

Qualitative Data Coding

simplypsychology.org/qualitative-data-coding.html

Saul McLeod, PhD

May 17, 2024

Coding is the process of analyzing qualitative data (usually text) by assigning labels (codes) to chunks of data that capture their essence or meaning. It allows you to condense, organize and interpret your data.

A code is a word or brief phrase that captures the essence of why you think a particular bit of data may be useful. A good analogy is that a code describes data like a hashtag describes a tweet.

	INTERVIEWER: So, what do you think of the current strategies to prevent road traffic collisions and neurotrauma in your city and country?	Codes
	PARTICIPANT: You know, the government has instituted things like traffic rules and all that, but these are not being enforced properly. There don't seem to be enough traffic police to cover the many areas in the city where there is a lot of traffic, and I suppose what you could call accident-prone areas. I have read on the internet that many other countries are having things like speed cameras and all that to prevent speeding. I know we have things like speed bumps and maybe the signs which show the speed limit. The problem is people are not following the rules. For that reason, it would be good to have these speed cameras, for example, so that people will realise that some enforcement is being carried out. The other thing is that	Gov. role Lack of compliance Lack of enforcement Accident reduction strategies Lack of compliance Accident reduction strategies

Codes usually are attached to 'chunks' of varying size-words, phrases, sentences, or whole paragraphs. They can take the form of a straightforward descriptive label or a more complex interpretive one (e.g. metaphor).

Coding is an iterative process, with researchers refining and revising their codes as their understanding of the data evolves.

The ultimate goal is to develop a coherent and meaningful coding scheme that captures the richness and complexity of the participants' experiences and helps answer the research questions.

Step 1: Familiarize yourself with the data

- Read through your data (interview transcripts, field notes, documents, etc.) several times. This process is called immersion.
- Think and reflect on what may be important in the data before making any firm decisions about ideas, or potential patterns.

Step 2: Decide on your coding approach

- Will you use predefined deductive codes (based on theory or prior research), or let codes emerge from the data (inductive coding)?
- Will a piece of data have one code or multiple?
- Will you code everything or selectively? Broader research questions may warrant coding more comprehensively.

If you decide not to code everything, it's crucial to:

1. Have clear criteria for what you will and won't code
2. Be transparent about your selection process in research reports
3. Remain open to revisiting uncoded data later in analysis

Step 3: Do a first round of coding

Start identifying preliminary codes which highlight important features of the data and may be relevant to the research question.

- Go through the data and assign initial codes to chunks that stand out
- Create a code name (a word or short phrase) that captures the essence of each chunk
- Keep a codebook – a list of your codes with descriptions or definitions
- Be open to adding, revising or combining codes as you go

First level coding mainly uses these descriptive, low inference codes, which are very useful in summarising segments of data and which provide the basis for later higher order coding.

Descriptive codes

1. **In vivo coding / Semantic coding:** This method uses words or short phrases directly from the participant's own language as codes. It deals with the surface-level content, labeling what participants directly say or describe. It identifies keywords, phrases, or sentences that capture the literal content.
Participant: "I was just so overwhelmed with everything."
Code: "overwhelmed"
2. **Process coding:** Uses gerunds ("-ing" words) to connote observable or conceptual action in the data.
Participant: "I started by brainstorming ideas, then I narrowed them down."
Codes: "brainstorming ideas," "narrowing down"
3. **Open coding:** A form of initial coding where the researcher remains open to any possible theoretical directions indicated by the data.
Participant: "I found the class really challenging, but I learned a lot."
Codes: "challenging class," "learning experience"
4. **Descriptive coding:** Summarizes the primary topic of a passage in a word or short phrase.
Participant: "I usually study in the library because it's quiet."
Code: "study environment"

Step 4: Review and refine codes

Later codes may be more interpretive, requiring some degree of inference beyond the data.

- Look over your initial codes and see if any can be combined, split up, or revised
- Ensure your code names clearly convey the meaning of the data
- Check if your codes are applied consistently across the dataset
- Get a second opinion from a peer or advisor if possible

Interpretive (latent) codes

Interpretive codes go beyond simple descriptions and reflect the researcher's understanding of the underlying (latent) meanings, experiences, or processes captured in the data.

These codes require the researcher to interpret the participants' words and actions in light of the research questions and theoretical framework.

They often draw on existing theories or concepts to interpret the data, providing a more conceptual "take" on what the participants are saying.

Latent codes require the researcher to dig beneath the surface and make inferences based on their expertise and knowledge. Latent coding requires more experience and theoretical knowledge than semantic coding.

For example, latent coding is a type of interpretive coding which goes beyond surface meaning in data. It digs for underlying emotions, motivations, or unspoken ideas the participant might not explicitly state

Latent coding looks for subtext, interprets the “why” behind what’s said, and considers the context (e.g. cultural influences, or unconscious biases).

Example: A participant might say, “Whenever I see a spider, I feel like I’m going to pass out. It takes me back to a bad experience as a kid.” A latent code here could be “Feelings of Panic Triggered by Spiders” because it goes beyond the surface fear and explores the emotional response and potential cause.

It’s useful to ask yourself the following questions:

- What are the assumptions made by the participants?
- What emotions or feelings are expressed or implied in the data?
- How do participants relate to or interact with others in the data?
- How do the participants’ experiences or perspectives change over time?
- What is surprising, unexpected, or contradictory in the data?
- What is not being said or shown in the data? What are the silences or absences?

Theoretical codes

Theoretical codes are the most abstract and conceptual type of codes. They are used to link the data to existing theories or to develop new theoretical insights.

Theoretical codes often emerge later in the analysis process, as researchers begin to identify patterns and connections across the descriptive and interpretive codes.

Examples

1. **Structural coding:** Applies a content-based phrase to a segment of data that relates to a specific research question.

Research question: What motivates students to succeed?

Participant: "I want to make my parents proud and be the first in my family to graduate college."

Interpretive Code: "family motivation"

Theoretical code: "Social identity theory"

2. **Value coding:** This method codes data according to the participants' values, attitudes, and beliefs, representing their perspectives or worldviews.

Participant: "I believe everyone deserves access to quality healthcare."

Interpretive Code: "healthcare access" (value)

Theoretical code: "Distributive justice"

Pattern codes

Second level coding tends to focus on pattern codes. A pattern code is more inferential, a sort of "meta-code."

Pattern codes pull together material into a smaller number of more meaningful units... a pattern code is a more abstract concept that brings together less abstract, more descriptive codes.

Pattern coding is often used in the later stages of data analysis, after the researcher has thoroughly familiarized themselves with the data and identified initial descriptive and interpretive codes.

By identifying patterns and relationships across the data, pattern codes help to develop a more coherent and meaningful understanding of the phenomenon and can contribute to theory development or refinement.

For Example

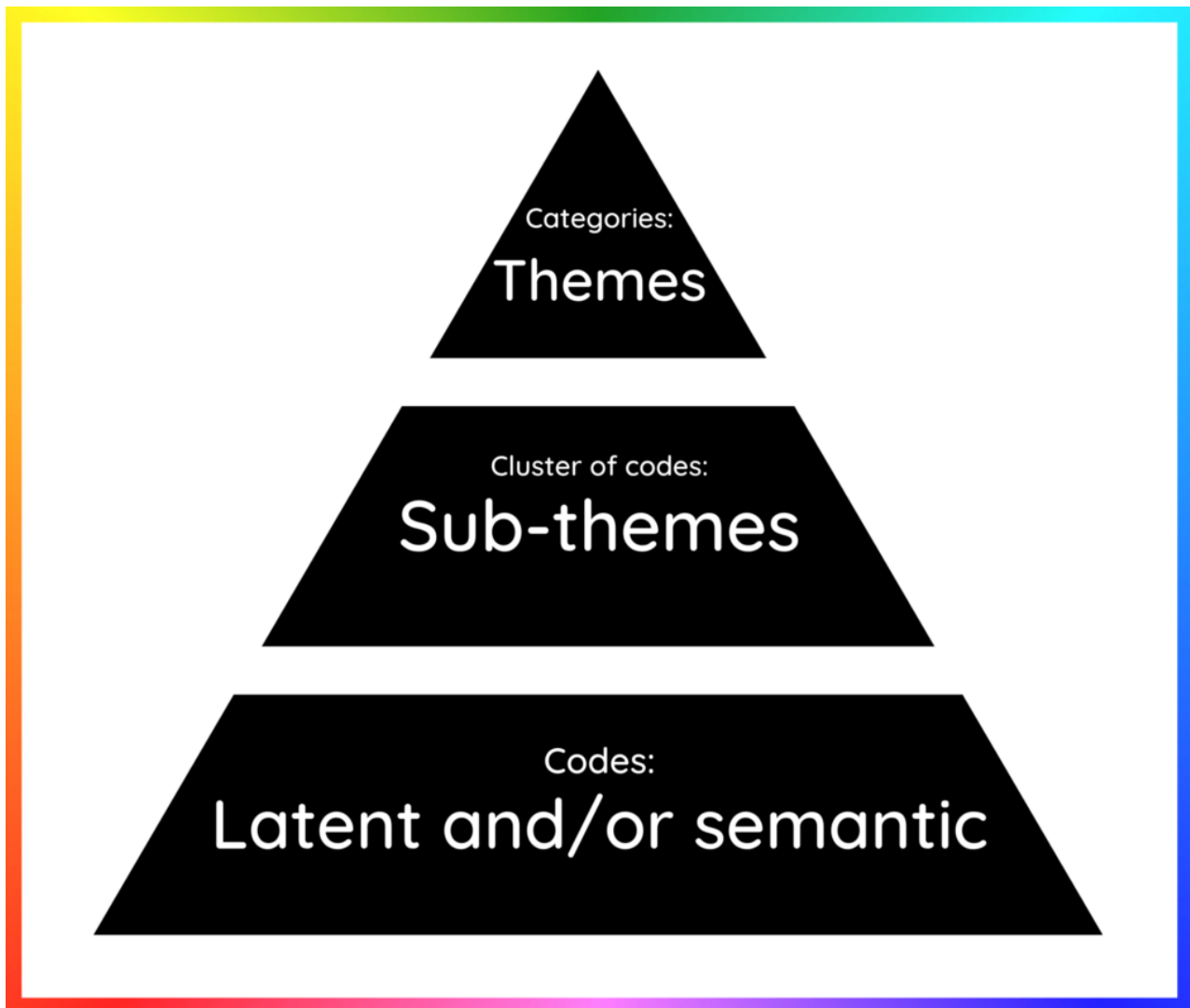
Let's say a researcher is studying the experiences of new mothers returning to work after maternity leave. They conduct interviews with several participants and initially use descriptive and interpretive codes to analyze the data. Some of these codes might include:

- "Guilt about leaving baby"
- "Struggle to balance work and family"
- "Support from colleagues"
- "Flexible work arrangements"

- “Breastfeeding challenges”

As the researcher reviews the coded data, they may notice that several of these codes relate to the broader theme of “work-family conflict.”

They might create a pattern code called “Navigating work-family conflict” that pulls together the various experiences and challenges described by the participants.



Codes are grouped into categories (sub-themes) based on similarities and relationships between them. Categories are then further analyzed and combined to identify overarching themes that capture the essential meanings, patterns, or concepts in the data. This process involves continual refinement, comparison, and abstraction of the categories. Researchers use their interpretive skills to identify the central ideas or recurring motifs that the categories represent, which become the themes that provide a higher-level understanding of the qualitative data.

Bibliography

Auerbach, C., & Silverstein, L. B. (2003). *Qualitative data: An introduction to coding and analysis*. New York, NY: New York University Press.

Bernard, H. R., Wutich, A., & Ryan, G. W. (2016). *Analyzing qualitative data: Systematic approaches*. Los Angeles, CA: SAGE.

Corbin, J., & Strauss, A. L. (2015). *Basics of qualitative research techniques and procedures for developing Grounded Theory* (4th ed.). Los Angeles, CA: SAGE.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). London, UK: SAGE.

Saldaña, J. (2009). *The coding manual for qualitative researchers*. London: Sage Publications

© 2025 Simply Psychology