

1 An Introduction to Codes and Coding

Chapter Summary

This chapter first presents the purposes and goals of *The Coding Manual for Qualitative Researchers*. It then provides definitions and examples of codes and categories and their roles in qualitative data analysis. The procedures and mechanics of coding follow, along with discussions of analytic software and team collaboration. The chapter concludes with reflections on necessary researcher attributes and the role of method in coding.

Purposes of the Manual

From Johnny Saldana (2016). *The Coding Manual for Qualitative Researchers* (3rd ed.) London, UK: Sage.

The three primary purposes of the manual are:

- to 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 with sources, descriptions, recommended applications, examples, and exercises for coding and further analyzing qualitative data.

This manual serves as a reference to supplement existing works in qualitative research design and fieldwork. It 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 2](#)), for example, is clearly profiled, streamlined, and re-envisioned in Kathy Charmaz's (2014) *Constructing Grounded Theory*. Graham R. Gibbs's (2007) *Analyzing Qualitative Data* provides an elegant survey of basic analytic processes, while Miles, Huberman, and Saldaña's (2014) *Qualitative Data Analysis: A Methods Sourcebook* offers a more detailed compendium.

The manual does not subscribe to any one specific research genre or methodology. Throughout this book you will read a breadth of perspectives on codes and coding, sometimes purposely juxtaposed to illustrate and highlight diverse opinions among scholars in the field. The following demonstrates just two examples of such professional divergence:

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. (Strauss, 1987, p. 27)

But the strongest objection to coding as a way to analyze qualitative research interviews is not philosophical but the fact that it does not and cannot work. It is impossible in practice. (Packer, 2011, p. 80)

No one, including myself, can claim final authority on the utility of coding or the “best” way to analyze qualitative data. In fact, I take moderate liberty in adapting and even renaming selected prescribed coding methods for clarity or flexibility’s sake. I do this not to standardize terminology within the field, but simply to employ consistency throughout this particular manual.

I must also emphasize at the very beginning that *there are times when coding the data is absolutely necessary, and times when it is most inappropriate for the study at hand*. All research questions, methodologies, conceptual frameworks, and fieldwork parameters are context-specific. Also, whether you choose to code or not depends on your individual value, attitude, and belief systems about qualitative inquiry. For the record, here are mine, from *Fundamentals of Qualitative Research*:

Qualitative research has evolved into a multidisciplinary enterprise, ranging from social science to art form. Yet many instructors of research methods vary in their allegiances, preferences, and prescriptions for how to conduct fieldwork and how to write about it. I myself take a pragmatic stance toward human inquiry and leave myself open to choosing the right tool for the right job. Sometimes a poem says it best; sometimes a data matrix does. Sometimes words say it best; sometimes numbers do. The more well versed you are in the field’s eclectic methods of investigation, the better your ability to understand the diverse patterns and complex meanings of social life. (Saldaña, 2011b, pp. 177–8)

Coding is just *one* way of analyzing qualitative data, not *the* way. Be cautious of those who demonize the method outright. And be equally cautious of those who swear unyielding affinity to codes or what has been colloquially labeled “coding fetishism.” I prefer that you yourself, rather than some presumptive theorist or hardcore methodologist, determine whether coding is appropriate for your particular research project.

General introductory texts in qualitative inquiry are so numerous and well written that it becomes difficult not just to find the best one to use, but which one of such quality works to select as a primary textbook for qualitative research courses. 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 with a useful 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 this manual an exhaustive resource. I deliberately exclude such discipline-specific methods as psychotherapy’s Narrative Processes Coding System (Angus, Levitt, & Hardtke, 1999), 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.

This manual serves primarily as a reference work. It is not necessarily meant to be read from cover to cover, but it certainly can be if you wish to acquaint yourself with all 33 coding methods’ profiles and their analytic possibilities. Several principles related to coding matters not discussed in the first two chapters 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 scan the manual to explore which coding method(s) might be appropriate for your particular study, read the profiles’ Description and Applications sections to determine whether further reading of the profile is merited, or check the glossary in [Appendix A](#). I doubt you will 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 do not 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, open-ended survey responses, drawings, artifacts, photographs, video, Internet sites, e-mail correspondence, academic and fictional literature, and so on. The portion of data coded during first cycle coding processes can range in magnitude from a single word to a full paragraph, an entire page of text or a stream of moving images. In second cycle coding processes, the portions coded can be the exact same units, longer passages of text, analytic memos about the data, and even a reconfiguration of the codes themselves developed thus far. Charmaz (2001) describes coding as the “critical link” between data collection and their explanation of meaning.

Do not confuse the use of *code* in qualitative data analysis with the use of *code* in the field of semiotics, even though slight parallels exist between the two applications. In semiotics, a code relates to the interpretation of symbols in their specific social and cultural contexts. And while some code choices by the analyst may appear metaphoric, most codes are not metaphors (according to the principles established by Lakoff & Johnson, 2003).

In qualitative data analysis, a code is a researcher-generated construct that symbolizes or “translates” data (Vogt, Vogt, Gardner, & Haeffele, 2014, p. 13) and thus attributes interpreted meaning to each individual datum for later purposes of pattern detection, categorization, assertion or proposition development, theory building, and other analytic processes. Just as a title represents and captures a book, 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 a Descriptive Code, which summarizes the primary topic of the excerpt that follows the same superscript number:

¹ 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.”

¹ 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 an In Vivo Code:

¹ He cares about me. He has never told me but he does. ² 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. ³ I really feel comfortable around him.

¹ sense of self-worth

² stability

³ "comfortable"

Did you agree with the codes? Did other words or phrases run through your mind as you read the data? It is all right if your choices differed from mine. Coding is not a precise science; it is primarily an interpretive act. Also be aware that a code can sometimes *summarize*, *distill*, or *condense* data, not simply *reduce* them. Madden (2010) notes that such analytic work does not diminish but "value adds" to the research story (p. 10).

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 Eclectic Coding:

¹ My son, Barry, went through a really tough time about, probably started the end of fifth grade and went into sixth grade. ² When he was growing up young in school he was a people-pleaser and his teachers loved him to death. ³ Two boys in particular that he chose to try to emulate, wouldn't, were not very good for him. ⁴ 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. ⁵ 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.

¹ middle-school hell

² teacher's pet

³ bad influences

⁴ tween angst

⁵ 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

A pattern is repetitive, regular, or consistent occurrences of action/data that appear more than twice. "At a basic level, pattern concerns the relation between unity and multiplicity. A pattern suggests a multiplicity of elements gathered into the unity of a particular arrangement" (Stenner, 2014, p. 136).

As qualitative researchers, we seek patterns as somewhat stable indicators of humans' ways of living and working to render the world "more comprehensible, predictable and tractable" (p. 143). They become more trustworthy evidence for our findings since patterns demonstrate habits, salience, and importance in people's daily lives. They help confirm our descriptions of people's "five Rs": routines, rituals, rules, roles, and relationships. Discerning these trends is a way to solidify our observations into concrete instances of meaning.

In the examples presented thus far, each unit of data was assigned its own unique code, 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:

¹ 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. ² Mrs. Jackson looks at them and says, "No talking, save it for the cafeteria. ³ Row Two." Five children seated in the second row of desks rise and walk to the children already standing in line.

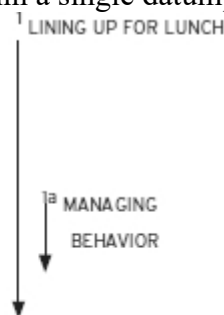
¹ lining up for lunch

² managing behavior

³ lining up for lunch

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):

¹ 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. ^{1a} Mrs. Jackson looks at them and says, "No talking, save it for the cafeteria. ¹ Row Two." Five children seated in the second row of desks rise and walk to the children already standing in line.



Take note of some important caveats when it comes to understanding patterns and regularity: idiosyncrasy *is* a pattern (Saldaña, 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 experiences and 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 share something in common – even if, paradoxically, that commonality consists of differences.

For example, each one of us may hold a strong opinion about who should lead our country. The fact that we each have an individual opinion about that issue is what we have in common. As for *who* we each believe should lead the country, that is where the differences and variations occur. Acknowledge that a confounding property of category construction in qualitative inquiry is that data cannot always be precisely and discretely bounded; they are within “fuzzy” boundaries at best (Tesch, 1990, pp. 135–8). That is why Simultaneous Coding is an option, when needed. 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)

Alvesson and Kärreman (2011), however, caution that a narrow focus on codification for pattern making with qualitative data can oversimplify the analytic process and hamper rich theory development: “Incoherencies, paradoxes, ambiguities, processes, and the like are certainly key aspects of social reality and worth exploring – both as topics in their own right and as a way of getting beyond premature pattern-fixing and the reproduction of taken-for-granted assumptions about specific patterns” (p. 42). Their advice is well taken, for it is not always the regularities of life but its anomalies and deviations that intrigue us, that stimulate us to question and to investigate why they exist concurrently with the mundane and normative – a process called “abductive analysis” (Tavory & Timmermans, 2014). As you code, construct patterns, certainly – but do not let those one or two codes that do not quite seem to fit anywhere frustrate you or stall your analytic work. Use these fragments as stimuli for deep reflection on the reason for their existence – if not their purpose – in the larger social scheme of things.

Coding lenses, filters, and angles

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 and from which angle you view the phenomenon. 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:

↓ There’s just no place in this country for illegal immigrants. Round them up and send those criminals back to where they came from.

↓ “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:

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

↑ 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:

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

↑ xenophobia

The collection of coding methods in this manual offers a repertoire of possible lenses, filters, and angles 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 – positions or angles how you perceive, document, and thus code your data (Adler & Adler, 1987). Filters influence the types of questions you ask and the types of responses you receive during interviews (Brinkmann & Kvale, 2015) and the detail and structuring of your field notes (Emerson, Fretz, & Shaw, 2011). Lenses refer to the gender, social class, and race/ethnicity of your participants – and yourself (Behar & Gordon, 1995; Saldaña, 2015; Stanfield & Dennis, 1993), and whether you collect data from adults or children (Greene & Hogan, 2005; Tisdall, Davis, & Gallagher, 2009; Zwiers & Morrissette, 1999).

Merriam (1998) states that “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 or genre of qualitative inquiry (e.g., case study, ethnography, phenomenology) and ontological, epistemological, and methodological issues that influence and affect your coding decisions (Creswell, 2013; 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). Like the characters in director Akira Kurosawa’s classic film, *Rashōmon*, multiple realities exist because we each perceive and interpret social life from different points of view.

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. 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 or algorithms to follow. Codes are significant phrases that “make meaning ... they are something that happens that make something [else] happen” (Fuller & Goriunova, 2014, p. 168) – they initiate a rigorous and evocative analysis and interpretation for a report. Plus, 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, 2013, p. 154).

Coding is a cyclical act. Rarely is the first cycle of coding data perfectly attempted. The second cycle (and possibly the third and fourth, etc.) 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). Locke, Feldman, and Golden-Biddle (in press) conceptualize the coding process as a “live” rather than inert action. Coding “is organic in which coding, codes and data shape each other; they are interdependent and inseparable” (p. 6). Once a code is applied to a datum during first cycle analysis, it is not a fixed representation but a dynamic and malleable process “through which to consider and interact with further observations and ideas” (p. 6). Indeed, heuristic fluidity is necessary to prioritize insightful qualitative analytic discovery over mere mechanistic validation.

Dey (1999) critically posits that “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 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 (i.e., a pattern), actively facilitate the development of categories and thus analysis of their connections. Ultimately, I like one of Charmaz’s (2014) 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. 113).

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 you apply and reapply codes to qualitative data, you are codifying – a process that permits data to be divided, grouped, reorganized and linked in order to consolidate meaning and develop explanation (Grbich, 2013). Bernard (2011) 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. 338). Coding 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 of Pattern Coding and Focused Coding in [Chapter 5](#)). 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

Synthesis combines different things in order to form a new whole, and it is the primary heuristic for qualitative data analysis – specifically, the transition from coding to categorizing (and from categorizing to other analytic syntheses). A quantitative parallel is determining the mean or average of a set of numbers. You take, say, 10 different test scores varying in range from a perfect score of 100 to the lowest achieved score of 62. Add each score (totaling 872), divide by the number of scores (10), and the mean is calculated (87.2). You have synthesized 10 different test scores into one new whole or symbol of meaning. But does qualitative data analysis have a heuristic equivalent? No and yes.

How do you “average” 10 different but somewhat comparable codes to arrive at a category? There is no qualitative algorithm or formula that adds up the words and calculates their mean. But there are

methods for synthesizing the collective, not to arrive at a reduced answer but to move toward *consolidated meaning*. That meaning may take the symbolic form of a category, theme, concept, or assertion, or set in motion a new line of investigation, interpretive thought, or the crystallization of a new theory. I blithely offer: “Quantitative analysis calculates the mean. Qualitative analysis calculates meaning.”

For example, in Harry, Sturges, and 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” (Silver & Lewins, 2014) – 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, Eastman’s (2012) ethnographic study, “Rebel Manhood: The Hegemonic Masculinity of the Southern Rock Music Revival,” employed grounded theory’s Initial, Focused, and Axial Coding to develop categories of “identity work strategies [southern U.S.] rebel men use to compensate for their lack of the economic resources and authority higher class men use to signify their hegemonic manhood” (p. 195). One major conceptual category was **Rebel Manhood as Protest Masculinity**, with its three subcategories *Protesting Education and Rejecting Cultural Capital*, *Protesting Work and Career*, and *Protesting Economic Authority*. Another conceptual category was **Compensatory Rebel Manhood Acts**, with its three subcategories *Drinking Alcohol and Violence*, *Drug Use*, and *Protesting Authority and Risk Taking*.

Maykut and 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’ve lost ten pounds on this new diet.”

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.

There are exceptions to every rule, however. Harding (2013) promotes that “codes can be placed in more than one category or subcategory” if you feel that the multiple classification is justified (p. 102). This tactic is incompatible with analytic methods such as Domain and Taxonomic Coding and analysis (see [Chapter 3](#)), but quite logical within the paradigm of “fuzzy sets,” which acknowledges that categories are not always discretely bounded but oftentimes overlap (Bazeley, 2013, p. 351). I prefer to keep my codes singular and clustered into their most appropriate categories for analysis. Yet it is good to know that, *if and when needed*, a code can get subsumed into more than one category. Too much of this, though, may suggest that the codes and/or the categories may not be as clearly defined as necessary, for there is a big difference between “fuzzy” category boundaries and “messy” ones.

Recoding and recategorizing

Rarely will anyone get coding right the first time. Qualitative inquiry demands meticulous attention to language and images, 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. Punch (2009), researching childhoods in Bolivia, describes how her codes, categories, and themes (as she defines them) developed and subdivided during her ethnographic fieldwork and concurrent data analysis:

[O]ne of my initial large codes was “home”. Everything relating to life at home was coded under this category and then subdivided into three themes: gender roles; child/adult work roles in the household; power and discipline. On reading through this latter category, I realized not only did it concern adult power over children, but also children’s strategies for counteracting adult power. After reorganizing these two sub-sections, I decided to split up the theme of children’s strategies into different types: avoidance strategies, coping strategies, and negotiation strategies. Finally, on browsing again through the sub-theme of negotiation strategies I found that I could further sub-divide it into child-parent negotiations and sibling negotiations. These data then formed the basis for structuring my findings on children’s lives at home. (pp. 94–5)

If you extract the coding scheme described in Punch’s narrative above, and transform it into an outline format or a hierarchical tree, it might appear thusly:

I. HOME

1. **Gender Roles**
2. **Child/Adult Work Roles in the Household**
- 3.

Power and Discipline

1. *Adult Power over Children*
- 2.

Children's Strategies for Counteracting Adult Power

1. Avoidance Strategies
2. Coping Strategies
- 3.

Negotiation Strategies

1. *Child/Parent Negotiations*
2. *Sibling Negotiations*

As you code and recode, expect – or rather, strive for – your codes and categories to become more refined and, depending on your methodological approach, more conceptual and abstract. Some of your first cycle codes may be later subsumed by other codes, relabeled, or dropped altogether. As you progress toward second cycle coding, you might rearrange and reclassify 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 (Saldaña, 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 by pushing them away, 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 young participants, oppression was not about the body and voice; it was about “force” and “feelings.” The three initial categories were eventually reduced to two during second cycle coding, 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

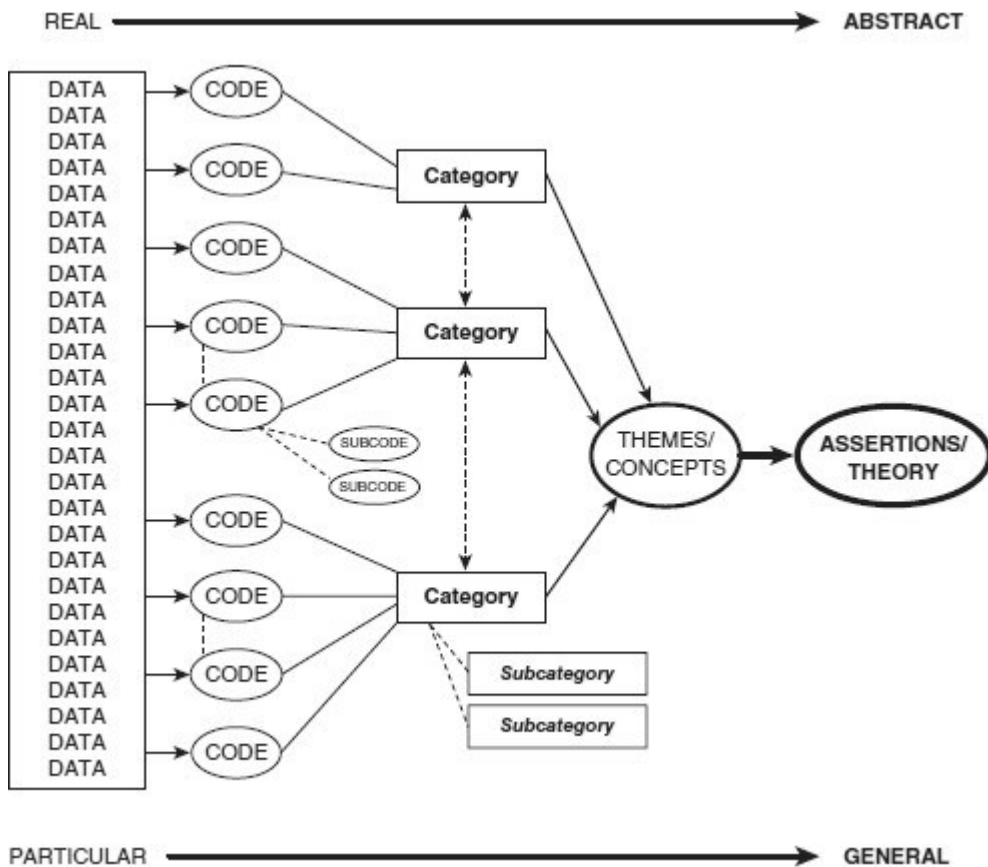
Also note how the subcodes themselves are specific, observable types of *realistic* actions related to the codes, while the two major categories labeled **Oppression** are more *conceptual* and *abstract* in nature.

See the Domain and Taxonomic Coding profile in [Chapter 3](#) for an extended discussion of this case, the Initial and Focused Coding examples in [Chapters 3](#) and [5](#) respectively, and the techniques of code mapping and code landscaping in [Chapter 4](#) to learn 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 you compare major categories with each other and consolidate them in various ways, you transcend the “particular reality” of your data and progress toward the thematic, conceptual, and theoretical. As a very basic process, codifying usually follows the ideal and streamlined scheme illustrated in [Figure 1.1](#).

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



Keep in mind that the actual act of reaching theory is much more complex than illustrated. Richards and Morse (2013) 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. 173). Our ability to show how these themes and concepts systematically interrelate leads toward the development of theory (Corbin & Strauss, 2015), 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 inquiry, but acknowledge that pre-existing 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, I constructed two major categories 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) – a statement that proposes a summative, interpretive observation of the local contexts of a study – that we developed and put forth 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 terminology waters. A theme can be an *outcome* of coding, categorization, or analytic reflection, but it is 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 and 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 denial means a false sense of security can be a theme.

Qualitative researchers are not algorithmic automatons. If we are carefully reading and reviewing the data before and as we formally code them, we cannot 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 2](#)) 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 of cycles there are other rich discoveries to be made with specific coding methods that explore such phenomena as participant processes, emotions, and values.

What Gets Coded?

Richards and Morse (2013) humorously advise for analytic work, “If it moves, code it” (p. 162). But what exactly *gets* coded in the data?

Units of social organization

Lofland, Snow, Anderson, and Lofland (2006) 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 will not 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)

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 3](#)) 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 Causation Coding, Pattern Coding, or grounded theory’s second cycle coding methods (see [Chapter 5](#); and Maxwell, 2004).

The coded researcher

In the coding examples profiled in [Chapters 3–5](#), you will notice that the interviewer’s questions, prompts, and comments are not coded. This is because the researcher’s utterances are more functional than substantive *in these particular cases* and do not merit a code. Also, I prioritize the participants’ data when analyzing interviews since I am studying their perceptions, not mine. My interpretations of their narratives via coding is my contribution to the meaning-making enterprise.

But if the exchanges between an interviewer and interviewee are more than just information gathering – if the interactions are significant, bidirectional dialogic exchanges of issues and jointly constructed meanings – then the researcher’s contributions could be appropriately coded alongside the participant’s. Certainly, the researcher’s participant observation field notes, authored from a first-person perspective, merit codes since they both document naturalistic action and include important interpretations of social life and potentially rich analytic insights.

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, mundane life that we might generate significant social insight. Others (e.g., Guest, MacQueen, & Namey, 2012; Morse, 2007; Seidman, 2013), if not most, feel that only the

most salient portions of the corpus related to the research questions merit examination, and that even up to one-half to two-thirds of the total record can be summarized or “deleted,” leaving the remainder for intensive data analysis. Paulus, Lester, and Dempster (2014) suggest that available digital tools for qualitative research may have made complete transcription of materials obsolete, with “gisted” or “essence” transcription as preferred approaches with programs such as Transana, which maintains the full audio/video data record for future researcher reference, as needed (pp. 94–100).

The potential hazard is that the portions deleted might contain the as yet unknown units of data that could pull everything together, or include the negative or discrepant case that motivates a rethinking of a code, category, theme, concept, assertion, or theory. Postmodern perspectives on ethnographic texts consider all documentation and reports partial and incomplete anyway, so the argument for maintaining and coding a full or condensed 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 does one feel more secure knowing and feeling what is important in the data record and what is not; I therefore code only what rises to the surface – “relevant text” as Auerbach and Silverstein (2003) label it. Sullivan (2012) identifies his significant passages of data from the corpus as “key moments,” and the reconstructed assembly of same-topic interview passages from different participants as “cherry-picked” dialogic “sound bites” for intensive thematic or discourse analysis. Everything else, like in a twentieth-century film editing studio, falls to the cutting room floor.

The beginning of my fieldwork career 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, but do not feel bound by that recommendation. You, too, will eventually discover from experience what matters and what does not in the data corpus. Code smart, not hard. (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 2](#)) 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 or cultural meanings: that is, by intentions, motives, beliefs, rules, discourses, and values” (Hammersley & Atkinson, 2007, 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, categories, analytic memos, and graphical summaries are “metadata activities” (MacQueen & Guest, 2008, p. 14).

Coding Techniques

Preparing data for coding gives you a bit more familiarity with the contents and initiates a few basic analytic processes. It is comparable to “warming up” before more detailed work begins.

Data layout

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 running data together as long unbroken passages, separate the text into short paragraph-length units with a line break 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, 2011, p. 192)). Gee, Michaels, and 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 computer-assisted qualitative data analysis software (CAQDAS) programs (discussed later).

Below is an excerpt from a word-processed interview transcript without any breaks in the text. The participant, a white male PhD student, reflects on the midpoint of his doctoral program of study:

PARTICIPANT: I’m 27 years old and I’ve got over \$50,000 in student loans that I have to pay off, and that scares the hell out of me. I’ve got to finish my dissertation next year because I can’t afford to keep going to school. I’ve got to get a job and start working.

INTERVIEWER: What kind of job do you hope to get?

PARTICIPANT: A teaching job at a university someplace.

INTERVIEWER: Any particular part of the country?

PARTICIPANT: I’d like to go back to the east coast, work at one of the major universities there. But I’m keeping myself open to wherever there’s a job. It’s hard listening to some of the others [in the current graduating class] like Jake and Brian interviewing for teaching jobs and being turned down. As a white male, that lessens my chances of getting hired.

INTERVIEWER: I think most employers really do look for the best person for the job, regardless of color.

PARTICIPANT: Maybe. If I can get some good recs [letters of recommendation], that should help. My grades have been real good and I’ve been getting my name out there at conferences.

INTERVIEWER: All of that’s important.

PARTICIPANT: The prospectus is the first step. Well, the IRB [Institutional Review Board approval] is the first step. I’m starting the lit review this summer, doing the interviews and participant observation in the fall, writing up as I go along, and being finished by spring.

INTERVIEWER: What if more time is needed for the dissertation?

PARTICIPANT: I’ve got to be finished by spring.

An unformatted excerpt such as the above could be entered into a CAQDAS program as is. But for manual coding, and even for some preliminary formatting for selected CAQDAS programs, the interview text can be divided into separate units or stanzas when a topic or subtopic shift occurs. Each

stanza, with a noticeable line break in between, could conceivably become a unit that will receive its own code. Other necessary formatting, such as truncating names or placing non-coded passages in brackets, can be taken care of at this layout stage of data preparation:

P: I'm 27 years old and I've got over \$50,000 in student loans that I have to pay off, and that scares the hell out of me. I've got to finish my dissertation next year because I can't afford to keep going to school. I've got to get a job and start working.

[I: What kind of job do you hope to get?]

P: A teaching job at a university someplace.

[I: Any particular part of the country?]

P: I'd like to go back to the east coast, work at one of the major universities there. But I'm keeping myself open to wherever there's a job.

It's hard listening to some of the others [in the current graduating class] like Jake and Brian interviewing for teaching jobs and being turned down. As a white male, that lessens my chances of getting hired.

[I: I think most employers really do look for the best person for the job, regardless of color.]

P: Maybe.

If I can get some good recs [letters of recommendation], that should help. My grades have been real good and I've been getting my name out there at conferences.

[I: All of that's important.]

P: The prospectus is the first step. Well, the IRB [Institutional Review Board approval] is the first step. I'm starting the lit review this summer, doing the interviews and participant observation in the fall, writing up as I go along, and being finished by spring.

[I: What if more time is needed for the dissertation?]

P: I've got to be finished by spring.

The interview excerpt above will be coded and analyzed in [Chapter 4](#)'s profile, Eclectic Coding.

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 (2013, p. 205) 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, and serve as illustrative examples throughout your report (Booth, Colomb, & Williams,

2008; Erickson, 1986; Lofland et al., 2006). 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.

Bernard and Ryan (2010) recommend that rich text features of word processing software can also enable initial coding and categorization *as* data are transcribed. In a health study related to participants talking about their experiences with the common cold, “*Signs and symptoms are tagged with italics*; treatments and behavioral modifications are tagged with underlining; and **diagnosis is tagged with bold type**” (p. 91, rich text features added). Field notes can also employ rich text features for “at a glance” separation before coding and analytic review:

- Descriptive, narrative passages of field notes are logged in regular font.
- **“Quotations, things spoken by participants, are logged in bold font.”**
- *Observer’s comments, such as the researcher’s subjective impressions or analytic jottings, are set in italics.*

Preliminary jottings

Start coding *as* you collect and format your data, not after all fieldwork has been completed. When you write up field notes, transcribe recorded interviews, or file 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 do not have to be accurate or final at this point, just ideas for analytic consideration while the study progresses. Do not rely 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 and 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	COLUMN 2	COLUMN 3
Raw Data	Preliminary Codes	Final Code
¹ The closer I get to retirement age, the faster I want it to happen. I’m	“retirement age”	¹ RETIREMENT ANXIETY

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.

financial obligations

*dreams of early
retirement*

Some of my students, during preliminary stages of analysis, devote the right-hand margin to tentative codes for specific data units, while the left-hand margin includes broader topics or interpretive jottings for later analytic memo writing (see [Chapter 2](#)).

Virtually all methodologists recommend initial and thorough readings of your data while writing analytic memos or jottings 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 and 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 et al. (2011) advise a general list of questions to consider when coding field notes, 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?
- How is what is going on here similar to, or different from, other incidents or events recorded elsewhere in the fieldnotes?
- What is the broader import or significance of this incident or event? What is it a case of? (p. 177)

I would add to this list the question I ask myself during all cycles of coding and data analysis: “What strikes you?” Sunstein and Chiseri-Strater (2012) expand on this by suggesting that fieldworkers, during all stages of a project, ask themselves:

- What surprised me? (to track your assumptions)
- What intrigued me? (to track your positionality)
- What disturbed me? (to track the tensions within your value, attitude, and belief systems) (p. 115).

Coding contrasting data

If you work 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 and Jackson (2013) recommend 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" (pp. 69–70). 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 keep you apprised 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 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 comes 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 shared 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 excerpt – a broad brush-stroke representation called Holistic Coding:

¹ 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.

¹ "a lot to learn"

The method above is called “lumper” coding. The opposite is someone who codes as a “splitter” – one who splits the data into smaller codable moments (Bernard, 2011, p. 379). 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 ¹ reality for me. ² I thought I was so ready for this population because I had taught other groups of kids. But this is such a ³ 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 ⁴ inner city schools. ... ⁵ I really had to learn about the kids. I had to learn about ⁶ 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 ⁷ a lot to learn that I had never even thought about.

¹ “reality”

² “i thought i was so ready”

³ “unique situation”

⁴ “inner city schools”

⁵ “i really had to learn”

⁶ “the culture”

⁷ “a lot to learn”

Now this 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 (see the discussion of Concept Coding in [Chapter 3](#)), while fine-grained splitting of data may overwhelm the analyst when it comes time to categorize the codes. Perspectives vary within the professional literature. Stern (2007) admits: “I never do a line-by-line [coding] analysis because there is so much filler to skip over. Rather, I do a search and seizure operation looking for cream [that rises to the top] in the data” (p. 118). But Charmaz (2008) advises that detailed line-by-line coding promotes a more trustworthy analysis that “reduces the likelihood of imputing your motives, fears, or unresolved personal issues to your respondents and to your collected data” (p. 94).

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 will discover which approach works best for you, your particular study, and your particular research goals.

The quantities of qualities

Harding (2013) openly acknowledges that his advice is subjective, yet he recommends that a code shared by approximately one-fourth of the study’s respondents merits consideration in the analysis and a possible contribution to the research findings. He also advises that roughly three-fourths of the total number of participants should share a similar code between them (related to an experience or opinion found in their data) for a “commonality” to be established, such as a category or theme. But my own experience has taught me that, *in some cases*, that unique instance of a code that appears just once and nowhere else in the data corpus, or a code that appears just two or three times across different cases or time periods, may hold important meaning for generating a significant insight in later analysis. Unfortunately, that same number of just one, two, or three instances of a code may also

suggest something unimportant, inconsequential, and unrelated to your research questions and purpose. The analyst must reconcile which one of these possibilities is at work.

Friese (2014) prescribes that qualitative research projects should never venture into the thousands for a final number of codes; between 50 and 300 different codes total are recommended (p. 92, 128). Lichtman (2013) projects that most qualitative studies in education will generate 80–100 different codes that will be organized into 15–20 categories and subcategories which eventually synthesize into five to seven major concepts (p. 248). Creswell (2013) begins his analyses with a shortlist 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 or six major themes (pp. 184–5). Other disciplines and varying approaches to qualitative inquiry may prescribe different sets of numbers as general guidelines for analysis, but MacQueen, McLellan, Kay, and Milstein (2009) observe that “For the most part, coders can only handle 30–40 codes at one time” for a study, especially if they use a system developed by someone else (p. 218).

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 or 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.

“Quantitizing” the qualitative

Some researchers may wish to transform their qualitative data and/or codes into quantitative representations for exploratory review or statistical analysis. Sandelowski, Voils, and Knafl (2009) posit that “the rhetorical appeal of numbers – their cultural association with scientific precision and rigor – has served to reinforce the necessity of converting qualitative into quantitative data” (p. 208). They propose that quantitizing, or the transformation of non-numeric data into a counted form of some kind, is “engineering data” to create different indices of meaning. Just as codes are symbolic summaries of larger excerpts of data, numbers are symbolic summaries of a measured outcome.

Quantitizing qualitative data is done for varying reasons, but several methodologists I have consulted advised me, “Ask yourself *why* you’re changing qualitative data into numbers in the first place.” If you are transforming words into numbers solely for what you believe may be more persuasive results and case-making, you may be doing it for the wrong reason. Instead, quantitizing may be better applied to content analytic studies, mixed methods studies, and field experiments that test for differences between treatment and second-site (i.e., control) groups, or differences between time periods with a single participating group. Mixed methods texts (Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 2010; Vogt et al., 2014) provide excellent discussions of rationales and procedures for data mixing and transformation. Here I address in detail only one – a purpose I label *paradigmatic corroboration*.

Assuming that quantitative and qualitative research, with their distinctive symbol systems of meaning, are two separate approaches to inquiry, it is possible to achieve comparable types of results when each approach examines the same local phenomenon or data set. As an example, I analyzed survey data that collected both quantitative ratings from respondents to close-ended prompts, plus written responses to related, open-ended, follow-up prompts (McCammon, Saldaña, Hines, & Omasta, 2012). A sample question from the e-mail survey administered to adults of varying ages is:

1. a. My participation in high school speech and/or theatre has affected the adult I am now
[choose one:]

4 – Strongly Agree

3 – Agree

2 – Disagree

1 – Strongly Disagree

1. b. In what ways do you think your participation in speech and/or theatre as a high school student has affected the adult you have become?

Paradigmatic corroboration occurs when the quantitative results of a data set do not simply harmonize or complement the qualitative analysis but corroborate it. In other words, the quantitative analytic results “jive” with or appear to correspond with the qualitative analytic outcomes. Part of the survey study analyzed differences between two or more configurations of respondents (e.g., between male and female respondents; between younger, mid-life, and older respondents) to observe whether quantitative analysis showed any statistically significant differences between groups, *in addition to* whether any noticeable qualitative differences in codes and categories were generated from the open-ended response data.

If there was no significant difference ($p < .05$) between male and female participants to the inferential statistical testing of their numeric ratings, then the codes and categories of the respective groups’ qualitative responses should also show no substantive differences between groups. But when there *was* a statistically significant difference between groups – for example, between respondents in their twenties and older respondents aged 50 plus – their respective qualitative codes and categories also revealed differences. For example, younger respondents offered more intrinsic benefits (friendships, empathy, artistic growth, etc.) of their high school theatre and speech participation; older adults referred to more extrinsic rewards and accomplishments (roles in plays, awards, career advancement). The numeric outcomes helped support the qualitative analytic results. Quantity corroborated quality, and vice versa.

Paradigmatic corroboration provides the analyst a “reality check” of his or her analytic work. It also provides two sets of lenses to examine the data for a multidimensional and more trustworthy account. Magnitude Coding could serve as one way of transforming or “quantitizing” qualitative data. Hypothesis Coding – and, with some adaptation, Evaluation Coding – is designed to test differences between two (or more) participant groups (see [Chapter 3](#) for these coding methods). Significant frequency or evaluative differences between a set of major codes from two groups can be assessed, for example, through the *t*-test (for larger samples) or the nonparametric Mann–Whitney *U*-test (for smaller samples or ordinal data). Most CAQDAS programs include statistical capabilities with your codes, revealing results such as frequencies, cluster analyses, correlations, and so forth. Software such as Dedoose compares statistical information sorted by participant demographics or other “descriptors.”

By no means should you infer that I advocate the transformation or quantitizing of qualitative data for all studies. It is an option available to you if and only if it will help meet your analytic goals and provide the best answers for your particular research questions. If you are in doubt, as some of my colleagues advised, ask yourself *why* you are changing qualitative data into numbers in the first place.

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, 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 work 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 (see Coding Collaboratively below). Bernard and Ryan (2010, p. 99) advise that, for some studies with a more compact number of codes, each item in the codebook can specify its:

- *short description* – the name of the code itself
- *detailed description* – a 1–3 sentence description of the coded datum’s qualities or properties
- *inclusion criteria* – conditions of the datum or phenomenon that merit the code
- *exclusion criteria* – exceptions or particular instances of the datum or phenomenon that do not merit the code
- *typical exemplars* – a few examples of data that best represent the code
- *atypical exemplars* – extreme or special examples of data that still represent the code
- *“close, but no”* – data examples that could mistakenly be assigned this particular code

Bazeley and Jackson (2013) advise that each major code be subjected to a “job description” which includes an explanation of its “purpose” and “performance” in the analytic scheme (p. 256).

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. Like other instructors, I 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 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 that “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). I would add to this the research goals of the

enterprise and the emergent satisfaction with the electronic coding system. Gallagher (2007) and her research team began a multi-site ethnography with CAQDAS, yet they soon learned that their software choice

was effective for data management, but inadequate for the nuanced and complex work of data analysis. [The software package] gave us style, but not substance; it sacrificed the attention to, and containment of, complexity we were after. ... In effect, we returned to a manual [coding] system that respected the sheer quantity and complexity of qualitative data and the surrounding contexts. (pp. 71, 73)

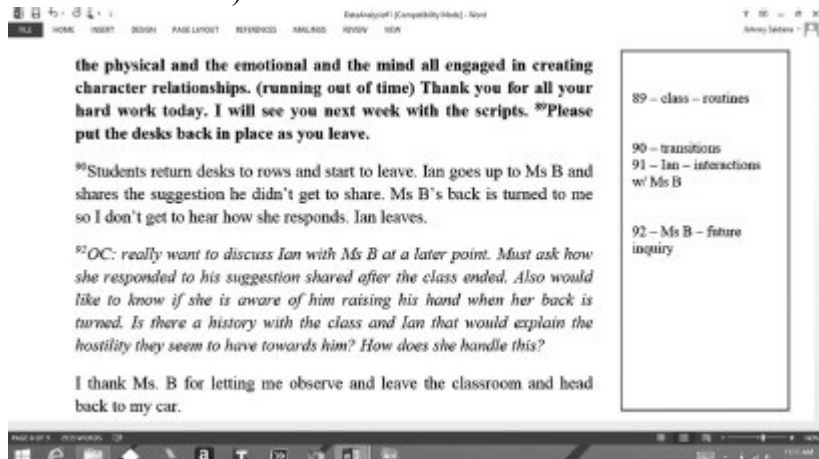
Coding collaboratively with hard-copy data is difficult enough for a research team. The task exponentially increases in complexity if CAQDAS files are shared and accessed at different times among individual team members.

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, researchers. 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. There is something about manipulating qualitative data on paper and writing codes in pencil that gives you more control over and ownership of the work. Perhaps this advice stems from my old-school ways of working that have become part of my “codus” operandi.

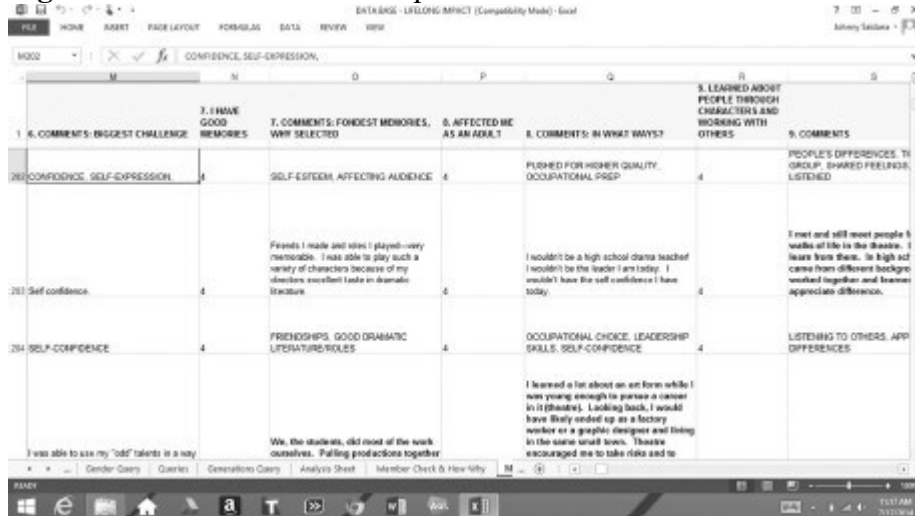
But for those with software literacy, a few of Microsoft Word’s basic functions can code directly onto data. Some will select a passage of text and insert a comment, which contains the code for the datum. Others might insert a vertical text box running along the right-hand margin and insert the codes aligned with the data (see [Figure 1.2](#)). Researchers with smaller data sets needing just three to ten major codes and/or categories total can assign a specific colored font to text passages that belong in the same category.

Figure 1.2 A Microsoft Word field notes document with codes in a right-margin text box (courtesy of Teresa Minarsich)



One of my mixed methods survey projects employed Microsoft Excel as a repository for the database because there were 234 surveys returned, and the software provided excellent organization with individual cells holding thousands of entries and their accompanying codes (see [Figure 1.3](#)). Each row represented an individual participant's survey data, while each column held the responses to a specific survey question. An additional row below each individual respondent contained the codes for his or her data. Excel also enabled me to calculate survey ratings into means and to conduct *t*-tests for subgroup comparisons. The software's CONCATENATE function merges qualitative data from cells you specify, making the extraction of codes into one single cell a speedy task if they have been properly formatted in advance.

Figure 1.3 A Microsoft Excel spreadsheet with mixed methods data and codes in its cells



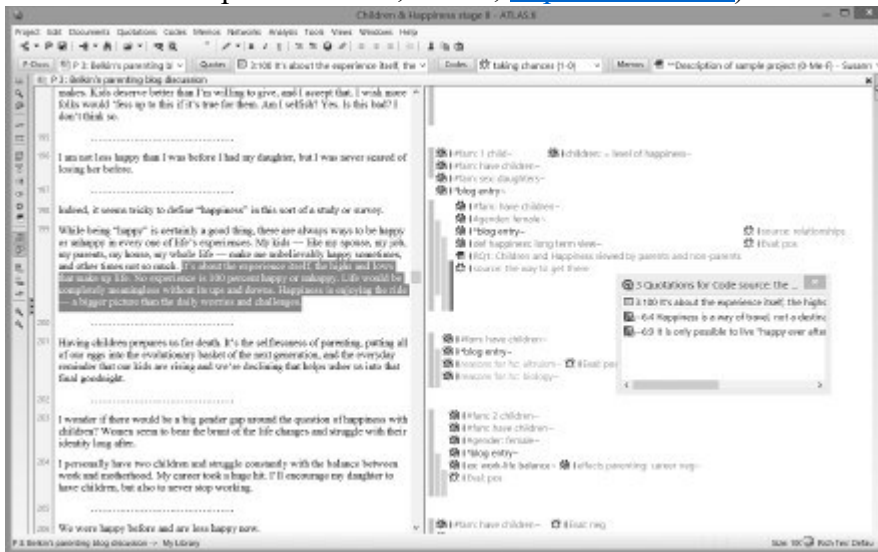
Nevertheless, there is something to be said for a large area of desk or table space with each code written on its own index card or “sticky note,” or multiple pages or strips of paper, spread out and arranged into appropriate clusters to see the smaller pieces of the larger puzzle – a literal, “old-school” 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, 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.

[Figure 1.4](#) illustrates a screen shot from the CAQDAS software program ATLAS.ti. Notice how the data are displayed in the left pane, with their corresponding codes and related notations in the right pane. Some specialty programs, like Transana, enable coding of digital audio and video documents stored in their files. Many of the programs enable both qualitative *and* quantitative and thus mixed methods analyses of the data. 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 text-edited work on a computer, back up your original files as a precautionary measure.

Figure 1.4 A screenshot from ATLAS.ti CAQDAS software (courtesy of ATLAS.ti Scientific Software Development GmbH, Berlin; <http://atlasti.com/>)



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

- AnSWR: <http://www.cdc.gov/hiv/library/software/answr/index.html>
- AQUAD: www.aquad.de/en
- ATLAS.ti: www.atlasti.com
- CAT (Coding Analysis Toolkit): cat.ucsur.pitt.edu/
- Dedoose: www.dedoose.com
- DiscoverText: www.discovertext.com
- HyperRESEARCH: www.researchware.com
- INTERACT: www.mangold-international.com
- MAXQDA: www.maxqda.com
- NVivo: www.qsrinternational.com
- QDA Miner: www.provalisresearch.com
- Qualrus: www.qualrus.com
- Quirkos: www.quirkos.com
- Transana (for audio and video data materials): www.transana.org
- V-Note (for audio and video data materials): www.v-note.org
- Weft QDA: www.pressure.to/qda/
- WordStat: www.provalisresearch.com

Selected CAQDAS programs come in both PC and Mac versions, and a few are available for Android. Programs such as AnSWR, AQUAD, and Weft QDA are available free of charge. Refer to Bazeley

and Jackson (2013), Edhlund (2011), Friese (2014), Silver and Lewins (2014), and Richards (2015) for accompanying literature on the major commercial programs. Also see Hahn (2008) and La Pelle (2004) for qualitative data analysis with basic text editing software and office suites; Brent and Slusarz (2003) and Meyer and Avery (2009) for advanced computational strategies with software; Paulus et al. (2014) for assorted digital tools available for qualitative data collection and analysis; Davidson and di Gregorio (2011) for Web 2.0 tools like DiscoverText and the Coding Analysis Toolkit; and Richards and Morse (2013) for what selected CAQDAS programs can and cannot do. Many CAQDAS programs are discussed and reviewed at an online forum for users: <http://caqdas.soc.surrey.ac.uk/>. And the International Institute for Qualitative Methodology presents a recorded webinar on selected CAQDAS programs at: <https://connect.srv.ualberta.ca/p6mk4z9ncsl/?launcher=false&fcsContent=true&pbMode=normal>. It is impractical to advise or prescribe which software program is “best” for particular qualitative studies and even for individual researchers. You are the best judge of your own software needs for your data, your available financial resources, and your personal preferences for user friendliness, so explore several of the programs available to you on your own at the web addresses provided above to make an informed decision. I have learned, however, that peer and instructor mentorship with a CAQDAS program is vital and more effective than just reading its software manual on your own. If you can enroll in workshops or classes in CAQDAS facilitated by master teachers, I highly recommend them. Alternatives consist of online video demonstrations, tutorials, and webinars offered by selected CAQDAS companies. Several of these have uploaded multiple short films about their products and features on YouTube (search for the clips by product name).

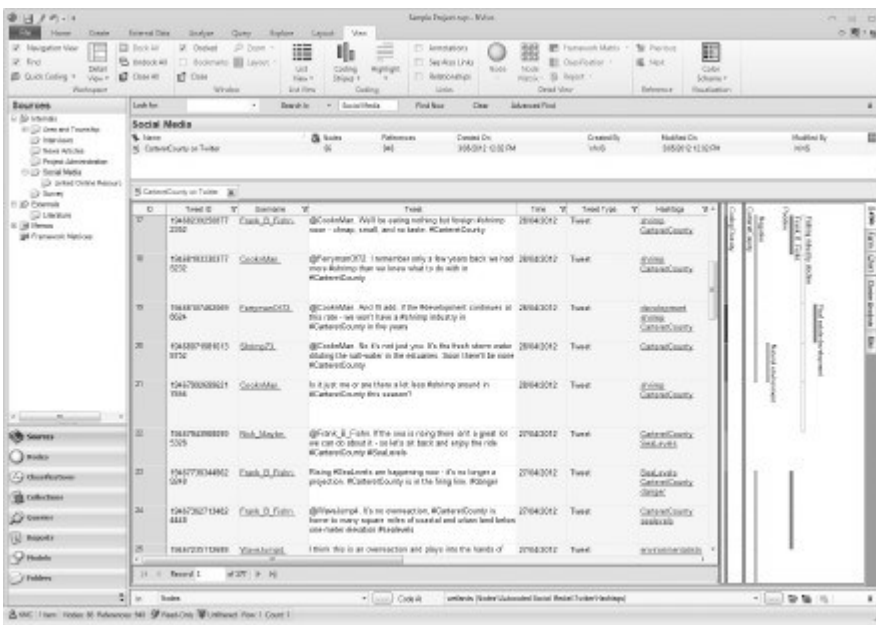
At the time of this writing, new technological tools exclusively designed or adaptable for qualitative data management and analysis seem to appear more and more frequently. Some programs, like ATLAS.ti, are accessible on an iPad; other programs, like NVivo, can import (or “capture”) and analyze social media data from platforms such as Twitter, Facebook, and YouTube. It becomes virtually impossible to keep up with all the electronic, software, and Internet resources available to researchers. My only recommendation is to gain as much general technological literacy as you can to make yourself aware of all your options, but to select your final tools wisely so that they help, rather than hinder, your analytic efforts. “To the software, a code is just an object that can be attached to various other objects and whose content can be searched and retrieved. Everything else is up to you” (Friese, 2014, p. 211).

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. Most commercial programs 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 (Silver & Lewins, 2014, p. 50).

One of the best features of some CAQDAS programs is their ability to display code labels themselves in various colors for “at a glance” reference and visual classification. [Figure 1.5](#) shows a sample screenshot from the most current version (10) of NVivo. Note how the Twitter data are accompanied with codes and “coding stripes,” which delineate which portion of data is assigned a particular code.

Figure 1.5 A screenshot from NVivo illustrating coding stripes (courtesy of NVivo / QSR International Pty Ltd.; www.qsrinternational.com)



Programs like Quirkos include a user-assigned color coding feature – a function that provides a unique color to each coding stripe and accompanying bin. Reviewing similarly color-coded data during second cycle coding makes it easier to refine first cycle codes and to create new or revised categories.

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 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 annotations, comments, or 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, clusters, and networks for “at a glance” user reference. [Figure 1.6](#) shows a screenshot from the most recent version of Quirkos, which displays both the data’s color-coded stripes in the right pane, with their corresponding color-coded bins in the left pane.

Figure 1.6 A screenshot from Quirkos illustrating coordinated coding stripes and bins (courtesy of Quirkos Software, www.quirkos.com)



Though I stated above that software does not code for you, there is a utilitarian function called “auto coding” available in most CAQDAS programs, which can alleviate some of the repetitiveness of manually coding similar passages of text, especially those gathered from surveys or structured interviews. Passages have to be formatted in prescribed ways and contain the same root 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 Silver and Lewins (2014) suggest that researchers should not feel obligated to use auto coding just because it is available.

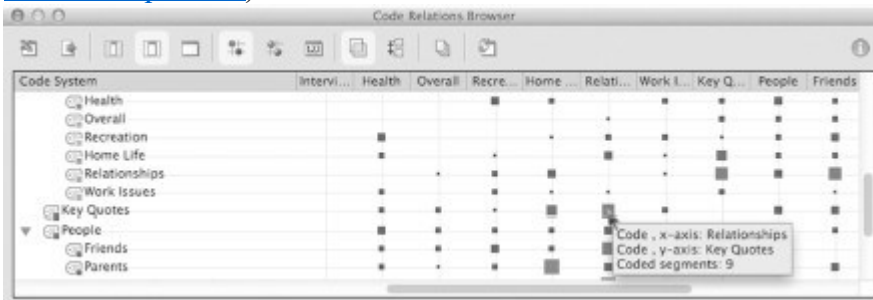
Searches and queries with CAQDAS

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 (Bazeley, 2013; Silver & Lewins, 2014). [Figure 1.7](#) illustrates Dedoose’s multiple and simultaneous representations of a data set quantitatively, qualitatively, and visually. [Figure 1.8](#) shows 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. Double-clicking one of the squares inside the Code Relations Browser brings up all the text segments with overlapping codes.

Figure 1.7 A screenshot from Dedoose illustrating its multiple, simultaneous representations of a data set (courtesy of SocioCultural Research Consultants, www.dedoose.com)



Figure 1.8 A Code Relations Browser screenshot from MAXQDA (courtesy of Verbi Software, www.maxqda.com)



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. And when the magnitude of a qualitative database on hard copy becomes overwhelming, the elegant data and coding displays of selected programs can provide the analyst with a sense of necessary order and organization, and enhance one’s cognitive grasp of the work in progress.

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, 2013) – but larger fieldwork projects may involve a team.

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; Schreier, 2012). 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 get posed for consideration that could possibly generate new and richer codes (Olesen, Drees, Hatton, Chico, & Schatzman, 1994). Ultimately, team members must coordinate and insure that their sometimes individual coding efforts harmonize, particularly if a central database and multi-user CAQDAS system are employed. MacQueen, McLellan-Lemal, Bartholow, and Milstein (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, 2014). Northcutt and 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, 2008; Heiligman, 1998; Warren, 2000). Haw and Hadfield (2011) and Heath, Hindmarsh, and Luff (2010) hold “data sessions” where informed colleagues and sometimes participants themselves are invited to preview and review video fragments from fieldwork to collaboratively interrogate and discuss relevant multiple dimensions of the research issues suggested. This dialogic exchange of ideas in a workshop and collegial atmosphere attunes the research team to new and varying perspectives before more intensive scrutiny and formal video analysis begin.

Team members can both code their own and others’ data gathered in the field to cast a wider analytic net and provide a “crowd-sourcing 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 (for formulas and discussions see Bernard, 2011, pp. 447–9; Boyatzis, 1998, pp. 144–59; DeCuir-Gunby, Marshall & McCulloch, 2011; Hruschka et al., 2004; and Krippendorff, 2009). There is no standard or base percentage of agreement among qualitative researchers, but the 80–90% range seems a minimal benchmark to those most concerned with an evidentiary statistic. Selected CAQDAS programs include such measures as the kappa coefficient, Pearson’s r , and other coding comparison queries as calculation functions for intercoder agreement.

Some methodologists question the utility and application of intercoder agreement for qualitative data analysis since the entire process is an interpretive enterprise. Thus, research teams may wish to dispense with such quantitative measures altogether and rely on intensive group discussion, “dialogical intersubjectivity,” coder adjudication, and simple group consensus as an agreement goal (Brinkmann & Kvale, 2015; Harry et al., 2005; Sandelowski & Barroso, 2007).

Coding by committee can range from a time-saving democratic effort, to a frustrating enterprise filled with road blocks, depending on the amount and complexity of data and – to be honest – the researcher personalities involved. Group dynamics suggest that a team meeting regularly to collectively code data should consist of no more than five people. More than five individuals makes problem-solving and decision-making exponentially more difficult. It may also be wise to develop strategies and

contingency plans ahead of time for what to do in case coding progress stalls or if professional disagreements occur and an executive decision needs to be made. I myself prefer to be the “lone wolf coder” when it comes to working with colleagues on a research project, but my team members are given copies of my coded data to review at all stages, and are encouraged to function as rigorous examiners and auditors of my analyses.

Coding solo

If you work 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 called “member checking”) as a way to validate the findings thus far. Even if you and other members of a research seminar each work on different projects, share coded field note excerpts and discuss your dilemmas about coding and analysis to generate peer support and to help you and others find better connections between categories in progress (Burant, Gray, Ndaw, McKinney-Keys, & Allen, 2007; Strauss, 1987). Discussion provides opportunities not only to articulate your internal thinking processes, but also to clarify your emergent ideas and possibly make 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 his or her account: (1) initially code *as* you transcribe interview data; (2) maintain a reflective journal on the research project with copious analytic memos; and (3) check your interpretations developed thus far with the participants themselves.

Necessary Personal Attributes for Coding

Aside from such cognitive skills as induction, deduction, abduction, retroduction, 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 do not. 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 a tightly organized framework for qualitative analysis; in fact, organization *is* 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 back-ups.

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*. 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 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 the emergent answers you need, be flexible with your approach and try a modified or different method altogether. Virtually all researcher-developed coding schemes are never fixed from the beginning – they evolve as analysis progresses.

Fifth, you need to be *creative*. There is a lot of art to social science. The 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 is a choice from a wide range of possible options. Creativity also means the ability to think visually, to think symbolically, to think in metaphors, and to think of as many ways as 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*. The precision of quantitative research rests with numeric accuracy. In qualitative research, our precision rests with our word choices. For example, there are subtle interpretive differences between something that “may,” “could,” “can,” “probably,” “possibly,” and “seemingly” happen; and a wide interpretive difference between something that happens “frequently,” “usually” and “often” (Hakel, 2009). Is a custard pie thrown in somebody’s face in a television situation comedy episode coded as juvenile violence or slapstick comedy? An unabridged dictionary and thesaurus become vital reference tools to find just the right words for your codes, categories, themes, concepts, assertions, and theories. Explore the origins of key words in an unabridged dictionary to find surprising new meanings (e.g., 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 Saldaña (2015) and the exercises and simulations in [Appendix D](#).

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. Also, 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 do not think most of the academic community minds (cf. Stewart, 1998). I am 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. Constas, 1992). When you invite important guests to your home for dinner, you do not 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 have worked so hard to prepare.

Yet, analogy aside, please do not refer to or consider this manual as a “cookbook” for your raw data. That suggests that the methods profiled here are like tested recipes guaranteed to produce successful dishes every time. Method “is just a way of ordering our capacity for insight – but does not produce it” (Ruthellen Josselson, in Wertz et al., 2011, p. 321). Most methodologists concur that coding schemes are customized to the specific contexts of a study; your data are unique, as are you and your creative abilities to code them. I do not have the answers to your questions, but you and your data do. In good faith, I guarantee you some guidance and, if we are both lucky, perhaps some insight.

Critiques against coding

There have been some legitimate critiques against coding, some of them philosophical and some of them methodological. Yet when I hear these criticisms I am inclined to think that my colleagues’ reservations originate from what used to be earlier, positivist approaches to coding – mechanical and technical paradigms that did indeed make the enterprise sheer drudgery and the outcomes often little more than topic-driven lists. Below are some of the most frequent criticisms I have heard against coding and my responses to those perceptions.

Coding is reductionist. Coding is what you perceive it to be. If you see it as reductionist, then that is what it will be for you. But recall that my definition of coding approaches the analytic act as one that assigns rich symbolic meanings through essence-capturing and/or evocative attributes to data. The 33 coding profiles in this book present an array of methods. And by design or necessity, a few are indeed meant to assist with nothing more complicated than descriptive, topical indexing, and even fewer are formulaic and prescriptive because that is how their developers intended them. But most of these methods generate discovery of the participant’s voice, processes, emotions, motivations, values, attitudes, beliefs, judgments, conflicts, microcultures, identities, life course patterns, etc. These are not “reductionist” outcomes but multidimensional facets about the people we study.

Coding tries to be objective. Somewhat and no. This could become an extended discussion about the ontological, epistemological, and methodological assumptions of inquiry, but let me bypass those in favor of a quick response. Intercoder agreement in team coding does indeed seem as if “objectivity” is the driving analytic force due to the need for two or more researchers to independently corroborate on the meaning of each datum. But in reality, the process is not so much objectivity as it is simply achieving similar results between two or more people.

For the individual researcher, assigning symbolic meanings (i.e., codes) to data is an act of personal signature. And since we each most likely perceive the social world differently, we will therefore experience it differently, interpret it differently, document it differently, code it differently, analyze it differently, and write about it differently. Objectivity has always been an ideal yet contrived and virtually impossible goal to achieve in quantitative research. So why should qualitative inquiry carry their baggage? We do not claim to be objective because the notion is a false god.

Coding is mechanistic, instrumentalist, and distances you from your data. If you are doing your job right as a qualitative researcher, nothing could be further from the truth. Coding *well* requires that you reflect deeply on the meanings of each and every datum. Coding *well* requires that you read, reread and reread yet again as you code, recode, and recode yet again. Coding *well* leads to total immersion in your data corpus with the outcome exponential and intimate familiarity with its details, subtleties, and nuances. When you can quote verbatim by memory what a participant said from your data corpus and remember its accompanying code, I do not understand how that action has “distanced” you from your work.

Coding is nothing more than counting. In traditional content analysis studies, counting the number of times a particular set of codes occurs is indeed an important measure to assess the frequency of items or phenomena. But one of the caveats I propose later in this manual is that frequency of occurrence is not necessarily an indicator of significance. The analytic approaches for most of these coding methods do not ask you to count; they ask you to ponder, to scrutinize, to interrogate, to experiment, to feel, to empathize, to sympathize, to speculate, to assess, to organize, to pattern, to categorize, to connect, to integrate, to synthesize, to reflect, to hypothesize, to assert, to conceptualize, to abstract, and – if you are really good – to theorize. Counting is easy; thinking is hard work.

Coding is “dangerous,” “violent,” and “destructive.” I have difficulty understanding why words such as these have been chosen to describe the act of coding. I associate these words with natural disasters, criminals, and war, not with qualitative data analysis. I feel these monikers are sensationalist hyperbole in a culture of fear, and I question their legitimacy and accuracy for describing their critics’ intended concerns. In other words, these are, to me, poor word choices for an argument. And poor word choosers make bad coders.

Coding is an outdated method for qualitative data analysis. Coding qualitative data has over a half-century of use, and a substantive track record in many disciplines and scholarly publications. The technology needed for the enterprise has most certainly evolved through time, as have the methodologies and methods. But the core process of coding remains to this day a legitimate option for qualitative researchers. It is a tradition that has endured, not out of mindless adherence to established protocols, but due to its successful utility as a purposeful analytic approach to voluminous amounts of data.

There has been a recent trend in some circles of scholarship to dismiss and discount coding outright as an old-fashioned, positivist approach that does not harmonize with more theory-based analytics (e.g., inspired by Derrida, Foucault, Deleuze, Butler). Coding does not preclude or push theory and theorists aside. You as the analyst can still weave theory into your thinking through analytic memo writing and in the final report itself. Coding is neither a philosophy nor a way of viewing the world; it is simply a heuristic for achieving some sense of clarity *about* the world from your data and your deep reflections on them.

There’s more to data analysis than just coding. I absolutely agree. The more than 40 analytic approaches documented in [Appendix B](#) alone support this perception. This manual advocates that coding is a heuristic – a method of discovery that hopefully stimulates your *thinking* about the data you have been given and have collected. And in case you forgot two very important principles stated at the beginning of this chapter, here they are again:

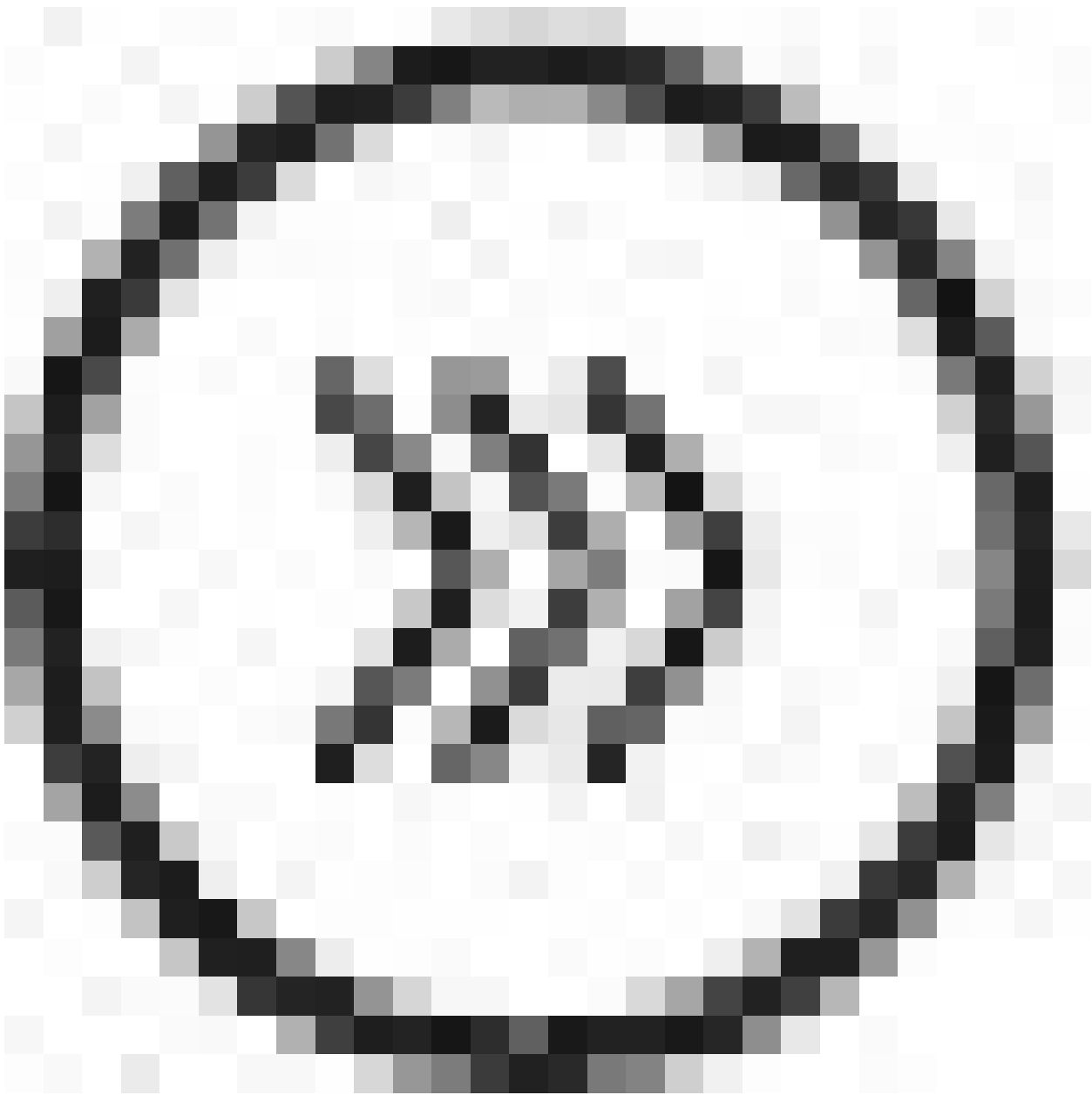
- Coding is just *one* way of analyzing qualitative data, not *the* way.
- There are times when coding the data is absolutely necessary, and times when it is most inappropriate for the study at hand.

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, 2011). My own qualitative research projects, in fact, have ranged from the realist to the literary and from the confessional to the critical (Van Maanen, 2011). 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 is 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. Corbin and Strauss (2015) wisely advise, “The best approach to coding is to relax and let your mind and intuition work for you” (p. 219).

This heightened, ever-present awareness of craft, of “how to,” transfers into my research work ethic. I am both humbly and keenly aware not only of what I do but why I do it. A metacognition of method, even in an emergent, intuitive, 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 qualitative 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.



Don't forget to visit <https://study.sagepub.com/saldanacoding3e> to access a wealth of resources including CAQDAS links, codes lists and coding examples, select free SAGE journal articles, sample interview transcripts, and group exercises and activities.