Thematic Analysis

Andrew J. B. Fugard, Henry W. W. Potts
(a.fugard@bbk.ac.uk)

To appear in *SAGE Research Methods Foundations*

The idea of themes is pervasive in everyday life. Consider, for example, shelves in a library (fiction, politics, poetry), sections in a supermarket (toilet paper, confectionary), or stages of a story (flirting on Tinder, first date, falling in love). Thematic analysis is a family of qualitative social research methods that formalize, to varying degrees, the process of developing themes. Definitions are difficult and depend on the scale of researchers' interest. One attempt is that a theme "represents some level of *patterned* response or meaning within the data set" (Braun & Clarke, 2006, p. 82, emphasis in original). This entry further explores thematic analysis through examples and by looking at various types of data used with thematic analysis, how to carry out such an analysis, how researchers and coders can define themes, and how to determine when a thematic analysis has been successfully conducted.

Each particular kind of thematic analysis depends on a *research question* to be investigated, which may evolve as a study progresses; *material* of some kind, for instance transcribed interviews or found text; and a particular *standpoint*, for instance the experiences of the researcher or a theory. The different approaches have in common a goal to articulate what it is that the different fragments of material are an instance of. In other words, expressions of a theme are found in the material and the analyst's task is to develop themes to characterize these expressions (Ryan & Bernard, 2003).

Different people can have different ideas of what constitutes a theme or what themes are important. This is highlighted by Let Toys Be Toys (http://www.lettoysbetoys.org.uk),

1

which campaigns to stop organizing toys by gendered stereotypes. Rather than "Girls" and "Boys" sections in shops, the campaigners argue that toys should be organized by what they do or the ages of children for which they are suitable. It can feel that themes "emerge" from the data, especially when reading material with an open research question and little theory. However, what exactly emerges may depend on the person performing the analysis and whatever influences, not necessarily articulated, they have had. Hence, it is preferred to emphasize the active construction of themes (Braun & Clarke, 2006).

Sources of Data

Thematic analysis is often performed in conjunction with a data collection exercise. For example, in studying the experiences of parents of children with cancer, a series of interviews may be carried out with the parents. Transcripts of the interviews are produced and a thematic analysis performed. Unlike in most quantitative research where data collection is expected to be completed before analysis is performed, it is common for thematic analysis to start while data collection is ongoing and to influence future data collection, informing what questions are asked or determining when sufficient interviews have been carried out.

While interviews are the most common data collection method used for thematic analysis, thematic analysis can be applied to data from a wide variety of sources. The researchers may have carried out focus groups or observations, or given participants diaries to complete. Thematic analysis can also be applied to mixed data of various forms.

Sometimes, thematic analysis is applied to the analysis of pre-existing or "found" materials. For example, if a researcher wishes to study people's experience of using

pregnancy apps, the researcher could carry out an interview study of users of pregnancy apps and ask them about the apps they have used. However, the researcher could also collate reviews people have left on an app store of such apps. A review by an individual posted online may be similar to their responses in an interview.

As with any analytic method, it is important to bear in mind the source of the data. If a researcher is collecting the data, they can control the question asked. With an online review, this is less clear. Who posts online reviews may be a biased sample of all app users, but that may also be true of the people who agree to be in the interview study.

Data do not have to be verbal or textual. An interview can be carried out in sign language. Data do not even have to be strictly linguistic. The use of images in tumblr accounts of individuals with chronic pain can be analyzed for themes. Certain patterns of behavior in observations can constitute a theme.

Carrying Out an Analysis

Thematic analysis involves creating themes and coding the data with respect to those themes. The former entails constructing themes using the data, plus the researcher's understanding, intuition, and theory. It is a process of making sense of the data and abstracting broader ideas than the explicit words on the paper. In some cases, a few discrete themes may seem obvious in the data. In other cases, themes interrelate in complex ways and are intuited from what participants have said.

Thematic analysis is not just about counting words or phrases. There are increasingly popular quantitative methods that do this, as with Linguistic Inquiry and Word Count

(LIWC), sentiment analysis, or machine learning with natural language processing. Instead, thematic analysis involves drawing connections at a deeper level, where two fragments of text using different words can be seen to be related at the level of meaning or a common phenomenon. The analysis is a creative process.

Thematic analysis requires familiarity with the data. If the researcher was involved in data collection or transcribing, this often serves as a first pass of the data. When using pre-existing material, a researcher reads through the material first. However, thematic analysis is also an iterative process. A researcher does not read the data once and then analyse it. Themes develop as the researcher goes through the material multiple times. Themes are developed alongside the coding of the data. Analysis may even iterate with additional data collection, such that it runs alongside interviews being conducted.

Central to the process is comparison, one approach to which is known as the constant comparative method (Boeije, 2002). Ideas in a section of one participant's interview (say) are compared to ideas in another section of the interview. Ideas are also compared with those in a second interview by the same person, or in interviews by other people. Comparison continues at all levels (taking into account any relationships between participants, for instance, whether they know each other). Each part of the data is thus analyzed, interpreted, with respect to the whole, and the whole of the data is analyzed with respect to its many parts.

A structure of themes and subthemes is common. More complex interrelationships can be noted, for example, how themes might overlap or complicate each other. A theme can encompass disagreement. In interviews on the experiences of children of recent immigrants at a school, one parent may say they value bilingualism, while the other may value a monolingual experience. Both of these may come under a single theme around the

4

mono/bilingualism contrast.

Thematic analysis can take a more inductive or deductive approach. In an inductive approach, the researcher is more open to being influenced by the data. In a deductive approach, there is more reliance on existing theory and there may even be an existing coding scheme. This may then be developed with reference to the data, or be tested against the data. These approaches are the ends of a spectrum. Many analyses will start with some theory and perhaps some ideas for themes, but then be led by the data.

Thematic analysis can be described as a sense-making exercise. In some cases, themes and the analysis are largely descriptive. However, thematic analysis can also propose meaning behind the surface expressions of the data. This meaning may relate to some theory, be it articulated or not. A thematic analysis can also seek an empathetic understanding of individuals' experiences (sometimes referred to as *Verstehen*). An important way to achieve this is through use of quotation and "thick" description, which uses narrative tools to affect the reader so they may feel something of what the participants feel, beyond mere description (Ponterotto, 2006).

Coding Schemes

The output of a thematic analysis is usually a set of themes, but also thus a coding scheme to support the analysis of the data used or future material. The output of a thematic analysis explains the themes posited, their relationship, and their meaning. A coding scheme is a complementary tool to be used when analyzing data for these themes.

Software

Thematic analysis can be done by hand, using just pen and paper, or a word processing file. However, software exists to aid the process. The software allows for the text to be marked up and then for the researcher to find all passages identified to a particular theme.

What Is a Theme?

There is ongoing epistemological debate in the literature around the ontological nature of the theme. What is a theme? Does a theme exist in the data, waiting to be discovered? Or is a theme a construction of the researcher, semi-detached from the data? The researcher's experience of thematic analysis as one or the other may relate to the context of the study. In some cases, themes seem like obvious summaries of patterns in the data. In other cases, themes clearly depend more on theory and other context the researcher brings to the analysis, with the coding process a more difficult process. These different perspectives have implications for questions of the reliability and generalisability of the analysis, discussed in the following section. If the themes are to be "unearthed" in the data, then two independent coders should find the same themes. However, if thematic analysis is an act of construction, it cannot necessarily be assumed that two researchers should produce the same answers or even that doing so would be desirable.

When Is Analysis Good Enough?

A good thematic analysis can be defined in a number of ways, including how many participants were (or how much material was) included and the process of analysis. In contrast to quantitative approaches, the former receives less attention. With thematic analysis often iteratively intertwined with data collection, a common approach to the answer of how many is the idea of theoretical saturation.

Theoretical Saturation

The idea of saturation is intuitively simple. Saturation has been achieved if additional data do not alter the conclusions of the analysis. No further data are therefore needed. In practice, determining when saturation has been achieved can entail some debate and uncertainty.

When a number of participants (or items of material) is needed from the outset, several authors have suggested rules of thumb (e.g., Braun & Clarke, 2013), while calculation-based approaches also exist (Bernard, Wutich, & Ryan, 2017). Kirsti Malterud and colleagues (2015) discuss a number of factors that should impact the decision: whether the study's aim is narrow (fewer participants needed) or broad (more needed); whether the sample is very specific to the study aim (fewer needed); whether existing theory was used (fewer needed); the quality of the dialogue that produced the material under analysis (the richer, the fewer participants needed); and whether the research is focused on describing individual experiences or a cross-case perspective.

Checking What Participants Think

A common approach to checking validity in thematic analysis is member checking. The analysis can be provided back to participants, usually with a copy of their own data (like an interview transcript, if not already shared at an earlier stage), to see if they agree with the themes. This may be replaced or supplemented by checking against other individuals similar to the participants, or other stakeholders. It is also possible for participants themselves to be involved in constructing themes. The boundaries between researcher and participant can become blurred, and this is desirable for participatory research (Rose, 2018).

Reliability

The number of coders involved in the thematic analysis should be described, although more or fewer is not necessarily better. Thought should be given to how multiple coders can work together. For example, do they jointly develop themes, or is there a lead researcher who develops a coding scheme for others to largely follow? Another option is for one main coder, with a second coder taking the finished coding scheme and using it on a subset of the data. It is advisable for coders to meet regularly. Measures of inter-rater reliability are a common approach in quantitative analyses, but only sometimes presented with thematic analysis, where there is more emphasis on reaching a consensus.

The relevance of reliability divides researchers into those thought to be "positivist" or critical realist view versus those taking a more constructionist stance to thematic analysis (Clarke & Braun, 2018). The idea is intuitive: different people following the same procedure should make the same decisions about themes if that procedure is sufficiently well articulated. However, taking a more constructionist stance, it is not necessarily a strength of

an analysis to say that two different people agree. Indeed, by analogy with the Let Toys Be Toys example, there may be good reasons for preferring one analysis over another, even if this preferred analysis is less common.

Generalization and Causation

Having conducted a thematic analysis on some sample, the question arises whether the results generalize or transfer to other individuals and other contexts. Much thematic analysis focuses on representing the positions of those particular people who took part in the research and retreats from questions of generalizability that can be seen as the territory of quantitative research. However, themes can concern causal relations and those relations might be expected to generalize. For instance, participants asked why they think a psychological therapy helped them can provide examples of the insights they had or steps they took to change their lives. Proponents of a *causal mosaic* approach argue that qualitative research focuses on a detailed level of analysis which can uncover complex causal processes that are obscured by the averaging of crude measures used in quantitative research (Johnson, Russo, & Schoonenboom, 2017). This contributes to more general understandings.

FURTHER READINGS

Bernard, H. R., Wutich, A., & Ryan, G. W. (2017). *Analyzing qualitative data: systematic approaches* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Braun, V., & Clarke, V. (2013). *Successful qualitative research: a practical guide for beginners*. London: SAGE Publications.

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied Thematic Analysis*. Thousand Oaks, CA: SAGE Publications.

REFERENCES

Bernard, H. R., Wutich, A., & Ryan, G. W. (2017). *Analyzing qualitative data: systematic approaches* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Boeije, H. (2002). A Purposeful Approach to the Constant Comparative Method in the Analysis of Qualitative Interviews. *Quality and Quantity*, *36*, 391–409. doi: 10.1023/A:1020909529486

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77–101. doi: 10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2013). *Successful qualitative research: a practical guide for beginners*. London: SAGE Publications.

Clarke, V., & Braun, V. (2018). Using thematic analysis in counselling and psychotherapy research: A critical reflection. *Counselling and Psychotherapy Research*, *18*, 107–110. doi: 10.1002/capr.12165

Johnson, R. B., Russo, F., & Schoonenboom, J. (2017). Causation in Mixed Methods Research: The Meeting of Philosophy, Science, and Practice. *Journal of Mixed Methods Research*. doi: 10.1177/1558689817719610

Malterud, K., Siersma, V. D., & Guassora, A. D. (2015). Sample Size in Qualitative

    Interview Studies: Guided by Information Power. *Qualitative Health Research*. doi:

    10.1177/1049732315617444

Ponterotto, J. G. (2006). Brief note on the origins, evolution, and meaning of the qualitative

    research concept "thick description." *The Qualitative Report*, *11*, 538–549.

Rose, D. (2018). Participatory research: real or imagined. *Social Psychiatry and Psychiatric*

    Epidemiology, *53*, 765–771. doi: 10.1007/s00127-018-1549-3

Ryan, G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field Methods*, *15*,

    85–109. doi: 10.1177/1525822X02239569