09 Introduction to Qualitative Data Analysis

1. Qualitative Data

Most data in qualitative research are spoken or written words. For example, one may ask participants to describe their experience attending integrated schools for the first time during the late 1960s and early 1970s (spoken data); or, one ask questionnaire respondents to describe which aspects of their job they found most frustrating (written data). Of course, there are other forms of data, but they too are often converted to words for analysis. The question, then, is how does one analyze such data?

2. Generic Steps for Qualitative Data Analysis (QDA)

LeCompte (2000) likens QDA to assembling pieces of a jigsaw puzzle.

- Many pieces to the puzzle – the raw text of responses to open-ended items
- Sort pieces into common piles – read responses and identify common responses
- Form themes of puzzle (e.g., sky, barn, water, flowers) – do the same for responses (e.g., anxiety, confidence, frustration)
- Find linking pieces of puzzle to connect themes – determine how response themes relate (e.g., when I experience frustration and I also experience anxiety)

2a. Data Preparation

Since most qualitative data are in the form of words, it is important that interviews, field notes, documents, etc. be transcribed and recorded in such a way that can be easily accessed and read.

First note that data analysis in qualitative research is often cyclical and may, perhaps should, begin once data collection commences. The cycle of collecting data and analyzing data during the data collection phase is known as interim analysis (analyzing data during the interim while data collection continues). Beginning data analysis early can help identify important themes or areas that should be explored.

At this initial stage researchers should read all their data carefully, and then re-read, then repeat again (and again). Why? The more familiar researchers are with their data, the more easily they can begin spotting or identifying important concepts in those data and see connections between concepts. With each reading researchers should record their impressions of the data, record their thoughts and interpretation of the data. These recordings will help build one's memory and provide insight when sorting/collecting data into broad categories and concepts.

LeCompte (2000, p. 148) suggests one use the following in preparation for QDA (if not using computer analysis systems):

- Make copies of all data so none is lost or ruined when memo-ing (adding researcher comments/notes to data)
- Put all notes and interviews in files by date of creation
- Create other files based on
    - types of data (e.g., interviews, questionnaires, field notes, artifacts),
    - participants (e.g., students, teachers, staff),
    - organizations (e.g., health agencies, foundations, schools)
    - subject or topic (e.g., recruitment of students, parent involvement);

    - do the above based upon needs and what seems reasonable.
- Catalog and store all documents and artifacts
- Label all files and boxes according to their contents.
- Create index or table for all contents for all data.
- Review research questions comparing them against data collected to ensure each question is addressed.
- Identify holes in data collection and address missing data so research questions can be answered.
- Collect additional data if needed.

2b. Develop Initial Codes and Code Data

At this stage the researcher will begin coding data; this means labeling relevant or important data points with unique labels to help separate data into unique and meaningful components. The researcher, when coding, is attempting to identify key ideas, behaviors, interactions, incidents, and terminology/phrases available in the data. In short, coding is labeling or naming things found in one's data.

Codes used for labeling data may be derived in several ways:

> **Deductive/A priori/Preset Codes** – Researcher develops a classification scheme of codes prior to collecting data. This approach may not allow important new information to be identified; probably few qualitative researchers employ this approach although can be a useful approach is one is interested in theory testing.
>
> **Inductive/Post hoc/Emergent Codes** – Codes for classifying data are developed while reading and coding the data. This approach allows data to speak and potentially enables the richness of the data to be revealed.
>
> **Mixed Preset and Emergent Codes** – This approach represents a combination of the two in which researchers develop an initial classification scheme with codes, but adds to these codes as new information is learned. Likely a common approach for many researchers.

Coding data and developing codes is an iterative process and requires much time and effort. When data from multiple interviews or long interviews are used, one can expect this coding process to last many hours or even days. In some types of studies (e.g. grounded theory), one does not stop collecting data until a saturation point is reached, which means collecting additional data provides little or no new information. One won't know this unless coding occurs simultaneously with data collection.

LeCompte (2000, p. 148) writes that researchers usually use three approaches to identifying things to code or name:

- Frequency – items are coded because they appear often (e.g., how many students expressed some form of anxiety, or how many students indicated the instructor is disorganized)
- Declaration – items are important because participants tell us they are important (e.g. students tell us the instructor's videos were very helpful)
- Omission – something expected did not occur, why and what does this mean (e.g., students never mention being assessed or tested); this approach probably only works when using some frame of reference to set expectations

Code Sheet Example:

http://www.bwgriffin.com/gsu/courses/edur9131/activities/sample_code_sheet_open_ended_authorship.pdf

This sheet was printed twice for each returned for questionnaire, and used by two coders separately and independently to code responses. Once completed, both were attached to the questionnaire and then responses were compared to assess inter-coder agreement levels.

Coded Examples:

http://www.bwgriffin.com/gsu/courses/edur9131/activities/open_ended_coding_example_1140.pdf

http://www.bwgriffin.com/gsu/courses/edur9131/activities/open_ended_coding_example_420.pdf

2c. Organize Data into Categories

At this stage most data will be identified via codes (although the process is iterative so new codes may be identified still or data may be labeled or relabeled with existing codes), so now the process of combining like codes into categories begins. Here one attempts to identify redundancies in codes and create subsets of codes to form broader categories of data. This reduction process helps to bring meaning to data; it allows one to more succinctly grasp key ideas found in the data.

One approach to determine unique codes is to compare and contrast data, and to sort items (units of data) into similar and dissimilar groupings.

2d. Further Refinement: Categories to Themes/Concepts/Taxonomies

In many cases one will be able to organize categories into still boarder themes/concepts. Sometimes this may not be possible, or categories may be themes/concepts (the two overlap). The notion, however, is that if there are many categories of data, it may be possible to further combine these into more general concepts that better reveal important information or meaning in the data. At this point some categories may be discarded as unimportant or because these categories provide little relevant, helpful information for telling the story of this research.

2e. Find Relations among Concepts and Categories/Themes/Taxonomies

Often one may be able to identify how various themes interrelate for study participants and researchers. This can lead to significant meaning and reveal important findings.

2f. Displaying Results

Textual Display
Most qualitative researchers present results in textual format; they describe the study setting, their perspectives and biases, summary of what they found, and often supplement this with quotations. I illustrate some of this below in "**3. Illustrated Example of Data Analysis**" especially in **3d** and **3e**.

Graphical Display
Sometimes textual results are also coupled with graphical displays. Kodish and Gittelsohn (2011) present a graphical display of data results from QDA showing linkages found in diabetes study.

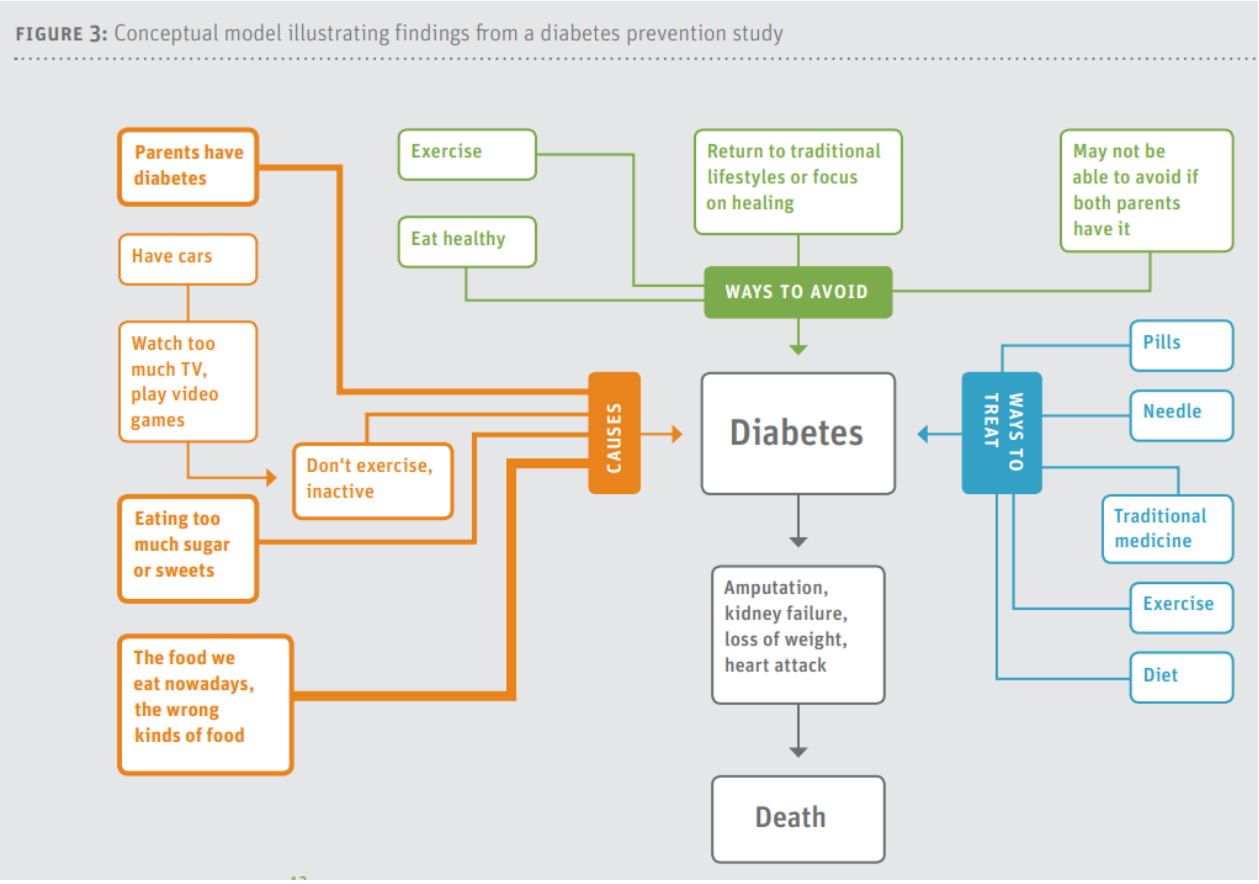
In this example there are Categories which form Themes:

Categories –

- Items displayed on the outside, examples
  - Parents have diabetes
  - Exercise
  - Pills

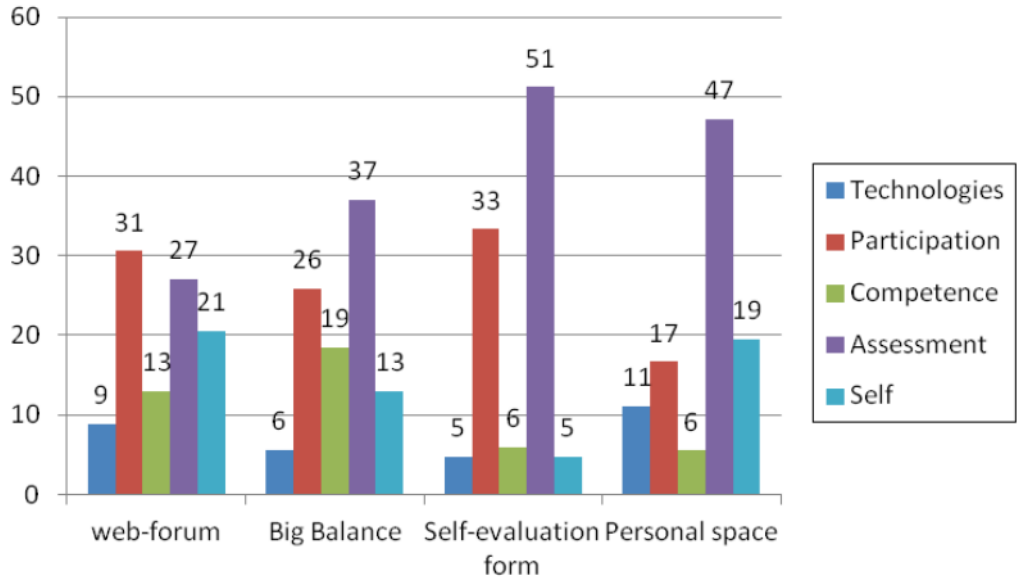
Themes –

- Causes
- Don't Exercise/Inactive
- Ways to Avoid
- Ways to Treat



Impedovo, Ritella, and Ligorio (2013) provide the following bar chart showing frequency of themes for different sections of e-portfolios examined. The X-axis contains four sections of the e-portfolio and the labels to the right are the data themes.

Figure 2
Distribution of the Themes Through the Sections of ePortfolios



Tabular Display

Another approach to displaying data is in tables. Below is an example from Moore and Griffin (2006) who asked participants to identify the benefits of co-authoring research.

Table 2: Perceived Benefits of Coauthored Publications

	Percentage of Respondents [1]	Number of Times Category Referenced [2]
Quality of Work/Ideas	65.0 (39)	
Diversity of Perspective in Work/Ideas		20
Clearer Thinking/Stronger Presentation/Better Written Work		17
Coauthor Peer Review of Work/Ideas		9
Other Quality of Work/Ideas		4
Synthesis of Ideas		3
Division of Labor/Workload	41.7 (25)	
Synthesis of Specialist Skills/Complementary Contributions of Authors		16
Shared Responsibility		2
Other Division of Labor/Workload		9
Collaboration	38.3 (23)	
Sharing of Ideas		8
Builds Community among Academics/ Interaction Among Colleagues		5
Emotional Support		4
General Enjoyment of Collaboration		3
Enables More Extensive Research		2
Motivation to Complete Task		2
Other Collaboration		5
Professional Development	30.0 (18)	
Mentor Novice Writers		9
Learn from Experienced Professionals		5
Enhanced Vita with Less Work		4
Other Professional Development		2

Note: The "Other" category of responses represents responses that could be classified into a given main grouping (such as Professional Development, Collaboration, etc.), but could not be determined to fit within one of the sub-categories for that grouping.

[1] Numbers in parentheses indicate the number of respondents out of 60 who provided a response that fit within a main grouping, e.g., 18 respondents indicated that some aspect of "Professional Development" was used to determine coauthorship.

[2] This column is a simple count of the number of times a specific reason was given for recognition of coauthorship. This column may sum to more than 60 since multiple reasons were often listed by each respondent.

3. Illustrated Example of Data Analysis

Below is an example showing how data from an interview may be coded to identify important concepts.

3a. Data Preparation

The interview between a researcher and teacher is transcribed and presented below.

Interviewer:

> "Please tell me what it was like to work under your previous principal, and how is it different with your current principal."

Teacher:

> "Wow, the difference is like night and day. Mr. Sykes was so controlling. He had to approve . . . like everything we did. If I wanted to try something new, maybe a new computer program with my students, I had to get his permission. Uh, I remember once . . . I wanted to try a new workbook that I downloaded from the internet. One day he was observing me teach, and he asked about it in kind of a disapproving way. I don't know, but it just seems he was so oppressive and didn't want us to try new things. I kind of lost interest in teaching when he was here."
>
> "Now with Mr. Rosen things are so different. When he first arrived, he said he wanted us to explore new ways of teaching, you know, to try different things in the classroom. I wasn't sure I believed him. So, I asked him about using some stuff I found on the web and he said 'Sure, go for it and let me know if it works.' So, I did and now I am constantly trying new things to help my students. It is much more exciting now to be in the classroom because I can teach the way I want."

3b. Develop Initial Codes and Code Data

Using an emergent design, note following codes within brackets [ ] and highlighted in yellow were added to the teacher's transcribed response. This is known as open-coding (reading through data and applying/developing codes).

> "Wow, the difference is like night and day. Mr. Sykes was so controlling. [authority control] He had to approve . . . like everything we did. [authority approval] If I wanted to try something new, maybe a new computer program with my students, I had to get his permission. [authority permission] Uh, I remember once . . . I wanted to try a new workbook that I downloaded from the internet. One day he was observing me teach, and he asked about it in kind of a disapproving way. [authority questioning] I don't know, but it just seems he was so oppressive and didn't want us to try new things. [authority oppressive] I kind of lost interest in teaching when he was here. [teacher interest lost]"
>
> "Now with Mr. Rosen things are so different. When he first arrived, he said he wanted us to explore new ways of teaching, [authority explore] you know, to try different things in the classroom. [classroom experiment] I wasn't sure I believed him. [teacher disbelief] So, I asked him about using some stuff I found on the web and he said 'Sure, go for it and let me know if it works.' [authority approval/flexibility] So, I did and now I am constantly trying new things to help my students. [teacher explore] It is much more exciting now to be in the classroom because I can teach the way I want. [teacher excited/freedom]"

3c. Organize Data into Categories

Below I attempt to organize the codes identified above into categories. Note there is some overlap and also one code is not used, [teacher disbelief], because it does not seem relevant to the categories that are emerging.

Authority

- authority control
- authority approval
- authority permission
- authority questioning
- authority oppressive
- authority approval/flexibility

Autonomy

- teacher explore
- authority approval/flexibility
- authority control

Motivation (my interpretation given data, but not said)

- teacher interest lost
- teacher excited/freedom

3d. Further Refinement: Categories to Themes/Concepts

As I think about this teacher's responses and descriptions, and review those considering the codes and categories presented above, it seems there are two general concepts presented in these data:

> **Teacher Autonomy Support** – In what ways does this teacher believe she is supported to be autonomous in her classroom? In what ways does she believe her level of classroom autonomy is being suppressed or controlled?
>
> **Teacher Motivation** – While this is a category identified in the previous step, I think it also represents an important concept for this teacher. Does this teacher demonstrate motivation or de-motivation to teach in her classroom?

3e. Find Relations among Concepts and Categories

Given this teacher's responses it appears that the level of autonomy support offered by the principals directly influenced her motivation to teach. As evidence, consider her statements:

> "Mr. Sykes was so controlling. He had to approve . . . like everything we did."
>
> "I kind of lost interest in teaching when he was here."
>
> "Now with Mr. Rosen things are so different. When he first arrived, he said he wanted us to explore new ways of teaching, you know, to try different things in the classroom."

"It is much more exciting now to be in the classroom because I can teach the way I want."

4. Codebook

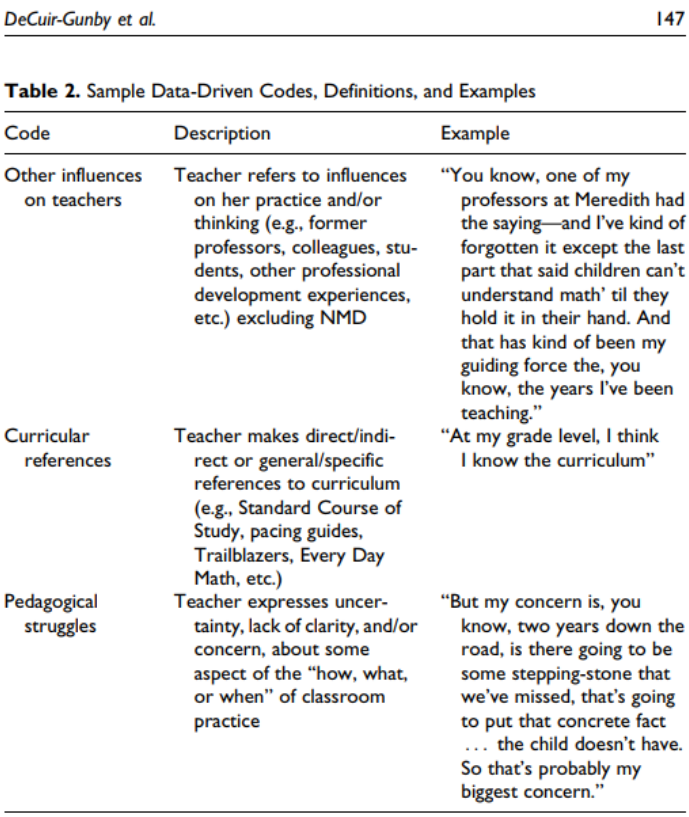
When developing codes, it is important to create a codebook to help clarify what each code represents. The codebook should contain at least three things:

- Code (e.g., Authority Control)
- Description of Code (e.g., Authority Control – individual in authority exerts control by …)
- Example of Code (e.g., Provide quotation from data that illustrates Authority Control)
- Parameters for Use (e.g., this explains when it should and should not be used; Authority Control is not applicable if someone who is not an authority attempts to control a teacher's behavior such as a peer teacher.)
- Number or Abbreviated Letters (shorthand coding, e.g., Authority Control = AC or 3.10, etc.) – this simply makes coding large chunks of data faster as compared with writing the code label repeatedly.

Such a codebook helps make clear to all involved in analyzing data the precise definition of each code, and this helps with increase inter-coder agreement when multiple people are involved in the coding process.

Codebook development can be time intensive because it often involves multiple revisions as codes are often changes throughout the coding process.

Below is an example from DeCuir-Gunby, Marshall, and McCulloch (2011).



DeCuir-Gunby et al. 147

Table 2. Sample Data-Driven Codes, Definitions, and Examples

Code	Description	Example
Other influences on teachers	Teacher refers to influences on her practice and/or thinking (e.g., former professors, colleagues, students, other professional development experiences, etc.) excluding NMD	"You know, one of my professors at Meredith had the saying—and I've kind of forgotten it except the last part that said children can't understand math' til they hold it in their hand. And that has kind of been my guiding force the, you know, the years I've been teaching."
Curricular references	Teacher makes direct/indirect or general/specific references to curriculum (e.g., Standard Course of Study, pacing guides, Trailblazers, Every Day Math, etc.)	"At my grade level, I think I know the curriculum"
Pedagogical struggles	Teacher expresses uncertainty, lack of clarity, and/or concern, about some aspect of the "how, what, or when" of classroom practice	"But my concern is, you know, two years down the road, is there going to be some stepping-stone that we've missed, that's going to put that concrete fact . . . the child doesn't have. So that's probably my biggest concern."

Kodish and Gittelsohn (2011) present the following example of a codebook.

TABLE 1: Excerpt of a codebook developed for inductive QDA of data collected at Kakuma Refugee Camp, Kenya.

Mnemonic or numeric "Brief" Code	Full Description of Code	When to use/not to use the code
2.0 Life in Kakuma	Refugee experiences residing in KRC	Use this family of codes when the CL or MNP beneficiary discusses his or her life as a refugee at KRC.
2.1 Hardship	Hardships faced while living in Kakuma, related to security, violence, tribalism, etc	Use this code for the array of hardships refugees discuss at KRC unrelated to illness experiences. Illness is mentioned a lot but use 2.2.
2.2 Illness	Illness experiences of the individual or of his or her family and/or community	Use this umbrella code for any health-related experience related to life in KRC. It can be related to anemia or another illness. Codes 2.2.1 and 2.2.2 will be used to distinguish between the types of illness discussed.
2.2.1 Ill. Anemia	Experiences with anemia or malnutrition, specifically	Use this code for health-related experiences, in particular those related to anemia and/or malnutrition. Also, "lack of blood" should be included in this code as it's referring to anemia.

Source: Adapted from one of the authors' projects

Below is a section of the codebook Kilby (2014) developed for a review of general practitioners in Norway.

SN	Code	Description	When to use
1.0	Prior	Reviewer's experiences with the GP and their offices prior to the consultation(s)	Use for references to interaction with other staff at the GP's office, for example receptionists, the office environment itself, including the waiting room, or any factor directly related to the patient's experience prior to the consultation
1.1	Convenience	Reviewer experiences convenience of factors or activities associated with the offices and consultation	Use for references to convenience of activities such as booking appointments and waiting times, office parking etc.
1.2	Inconvenience	Reviewer experiences inconvenience of factors or activities associated with the offices and consultation	Use for references to inconvenience of activities such as booking appointments and waiting times, office parking etc.
1.3	Clinic staff	Reviewer's experience in interacting with other staff at the clinic	Use for references to receptionists, secretaries, and any other staff at the GP's clinic apart from the GP
1.4	Availability	Reviewer's experience with trying to access the clinic or book a timely appointment	Use for references to how long patients must wait before an appointment with their GP is available, whether they can get an appointment on short notice, and to perceptions of the GP's general availability
1.5	Waiting times	Reviewer's comments on waiting times at the clinic	Use for references to time spent in the waiting room before a consultation with their GP, with or without an appointment
1.6	Phone service	Reviewer's experiences in contacting the clinic via telephone	Use for references to the degree of in/convenience in contacting the clinic by phone, being put through to receptionists or receiving recorded messages, waiting times on the phone etc.
2.0	Consultation	Reviewer's experiences with their face-to-face consultation(s) with the GP	Use for references to the reviewer's experiences during their consultation/s with their GP, including the reviewer's perception of their GP's interpersonal manner, communication skills, and clinical competence
2.1	Personality factors	Reviewer highlights certain positive or negative personality traits of their GP	Use for reviewer references to their GP's personality traits, both without context, and to explain how such qualities affected their satisfaction with the consultation
2.1.1	Sympathetic/empathetic/ compassionate	Reviewer's experiences with their GP's sympathy, empathy, or compassion	Use for direct patient references to their judgement of their GP's sympathy, empathy or compassion, either in specific cases or as a general character trait
2.1.2	Understanding	Reviewer feels that their GP does or does not understand them	Use for references to the GP's ability and effort in understanding the patient's feelings or medical concerns, as well as whether or not the reviewer feels understood by their GP

References

DeCuir-Gunby, J.T., P.L. Marshall, A.W. McCulloch. (2011). Developing and using a codebook for the analysis of interview data: an example from a professional development research project, Field Methods 23, pp. 136–155.

Impedovo, M.A., Ritella, G., & Ligorio, M.B. (2013). Developing codebooks as a new tool to analyze students' ePortfolios. International Journal of ePortfolio, 3, 161-176.

Kilby, G. (2014). A qualitative content analysis of onlne GP reviews in Norway. Master's Thesis, Department of Health Management and Health Economics, University of Oslo.

LeCompte, M.D. (2000). Analyzing qualitative data. Theory Into Practice, 39(3), 146–154.

Moore, M. T., and B. W. Griffin. 2006. Identification of Factors that Influence Authorship Name Placement and Decisions to Collaborate in Peer-Reviewed, Education-Related Publications. Studies in Educational Evaluation 32 (2): 125–35.

Stephen Kodish & Joel Gittelsohn (2011). Systematic Data Analysis in Qualitative Health Research: Building Credible and Clear Findings. Sight and Life.