(How to …) Analyse qualitative data.

Susan Jamieson

Director of the Health Professions Education programme at the University of Glasgow, UK.

Wolfson Medical School Building (R307),

School of Medicine, Dentistry & Nursing,

College of Medical, veterinary & Life Sciences,

University Avenue,

Glasgow, G12 8QQ.

Tel: +44-141-330-8488 (direct)

Email: susan.jamieson@glasgow.ac.uk

Key words: qualitative data analysis, thematic analysis, textual data, methodology

Qualitative research and qualitative data.

For many who embark upon medical education research, this is often through small-scale studies in their local context, possibly as they undertake masters-level study in medical education. In this case, the masters' dissertation project may be their first foray into qualitative research, which is research that aims to explore phenomena, or understand social situations; as opposed to quantitative research, which is concerned with quantification of phenomena, and correlational or causal relationships between entities. Qualitative data is most commonly taken to mean text: that is, written documents or type-written representations of the spoken word. Pictures, videos, art work and music are some other forms of qualitative data, and interested readers might like to browse the Museum of Qualitative Data website (see Resource Box), but the focus of this article is qualitative analysis of textual data.

Sources of textual data.

Obviously, before you can analyse data, you need to collect it. For novice researchers, the data collection methods of choice are frequently semi-structured interviews or focus groups, each of which generates textual data, and each of which is popular with the wider medical education research community. Other sources of textual data include written assessments, narratives, discourse, reflective logs, field notes, documents, and 'think aloud', where the participant is asked to recite what is going through his or her mind in a specific situation. Sometimes 'think aloud' is followed by 'stimulated recall', where the participant is asked to recite retrospectively what she or he was thinking at a specific moment, and they may be prompted to do this by the interviewer or, for example, by reviewing a video of a social situation, such as a consultation. Each of these methods gives rise to written text, or audio files that may be transcribed into written text.

Issues with qualitative data analysis.

Several issues may contribute to the difficulty new researchers experience in undertaking qualitative data analysis, and these include the practicalities of how to begin; the terminology; and the specific purpose of the research (see Table 1).

Regardless of how you arrive at having a body of text to analyse, you need to think in advance about how you are going to analyse that data, for reasons that will become apparent. However, it is difficult to find clear, practical advice on how to go about qualitative data analysis. This is exacerbated by the fact that the process tends not to be well-described in primary research articles. Possibly, this lack of methodological detail in published articles has contributed to the perception that qualitative research is easy and necessarily lacking in rigour. Neither is the case!

Whilst several good resources exist, including textbooks, methodological articles, and even YouTube videos (see Resource Box), they often use terms that are not well-defined, or they may use different terms for the same entity. For example, in referring to ideas suggested by the data, some authors describe these as themes, whilst others describe them as categories.

Another issue with qualitative data analysis is that the method you use may depend on the purpose of your research, and hence your purpose in undertaking the data analysis. Qualitative data analysis may be used to test a hypothesis or model, to understand phenomena, or to generate new theory, which we define here as a high-level explanation for a social phenomenon. Normally, we might associate hypothesis-testing with quantitative research, but it is possible to form hypotheses about social phenomena and test these using

qualitative methods; for example, analysis of discourse during a problem-based learning (PBL) tutorial could be used to test a hypothesis that mature students are more likely to apply critical thinking to a problem, than school-leavers. Most commonly, in qualitative data analysis we are seeking to explore or understand phenomena and social situations; for example, we might analyse narratives to understand how specific health professionals developed their professional identity, or we might analyse interviews to elicit student perceptions about the clinical learning environment.

Approaches to qualitative data analysis.

Your approach to data analysis reflects the extent to which you come to the process with pre-conceived ideas, or whether you aim to identify new ideas and explanations. To some degree, the purpose of your data analysis dictates the most appropriate approach. Table 1 shows approaches that - arguably - might best align with specific purposes. [Insert Table 1 here]. These approaches have specific philosophical implications, and their use potentially tells us something about the researchers' views on the nature of knowledge (their epistemology) and on social reality (their ontology).

Content analysis.

Content analysis was first used by newspaper advertisers in the 1940s; when they wanted to find out what interested a particular readership, they simply counted references to specific topics. This form of qualitative data analysis has been called quasi-quantitative, because it involves counting, but it is qualitative in the sense that it is working with text and aiming to understand what is meaningful to people. It is a deductive approach to data analysis, meaning

that the researcher has a preconceived idea about what he is seeking in the data and looks for evidence that this is present.

Thematic analysis.

Thematic analysis is probably the most common form of qualitative data analysis, especially for those aiming to explore or understand phenomena and social situations. In thematic analysis, you seek to identify basic ideas about the phenomenon under study, or the situation, and place your own interpretation on the meaning of participants' words. Thematic analysis is generally regarded as inductive, meaning that themes arise from the data, but often thematic analysis may be a mixture of deductive and inductive. So, for example, a researcher who is seeking to understand colleagues' perceptions about professionalism may be aware of themes reported in the literature and may look for evidence of these in his or her data, but will likely be interested to identify novel themes, or contradictions. Possibly, because researchers often combine deductive and inductive data analysis, this contributes to the perceived 'murkiness' of the methodology.

Grounded theory.

Arguably the most well-known form of qualitative data analysis is grounded theory, in which themes *and* theory are generated inductively, from the data. However, grounded theory is considered difficult to do well, and is potentially very time-consuming, because of the need for data saturation, which refers to the situation where no new themes arise when the data set is expanded (e.g., when more interviews are undertaken). Grounded theory is generally regarded as out-with the scope of masters-level projects, and not advised for newcomers to medical education research.

Another term that you may come across is Framework Analysis [1,2], but rather than an implied philosophical approach, this is effectively a set of practical steps to carry out qualitative data analysis in a reproducible fashion. This data analysis method was devised to inform rapid development of social policy. It is particularly popular with some health professionals, possibly because aspects of the method contribute to the perceived reliability - or dependability - of the findings; thus specific steps are followed and the data is coded by multiple researchers. For those interested in reading further, see the worked example by Ward et al [2].

The process of qualitative data analysis.

Because of the likelihood that novice medical education researchers will choose to undertake exploratory studies and hence use thematic analysis, this is our focus in considering the process of qualitative data analysis. Figure 1 shows five steps in the process: data preparation, immersion in the data, coding, generation of themes, and abstraction to over-arching themes. Some sources give five steps [3]; others give four, including the excellent YouTube videos by Graham Gibbs (see Resource Box), which he acknowledges as based on Bryman's classic text on social research methods [4]. However, the process these and other authors describe is essentially the same.

[Insert Figure 1 here]

Data preparation.

Data preparation is the term used to describe the conversion of written text or audio files into a form that you can work with. In terms of written text, an example might be a handwritten dairy, which would need to be word-processed to facilitate data analysis. Where audio-

recorded data is collected via interviews and focus groups, the data must be transcribed, which means that a word-processed document is created from close listening to the audio-recording. It is possible to buy software that will slow down the playback of audio files, or you may have funds to pay for secretarial support to have files transcribed. Arguably, it is better to transcribe the data yourself, since this helps to familiarise yourself with the data, but it can be very laborious, and it may be a conservative estimate that one hour's worth of recording can take ten hours of typing time.

Data preparation also includes the formatting of text and other aspects of editing. For example, Ruona [3] describes the use of word-processing software to format the text into tables that can subsequently be used in the data analysis process. If you plan to use software such as NVIVO to organise and analyse your data, you will need to upload the prepared transcripts.

If you pay for transcription, it may be necessary to correct inaccuracies in specialist vocabulary; an example from my own work with medical students talking about biomedical aspects of PBL scenarios, was when the word 'interleukin' was regularly transcribed as 'interlocking'. If you know your data is likely to contain technical terms with which the transcriber will be unfamiliar, it may be worthwhile providing them with a glossary. This potential difficulty with vocabulary is of course another reason for transcribing your own data. A further aspect of editing may be to remove the 'uhms' and 'ahs' from the text; however, although it is beyond the scope of this article, you should be aware that for certain methodological approaches, such as conversation analysis, these are an important part of the data.

A final, important aspect of data preparation is to anonymise the data by giving the participants individual code names or pseudonyms, and creating a separate, secure document linking these codes names or pseudonyms to actual participants. Ruona [3] recommends italicising the interviewer's text, and also emphasises that it is critical to keep a master copy of the transcript.

Immersion in the data.

How do you achieve immersion in textual data? Read, read and read it again. And at the same time, *think* about your data. In fact, your initial familiarity with the data may come from having been present during data collection, if you were the interviewer. Or you may choose to listen to audio-recorded data before reading the transcript. If you transcribe audio files yourself, you will almost certainly become very familiar with the data, from playing it back to check the accuracy of your transcription.

During this step, you should form a general impression of themes contained in the data, which may be themes you expected to find, or novel themes. For grounded theory, the expectation is that you should come to the data with no pre-conceived ideas, but for thematic analysis, you may expect certain themes based on prior reading of the literature; or from previous research, such as a survey on the same topic; or from your experience. In this last case, you must be careful to avoid introducing bias. Since it is generally recommended that you begin qualitative data analysis before you have finished data collection (except in the case of some authors writing about Framework Analysis) it is likely that for later transcripts you will seek and recognise themes that arose in early transcripts.

Coding.

Coding is when you begin to mark up your transcript, and this is when you really begin to analyse your data: you break it down and looking for meaning. A code is a label or tag that relates to a particular theme (otherwise called a category); a code therefore ascribes meaning to the coded text. The entity that you choose to code - the word, phrase, sentence or paragraph - is called the unit of meaning, or unit of analysis. You should have a firm rationale for choosing a particular unit of analysis. For example, in my own research into critical thinking in the context of a PBL tutorial, I coded discourse on the basis of often lengthy, successive utterances by group members, since I was interested in the context in which statements were made; and in the impact of what one student said, on the thinking of another.

How do you know what to code? You should look for something that makes sense to you, in relation to the research questions or topic. Ryan and Bernard [5] offer excellent suggestions for identifying themes, including repetitions, since this implies importance; indigenous typologies, which means categories that people use locally, including slang; metaphors and analogies; transitions from one speaker or topic to another; similarities and differences; linguistic connectors, which are words that are meaningful to the research topic – an example would be the coding of 'because' as an indicator of hypothesis formation; missing data; and theory-related data.

As new themes are identified in later transcripts, you should go back and re-code initial transcripts to look for evidence of the new themes: this is termed 'constant comparison'. Thus your coding scheme, or final list of codes, will evolve over time. One final activity you may undertake during coding is memo'ing, which is essentially writing down your thought

processes during coding: why you made certain coding decisions, and ideas that occurred for themes. Sometimes memos are used as a source of textual data in their own right.

From a practical point of view, some find software such as NVIVO to be immensely useful in the coding process. However, independently from Ruona [3], I developed my preferred practice of tabulating utterances in a Word document, and writing the code and the rationale for the coding decision in adjacent columns. Text can be highlighted in different colours, to illustrate related themes visually.

Generation of themes.

Now comes the step at which you aim to reduce your data by grouping codes into themes, or categories. You are also analysing the data at a deeper level, looking for patterns and relationships between themes, and interpreting the data, often in relation to the literature. For example, in discourses from small group tutorials, you may have coded text that exemplified 'questioning' and 'summarising', and these could be combined into a theme of 'facilitation skills'; reference to the literature would suggest additional facilitation skills, prompting you to revisit your data and code additional text to this theme.

Abstraction/generation of theory.

You may go on to achieve further abstraction by synthesising your findings; or you may generate an explanatory theory, although this is particularly specific to a grounded theory approach.

Putting it into practice.

In teaching these analytical skills to masters' students, for ethical reasons I am unable to let them practise with actual research data. Rather than prepare dummy transcripts, I generally direct students to a series of articles written by people with dementia, or their carers (see Resource Box), and ask them to use these articles in lieu of interview transcripts, to answer the research question: 'What are the issues involved in having dementia, or in caring for someone with dementia?". The students undertake thematic analysis and compare derived codes and themes with each other and myself. You might like to try this for yourself.

In terms of undertaking qualitative data analysis 'for real', there are some further practical points to consider. Timing of data analysis, and materials for data analysis, such as specific software, have been discussed already. Other issues are the amount of data and whether coding should be an individual or team activity. Masters' students from a clinical background often worry about the optimal number of interviews; whilst sample size or composition may be important in terms of answering specific research question(s), in terms of data analysis, it is the *volume* of data that is the main consideration. A one-hour interview can results in pages of transcript, and take hours or days to analyse. Regarding whether coding should be an individual or team activity, there are alternative perspectives. Thus in content analysis and Framework Analysis, great emphasis is placed on inter-rater reliability and having the team confirm one another's coding decisions; but naturalists might argue that interpretations are likely to be subjective and there should be no expectation of agreement.

Establishing rigour.

So how *do* you ensure that your data analysis is rigorous, or trustworthy? You can employ measures that are well-described in social [4] or education [6] research methods texts. Ruona [3] suggests that you should be concerned about three key aspects: ways in which your approach to qualitative data analysis ensures the credibility, dependability and transferability of your findings.

Credibility.

Credibility is similar to the concept of internal validity, and may be assisted by the process of 'member-checking', whereby participants are asked to confirm the authenticity of transcriptions and/or derived themes; an alternative is peer review, where coding decisions are justified to experienced researchers.

Dependability.

Dependability is similar to the notion of reliability in quantitative research, but in qualitative research it may be achieved by measures such as the use of an interview schedule; by clear examples of codes; and by reflexivity, which is where the researcher is open about his or her influence on the research, and vice versa.

Transferability.

The contextual, often small-scale nature of qualitative research means that its findings cannot be generalised. However, they may be more or less transferable to, or applicable in, other contexts. This is ultimately for the reader to decide, but you can aid them by providing clear

descriptions of your methodology and your decision-making process during coding; the latter is supported by memo'ing and providing an audit trail (what you did, when you did it, why you did it).

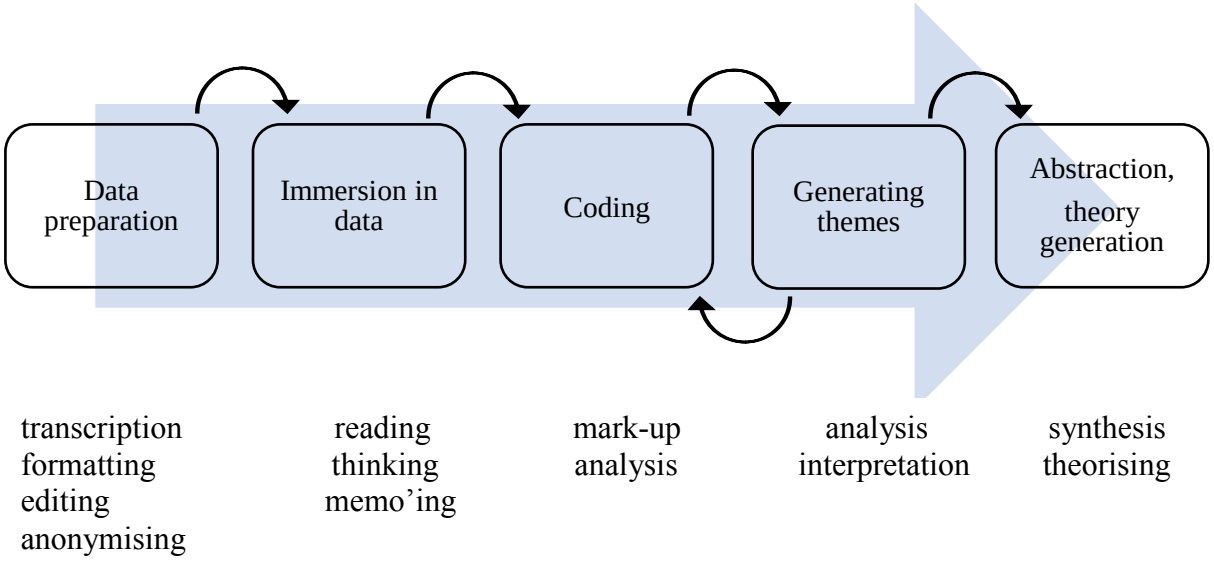
Ultimately, there is no single way of undertaking qualitative data analysis, but the process described and the tips and resources provided should encourage you to test the waters!

References.

1. Smith J, Firth J. Qualitative data analysis: the framework approach. Nurse Res. 2011; 18:52-62.
2. Ward DJ, Furber C, Tierney S, et al. Using framework analysis in nursing research: a worked example. J. Adv. Nurs. 2013;69:2423-2431.
3. Ruona WE. Research in organizations: Foundations and methods of inquiry. California: Berrett-Koehler Publishers; 2005. Chapter 14, Analyzing qualitative data; p.223-263.
4. Bryman A. Social research methods. Oxford: Oxford University Press; 2015.
5. Ryan GW, Bernard HR. Techniques to identify themes. Field Methods. 2003;15:85-109.
6. Cohen L, Manion L, Morrison K. Research methods in education. London: Routledge; 2013.

Table 1: Aligning purpose and approach in qualitative data analysis		
Purpose	**Approach**	**Deductive/inductive**
Testing a hypothesis or model	Content analysis	Deductive
Understanding phenomena	Thematic analysis	Deductive *and* inductive
Generating theory	Grounded theory	Inductive

Figure 1: The process of qualitative data analysis.



Resources box *(all resources accessed 290616)*

Museum of qualitative data website

http://museumofqualitativedata.info/

YouTube videos

Gibbs, G (2010). Bryman's 4 stages. Part 1. University of Huddersfield.

https://www.youtube.com/watch?v=7X7VuQxPfpk

Gibbs, G (2010). Thematic coding, Part 2. University of Huddersfield.

https://www.youtube.com/watch?v=B_YXR9kp1_o

Gibbs, G (2010). What can codes be used for? Part 3. University of Huddersfield.

https://www.youtube.com/watch?v=3oo8ZcBJIEY

Gibbs, G (2010). What is coding for? Part 4. University of Huddersfield.

https://www.youtube.com/watch?v=5xM-9yuBhMc

Gibbs, G (2010). Code list and code hierarchy. Part 5. University of Huddersfield.

https://www.youtube.com/watch?v=DVpkuTdkZvA

Patient's/carers' stories to use as practice transcripts

http://www.alzheimer-europe.org/Living-with-dementia/Personal-experiences-of-living-with-dementia