or table space with multiple pages or strips of paper spread out to see the smaller pieces of the larger puzzle – a literal perspective not always possible on a computer's monitor screen. After you feel the codes are fairly well set from your initial hard-copy work, then transfer your codes onto the electronic file. But first, "Touch the data... Handling the data gets additional data out of memory and into the record. It turns abstract information into concrete data" (Graue & Walsh, 1998, p. 145). Even proponents of CAQDAS recommend that hard-copy printouts of code lists and coded data be generated occasionally to permit you to work with traditional writing materials such as red pens and highlighters to explore data in fresh ways.

### ***Coding electronically***

After you have gained some experience with hard-copy coding and have developed a basic understanding of the fundamentals of qualitative data analysis, apply that experiential knowledge base by working with CAQDAS. Keep in mind that CAQDAS itself does not actually code the data for you; that task is still the responsibility of the researcher. The software efficiently stores, organizes, manages, and reconfigures your data to enable human analytic reflection.

Some programs even enable coding of digital audio and video documents stored in their files. I advise that you work with a smaller portion of your data first, such as a day's field notes or a single interview transcript, before importing the data corpus into the program. As with all word-processed work on a computer, backup your original files as a precautionary measure.

Three major CAQDAS programs to explore, whose commercial websites provide online tutorials and demonstration software/manual downloads of their most current versions, are:

- ATLAS.ti: [www.atlasti.com](http://www.atlasti.com)
- MAXQDA: [www.maxqda.com](http://www.maxqda.com)
- NVivo: [www.qsrinternational.com](http://www.qsrinternational.com)

Refer to Lewins & Silver (2007) and Bazeley (2007) for accompanying literature on these programs. Also see Richards & Morse (2007, pp. 85–90) for what selected CAQDAS programs can and cannot do; Hahn (2008) and La Pelle (2004) for qualitative data analysis with basic word-processing software; and Brent & Slusarz (2003) for advanced computational strategies with software. Other CAQDAS programs, such as HyperRESEARCH and QDA Miner, are discussed and reviewed at an online forum for users: <http://caqdas.soc.surrey.ac.uk/>.

### ***Data formatting for CAQDAS***

The heading and paragraph formats of qualitative data such as field notes and, in particular, interview transcripts, need to conform consistently with the particular software package's prescriptions for text layout. This becomes vital for its coding and retrieval functions to work consistently and reliably. ATLAS.ti, MAXQDA, and NVivo all import and handle documents saved in rich text format, enabling you to employ supplemental “cosmetic” coding devices such as colored fonts, bolding, and italicizing in your data (Lewins & Silver, 2007, p. 61). One of the best features of some CAQDAS programs is their ability to display code labels themselves in various user-assigned colors for “at a glance” reference and visual classification. Figure 1.2 illustrates a sample screen shot from the most current version of NVivo. Note how the video data and its transcript are accompanied with codes and “coding stripes,” which delineate which portion of data is assigned a particular code.

### ***Coding capabilities with CAQDAS***

Selected qualitative data analysis programs permit you to do what you can do manually, such as: apply more than one code to the same passage or sequential

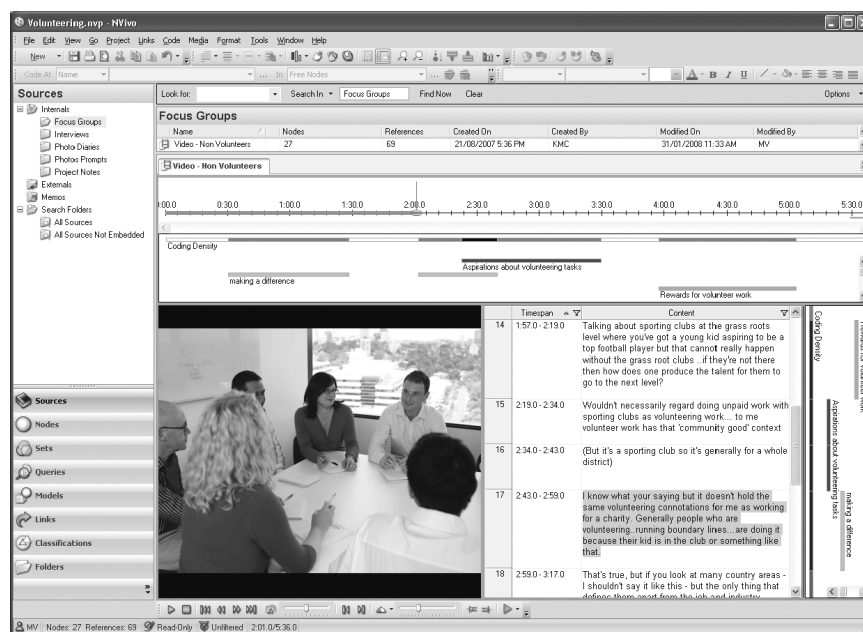


FIGURE 1.2 A Video Coding screenshot from NVivo, v. 8 (NVivo 8 is designed and developed by QSR International Pty Ltd. NVivo is a trademark or registered trademark of QSR International. Patents pending. [www.qsrinternational.com](http://www.qsrinternational.com))

passages of text (variously labeled in the methods literature as “simultaneous coding,” “double coding,” “co-occurrence coding,” “multiple coding,” or “overlap coding”); code a smaller portion of text within a larger portion of coded text (“subcoding,” “embedded coding,” or “nested coding”); and subsume several similarly coded passages under one larger code (“pattern coding,” “meta coding,” “umbrella coding,” or “hierarchical coding”); along with the ability to instantly and conveniently insert analytic memos related to a specific datum or code. Each CAQDAS program will employ its own distinct set of terms for its coding functions and operations, so refer to the user’s manual for specific ways of working.

CAQDAS, unlike the human mind, can maintain and permit you to organize evolving and potentially complex coding systems into such formats as hierarchies and networks for “at a glance” user reference. Figure 1.3 illustrates a sample window excerpt from ATLAS.ti’s Code Manager function, which lists each code by name and provides descending frequency counts of the codes in progress (serendipitously, the codes are extracted from work on a document about grounded theory).

Though I stated above that software does not code for you, there is an interesting utilitarian function called “auto coding” available in most CAQDAS



**Code Manager [HU: Grounded Theory - A Glossary]**

Codes Edit Miscellaneous Output View

Name	Grounded	Density	Author	Created	Modified
core category	17	0	Lea	05/31...	11/12...
category	17	0	Lea	05/31...	11/12...
theoretical sampling	15	0	Lea	05/31...	11/12...
research phases	14	0	Lea	05/31...	11/12...
provisional, revision	13	0	Lea	05/31...	11/12...
integration	13	0	Lea	05/31...	11/12...
concept	12	0	Lea	05/31...	11/12...
publication	12	0	Lea	05/31...	11/12...
theoretical sorting	12	0	Lea	05/31...	11/12...
memo	12	0	Lea	05/31...	11/12...
conceptual density	11	0	Lea	05/31...	11/12...
open coding	11	2	Lea	05/31...	01/21...
theoretical saturation	11	0	Lea	05/31...	11/12...
comparative analysis	10	0	Lea	05/31...	11/12...
verification	9	0	Lea	05/31...	11/12...
scope	8	0	Lea	05/31...	11/12...
grounded theory	8	0	Lea	05/31...	11/12...
coding	8	6	Lea	05/31...	01/21...
bias	7	0	Lea	05/31...	11/12...
generative question	7	0	Lea	05/31...	11/12...
conditions-consequences-tactic-strategies-interactions	7	1	Lea	05/31...	01/21...
conceptual linkage, relation (-ship)	7	1	Lea	05/31...	01/21...
hypothesis	7	0	Lea	05/31...	11/12...
variation	6	0	Lea	05/31...	11/12...
comparison	6	0	Lea	05/31...	11/12...
dimension	6	0	Lea	05/31...	11/12...
data collection	6	3	Lea	05/31...	01/21...
property, feature	6	0	Lea	05/31...	11/12...
theoretical sensitivity	6	0	Lea	05/31...	11/12...
generalizability	5	0	Lea	06/05...	11/12...
dimensionalizing	5	0	Lea	05/31...	11/12...
empirical grounding	5	0	Lea	05/31...	11/12...
in vivo code	4	1	Lea	05/31...	01/21...
selective coding	4	2	Lea	05/31...	01/21...
conceptualization	4	0	Lea	05/31...	11/12...
coding paradigm	4	0	Lea	05/31...	11/12...
systematic	4	0	Lea	05/31...	11/12...
scientific construct	4	1	Lea	05/31...	01/21...
relevance	4	0	Lea	05/31...	11/12...
diagram(ing)	4	0	Lea	05/31...	01/21...
explanatory power	4	0	Lea	05/31...	11/12...
experiential data	4	0	Lea	05/31...	11/12...
precision, specificity	4	0	Lea	06/05...	11/12...
labeling of codes	4	0	Lea	05/31...	11/12...
memo writing	3	3	Lea	05/31...	01/21...
evaluation criteria	3	0	Lea	06/05...	11/12...
indicator-concept-model	3	0	Lea	06/05...	11/12...
triad of analytic operations	3	3	Lea	05/31...	01/21...
completeness	3	0	Lea	05/31...	11/12...
code	2	2	Lea	05/31...	01/21...
axial coding	2	3	Lea	05/31...	01/21...

67 Codes

FIGURE 1.3 A Code Manager screenshot excerpt from ATLAS.ti, v.5.2 (courtesy of ATLAS.ti)

programs, which can alleviate some of the repetitiveness of manually coding similar passages of text. Passages have to be formatted in prescribed ways and contain the exact same word or phrase, however, for this function to work accurately. The ATLAS.ti handbook strongly recommends a manual review after auto coding has been performed to verify the software's coding assignments, and Lewins & Silver (2007) suggest that researchers should not feel "compelled to make use of auto coding just because it is available" (p. 21).

### **Searches and queries with CAQDAS**

Another one of CAQDAS's advantages over manual paper-and-pencil coding and analysis is its search and querying abilities to quickly collect and display key words and phrases and similarly-coded data for examination. Searches or queries of coded passages can even find where particular codes co-occur, overlap, appear in a sequence, or lie in proximity to each other. These search functions can perform such actions as retrieve, filter, group, link, and compare, enabling the researcher to perform such human actions as infer, make connections, identify patterns and relationships, interpret, and build theory with the data (Lewins & Silver, 2007, p. 13). Figure 1.4 illustrates a sample MAXQDA Code Relations Browser window, which enables you to determine *possible* interrelationships among coded data (Kuckartz, 2007). The varying sizes of the squares within the matrix indicate the relative frequency of such matches.

CAQDAS also permits the researcher to shift quickly back and forth between multiple analytic tasks, such as coding, analytic memo writing, and exploring patterns in progress. Add to this the software's ability to recode, uncode, rename, delete, move, merge, group, and assign different codes to shorter and longer passages of text with a few mouse clicks and keystrokes during Second Cycle Coding, and the advantages of CAQDAS over paper and pencil soon become apparent.

Rather than presenting in this section an extended discussion of CAQDAS's specific applications with coding and data analysis, additional references will be made on an "as relevant" basis throughout the rest of this manual. Since most readers of this book are more than likely newcomers to qualitative data analysis, I assume that manual coding will be the first method you employ. Thus, I present the coding profiles with that assumption in mind. Those with experience or expertise in CAQDAS programs can adapt the coding principles described in this manual into their particular software package's active files and documents.

### **Solo and Team Coding**

Coding in most qualitative studies is a solitary act – the "lone ethnographer" intimately at work with her data (Galman, 2007) – but larger fieldwork projects may involve a team.

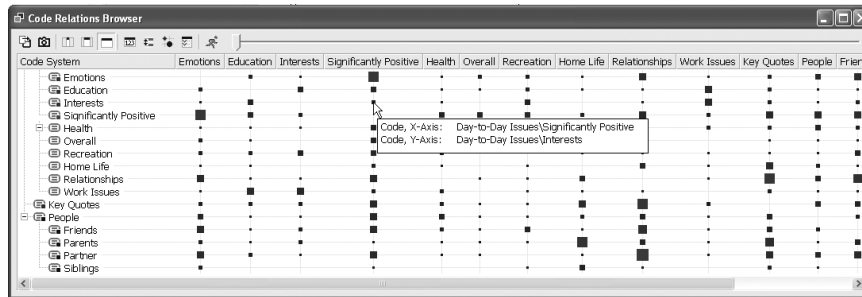


FIGURE 1.4 A Code Relations Browser screenshot from MAXQDA, v. 2007 (courtesy of MAXQDA/VERBI Software)

### **Coding collaboratively**

Writers of joint research projects advocate that coding in these cases can and should be a collaborative effort (Erickson & Stull, 1998; Guest & MacQueen, 2008). Multiple minds bring multiple ways of analyzing and interpreting the data: “a research team builds codes and coding builds a team through the creation of shared interpretation and understanding of the phenomenon being studied” (Weston et al., 2001, p. 382). Provocative questions are posed for consideration that could possibly generate new and richer codes (Olesen et al., 1994). Ultimately, team members must coordinate and insure that their sometimes individual coding efforts harmonize, particularly if a central data base and CAQDAS system are employed. MacQueen et al. (2008, p. 132) strongly advise that one member of the team be assigned primary responsibility as “codebook editor” – the one who creates, updates, revises, and maintains the master list for the group.

Those conducting action or community-based research can invite the study’s participants/stakeholders themselves into the analytic process as a collaborative venture to provide a sense of ownership and investment in data analysis and its consequent recommendations for social change (Stringer, 1999). Northcutt & McCoy (2004) label focus group development of their own categories of interest “affinities.” Children and adolescents, too, can be taught to investigate and analyze issues that relate to their social worlds (Alderson, 2000; Heiligman, 1998; Warren, 2000).

Team members can both code their own and others’ data gathered in the field to cast a wider analytic net and provide a “reality check” for each other. For these types of collaborative ventures, *intercoder agreement* or *interpretive convergence* – the percentage at which different coders agree and remain consistent with their assignment of particular codes to particular data – is an important part of the process (see Bernard, 2006, pp. 512–15; Boyatzis, 1998, pp. 144–59; Hruschka et al., 2004; and Miles & Huberman, 1994, p. 64 for simple formulas). There is no standard or base percentage of agreement among qualitative

researchers, but the 85%–90% range seems a minimal benchmark to those most concerned with an evidentiary statistic. Other research teams may wish to dispense with such quantitative measures all together and rely on intensive group discussion and simple group “consensus” as an agreement goal (Harry, Sturges, & Klingner, 2005, p. 6).

### **Coding solo**

If you're working as a lone ethnographer, shop talk with a colleague or mentor about your coding and analysis as you progress through them. Both solo and team coders can even consult the participants themselves during analysis (a process sometimes called “member checking”) as a way of validating the findings thus far. Even if you and other members of a research seminar are each working on different projects, sharing coded field note excerpts and discussing your “dilemmas” about coding and analysis generate peer support and may even help you and others find better connections between categories in progress (Burant et al., 2007; Strauss, 1987). Discussion provides not only an opportunity to articulate your internal thinking processes, but also presents windows of opportunity for clarifying your emergent ideas and possibly making new insights about the data.

Ezzy (2002, pp. 67–74) recommends several strategies for checking the progress of your analysis while still in the field. Though applicable for team researchers as well, the lone researcher can benefit most from these recommendations to assess the trustworthiness of her account: (1) check your interpretations developed thus far with the participants themselves; (2) initially code *as* you transcribe interview data; and (3) maintain a reflective journal on the research project with copious analytic memos.

### **Necessary Personal Attributes for Coding**

Aside from such cognitive skills as induction, deduction, abduction, synthesis, evaluation, and logical and critical thinking, there are seven personal attributes all qualitative researchers should possess, particularly for coding processes.

First, you need to be *organized*. This is not a gift that some people have and others don't. Organization is a set of disciplined skills that can be learned and cultivated as habits. A small-scale qualitative study's word count of data will range in the tens- and sometimes hundreds-of-thousands of words. The multiple codes you generate will need an organized framework for qualitative analysis. And despite the electronic filing systems of hard drives and CAQDAS, you will still encounter and manipulate many pages of paper in qualitative work. Date and label all incoming data and keep multiple digital and hard copies as backup.

Second, you need to exercise *perseverance*. Virtually every writer of qualitative research methods literature remarks that coding data is challenging and time-consuming. Some writers also declare how tedious and frustrating it can be. Take breaks from your work when you need to, of course – this will keep you refreshed and alert. But cultivate a personal work ethic and create an environment and schedule that enable you to sustain extended periods of time with analytic tasks requiring your full concentration.

Third, you need to be able to *deal with ambiguity*. The acts of coding and codifying are not precise sciences with specific algorithms or procedures to follow. Yes, occasionally answers may suddenly and serendipitously crystallize out of nowhere. But at other times, a piece of the analytic puzzle may be missing for days or weeks or even months. Rich ideas need time to formulate, so have trust and faith in yourself that these may emerge in due time. But remember that you can accelerate the process through analytic memo writing.

Fourth, you will need to exercise *flexibility*. Coding is a cyclical process that requires you to recode not just once but twice (and sometimes even more). Virtually no one gets it right the first time. If you notice that your initial methods choices may not be working for you or not delivering emergent answers you need, be flexible with your approach and try a modified or different method all together. Virtually all researcher-developed coding schemes are never fixed from the beginning – they evolve as analysis progresses.

Fifth, you need to be *creative*. There's a lot of art to social science. Noted ethnographer Michael H. Agar (1996) asserts that the early stages of analysis depend on "a little bit of data and a lot of right brain" (p. 46). We generally advocate that qualitative researchers remain close to and deeply rooted in their data, but every code and category you construct or select are choices from a wide range of possible options. Creativity also means the ability to think visually, to think in metaphors, and to think of as many ways possible to approach a problem. Creativity is essential for your data collection, data analysis, and even for your final written report.

Sixth, you need to be *rigorously ethical*. Honesty is perhaps another way to describe this, but I deliberately choose the phrase because it implies that you will always be: rigorously ethical with your participants and treat them with respect; rigorously ethical with your data and not ignore or delete those seemingly problematic passages of text; and rigorously ethical with your analysis by maintaining a sense of scholarly integrity and working hard toward the final outcomes.

The seventh and arguably most important skill you need for coding is an *extensive vocabulary*. Quantitative research's precision rests with numeric accuracy. In qualitative research, our precision rests with our word choices. An unabridged dictionary and thesaurus become vital reference tools to find just

the right words for your codes, categories, themes, concepts, and theories. Explore the origins of key words in an unabridged dictionary to find surprising new meanings (for example, did you know that the root word of *hypocrite* is “actor”?). A thesaurus review of a key word chosen as a code or category may introduce you to an even better – and more precise – word for your analysis.

For an applied introduction to the cognitive skills and personal attributes necessary for coding and qualitative data analysis, see Appendix A’s exercises and simulations.

### On Method

Thorough – even cursory – descriptions about the researcher’s code development and coding processes rarely make it into the methods section of a final report (but a dissertation writer should consider including his or her codebook as an appendix to the study). The majority of readers would most likely find the discussion tedious or irrelevant compared to the more important features, such as the major categories and findings. Plus, scholarly journals place length restrictions on article manuscripts, so some components of the research story must be left out and, more often than not, codes and coding fall by the wayside. But in all honesty, I don’t think most of the academic community minds (cf. Stewart, 1998). I’m not advocating that published research should include what most feel is a behind-the-scenes matter. Just acknowledge that the long time and rigorous effort you put into, and joyous personal analytic growth you experience through, coding and analytic memo writing are private affairs between you and your data (cf. Conostas, 1992). When you invite important guests to your home for dinner, you don’t ask them to appear two or three hours before the scheduled serving time to watch you cook in the kitchen. They arrive just before the meal to feast on and enjoy what you’ve worked so hard to prepare.

Yet, analogy aside, please don’t refer to or consider this manual a “cookbook” for your raw data. That suggests that the methods profiled here are like tested recipes guaranteed to produce successful dishes every time. Most methodologists concur that coding is “an idiosyncratic enterprise” (Glesne, 2006, p. 153) and the “search for one perfect method of data analysis is fruitless” (Coffey & Atkinson, 1996, p. 2). Each qualitative study is context-specific and your data are unique, as are you and your creative abilities to code them. I don’t have the answers to your questions, but you and your data do. In good faith, I guarantee you some guidance and, if we’re both lucky, perhaps some insight.

(I jokingly mused to myself whether this manual might be disparagingly tagged by some as “the *Cliff’s Notes* of qualitative data analysis” or *Coding for Dummies*. Either way, as a pragmatist I’ll take that as a *compliment* about the work.)

***Coding as craft***

I am very well aware of the interpretivist turn in qualitative inquiry and the movements toward narrative presentation and emancipatory social action through ethnographic fieldwork (Denzin & Lincoln, 2005). My own qualitative research projects, in fact, have ranged from the realist to the literary and from the confessional to the critical (van Maanen, 1988). But as a theatre practitioner, my discipline acknowledges that we must attend to both the art *and craft* of what we do to make our stage production work successful. And as a teacher educator, it's my job to teach how to teach. Hence, I must have an attunement to various methods of classroom practice because my professional responsibilities require that I do. Some methods are organizational, managerial, time-efficient, and related to carefully planned curriculum design. Yet I emphasize to my students that such processes as the creative impulse, trusting your instincts, taking a risk, and just being empathetically human in the classroom are also legitimate methods of teaching practice. Education is complex; so is social life in general and so is qualitative inquiry in particular.

This heightened, ever-present awareness of craft, of "how to," transfers into my research work ethic. I have become both humbly and keenly aware not only of what I'm doing but why I'm doing it. A metacognition of method, even in an emergent, inductive-oriented, and socially conscious enterprise such as qualitative inquiry, is vitally important. This awareness comes with time and experience (and trial and error), but development can be accelerated if you have some preparatory knowledge of "how to." I hope this manual smoothes your learning curve a bit and assists with your professional and personal growth as a researcher.

This introduction focused on codes and coding. There is an accompanying heuristic with this process – writing analytic memos, the subject of the next chapter.