TIPS & TOOLS #18: CODING QUALITATIVE DATA

This tip sheet provides an overview of the process of coding qualitative data, which is an important part of developing and refining interpretations in your interview, focus group or observational data. We begin with an overview of coding, and then the various aspects of coding are discussed. We also provide a sample interview transcript with codes and marginal remarks.

What Is Coding?

Coding is the process of organizing and sorting your data. Codes serve as a way to label, compile and organize your data. They also allow you to summarize and synthesize what is happening in your data. In linking data collection and interpreting the data, coding becomes the basis for developing the analysis. It is generally understood, then, that "coding is analysis."

Before we jump into the process of coding data, it is important to think about the big picture. One of the keys in coding your data, and in conducting a qualitative analysis more generally, is developing a storyline. Essentially, this element is primary to analyzing your data. This is the reason that thinking about the purpose of your evaluation—before, during and after data collection—is so critical. In thinking about it another way, the purpose of the study is your storyline and it is the analytic thread that unites and integrates the major themes of your evaluation. In this manner, it is the answer to the question: "What is this evaluation about?"

Creating a Storyline

The best way to develop your storyline is to write down a sentence or short paragraph that describes your evaluation in general terms. For instance, if you are interviewing multi-unit housing (MUH) owners and landlords concerning a potential smoke-free MUH policy, what are you trying to find out, and what do you want to convey with this information? Your coding scheme should be based on the story that you want to communicate to others. Many people start coding data without any idea of what they will end up writing or how they will write their evaluation, and as a result their coding scheme lacks coherence. In this case, coding becomes a waste of time. This is why end-use strategizing—thinking about the purpose of your study before you begin—is so important and a critical first step (see Tips and Tools #16(a)—Analyzing Qualitative Data—for more on end-use strategizing) in conducting all evaluations, and in this case, completing the coding process.

Why is developing a storyline so important to the coding process?

- It will help you decide what concepts and themes you want to communicate in your evaluation.

- It guides you in how your data should be organized and coded.

- It gives you the basic structure for your coding scheme.

Code Your Data

Coding can be done in any number of ways, but it usually involves assigning a word, phrase, number or symbol to each coding category. You will go through all your textual data (interview transcripts, direct notes, field observations, etc.) in a systematic way. The ideas, concepts and themes are coded to fit the categories.

Creating Codes

The process of creating codes can be both pre-set and open. We recommend a hybrid, using both these two models. Before beginning data collection and the coding process, it is good to begin with a "start list" of pre-set codes (often referred to as "a priori codes"). These initial codes derive from the conceptual framework, list of research questions, problem areas, etc. Your prior knowledge of the subject matter and your subject expertise will also help you create these codes. For instance, if you are interviewing MUH owners and managers, you may already think about the codes "economic issues" or "tenant smoking" or "common areas" (the list could go on and on. At a later time, the codes "economic issues" and "tenant smoking" may be collapsed into a larger code or theme of "barriers to policy."

Pre-Set Codes

A pre-set list can have as little as 10 codes or up to 40-50 codes. We recommend not creating too many codes because the person coding can become overwhelmed or make mistakes in the coding process if there are too many. In creating these codes, it is important to create a "code book," which is list of the codes and what they mean.

Emergent Codes

While it is good to begin data collection and coding with pre-set codes, another set of codes will emerge from reading and analyzing the data. These "emergent codes" are those ideas, concepts, actions, relationships, meanings, etc. that come up in the data and are different than the pre-set codes. For instance, in the aforementioned example of interviews of MUH owners and managers, the issue of tenants smoking medicinal marijuana may have come up. This may be seen as a tricky legal issue by the owners and managers. It may have been something not coded before data collection and coding began. So, the text discussing this issue could be

coded as "legal issue" (which was probably identified as a start code) and "medicinal marijuana." Because there's a good chance that medicinal marijuana was not a start code, it is added to the code book as an emergent code. In many cases, the "surprise" emergent codes form the basis of interesting stories and may indeed become part of the major storyline told in your evaluation.

Coding as a System of Organizing Your Data

One easy way to think about coding is to see it as a system to organize your data. In essence, it is a personal filing system. You place data in the code just as you would file something in a folder. A systematic way to code data is to ask yourself the following questions as you read the text:

- What is this saying? What does it represent?

- What is this an example of?

- What do I see is going on here?

- What is happening?

- What kind of events are at issue here?

- What is trying to be conveyed?

The word, number or symbol that you assigned to the item of data in answering such questions is a code. These are labels that classify items of information. We recommend using words or phrases as codes and in your marginal notes for later ease of analysis (sometimes numbers and symbols can be confusing).

Refining Your Codes

It is important to note that as your data are coded, the coding scheme will be refined. Meaning, you will add, collapse, expand and revise the coding categories. This is especially true of the pre-set codes. Oftentimes, what one expects to find in the data is not there. It happens. Moreover, some codes simply do not work or conflate other ideas from different codes. Alternatively, sometimes codes flourish in a way that there is too much data. In this case, the code needs to be broken down into sub-codes in order to better organize the data. The rule of thumb for coding is to make the codes fit the data, rather than trying to make your data fit your codes.

Coding "Notes"

Finally, as part of the process of coding, it is important to jot down notes of your reactions and ideas that emerge. These ideas are important and vital to the analytic process. These notes

may suggest new interpretations, as well as connections with other data. Moreover, if you are mindful of what is growing out of the data, your notes will usually point toward questions and issues for you to look into as you code and collect more data.

Illustration of Coding and Marginal Remarks

Figure 1 is an example of the initial coding and marginal remarks on a hard copy of interview transcripts. Typically, in an interview transcript like the one shown (Figure 1), you will read the entire interview from start to finish without coding. On the second read, you will want to jot down both codes and remarks on a hardcopy as you read it. As mentioned previously, the codes will derive from both those created prior to data collection ("pre-set codes," also referred to as "a priori codes"), as well as those that are created as data are collected and transcripts are reviewed (referred to as "emergent codes").

Figure 1: Illustration of Coding and Marginal Remarks

It was really good. There was a variety of activities, the overhead and information where they talked about it, and the opportunity to practice the activities together. I liked it. The 5 hours went really quickly. We had a good group, and felt very comfortable because everyone was open and sharing. And the lunch was wonderful. Having lunch was a good idea.

Yes, the structure helped my grasp the information, and I enjoyed the group size and variety of activities. 5.5 hours was good enough, and it went quickly. It all seemed to follow their outline, and it gave the opportunity to listen and then practice and get to know other people, because I was there by myself.

I think I would have gotten the same information either way with the overheads and printouts, but the interpersonal and opportunity to relate to other people and have back and forth and the ability to ask questions was more personal and enjoyable. In a webinar, I don't feel comfortable asking questions to someone I don't know, so the personal, face to face was better than if I did the webinar. I imagine I would have gotten the same information, but it wouldn't have been as enjoyable, and without the activities and other people's questions, I probably wouldn't remember as much.

1. What was your overall impression of the training in terms of its usefulness to your work?

It is great because I do survey development and work with people that develop surveys. It was really helpful. Afterward, I analyzed surveys and it made me wish I had attended the training before, because now I know there's better ways to do it. Yes, it has definitely met my expectations

2. Has your confidence level about designing or adapting surveys changed at all because of what you learned in the training?

I think my confidence increased because before I just did it, and now I know the reason why I should be doing such things. I am more confident in my abilities. I feel more knowledgeable. The one trap I know to be very careful of the double barreled questions and to use simpler words in the surveys so it's more easily understood. I also like the idea of putting similar kinds of questions together, like putting yes/no questions together and putting multiple choice questions together, and other formatting issues that make it pleasing to the eye and easy to follow. The double barrel one is the one that I have encountered in the past.

And another thing I've never done, and made me wish I had the training before I did a survey, was that I would have known to field test the survey first to make sure that the questions are clear and easily understood and you're getting a clear idea of what their thoughts are instead of things that can be ambiguous.

After the initial coding, Word files need to be created based on your codes. Think of this process as cutting and pasting the quotes on a poster board (how people did it on the good ol' days!). The marginal notes will also come in handy when thinking about how the codes fit together. Thus, in Figure 1, we have the codes "hands on activities" and "reaction." After going through a handful of the interviews and rereading our marginal notations, we see that these codes are part of a larger theme of "onsite learning vs. webinar." At a later point, we thus collapsed the codes into a larger theme and can discuss various aspects of onsite learning and contrast it with remote (webinar) learning. We can compare these two major ideas based on our data. Moreover, within "onsite learning" we may compare participants' views in order to show the complexity of the analysis. Our coding process will thus enable us to show the richness, complexities and contradictions of the social milieu we are evaluating, which is the basis of qualitative methods.

References:

Gibbs, Graham. 2007. *Analyzing Qualitative Data*.

Lofland, John, and Lofland, Lyn H. 1995. *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis.* Wadsworth Publishing Company: B elmont, CA.

Miles, Matthew B., and Huberman, Michael A. 1994. *Qualitative Data Analysis*.

Taylor, S.J. and Bogdan, R. (1998). *Introduction to Qualitative Research Methods*. Wiley and Sons: Hoboken, NJ.