Data Analysis in Qualitative Research

Formal qualitative research has been conducted since the early 20th century. Qualitative data analysis, however, is still a relatively new and rapidly developing branch of research methodology. Writing in 1984, in the first edition of their book entitled *Qualitative Data Analysis*, pioneers in qualitative data analysis Matthew Miles and Michael Huberman noted that "we have few agreed-on canons for qualitative data analysis" (p. 16). In 1994, in the second edition of their book, they noted, "Today, we have come far from that state of affairs. . . . Still, much remains to be done" (Miles & Huberman, p. 428). Over recent years, many qualitative researchers have realized the need for more systematic data analysis procedures, and they have started to write more about how to conduct qualitative research data analysis (e.g., Bazeley, 2013; Bernard & Ryan, 2010; Bryman & Burgess, 1994; Dey, 1993; Huberman & Miles, 1994; LeCompte & Preissle, 1993; Lofland & Lofland, 1995; Miles & Huberman, 1994; Miles, Huberman, & Saldaña, 2014; Patton, 1990; Silverman, 1993; Strauss & Corbin, 1990). In this chapter, we introduce you to the terminology surrounding qualitative data analysis, show you the basics of qualitative data analysis, and briefly discuss the use of computer software in the analysis of qualitative data.

Interim Analysis

Data analysis begins early in a qualitative research study, and during a single research study, qualitative researchers alternate between data collection (e.g., interviews, observations, focus groups, documents, physical artifacts, field notes) and data analysis (creating meaning from raw data). This cyclical or recursive process of collecting data, analyzing the data, collecting additional data, analyzing those data, and so on throughout the research project is called interim analysis (Miles & Huberman, 1994).

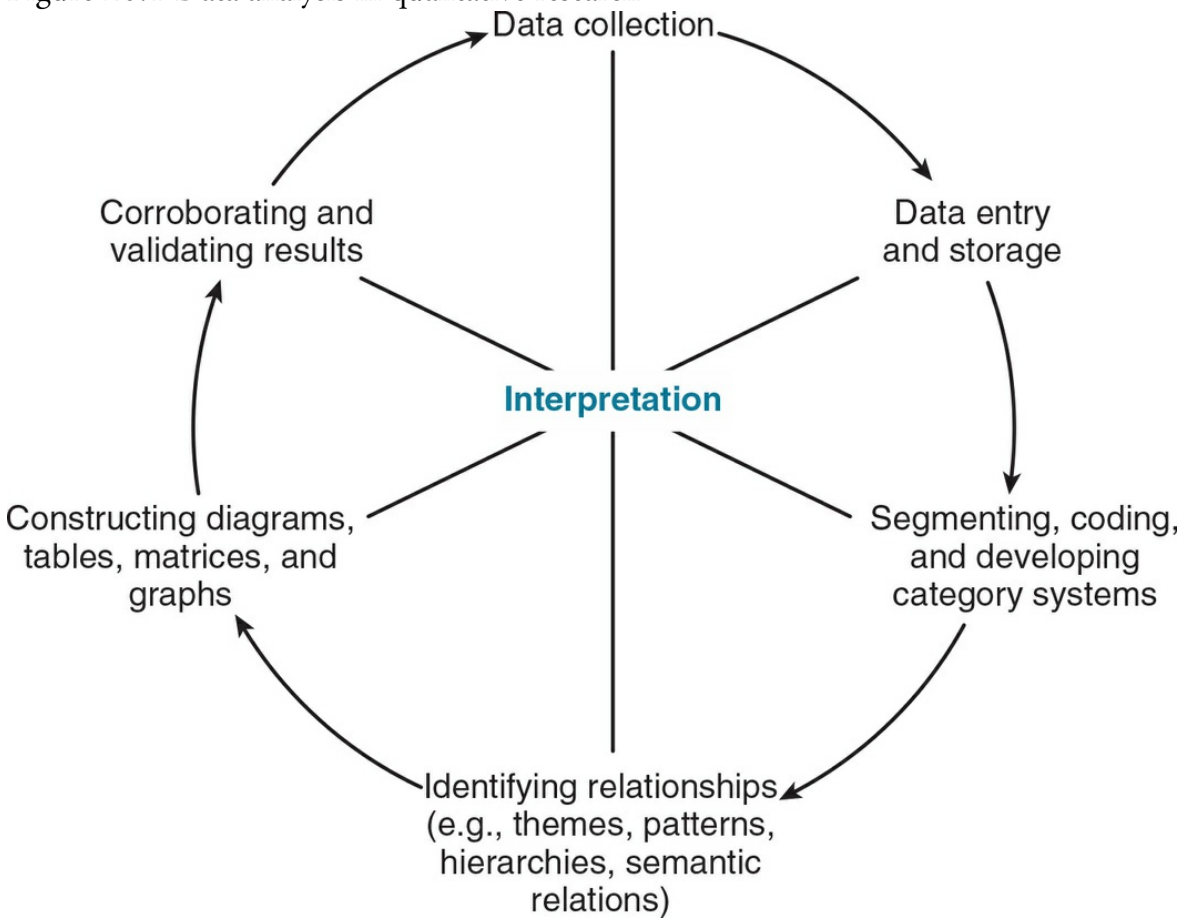
Interim analysis is used in qualitative research because qualitative researchers usually collect data over an extended time period and they continually need to learn more and more about what they are studying during this time frame. In other words, qualitative researchers use interim analysis to develop a successively deeper understanding of their research topic and to guide each round of data collection. This is a strength of qualitative research. By collecting data at more than one time, qualitative researchers are able to get data that help refine their developing theories and test their inductively generated hypotheses (i.e., hypotheses developed from examining their data or developed when they are in the field). Qualitative researchers basically act like detectives when they carefully examine and ask questions of their data and then re-enter the field to collect more data to help answer their questions. Interim analysis continues until the process or topic the researcher is studying is understood (or until the researcher runs out of resources!). Grounded theorists use the term *theoretical saturation* to describe the situation in which understanding has been reached and there is no current need for more data. We have summarized the qualitative data-collection process in Figure 20.1.

Interim analysis The cyclical process of collecting and analyzing data during a single research study

Memoing

A helpful tool for recording ideas generated during data analysis is memoing (writing memos). Memos are reflective notes that researchers write to themselves about what they are learning from their data. Memos can include notes about anything, including thoughts on emerging *concepts*, *themes*, and *patterns* found in the data; the need for further data collection; a comparison that needs to be made in the data; and virtually anything else. Memos written early in a project tend to be more speculative, and memos written later in a project tend to be more focused and conclusive. Memoing is an important tool to use during a research project to record insights gained from reflecting on data. Because qualitative data analysis is an interpretative process, it is important that you keep track of your ideas. You should try to record your insights as they occur so that you do not have to rely on your memory later.

Figure 20.1 Data analysis in qualitative research



Memoing Recording reflective notes about what you are learning from the data

Analysis of Visual Data

As you read about visual data, think about the old adage that "a picture is worth a thousand words." We will change the adage to "an image might be worth a thousand words" because visual data can include any type of image, such as photographs, art, pictures in books, video images, nonverbal expressions shown to you by your research participants, and any "signs" that are in the field for you to see. Researchers who rely on extensive visual data argue that it is a myth that it is necessary to present findings in written form, or what Collier (2002) called "the deceptive world of words" (p. 59). In many fields, such as cultural anthropology and media studies, visual data are primary sources of evidence. We mention here three approaches to visual data analysis: photo interviewing analysis, semiotic visual analysis, and visual content analysis.

Photo interviewing is a method of data collection (described in Chapter 9) in which researchers show images to research participants during formal or informal interviews. What is unique in this approach is that the researcher has the participant "analyze" the pictures shown to him or her; the researcher records the participant's thoughts, memories, and reactions as "results." In this approach, the pictures are the stimulus, and the participant is the analyst. The researcher reports these descriptive findings as the primary results. In addition to this photo interviewing analysis, the researcher can interpret the results further. In the remainder of this chapter, data analysis is considered to be conducted by the qualitative researcher or the qualitative researcher in combination with the participants after the initial data have been collected.

Semiotic visual analysis is based on the theory of semiotics. Semiotics is the study of signs and what they stand for in a human culture. A *sign* is something that stands for something else and may mean something different to people in different capacities. A researcher who conducts semiotic analysis is therefore very concerned with what the signs in visual images mean. Semiotic researchers are not concerned with finding images that are statistically representative of a large set of images. Rather, they are concerned with individual images that have conceptual meaning or with how meaning is produced by images.

Images often have layered meanings. From a semiotic perspective, images are denotative and connotative (Barthes, 1973). In the first layer, called *denotative* meaning, researchers simply want to know what is being depicted in the images. This layer assumes that we can only recognize what we already know, and this knowledge can be affected by verbal captions placed under photographs, for example, or by visual stereotypes in our cultures. The second semiotic layer, *connotative* meaning, builds on what researchers and participants know and explores the ways in which ideas and values are expressed and represented in images. This is what is so exciting and so exasperating about semiotic research. *Semiotics* explores myths, and nowhere is mythology as evident as in visual imagery.

Visual content analysis is different from semiotic analysis. Visual content analysis is based on what is directly visible to the researcher in an image or set of images. It differs from other methods of visual analysis in that it is more quantitative. For example, with visual content analysis, researchers might examine the relative frequencies of women or minorities in school texts or on websites that recruit college professors. Unlike more qualitative visual data analysis methods, visual content analysis concentrates on studying a representative sample rather than individual instances of images. It is less concerned with deep meaning and more concerned with prevalence. Visual content analysis begins with assertions or hypotheses that categorize and compare visual content. The categories are observable. The corpus (sample size or domain) of the study is decided ahead of time based on the research questions, how important it is to generalize the findings, and the statistical procedures to be employed. Visual content analysis is often limited to isolated content that represents particular variables under study. The variables are limited by clearly defined values that coders can classify consistently (reliably). For example, the variable *setting* takes on one or more of the values of office, domestic, public, religious, school, outside, or other.

Photo interviewing analysis Analysis is done by the participant, who examines and "analyzes" a set of visual images

Semiotic visual analysis The identification and interpretation of symbolic meaning of visual data

Semiotics The study of signs and what they stand for in human culture

Visual content analysis The identification and counting of events, characteristics, or other phenomena in visual data

There are numerous methods of visual analysis, using both qualitative (interpreting) and quantitative (counting) data analysis approaches. The visual data analyzed can include single images, such as a photograph or a drawing, or multiple images, such as time-sequenced images or videos. Visual data can be "analyzed" by participants (group members, informants) during data collection to report events and construct meaning, by expert coders to count the occurrence of particular concrete phenomena, and by individuals adept at interpreting cultural meaning as well as through a number of other approaches. Visual data also can be included in qualitative software, and they can be in their original form (e.g.,

pictures, photographs, video images). You can write up theoretical memos (i.e., your and your participants' interpretations and thoughts about your visual data) and include these memos along with your other transcribed materials in your qualitative research data set. In qualitative research, you often will have a set of multiple kinds of materials to analyze in order to learn about what you are studying.

Data Entry and Storage

To analyze qualitative data carefully, we recommend that you transcribe most of your data. Transcription is the process of transforming qualitative research data, such as audio recordings of interviews or field notes written from observations, into typed text. The typed text is called a transcript. If the original data source is an audio recording, transcription involves sitting down, listening to the tape recording, and typing what was said into a word processing file. If the data are memos, open-ended questionnaires, or observational field notes, transcription involves typing the handwritten text into a word processing file. In short, transcription involves transferring data from a less usable to a more usable form. After you transcribe your data, you should put your original data somewhere for safekeeping.

Transcription Transforming qualitative data into typed text

Some qualitative researchers use a voice recognition computer program, which can make transcribing relatively easy. These programs create transcriptions of data while you read the words and sentences into a microphone attached to your computer. Two popular programs are IBM's ViaVoice and Dragon's Naturally Speaking. The main advantage of voice recognition software is that it is easier to talk into a microphone than it is to type. Time savings are not currently large in comparison with typing, but the efficiency of these programs will continue to improve over time.

The principles discussed in this chapter also apply when your qualitative data do not directly lend themselves to text (e.g., videotapes of observations, still pictures, and artifacts). You cannot directly transcribe these kinds of data sources. What you can do, however, is use the principles of coding (discussed in the next section) and put the codes and your comments into text files for further qualitative data analysis.

Segmenting, Coding, and Developing Category Systems

Segmenting involves dividing the data into meaningful analytical units. When you segment text data, you read the text line by line and continually ask yourself the following kinds of questions: Do I see a segment of text that has a specific meaning that might be important for my research study? Is this segment different in some way from the text coming before and after it? Where does this segment start and end? A meaningful unit (i.e., segment) of text can be a word, a single sentence, or several sentences, or it might include a larger passage such as a paragraph or even a complete document. The segment of text must have meaning that the researcher thinks should be documented.

Segmenting Dividing data into meaningful analytical units

Coding is the process of marking segments of data (usually text data) with symbols, descriptive words, or category names. Here is how Miles and Huberman (1994) explained it:

> Codes are tags or labels for assigning units of meaning to the descriptive or inferential information compiled during a study. Codes usually are attached to "chunks" of varying size—words, phrases, sentences, or whole paragraphs. . . . They can take the form of a straightforward category label or a more complex one. (p. 56)

Coding Marking segments of data with symbols, descriptive words, or category names

When a researcher finds a meaningful segment of text in a transcript, he or she assigns a code or category name to signify or identify that particular segment. As you can see, segmenting and coding go hand in hand because segmenting involves locating meaningful segments of data and coding involves marking or labeling those segments with codes or categories.

An example of a coded interview transcript is shown in Table 20.1. The narrative in the transcript is from an interview with a college teacher (CT) by a researcher (R). You can see that the researcher read the text line by line and placed descriptive words or phrases in the left-hand margin next to the segments of text. The researcher also placed brackets around

the segments of data to make it clear where each segment started and ended. (Some other ways to mark segments are to use line numbers or to underline the relevant text.) In this example, a college teacher was asked about the experiences her students had when they visited elementary school classrooms as a course requirement in an educational psychology course. The teacher believed that the visitation experiences provided experiential learning by giving her students (potential future teachers) information that helped them make career choices.

As new codes are developed during coding, they must be added to the master list of codes if they are not already on the list. A master list is simply a list of all the codes used in the research study. The master list should include each code followed by the full code name and a brief description or definition of the code. A well-structured master list will enable other researchers working on the project to use the list readily.

Master list A list of all the codes used in a research study

During coding, the codes on the master list should be reapplied to new segments of text each time an appropriate segment is encountered. For example, one category from the master list for the data in Table 20.1 would be "career choice." Therefore, when the data analyst for this research study encountered another segment of data in which the same or a different person being interviewed made a comment about career choice, the researcher would reapply the label "career choice." Every time a segment of text was about career choice, the researcher would use the code "career choice" to refer to that segment.

TABLE 20.1 Example of Coded Text Data

	R: Well, let's start with the impact of early field experiences in the schools for undergraduate education majors. What kind of an impact do you see these experiences having on your students?
Book learning Experiential learning	CT: I think it gives them a needed view into the classroom in the real world. [It's one thing to read about teaching in books;] [it is another to actually go into a real classroom, with real students, and actually try to teach them something. Basically, I think that there is something to learning by experience.]
Classroom management Teaching strategies	[My students can try out the classroom management principles] and the [teaching strategies] I teach them about in my educational psychology course. My students can also learn that all elementary students are not alike.
Common student needs Individual student needs	[The kids have a set of common needs], but [they also have a set of needs unique to each individual in the classroom].
	R: Are there any other results from going into the classroom?
Career choice Timing of vocational learning	CT: Yes. Most of my students have not been in a real classroom since they were in school themselves. Things have changed in the schools in many ways since then. [The vocational experience of going into the classroom has helped some of my students decide that teaching really was not for them.] I hate to lose potential teachers, but [it is probably better that they decide now than wait until they have completed four years of education learning to be a teacher and then decide they don't want to be in a classroom.]

Here is an example of coding based on data from a consulting project done by one of this book's authors. The members of a public organization filled out an open-ended questionnaire in which one of the questions asked was, What are some specific problems needing action in your organization? The participants' responses are shown in Table 20.2. Take a look at the responses for a moment and decide whether you notice any meaningful categories of information. Then look at Table 20.3 and see how the data were coded. As you can see, the answers to the open-ended question are segmented into six categories. The codes are shown in the left-hand margin. The members of the organization listed a number of problems in their organization, and these problems fell into the categories of management issues, physical environment, personnel practices, employee development, intergroup and interpersonal relations, and work structure. These six categories were determined by examining the responses and sorting them into these inductive categories.

If you think that you or someone\ else might have coded the responses from the previous example differently, you are probably right. When you have high consistency among different coders about the appropriate codes, you have intercoder reliability. Intercoder reliability is a type of interrater reliability (discussed in Chapter 7; also see Miles & Huberman, 1994, p. 64). Intercoder reliability adds to the objectivity of the research, and it reduces errors due to inconsistencies among coders. Achieving high consistency requires training and a good deal of practice. Intracoder reliability is also important. That is, it is also important that each individual coder be consistent. To help you remember the difference between intercoder reliability and intracoder reliability, remember that the prefix *inter-* means "between" and the prefix *intra-* means "within." Therefore, intercoder reliability means reliability, or consistency, between or across coders, and intracoder reliability means reliability within a single coder. If the authors of qualitative research articles that you read address the issues of intercoder and intracoder reliability, you should upgrade your evaluation of their research.

TABLE 20.2 Unordered List of Responses to the Open-Ended Question, What are some specific problems needing action in your organization?

Participant Responses
There is not enough space for everyone.
Our office furniture is dated and needs replacing.
We need a better cleaning service for the office.
We need more objective recruitment and hiring standards.
We need objective performance appraisal and reward systems.
We need consistent application of policy.
There are leadership problems.
Nonproductive staff members should not be retained.
Each department has stereotypes of the other departments.
Decisions are often based on inaccurate information.
We need more opportunities for advancement here.
Our product is not consistent because there are too many styles.
There is too much gossiping and criticizing.
Responsibilities at various levels are unclear.
We need a suggestion box.
We need more computer terminals.
There is a lot of "us and them" sentiment here.
There is a lack of attention to individual needs.
There is favoritism and preferential treatment of staff.
More training is needed at all levels.
There needs to be better assessment of employee ability and performance so that promotions can be more objectively based.
Training is needed for new employees.
Many employees are carrying the weight of other untrained employees.
This office is "turf" oriented.
There is a pecking order at every level and within every level.
Communication needs improving.
Certain departments are put on a pedestal.
There are too many review levels for our product.
Too many signatures are required.
There is a lot of overlap and redundancy.
The components of our office work against one another rather than as a team.

If you want to code your own data and develop category names, you should start with words that describe the content of the segments of data. You will often want the category name to be more abstract than the literal text so that the same category name can be applied to other, similar instances of the phenomenon that you encounter as you read more text. For example, in Table 20.3, the category name "physical environment" was used

rather than “office furniture” so that other aspects of the physical environment, in addition to office furniture, could be included in the category. This ability to develop category names comes with practice. You might not get the best category name on your first try. If you don’t, all you have to do is generate a new category name and use the new category name on the transcripts. When you actually code some written text, you will find that this process of coding is easier than you might think.

TABLE 20.3 Categorization of Responses to the Open-Ended Question, What are some specific problems needing action in your organization?

Inductive Categories	*Participant Responses*
Management issues	There are leadership problems.
	We need a suggestion box.
	There is a lack of attention to individual needs.
	There is favoritism and preferential treatment of staff.
	Decisions are often based on inaccurate information.
	We need consistent application of policy.
Physical environment	We need a better cleaning service for the office.
	Our office furniture is dated and needs replacing.
	We need more computer terminals.
	There is not enough space for everyone.
Personnel practices	We need more objective recruitment and hiring standards.
	We need objective performance appraisal and reward systems.
	Nonproductive staff members should not be retained.
	There needs to be better assessment of employee ability and performance so that promotions can be more objectively based.
Employee development	More training is needed at all levels.
	Training is needed for new employees.
	Many employees are carrying the weight of other untrained employees.
	We need more opportunities for advancement here.
Intergroup and interpersonal relations	This office is "turf" oriented.
	There is a lot of "us and them" sentiment here.
	There is a pecking order at every level and within every level.
	Communication needs improving.
	There is too much gossiping and criticizing.
	Certain departments are put on a pedestal.
	Each department has stereotypes of the other departments.
Work structure	There are too many review levels for our product.
	Too many signatures are required.
	Responsibilities at various levels are unclear.
	The components of our office work against one another rather than as a team.
	There is a lot of overlap and redundancy.
	Our product is not consistent because there are too many styles.

Full descriptive words or phrases are not always used in coding. Some researchers prefer to use abbreviations of category names as their codes. Using abbreviations can save time compared to writing out full category names every time a category appears in the data. Other researchers develop complex symbol systems for coding their data. When you code

some data for yourself, you must decide whether you want to use full words, phrases, abbreviations, or a complex symbolic coding system.

An example of data coded using a symbolic coding system is shown in Table 20.4. The transcript is an excerpt from an observational study done by educational ethnographer Margaret LeCompte, who was studying norms in the elementary school classroom. LeCompte placed the time in the left column every 5 minutes or when an activity changed. She placed teacher talk in quotes and placed student talk and information recorded by the researcher in parentheses. The type of activity is indicated in the left margin. The code *R* stands for teacher talk that establishes rules, the code *T* stands for teacher talk focused on organizing a time schedule for the students, and the code *W* stands for teacher talk that is focused on student tasks or student work. Although the codes that are used in the table are not very clear to the outside reader, they had very precise meaning to LeCompte. LeCompte inductively developed her coding system early in her research study, and she used it in her later data analysis.

TABLE 20.4 Symbolic Coding System Used on Field Note Transcript

(Children are playing outside the classroom; a few are standing on the porch. The teacher arrives.)

8:55			"Come in, girls first." (There's some messing around before they line up.) (They come in and move toward their seats.) (T2A) "Mrs. Smith is ready to start." (She's sitting on the desk in the front of the room.) (R1A) "Mrs. Smith is waiting." (R2B) "I like the way Bernie is sitting down, and Atocha." (R1A) "Please, people, do not throw snowballs at one another." (R4B) "There isn't enough snow on the ground and you pick up rocks with it. If we have a lot of snow we'll have a snowball fight, but please don't throw the snow when there isn't much . . ." (R4A) "If you go along with me and don't throw now, as soon as there's good stuff we'll have a snowball fight." (R4B) "It isn't just that you hurt people, but you'll get in trouble too."
8:57		T2A	
	Getting	R1A	
	settled	R2B	
		R1A	
		R4B	
		R4A	
		R4B	
9:03		T2A	(T2A) "All right, the girls will go to bake cookies at recess." (W1B) "Boys, come back here if you aren't done; if you can't work alone you can go into Mrs. Dvorak's game
		W1B	
	Getting	W2B	room." (W2B) "I expect if you come in here to work
	organized	R1A	I expect you to work." (R1A) "I want everybody to bring a nickel by Monday." (Is it for the girl's surprise?) "No, it's for everybody"."

Source: From M. D. LeCompte and J. Preissle, *Ethnography and Qualitative Design in Educational Research,* p. 294, copyright © 1993 Academic Press. Reprinted by permission of Elsevier and the authors.

Note: Teacher talk is recorded in quotations; pupil talk and locational description are enclosed in parentheses.

Inductive and A Priori Codes

Because of the inductive nature of most qualitative research, qualitative researchers traditionally generate their codes or category names directly from their data. When you develop codes this way, you are actually generating inductive codes, which are defined as codes that are generated by the researcher by directly examining the data during the coding process. Inductive codes can be based on emic terms (terms that are used by the participants themselves). Codes that use the language and words of the participants are called in vivo codes. For example, high school students might use the emic term *jocks* to refer to students who play sports. Inductive codes can also be based on social science terms that a researcher is familiar with. For example, a social science term for *jocks* might be *athletic role.* Finally, inductive codes might be good, clear, descriptive words that most people would agree characterize a segment of data (e.g., we might agree that the segment of data refers to athletes).

Sometimes researchers bring an already developed coding scheme to the research project. These codes are called a priori codes or preexisting codes because they were developed before or at the very beginning of the current research study. A priori codes are used when a researcher is trying to replicate or extend a certain line of previous research. Researchers may also establish some a priori codes before data collection based on their relevance to the research questions. When researchers bring a priori codes to a research study, they come in with a start list of codes—an already developed master list that they can use for coding. During coding, however, the researcher should apply these codes only when they clearly fit segments of data. The codes should not be forced onto the data, and new codes should be generated when data segments are found that do not fit any of the codes on the list. In practice, many researchers employ both preexisting and inductive codes.

Inductive codes Codes that are generated by a researcher by directly examining the data

In vivo codes Codes that use the words of the research participants

A priori codes Codes that were developed before examining the current data

Co-Occurring and Facesheet Codes

In our discussion so far, we have used just one descriptive category for any given segment of data. If you code transcripts, however, it is very possible that the codes will overlap. In other words, more than one topic or category might be applied to the same set of data. If the categories are intertwined, you simply allow the codes to overlap naturally, and the result is what is called co-occurring codes. Co-occurring codes are sets of codes (i.e., two or more codes) that overlap partially or completely. Co-occurring codes might merely show conceptual redundancy in coding (i.e., the two codes mean basically the same thing). More interestingly, co-occurring codes might suggest a relationship among categories within a set of text for a single individual (e.g., an interview transcript) or across multiple sets of text for different individuals (i.e., across several interview transcripts).

Co-occurring codes Codes that overlap partially or completely

An example of co-occurring codes within an individual's transcript is shown in Table 20.5. If you look at the text in the table, you will see that "mood" is the category marking lines 8–13, "positive" is the category for lines 11–20, "like" is the category for lines 16–20, "don't like" is the category for lines 21–29, "miss" is the category for lines 30–40, and "they" is the category for lines 32–34. As you can see, some of these categories overlap. More specifically, lines 32–34 are coded with two co-occurring codes. The two codes "miss" and "they" co-occur for these three lines. Also, lines 16–20 are coded with the codes "like" and "positive." Therefore, these are also co-occurring codes. The key point to remember is that you can allow codes to overlap when coding data.

TABLE 20.5 Text With Overlapping Codes

P: Would you go to school if you didn't have to?	6 7	
N: Sometimes I would go to school if I was in the mood, and other times—when I wanted to go horseback riding or something like that—I wouldn't. But sometimes when I felt I was in the mood to go, I would.	8 9 10 11 12 13	Mood
P: Do you like school?	14	
N: Sort of, sometimes.	15	Positive
P: What are some of the things you like about school?	16 17	
N: Well, sometimes I like school because we have extra recess or art or something like that, and other times—when we have to go to music—I don't like it because I don't like music.	18 19 20	Like
P: What are some other things you don't like about school?	21 22 23 24	
N: When we have a whole lot of work—like English, spelling, and handwriting, all kinds of stuff in the morning and stuff. Then we have about forty-two problems in math, like we did that one time.	25 26 27 28 29	Don't like
P: If you didn't go to school, what would you miss?	30 31	
N: Sometimes—like if there was laws and stuff that we don't have to go to school and they changed things around—then I might miss art one day and something [else] the next day and something like that, and I wouldn't even know what day we are supposed to have them on.	32 33 34 35 36 37 38	"They" Miss
P: What do you think you'd miss the most about school?	39 40	

Source: From M. D. LeCompte and J. Preissle, *Ethnography and Qualitative Design in Educational Research*, p. 294, copyright © 1993 Academic Press. Reprinted by permission of Elsevier and the authors.

A researcher can also attach codes to an entire document, interview, or set of lines. For example, lines 6–40 in Table 20.5 (i.e., all the given lines) could have been given a code such as "school" because that was the topic of discussion in all of the lines. If you had several interview transcripts, you might decide to attach the code "female" or "male" to each transcript to signify the participant's gender. Codes that apply to a complete document or case (e.g., to an interview) are called facesheet codes. The origin of the term *facesheet* probably comes from researchers attaching a sheet of paper to each transcript with codes listed that apply to the whole transcript. Demographic variables are frequently used as facesheet codes (e.g., gender, age, race, occupation, school). Researchers might later decide to sort their data files by facesheet codes to search for group differences (e.g., differences between older and younger teachers) or other relationships in the data.

Facesheet codes Codes that apply to a complete document or case

First-Stage and Second-Stage Coding

We have now provided you with the basics of what can be called first-stage coding. It is similar to what grounded theorists call open coding. In first-stage coding you inductively and/or deductively code and categorize your transcribed data and show the results. It is called first-stage coding because it is the initial or first stage of your qualitative analysis. You should *not* end with first stage coding. After this process, you need to do something more with your codes, categories, and text. That is, you need to look for themes, enumerate them, and you should look for relationships among the codes/categories/themes. This later step is called second-stage coding, and it is where you make sense of your codes and complete the analysis. Before moving to the next section, we recommend that you browse the Chapter 20 Appendix (titled "Additional Types of Codes for Qualitative Research"). This appendix includes 13 additional kinds of codes that will add to your analytic repertoire and help you to become a better coder. We put the material in an appendix because it is not something to memorize but, instead, is a helpful resource for your research.

First-stage coding Initial coding of qualitative research data

Second-stage coding Follow-up coding to organize codes/categories and determine their interrelationships for the research report

Enumeration

We have talked about the importance of transcribing data, and we have shown you the basics of assigning codes to qualitative data. At this point, a data analyst might decide to determine how frequently words or coded categories appear in the data. This process of quantifying data is called enumeration. Enumeration helps qualitative researchers communicate concepts such as "amount" or "frequency" when writing up the results. Often a reader needs to know how much or how often, in addition to knowing that something happened. Weber (1990), for example, reported the word frequencies used in the 1980 Democratic and Republican platforms. The five most common words in the Democratic platform were *our* (430 occurrences), *must* (321), *Democratic* (226), *federal* (177), and *support* (144). The most common words in the Republican platform were *our* (347), *their* (161), *administration* (131), *government* (128), and *Republican* (126). Word or code frequencies can help researchers determine the importance of words and ideas. Listing frequencies can also help in identifying prominent themes in the data (e.g., What kinds of things did the participants say many times?).

Enumeration The process of quantifying data

When numbers are reported in qualitative research reports, you must always be sure to check the basis of the numbers being used, or you could be misled. For example, in the Democratic and Republican platform example, the basis was all words in the document (e.g., 144 of the words in the Democratic platform were *support*). A number such as this simply points out the emphasis placed on a word by the writer of the document. If several interview transcripts are analyzed, the basis of a reported number might be the number of words mentioned by all of the participants. If a word had a high frequency in this case, you might be inclined to believe that most of the participants used the word frequently. However, a high frequency of a particular word could also mean that a single participant used the particular word many times. In other words, a word might have a large frequency simply because one or two research participants used the word many times, not because a large number of different participants used the word. Enumeration can be helpful in qualitative data analysis, but always be careful to recognize the kinds of numbers that are being reported.

"Themeing the Data" and Creating Hierarchical Category Systems

Categories are the basic building blocks of qualitative data analysis because qualitative researchers make sense of their data by identifying and studying the categories that appear in their data. You can think of the set of categories for a collection of data as forming a classification system characterizing those data. Rather than having to think about each sentence or each word in the data, the researcher will, after coding the data, be able to focus on the themes and relationships suggested by the classification system. You learned earlier how to find categories in qualitative data, and you learned that you may want to count these categories for suggestive themes. Themes are frequently occurring codes/categories/words that occur frequently in your data and help summarize the findings. Themeing the data or thematic analysis is perhaps the most popular type of qualitative data analysis, and it is quite important. However, we recommend that you do not stop after identifying themes because this may result in your missing many important nuances and relationships in your data.

Another kind of analysis is hierarchical analysis. In hierarchical analysis, categories are organized into different levels, typologies, and hierarchical systems. A set of subcategories might fall beneath a certain category, and that certain category might itself fall under an even higher-level category. Think about the category called *fruit.* In this case, some possible subcategories are oranges, grapefruit, kiwi, apples, and bananas. These are subcategories of fruit because they are "part of" or "types of" the higher-level category called *fruit.* The category fruit may itself be a subcategory of yet a higher category called *food group.* Systems of categories like this are called hierarchies because they are layered or fall into different levels.

Theme A word, or more typically, a set of words denoting an important idea that occurs multiple times in your data

Thematic analysis Identification of themes in the research findings

Hierarchical analysis Search for potential hierarchical arrangement of inductively generated categories in qualitative data analysis

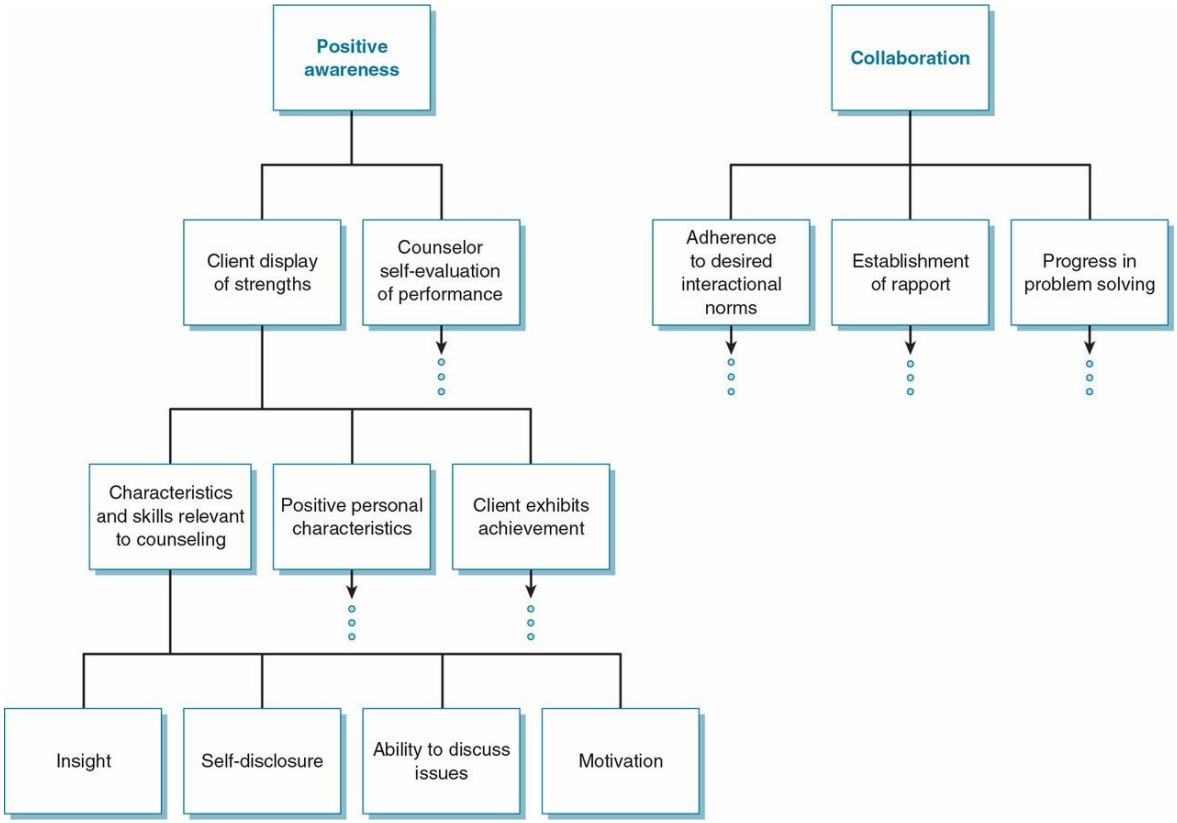
An example of a hierarchical classification system can be found in a research article by Frontman and Kunkel (1994). These researchers were interested in when and how counselors believed a session with a client was successful. They interviewed 69 mental health workers from various mental health fields, including counseling psychology, clinical psychology, marriage and family therapy, social work, and school psychology. After an initial session with a client, the participants filled out an open-ended questionnaire asking them to describe what they felt was successful in the session. A team of researchers analyzed the transcripts and came up with a rather elaborate hierarchical classification system. Frontman and Kunkel reported that they developed their hierarchy in a bottom-up fashion, which means that the lowest-level categories are the closest to the actual data collected in the study. This bottom-up, or inductive, strategy is the most common approach used by qualitative researchers (Weitzman & Miles, 1995).

We have reproduced a small part of Frontman and Kunkel's (1994) classification system in Figure 20.2 to give you a feel for hierarchical coding. When looking at the figure, be sure to realize that many of the categories in Frontman and Kunkel's hierarchical system are left out; the downward arrows indicate where additional levels and categories were excluded. All 44 categories in their full hierarchical classification system are given in the published article.

You can see that the higher levels of the hierarchy shown in Figure 20.2 are more general than the lower levels. That is, a higher-level category includes or subsumes the categories falling under it. The highest level of the hierarchy in Figure 20.2 includes the very general categories called positive awareness and collaboration. Frontman and Kunkel (1994) decided near the conclusion of their research project that these two general categories subsumed the sets of categories falling below them. At the second-highest level of the hierarchy in Figure 20.2, you can see that the researchers categorized counselors' construal of success into five categories. The five categories with brief explanations are as follows:

1. Client display of strengths ("The skills, actions, and characteristics expressed by the client that the counselor connotes as indication of success.")
2. Counselor self-evaluation of performance ("Counselor assesses success through evaluating the quality of his or her performance during the session.")
3. Adherence to desired interactional norms ("Success is determined by the presence of particular interactional patterns in the session.")
4. Establishment of rapport ("Success is defined as indication that rapport between counselor and client is being established.")
5. Progress in problem solving ("Success is attributed to client making progress toward establishing and implementing direct steps in solving a problem"; Frontman & Kunkel, pp. 498–499.)

Figure 20.2 Hierarchical categorization of counselors' construal of success in the initial session. The vertical ellipses show where we have left out some of Frontman and Kunkel's subcategories.



Source: From K. C. Frontman and M. A. Kunkel. A grounded theory of counselors' construal of success in the initial session. Journal of Counseling Psychology, 41(4), 492–499. Copyright © by the American Psychological Association. Reproduced with permission.

The first two of the five categories we just listed are part of positive awareness, and the last three are part of collaboration. At the second-lowest level of the hierarchy, we provide the categories falling under "client display of strengths." At the lowest level of the hierarchy, we show the categories falling under "characteristics and skills relevant to counseling." As you can see, there is a total of four levels in the hierarchy shown in Figure 20.2. We find Frontman and Kunkel's (1994) hierarchy interesting because it provides a direct picture of the hierarchical structure of their data. It is also interesting to see what counselors believe makes therapy successful.

Identifying Relationships Among Categories

In this section, we show you some ways to explore relationships in qualitative research data. *When qualitative researchers use the term* relationship, *it has a slightly different meaning than when quantitative researchers use the term*. You learned in earlier chapters that quantitative researchers focus their efforts on examining the relationships among variables. Qualitative researchers, however, attach a much broader meaning to the term *relationship.* The hierarchical system just shown in Figure 20.2 is one type of relationship. Another important type of relationship is potentially found when you check to see if your categories or themes *vary* by demographic or other grouping factors. For example, do males and females or different ethnic groups interpret the phenomenon (e.g., teaching math) differently and therefore act differently? This adds nuance and greater understanding to your general themes. We call this qualitative subgroup analysis.

Qualitative subgroup analysis Looking for relationships between categories/themes and demographic or other grouping factors

In short, qualitative researchers use the term *relationship* to refer to many different kinds of relations or connections between things, including but not limited to variables. This is not better or worse than quantitative analysis; it is just different. A summary of several additional kinds of relationships identified by one well-known qualitative researcher named James Spradley (1979) is given in Table 20.6. Take a moment to examine the nine relationships because you might identify some of these relationships when you are reading transcripts or when you are examining categories generated from your data. Spradley's list is not exhaustive, but it is suggestive. You will undoubtedly find additional kinds of relationships if you analyze some transcribed data.

Suppose you were reading an interview transcript and you came across the following text: "When I just ignore Johnny's acting out, he becomes more aggressive toward the other students in my classroom. But if I walk over and stand beside him, he will usually quiet down for a little while." This text suggests a possible causal process operating among several categories. (In Table 20.6, this is called a cause-effect relationship.) It is suggested, in particular, that ignoring Johnny's behavioral outbursts results in aggressive behavior and proximity results in less aggressive behavior. Obviously, two sentences like this in a transcript do not provide solid evidence of a general cause-and-effect relationship; however, statements like this do have a causal form, and they might suggest that you do additional analysis and data collection to explore the relationship further.

TABLE 20.6 Spradley's Universal Semantic Relationships

Title	*Form of Relationship*
1. Strict inclusion	*X* is a kind of *Y*.
2. Spatial	*X* is a place in *Y*; *X* is a part of *Y*.
3. Cause-effect	*X* is a cause of *Y*; *Y* is a result of *X*.
4. Rationale	*X* is a reason for doing *Y*.
5. Location for action	*X* is a place for doing *Y*.
6. Function	*X* is used for *Y*.
7. Means-end	*X* is a way to do *Y*.
8. Sequence	*X* is a step (stage) in *Y*.
9. Attribution	*X* is an attribute (characteristic) of *Y*.

Source: Adapted from J. P. Spradley, 1979, p. 111.

Now recall the hierarchical categorization that we showed you in Figure 20.2. If you look at the figure again, you will see that one of the categories was "characteristics and skills relevant to counseling." That category has four characteristics falling under it: insight, self-disclosure, ability to discuss issues, and motivation. You can view these four subcategories as following Spradley's strict inclusion relationship because they are "kinds of" characteristics or skills. Strict inclusion is a very common form of relationship in qualitative data analysis.

Educational researchers often use the term *typology* to refer to categories that follow Spradley's strict inclusion form of relationship. A typology is a classification system that breaks something down into its different types or kinds. A typology is basically the same thing as a taxonomy. You might remember what a taxonomy is from your high school or college biology class. (Okay, I know it has been a long time!) In biology, the levels of the animal taxonomy are kingdom, phylum, class, order, family, genus, and species. (Here's a memory aid: Kings Play Chess On Fiber Glass Stools.) Bailey (1994) pointed out that "the term *taxonomy* is more generally used in the biological sciences, while *typology* is used in the social sciences" (p. 6). Typologies are useful because they help make sense out of qualitative data.

Typologies can be simple or complex. You might, for example, be interested in the types of cliques in schools, types of teaching strategies used by teachers, or types of student lifestyles. These would be fairly simple, one-dimensional typologies. At a more complex level, you could view the hierarchical classification in Figure 20.2 as one big typology, showing the types of counselors' construals of success. To construct a typology, it is helpful to construct mutually exclusive and exhaustive categories. Mutually exclusive categories are clearly separate or distinct; they do not overlap. Exhaustive categories classify all the relevant cases in your data. Exhaustiveness of categories can be difficult in qualitative research because

some cases simply don't fit into a typology. However, the more cases there are that fit into your typology, the better.

Typology A classification system that breaks something down into different types or kinds

Mutually exclusive categories A set of categories that are separate or distinct

Exhaustive categories A set of categories that classify all of the relevant cases in the data

Another interesting typology was constructed by Patton (1990) when he was helping a group of high school teachers develop a student dropout prevention program. Patton observed and interviewed teachers, and here is what he found:

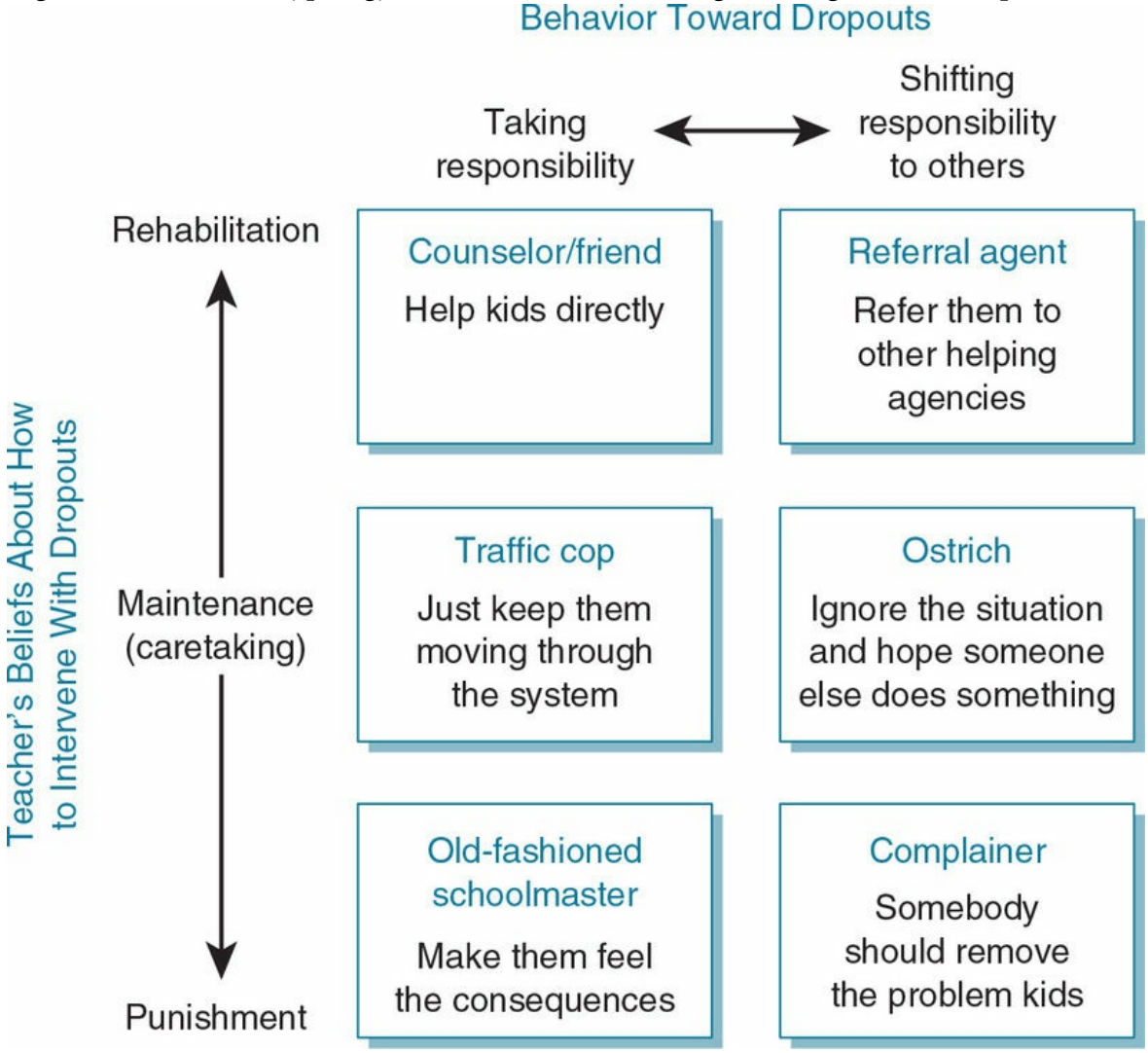
> The inductive analysis of the data suggested that teachers' behaviors toward dropouts could be conceptualized along a continuum according to the extent to which teachers were willing to take direct responsibility for doing something about the problem. This dimension varied from taking responsibility to shifting responsibility to others. The second dimension concerned the teachers' views about the effective intervention strategies. The inductive analysis revealed three perspectives among the teachers. Some teachers believed that a rehabilitation effort was needed to help kids with their problems; some teachers preferred a maintenance or caretaking effort aimed at just keeping the school running, that is, maintaining the system; and still other teachers favored finding some way of punishing students for their unacceptable and inappropriate behaviors, no longer letting them get away with the infractions they had been committing in the past. (pp. 411–412)

You can see from this quote that Patton found two simple or one-dimensional typologies that were related to dropout prevention: (1) teachers' beliefs about how to deal with dropouts and (2) teachers' behaviors toward dropouts.

Patton then decided to cross these two simple typologies in a two-dimensional matrix. When he did this, he found a typology that made a lot of sense to the teachers in the research study. The typology included six types of teacher roles in dealing with the high school dropout problem. The roles are shown in the six cells of the matrix in Figure 20.3.

The different types of teacher roles shown in the figure are counselor/friend, traffic cop, old-fashioned schoolmaster, referral agent, ostrich, and complainer. You might know some of these kinds of teachers at your own school. Remember, when analyzing qualitative data, you can sometimes find new and interesting information by cross-classifying two or more dimensions.

Figure 20.3 Patton's typology of teacher roles in dealing with high school dropouts



Source: Adapted from M. Q. Patton, How to Use Qualitative Methods in Evaluation, p. 413, copyright © 1990 by SAGE, Thousand Oaks, CA. Reprinted by permission of SAGE.

Now let's look at an example of Spradley's "sequence" type of relationship (Table 20.6). This example comes from an article titled "A Framework for Describing Developmental Change Among Older Adults" by Fisher (1993). Fisher pursued this research because he was interested in determining whether older adulthood could be categorized into a set of meaningful stages. He decided not to rely on the stages presented in popular developmental

psychology books because many of these lists were dated. Also, some of these lists lumped all older people into a single developmental stage called *old age*. Fisher decided that he wanted to explore the concept of old age using qualitative research.

Fisher conducted in-depth interviews with 74 adults whose ages ranged from 61 to 94 years. Using in-depth, open-ended interviews, he asked his participants what kinds of experiences they had in their lives. An interesting theme in his research findings was a tendency by all of the older adults toward adaptation to their life circumstances, no matter what the circumstances were. Fisher also generated five core categories from his data that could be ordered by time. These categories resulted in the following sequence of old age: (1) continuity with middle age, (2) early transition, (3) revised lifestyle, (4) later transition, and (5) final period. You can see the defining characteristics of each of these five stages in Table 20.7.

TABLE 20.7 Categories Ordered by Time

Category I: Continuity With Middle Age	
Characteristics:	Retirement plans pursued Middle-age lifestyle continued Other activities substituted for work
Category II: Early Transition	
Characteristics:	Involuntary transitional events Voluntary transitional events End of continuity with middle age
Category III: Revised Lifestyle	
Characteristics:	Adaptation to changes of early transition Stable lifestyle appropriate to older adulthood Socialization realized through age-group affiliation
Category IV: Later Transition	
Characteristics:	Loss of health and mobility Need for assistance and/or care Loss of autonomy
Category V: Final Period	
Characteristics:	Adaptation to changes of later transition Stable lifestyle appropriate to level of dependency Sense of finitude, mortality

Source: Adapted from Fisher, J. C. (1993). A framework for describing developmental change among older adults. *Adult Education Quarterly*, *43*(2), 81.

Drawing Diagrams

A useful tool for showing the relationships among categories is called diagramming (i.e., making diagrams). A diagram is "a plan, sketch, drawing, or outline designed to demonstrate or explain how something works or to clarify the relationship between the parts of a whole" (*The American Heritage Dictionary*, 2010). Figures 20.2 and 20.3, which we discussed in the previous section, are examples of diagrams. Diagrams are very popular with visually oriented learners and can be used to demonstrate relationships effectively for the readers of reports. The use of diagrams can also be helpful during data analysis when you are trying to make sense out of your data.

An easily understood example of a diagram showing a complex process appears in Figure 16.1 (page 459). This diagram depicts a grounded theory about how departmental chairpersons at universities facilitate the growth and development of their faculty members. The diagram shows that the career stage of the faculty member determines the type of faculty issue chairpersons are concerned with, and the faculty issue determines the specific strategy a chairperson uses in working with a faculty member. The diagram also lists the outcomes resulting from applying the strategies.

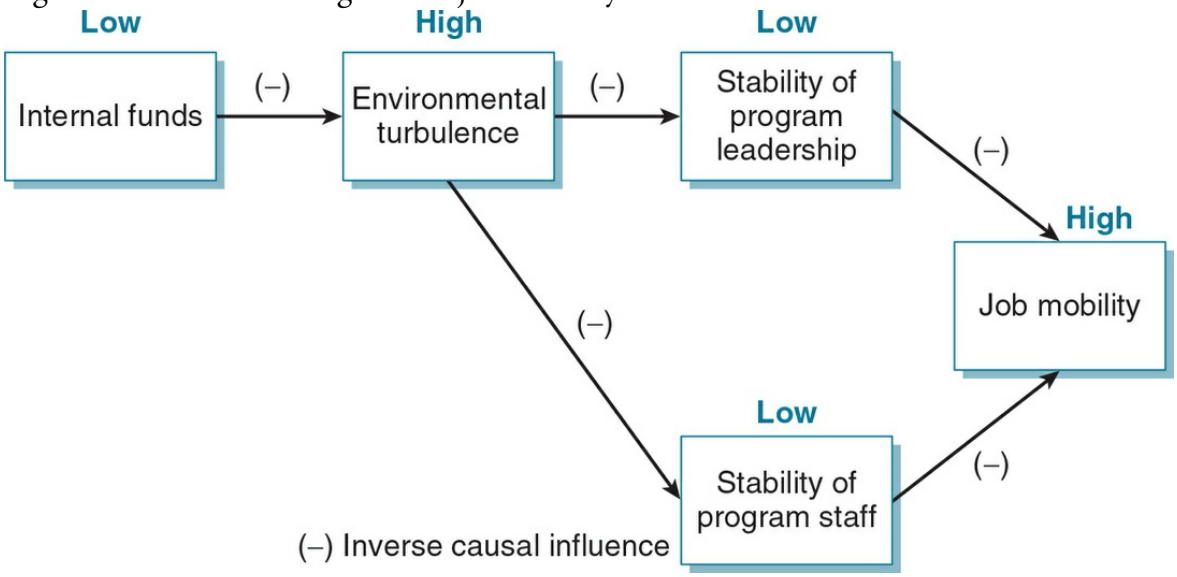
A similar type of diagram is called a network diagram. A network diagram shows the direct links between variables or events over time (Miles & Huberman, 1994). An example of a network diagram is the path analysis diagram that appears in Figure 14.4 (page 410). The path analysis diagram was based on quantitative research. However, network diagrams can also be based on qualitative data. Qualitative researchers often use these diagrams to depict their thinking about potential causal relationships. We have included an example of a part of a network diagram based on qualitative data in Figure 20.4. This diagram is based on a school innovation and improvement study by Miles and Huberman. According to the diagram, low internal funds in school districts resulted in high environmental turbulence in those districts (e.g., a shortage of money resulted in uncertain operating conditions for principals). This resulted in low stability for the leaders of the various school improvement programs and in low stability for the program staff. As a result of this instability, job mobility was high.

Diagramming Making a sketch, drawing, or outline to show how something works or to clarify the relationship between the parts of a whole

Network diagram A diagram showing the direct links between variables or events over time

To learn more about causal network diagrams, you should take a look at the work of Miles and Huberman (1994). They provide an extensive discussion of the issues surrounding causal analysis in qualitative research, and they discuss how to develop causal networks based on a single case or on multiple cases. If you are interested in cause-and-effect relationships with qualitative data, you should also review Chapter 11 on research validity in this book, especially the sections on internal validity. Miles and Huberman also discussed how to construct many different kinds of interesting matrices (i.e., classifications of two or more dimensions) to aid in the analysis and presentation of qualitative research data.

Figure 20.4 Network diagram for job mobility



Source: From M. B. Miles and A. M. Huberman, Qualitative Data Analysis: An Expanded Source Book, p. 231, copyright © 1994 by SAGE, Thousand Oaks, CA. Reprinted by permission of SAGE.

Corroborating and Validating Results

In Chapter 11, we discussed how to assess and promote the validity or trustworthiness of your qualitative research data (in the last section of the chapter). We recommend that you take a moment right now and review the five types of validity and the strategies that are used to promote qualitative research validity (they are shown in Table 11.2). It is essential that you think about validity and use the strategies throughout the qualitative data-collection, analysis, and write-up process whenever possible.

Computer Programs for Qualitative Data Analysis

Qualitative researchers are just beginning to capitalize on the possibilities for computer use in the analysis of qualitative data. Although qualitative researchers have been using word processors for transcribing and editing their data for quite some time, only during the last decade have a number of qualitative data analysis computer programs become readily available. The developers of these programs examined the procedures that qualitative researchers follow when making sense of their data and then developed programs that help automate these procedures. Before we examine the potential of using qualitative data analysis programs, we look at how qualitative researchers have traditionally made sense of their data without these programs.

Qualitative researchers traditionally use a filing system approach to data analysis. They begin their data analysis by transcribing their data and making copies of the various data documents. Then they hand-code the data in the left margin of these copies. After this, researchers make copies of the coded data and cut the data into segments of text with the marked codes. A filing system is created, with one folder for each code, and the segments of text are placed into their appropriate folders. If a segment of text has more than one code, then more than one copy of the segment is made, and a copy of the segment is placed in all the relevant folders. This way, all the folders contain all the appropriate data segments. At this point, researchers can reread the segments of text in each folder, looking for themes in the data.

More complex analyses require even more work when done by hand. For example, searching for two co-occurring codes typically requires making a folder with the two codes as its title, locating the two individual code folders, and then checking the text segments in those folders to see whether they include both codes in the left margin. If both codes are present, the segment of text is copied and placed into the new two-code folder.

As you can see, doing complex data analysis by hand can be time-consuming and quite difficult. Perhaps this is one reason why qualitative data analysis has not advanced as rapidly as has quantitative data analysis. Because of the increasing use of computer programs, however, we predict that the analysis of qualitative data will take a giant step forward during the next decade. One reason for our prediction is that procedures that are highly time-consuming when done by hand can be done with just a few keystrokes on the computer. So that you have a basic idea about the potential of computer data analysis, we list a few of the capabilities of qualitative data analysis programs.

Qualitative data analysis programs can be used to do virtually everything we have discussed in this chapter. They can, for example, be used to store and code your data. During coding, most programs allow complex hierarchical classification systems such as the one shown in Figure 20.2 to be developed. Most programs allow the use of many different kinds of codes,

including co-occurring and facesheet codes. Enumeration is easily done with just a few clicks of the computer mouse. Many programs allow you to attach memos or annotations to the codes or data documents so that you can record the important thoughts you have during analysis. Some programs will produce graphics that can be used in presenting the data. Finally, the heart and soul of most qualitative data analysis programs are their searching capabilities, the topic to which we now turn.

You can perform simple or complex searches with computer packages that use Boolean operators. Boolean operators are words used to create logical combinations based on basic laws of thought. We all use Boolean operators every day when we think and talk about things. Some common Boolean operators we all use are AND, OR, NOT, IF, THEN, and EXCEPT. Qualitative data analysis computer programs are written so that you can search your data or a set of codes using these and many other operators.

Boolean operators Words such as *and* and *or* that create logical combinations

You might, for example, search the codes or text in a set of interview transcripts concerning teacher satisfaction using the following string of words: "male AND satisfied AND first grade." The Boolean operator AND is called the intersection operator because it finds all intersections of the words or codes. This search would locate all instances of male, first-grade teachers who were satisfied. Similarly, you could search for female teachers using this string of words: "female AND teacher." You can find disconfirming cases (instances that do not have any of the characteristics) by adding the word NOT to the search command (e.g., "NOT teacher" will find all nonteacher instances).

Another operator is OR, also called the union operator. This operator finds all instances that take on any one of the provided words or codes. For example, if you searched a document with the command "female OR first grade" you would come up with instances that are either "female" or "first grade" or both. Another kind of search command is called FOLLOWED-BY in one popular program. Using this you can find instances in which two codes occur in a specific order in the data (e.g., punishment FOLLOWED-BY quiet behavior). As you can see, you can do a lot of different kinds of searches using Boolean operators.

Many qualitative and mixed research data analysis computer programs are available. The most popular programs are MAXQDA (http://www.maxqda.com), QDA Miner (http://provalisresearch.com/products/qualitative-data-analysis-software/), NVivo (http://www.qsrinternational.com/products_nvivo.aspx), and HyperRESEARCH (http://www.researchware.com/products/hyperresearch.html). Many others also work well, such as the new package Dedoose (http://www.dedoose.com) and older packages such as Ethnograph (http://www.qualisresearch.com) and ATLAS

(http://www.atlasti.com/qualitative-analysis-software.html). These and some additional links also are provided at the end of this chapter. Most of the companies will allow you to download a demonstration copy from their Internet site free of charge. If you decide that you are interested in a qualitative data analysis program, testing out a demonstration copy is an excellent way to find out which program best suits your particular needs.

We conclude by listing some of the advantages and disadvantages of using computer programs for the analysis of qualitative data. The advantages are that qualitative data analysis computer programs can help in storing and organizing data, they can be used for all of the analyses discussed in this chapter plus many more, they can reduce the time required to analyze data (e.g., an analysis procedure that takes a lot of time by hand may take virtually no time with a computer program), and they can make procedures available to you that are rarely done by hand because they are either too time-consuming or too complex. Some disadvantages are that computer programs can take time to learn, they cost money and require computer availability, and they can become outdated. The biggest disadvantage is startup time. Nonetheless, if you are planning on doing a lot of qualitative data analysis for an extended period of time, we recommend the use of computer programs.