

Module #4. Qualitative Data Analysis

In this module, we focus on the processes of description and discovery fundamental to qualitative analysis. The example we use in this section tracks the Lucretia Mott design team, which is seeking to innovatively rethink how to structure the student-counselor relationship to support college and career readiness. This example is grounded in the Carnegie design principle of positive youth development, and specifically on consistent student-adult relationships that communicate high expectations for student learning and behavior.

The sequence of sections in this module appears below:

1. Determine the best approach to analyze your data
2. Emergent coding
3. Established coding

In the course of your research and design process, you will likely have collected qualitative—non-numerical—data, whether through focus groups, interviews, open-ended survey questions, or artifacts. These data, often in the form of pages of transcripts or survey responses, can feel overwhelming. But by making meaning out of them, you not only begin to see themes in your findings, but also clarify and deepen your own understanding of the experiences of students and adults in your would-be school. This deepening of understanding is one important element of Carnegie's principle of continuous improvement of the school model.

To glean meaningful insight from rich and complex qualitative data, it serves to undertake a process called coding. Coding entails categorizing qualitative data to identify key themes and, in some instances, examining frequency of, and relationships between, different themes. By coding qualitative data, you identify salient themes and can apply these findings to your school design.

1. DETERMINE THE BEST APPROACH TO ANALYSIS FOR YOUR DATA

Key information

Forms of data analysis. There are two families of data analysis: qualitative analysis, which seeks to understand artifacts in the form of words, pictures, and videos; and quantitative analysis, which examines the relationships between data in numerical form. In this module, we focus on qualitative analysis.

-
- *Qualitative analysis seeks to understand artifacts in the form of words, pictures, and videos*
 - *Quantitative analysis examines the relationships between data in numerical form*
 - *Coding entails classifying pieces of qualitative data into categories that can be described or counted*
-

Coding qualitative data. After collecting qualitative data through interviews or open-ended survey questions,¹ the analysis process seeks to highlight similarities in an inherently rich and varied dataset. This process is called coding:

¹ Focus groups—unless you do more than one—do not typically require coding. Rather, it is best to have another person observe the focus group and also write up notes just after the group. You and your co-observer can then share notes and arrive at an agreement about the themes and ideas that were most salient.

classifying pieces of qualitative data into categories that can be described or counted. There are two broad categories of coding:

- Emergent coding: The codes used to categorize data emerge from the data; this process is used when your question is broad and exploratory.
- Established coding: The codes used to categorize data are established prior to beginning the coding process; this rigorous process is used when you have a clear and focused set of questions and hypotheses.

Your choice of coding

Your choice of coding process will influence how you report on your data. With thoughtful emergent coding, you will arrive at helpful examples and a preliminary sense of the relationships between the themes you identified. If you want to incorporate greater rigor into your coding—such as reporting frequency of certain themes and relationships between them—you will need to conduct established coding.

The choice between emergent and established coding depends on the phase of development of your research. In some instances, emergent coding is conducted prior to established coding when the researcher ultimately decides to take additional steps to quantify their qualitative data.

Your choice of coding process will influence how you report on your data. With thoughtful emergent coding, you will not conclude that certain themes are especially common or important, but rather will identify that some or many respondents endorsed a given theme. You will glean helpful examples and a preliminary sense of the relationships between the themes you identified. If, however, you want to incorporate greater rigor into your coding—such as reporting frequency of certain themes and relationships between them—you will need to conduct established coding.

For example, with emergent coding, you can report: Female students report that gender-matched relationships feel more comfortable and allow for more personal disclosure. With established coding, you can report: 25% of boys and 75% of girls comment on how gender-matched counseling allows for greater disclosure.

The amount of time required for emergent and established coding varies depending on the complexity of the coding scheme and the length of the transcript or artifact (note that we explain the distinctions between types of coding schemes in Section 3 of this module:

APPLICATION TO EXAMPLES

What type of coding should be used for qualitative data? For the purposes of this module, we maintain our focus on the principle of positive youth development, particularly the importance of consistent and caring student-adult relationships. The scenarios explored in this section focus on the Lucretia Mott design team's consideration of the possibility of setting up single-sex counseling groups during advisory:

Our research question is...	We will probably use....
Broad and exploratory: At this point, we are seeking to generally describe our topic of interest—single-sex counseling groups—to inform the design of our new school. We have not arrived at highly specific research questions and are not sure what answers will arise from our data.	Emergent coding
Based on research that says single-sex counseling groups create more supportive and safer environments for participants, we plan to have single-sex counseling groups at our school. We think conversations will look different in the different groups, but we are not sure how. What do these conversations look like for boys and girls?	
Focused and coupled with hypotheses: At this point, we have moved the general description stage, and have a clear sense of what we want to know and what findings we expect to reach. We have collected qualitative data on a few specific elements of our topic of interest, and have some idea of the answers we expect to arise from our data.	Established coding
We plan to have single-sex counseling groups that focus on family, peers, school, and the future. Under each of these headings, we have identified possible topics and what to get student feedback on how engaging these will be.	

2. EMERGENT CODING

Key Information

Emergent coding to identify broad themes. If you have opted to use emergent coding as your mode of analysis, you likely have collected data exploring a broad question or idea and are interested in identifying some findings around a broad set of themes. (Note that, as referenced in the previous section, focus groups—unless you do more than one—do not typically require coding. Rather, it is best to have another person observe the focus group and also write up notes just after the group. You and your co-observer can then share notes and arrive at an agreement about the themes and ideas that were most salient.)

The purpose of emergent coding

When you conduct emergent coding, themes emerge as you go through the data. Your goal is not to garner support for the importance of any particular theme within the general population, but to identify those themes that seem to emerge again and

again from the data. If you wish to see if themes and relationship you identified through the emergent coding process hold under more rigorous analyses, you could subject the same data to established coding.

When conducting emergent coding, the tool is your brain. Your brain will notice themes and connections. While the brain is a powerful tool, it also has limits. The process outlined here is intended to protect against the natural limitations of memory:

1. Initial review

- Read through a subset of the data (approximately 10%) with a general question in mind (such as, What kinds of relationships do students talk about in their interviews?). Starting with a small sub-set of data ensures you do not have to re-review the entire set of transcripts as new themes emerge.
- As you begin to see themes emerge from the data, highlight instances of them in a particular color, which will allow you to easily find examples when you report your findings.

2. Note themes

- Each time you identify a new theme, go to a “Notes” document and write up a description of the theme. Doing this when the theme first emerges helps you to differentiate between appearances of new themes and instances of previously-identified themes.
- Then, go back through earlier transcripts to identify this theme.
- After reviewing a sub-set of transcripts and before taking a break, refer to your notes document and make a note of:
 - What are you seeing? Does it affirm what you expected or surprise you?
 - What aren’t you seeing? What surprises arise? Why do you think they arose?
 - When you return to the data, read through these notes to remind yourself of where you were in the data analysis process. Do not ignore this step! Otherwise, you will find yourself doing duplicative work from day to day.

A note about note-taking

Since you are asking your brain to puzzle over something and identify trends, it is likely you will find yourself thinking about the task when you are not coding. Be sure to have some mechanism for capturing those thoughts, such as sending yourself e-mails from your phone or carrying around a little notebook.

Application to Example

Here, we present an illustration of the emergent coding process through excerpts from student interviews conducted by the Lucretia Mott design team focused on students' relationships with their counselors:

Research question: Based on research that says single-sex counseling groups create more supportive and safer environments for participants, we plan to have single-sex counseling groups at our school. We think conversations will look different in the different groups, but we are not sure how. What do these conversations look like for boys and girls?

Excerpt from Interview Transcripts

Questioner: Have you had helpful conversations with a counselor at school? Tell me about that.

Male Student 1: My counselor at school. He's really cool. He just listens really good, which is good because not a lot of my teachers do that, and that really makes a difference. And I didn't used to have anyone like that, but Joe, he sees me every week and he really cares. And what we talked about a lot the last month is what I'm going to do after because now I'm going to be in tenth grade next year and he's always saying it's important that I think about what's going to happen after I graduate.

Questioner: And what are the things in your conversations that are most helpful to you?

Male Student 1: Yeah, the college stuff and I'll do after I graduate. Cause you know what, I never thought that I was going to go to college, but Joe does and that is really—that makes me see things different than I used to. So it's really big that he talks about college and not just about my grades and my this and that.

Questioner: Have you had helpful conversations with a counselor at school? Tell me about that.

Male Student 2: My counselor is Wendy and I used to have no idea what different kinds of jobs were that were out there and she had me do this career inventory that opened up my eyes. It was the most useful thing I did in counseling. Hands down.

Questioner: Have you had helpful conversations with a counselor at school? Tell me about that.

Female Student 1: My counselor's Wendy. The most useful thing she does with me is, last year, when I was having some problems with this group of girls I used to hang out with and they were hassling me. I was in a bad place. They just were not the right group for me, but I didn't see that and Wendy helped me see that. But it's not like she tells you what you're supposed to do. Instead, she talked me through it and she, I guess she helped me realize...

Initial Review

During an initial review of this sub-set of transcripts, the design team highlights themes that jump out at them in different colors, as shown below. They also note the corresponding themes, with a brief description, in a separate

notes document (shown here next to the transcript for ease of reading).

Questioner: Have you had helpful conversations with a counselor at school?
Tell me about that.

Male Student 1: My counselor at school. He's really cool. He just listens really good, which is good because not a lot of my teachers do that, and that really makes a difference. And I didn't used to have anyone like that, but Joe, he sees me every week and he really cares. And what we talked about a lot the last month is what I'm going to do after because now I'm going to be in tenth grade next year and he's always saying it's important that I think about what's going to happen after I graduate.

Questioner: And what are the things in your conversations that are most helpful to you?

Male Student 1: Yeah, the college^[1] stuff and I'll do after I graduate. Cause you know what, I never thought that I was going to go to college, but Joe does and that is really—that makes me see things different^[2] than I used to. So it's really big that he talks about college and not just about my grades and my this and that.

[1] Focus: College as focus of discussions

[2] Effect: See things differently with respect to the future

Questioner: Have you had helpful conversations with a counselor at school?
Tell me about that.

Male Student 2: My counselor is Wendy and I used to have no idea what different kinds of jobs^[3] were that were out there and she had me do this career inventory^[4] that opened up my eyes. It was the most useful thing I did in counseling. Hands down.

[3] Focus: Career as focus of discussions

[4] Method: Written inventory

Questioner: Have you had helpful conversations with a counselor at school?
Tell me about that.

Female Student 1: My counselor's Wendy. The most useful thing she does with me is, last year, when I was having some problems with this group of girls^[5] I used to hang out with and they were hassling me. I was in a bad place. They just were not the right group for me, but I didn't see that and Wendy helped me see that. But it's not like she tells you what you're supposed to do. Instead, she talked me through it^[6] and she, I guess she helped me realize^[7] ...

[5] Focus: Problems with peer group

[6] Method: Open conversation

[7] Effect: Helped realize

Noting Themes

After tracking the themes they identified, the design team asks:

- What are we seeing? Does it affirm what we expected or surprise us?

- What aren't we seeing? What surprises us? Why do you think these surprises arose?

In their notes, they record the following, and will return to these notes when they continue review of additional transcripts:

Focus: Career/college seem to come up repeatedly, especially from male students. Female students mention peer group more often, which does not surprise us.

Methods/Effects: A student mentioned using a career inventory, but more focused on open-ended conversation with their counselor as helpful for arriving at new insights about themselves. This affirms our expectation that students value feeling that they are able to talk to an open, non-judgmental adult.

Surprises: Students have not mentioned values at all—we thought that might come up more. Perhaps students don't think about their conversations with counselors in terms of values, or perhaps counselors are not explicitly articulating the values—work ethic, responsibility, etc.—bound up with conversations about college and career, relationships with friends, and the like.

A variation: Finding an anchor

Another approach to emergent coding is to explore a single theme in-depth across your data set:

Consider that, during their data review, the Lucretia Mott design team noticed that girls matched with women counselors talked a lot about sharing information about their lives, while girls matched with male counselors emphasized talking about their career goals. They thought this was intriguing, so they examined their full data set, pulled out interviews with girls, and categorized them by whether they were matched with male or female counselors. They then skimmed through transcripts, looking for mentions of "career"—color-coded purple—and "life" or "friends" or "family"—color-coded blue.

They made a matrix to see if girls matched to women seemed to be talking about "life"-related topics more than career:

G-W Interview	Girls matched to women	Girls matched to men	G-M Interview
Interview 1	Line 1-3 Line 5-5 Line 15-17	Line 6-8 Line 8-9 Line 12-20	Interview 2
Interview 3	Line 2-4 Line 15-16 Line 18-18	Line 6-7 Line 8-16 Line 27	Interview 4

Interview 5	Line 4-7	Line 6-10	Interview 6
	Line 8-10	Line 13-14	
		Line 30-31	

The visualization above seems to affirm their expectation that girls tend to talk about “life”-related topics with female counselors more than with male counselors. However, as evidenced by Interview 3, even when a theme seems to play out, there will be examples that do not fit.

3. ESTABLISHED CODING

Key Information

Established coding is a more rigorous, less exploratory process. It is used when you:

- Know exactly what you are looking for; and
- Want more confidence in your coding so you can translate your findings into quantitative terms.

Established coding follows a series of steps:

1. Define a coding scheme. The first step in established coding is defining a coding scheme. A coding scheme includes:

- The specific codes expected to be found in the data;
- A definition of the code that makes clear what statements would fall under the code; and
- An example excerpt (added during subsequent steps).

Coding schemes can be generated two ways. First, if you have previously conducted emergent coding of your data, this process can yield a set of codes drawn from the themes you already identified. Second, if you are certain of the categories that will arise while coding—through preliminary research, for example—you can develop the coding scheme yourself.

Codes should be discrete. In addition, at this point—before the coding process has begun—codes should be defined in very general terms. As you code, add examples you find of each code to the scheme. Coding schemes can be revised during the coding process, so this coding scheme is a starting point. Consider starting with three to five codes; you can add more complex codes as you review your data.

2. Double-code. It is best to have two people code independently and then compare what they find for a sub-set of transcripts or artifacts. You would be surprised by how little agreement there can be among coders, even when codes seem relatively straightforward. Once you have generated a coding scheme, if at all possible, identify another person who will code a sub-set of the transcripts or artifacts along with you. While this step may seem duplicative or inefficient, it will help you ensure the reliability of your coding. The step therefore aligns with the Carnegie design

principle of operational efficiency through its purposeful use of time and people. Review the coding scheme with the person you identify.

How many transcripts should we double-code?

One helpful rule of thumb is to code one more transcript than code in your scheme. So, if you have developed a four-code scheme, code five transcripts on the first round of review.

Then, each of you should code the same set of transcripts or artifacts—around five to ten percent of the total amount. When deciding how much to code, bear in mind that you should code just enough transcripts to see a reasonable range of responses, but not so many that if you discover you are coding in significantly different ways, you have to re-code a huge number of transcripts.

After each person has coded the set of training transcripts, they should meet and compare their codes. Each code that you give should be compared to the code the other person gave to the same segment. Talk through each disagreeing code and share with each other what you noticed, especially if it appears that a code should be added. The person responsible for coding the remaining transcripts should ultimately decide on how to resolve a disagreeing code, as their reliability is being established by this double-coding process.

In addition, keep track of the proportion of segments for which you and your fellow coder had initial agreement. Once you have reached at least 80% agreement, you can feel comfortable that each of you has been sufficiently “trained” for solo coding using that coding scheme.

How long will it take for double-coders to reach agreement?

For easier schemes, one session of reflection on double-coding should typically afford sufficient time to arrive at 80% agreement on codes. For more complex schemes, you may need to meet two to three times to arrive at sufficient agreement.

3. Independently code and revise coding scheme. As you independently code data, you may find that different items are falling under the same code. For example, if you have a “talk about friends” code and you find that there are two sub-categories that feel different to you (such as talking about friends positively and talking about friends negatively) you should make a new code and go back through your transcripts to look for that code. When you add new codes, you may need to change old definitions.

4. Prepare for quantitative data analysis. After completing the established coding process, you may want to quantify your data by examining frequency of certain codes, comparing codes across sub-groups, and examining correlations. Our guide on quantitative analysis contains guidance around how to approach this aspect of the work.

Application to Example

Here, we present an illustration of the established coding process through excerpts from student interviews conducted by the Lucretia Mott design team focused on students' relationships with their counselors. At this point, the team has completed the emergent coding process for the previously-described set of transcripts, and now wish to conduct established coding so they can perform some quantitative analysis of our data. Their research question is:

Research question: We plan to have single-sex counseling groups. We expect there to be four topics that these will focus on—family, peers, school, and the future. We want to decide on 2-3 topics under each of these headings that will be of most value to students so that we can use these to anchor our curriculum.

Defining a coding scheme: Based on the emergent coding process, the design team arrives at a first draft of a coding scheme. As they are not yet sure what the sub-topics will be, they begin with codes focused on the four areas identified above. At this point, they do not include examples, since they will add in examples as they begin the coding process.

Code	Definition	Examples
Conversation with counselor about family	A mention in the interview of a conversation with the guidance counselor that is focused on family (the student's family or the counselor's family)	
Conversations with counselor about peers	A mention in the interview of a conversation with a guidance counselor that is focused on non-family peers (friends, acquaintances, fellow students with whom the student has a problem, etc.)	
Conversations with counselor about school	A mention in the interview of a conversation with a guidance counselor that is focused on academic topics (teachers, schoolwork, grades, etc.)	
Conversations with counselor about the future	A mention in the interview of a conversation with a guidance counselor that is focused on future career goals or hopes	

Double-coding. As the team double-codes, they realize that, while they are arriving at consensus around which segments of conversations fall into each category, they also want to split the "future" code into conversations focused on college and career, respectively. They split this code into two separate codes:

Conversations with counselor about college	A mention in the interview of a conversation with a guidance counselor that is focused on the college application process, selecting a college, the importance of college, etc.	
Conversations with counselor about career	A mention in the interview of a conversation with a guidance counselor that is focused on selecting a career, different career paths, etc.	

Independent coding and revision. As the team codes, they identify sub-codes for each broader code, such as:

Code		Definition	Examples
Conversation with counselor about family			
	Conflict with parents	A mention in the interview of a conversation with the guidance counselor that is focused on conflict with parents/guardians	
	Relationships with siblings	A mention in the interview of a conversation with the guidance counselor that is focused on relationships with siblings	