**INTRODUCTION TO THEMATIC ANALYSIS**

**by Simon Moss**

|  |
| --- |
| **Introduction** |

Thematic analysis is one of the most common techniques that researchers utilise to analyse qualitative data, such as answers during interviews. This term, although utilised to describe a range of techniques, often refers to a particular article, published by Braun and Clarke (2006; see also Braun & Clarke, 2016, 2018, 2019a, 2019b). This variant is sometimes called reflexive thematic analysis, because this approach demands considerably reflection. Reflexive thematic analysis is a flexible technique that can be applied in a range of circumstances. According to these authors, to conduct thematic analysis, researchers should

* first peruse and read the data in detail several times, while they record initial ideas or impressions
* second, assign codes—often comprising one to five words—to summarise particular segments of data, such as sentences or propositions
* third, identify overlapping or related codes to uncover themes, defined as recurring patterns or concepts in the data.
* fourth, review the data and codes again to evaluate and refine these themes
* fifth, write a few sentences to define the essence or features of these themes and collate data that illustrates these features
* finally, translate these themes into a meaningful and unified narrative

|  |
| --- |
| **Phase 1. Familiarise yourself with the data** |

During the first phase of thematic analysis, you should become familiar with your data—the transcripts, observations, polices, or other information. In practice, most researchers

* read the data at least twice and sometimes more often
* as they read the data, they record their insights in a journal
* these insights might include similarities or differences they noticed across the data;
* or these insights might include possible codes they could apply later—together with some notes about these codes

If conducting interviews, you should obviously transcribe these interviews. This activity also helps familiarise yourself with the data. Braun and Clarke recommend that researchers should transcribe the interviews as thoroughly as possible but they also recognize that transcription entails some choices and intuitions around how to translate the speech into text. To be thorough, obvious signals in the speech, such as shift in volume or pitch, could also be recorded.

|  |
| --- |
| **Phase 2: Code the data** |

During this phase, attempt to reduce segments of data—such as sentences, clauses, or propositions—into labels or codes that often comprise several words or short descriptions. These codes might be recorded in the margin or in some other format. If conducting research on the experience of PhD candidates, typical codes to represent sentences could be

* enjoyment of statistics
* strived to impress supervisor

The aim of this phase is primarily to simplify the data, enhancing the capacity of researchers to recognize patterns and themes. While researchers code the data, they tend to record notes in a journal, such as a separate Word file. They may, for example, record notes about

* possible interpretations of these codes, such as underlying assumptions or beliefs, sometimes called data complication rather than data reduction
* initial thoughts about patterns in these codes, such as the observation that one code tends to be more common in a particular circumstance
* choices as to why they combined various codes into one label
* possible relationships between codes

The kinds and range of codes that researchers apply depends on the research questions and topics. Nevertheless, many of the codes revolve around

* the goals or motives of individuals—that is, what individuals are striving to achieve
* the means or strategies these individuals are using to achieve these goals
* the assumptions or beliefs of individuals

Few researchers will merely code all the data once. Instead, most researchers will

* code a portion of data, such as one or two interviews
* then code these data again, while potentially blending, distinguishing, or clarifying codes
* then continue with another portion of data and so forth

Researchers continue this activity until they have coded the entire dataset. Indeed, proponents of thematic analysis deliberately attempt to devote the same attention to every segment of data—primarily to diminish the effects of biases or preconceptions about the topic.

**Reliability of coding**

Proponents of some variants of thematic analysis calculate inter-rater reliability (e.g., Guest et al., 2012; Joffe, 2011). These researchers

* may invite two or more individuals to code the data
* encourage these individuals to use a codebook—a series of codes they have already defined
* assess the degree to which these individuals assign the same codes—usually be calculating statistics like Cohen’s Kappa

In contrast, researchers who espouse the reflexive thematic analysis, as defined by Braun and Clarke (2006), recognise that coding is personal and intuitive, demanding contemplation and even imagination. Researchers do not attempt to assign the right code—but merely a code that seems to characterise the essence of a sentence or segment of text. Consequently, when researchers utilise this variant of thematic analysis, they seldom calculate inter-rater reliability. Indeed, in contrast to proponents of template analysis (e.g. King & Brooks, 2017) or framework analysis (e.g. Ritchie & Spencer, 1994), Braun and Clarke (2006) do not recommend that researchers utilise a codebook but instead should develop codes as they scrutinize their data.

**Differences in coding between thematic analysis and interpretative phenomenological analysis**

Most qualitative approaches entail coding. Yet, some of the recommendations around coding vary across these approaches. For example

* when researchers apply interpretative phenomenological analysis—another common approach-they commence the analysis with an activity in which they read the data and insert preliminary comments on segments of data; these comments are not really codes but commentaries that might later evolve into codes
* when researchers apply thematic analysis, they initially familiarise themselves with the data and then code the entire dataset, assigning a few words to each segment. In contrast, the first phase of interpretative phenomenological analysis can be regarded as midway between these two activities of familiarisation and coding.

Likewise, thematic analysis also differs from interpretative phenomenological analysis on whether researchers should code all, or only some, of the data before they start to identify themes. Specifically

* when researchers apply thematic analysis, they typically assign codes to all the data before they attempt to extract themes
* when researchers apply interpretative phenomenological analysis, they typically assign codes to the transcript of one participant, and then derive themes from these codes, before they shift to the next participant.

The reason that thematic analysis differs from interpretative phenomenological analysis revolves around the distinction between idiographic and nomothetic conclusions. Interpretative phenomenological analysis generates two classes of themes and conclusions.

* The first class are themes and conclusions that are specific to particular individuals, designed to underscore the distinct features and experiences of each person, called an idiographic approach.
* The second class are themes and conclusions that are meaningful across individuals, called a nomothetic approach

Thematic analysis adopts a nomothetic approach and does not generate themes that are relevant to only one or two participants, for example. Consequently, proponents of thematic analysis are not as inclined to generate themes that might be relevant to only one participant.

|  |
| --- |
| **Phase 3: Develop themes** |

During this phase, you attempt to develop themes—broader categories or concepts that blend multiple codes. Themes may revolve around the causes of some phenomenon, the underlying assumptions of participants, or other attempts to characterise the meaning that underpins the data (Braun & Clarke, 2018). If a project revolves around the experiences of PhD candidates, typical themes might be

* feeling proud while feeling like an imposter
* the motivations to complete change over time

To derive these themes

* you can scan the labels to identify codes that may be related to the same broader concept
* skim your journal entries to help guide these decisions

After completing this phase, you should have generated

* a series of candidate themes—often described using a couple of phrases or sentences—together with a list of codes that correspond to each theme
* some journal entries about the possible meaning or interpretation of these themes, the possible relationships between these themes, as well as questions to explore later

**Semantic versus latent codes and themes**

Researchers sometimes distinguish semantic codes and themes from latent codes and themes. Semantic codes and themes correspond to the surface meaning of data and do not refer to the underlying patterns, assumptions, or concepts. Latent codes do refer to the underlying, often intangible, patterns, assumptions, or concepts. For example, consider the following extract

*I tell my kids never to go out at night. Things aren’t like they used to*

A semantic code or theme might revolve around strict rules. A latent code or theme might revolve around the underlying belief the world is unsafe. To generate this latent code or theme, the researcher needed to contemplate the underlying meaning or assumption of these responses.

**Emergent versus superordinate themes**

Some qualitative approaches differentiate emergent and superordinate themes—a distinction that is not as pertinent to thematic analysis. For example, when researchers apply interpretative phenomenological analysis, they might recognise that a set of codes overlap or correspond to some pattern, generating a theme, called an emergent theme. Next, they might recognise that some of these emergent themes overlap, uncovering what is called a superordinate theme. After the transcripts of each participant have been analysed, the researcher might develop superordinate themes that apply to multiple participants. This distinction does not apply to thematic analysis. Specifically

* proponents of thematic analysis do not tend to distinguish between emergent and superordinate themes—but might distinguish themes and subthemes
* indeed, most proponents of thematic analysis do not refer to emerging themes at all (Braun & Clarke, 2006)

That is, references to emerging themes imply the themes emerge spontaneously from the data. Yet, according to most proponents of thematic analysis, researchers need to utilise their effort, intuition, and ingenuity to uncover these themes. Which themes are proposed depends on the researcher, the data, and the circumstances. The themes do not miraculously emerge as such.

**Number of themes**

The number of themes varies across projects. If the data are extensive, the report is long, and the research question is broad, the researcher might uncover six or more themes. In contrast, if the data are limited, the report is short, and the research question is narrow, the researcher might uncover only two themes—or even one theme coupled with several subthemes. Specifically

* a typical journal article or thesis chapter will present two to six themes—although most of these themes might contain subthemes
* if the article or chapter comprises more than six or so themes, these themes might not be discussed in enough depth; the nuances and features might be overlooked

Other approaches generate a different number of themes. For example, when researchers apply interpretative phenomenological analysis, they might generate many emergent themes—often more than 20 or 30—and 5 to 10 superordinate themes, for example; this number is midway between the number of codes and themes in thematic analysis

|  |
| --- |
| **Phase 4: Review and refine the themes** |

During this phase, you need to evaluate and to refine the themes. In particular, you need to ascertain the extent to which the themes represent patterns and nuances of the data. Specifically, you should review the data again; that is, determine the extent to which the themes encompass both the original text and the codes. If text or codes do not seem to cohere with these themes, some changes should be consider. Specifically

* you could modify the description or scope of these themes
* you might construct a few subthemes; each subtheme revolves around one facet of this theme
* you might combine multiple themes—an activity that can sometimes generate a more interesting account and is necessary if themes seem to overlap
* you could develop an additional theme or discard an unhelpful theme
* you should record these changes, as well as other insights, such as the association between themes, in your journal

|  |
| --- |
| **Phase 5: Defining and naming themes** |

Once you feel the themes represent the data and codes effectively, you need to characterise these themes as explicitly as possible. In particular, you should define the essence of this theme in a few sentences; that is, highlight the interesting or important features of this theme. For example, to define the theme around feeling proud while feeling like an imposter, you might write

|  |
| --- |
| PhD candidates felt proud of their status when they enrolled. This pride, however, sparked doubts about whether they have acquired the capacity to thrive in research. Consequently, they often shifted between feelings of pride and doubts |

Besides these sentences, you should also

* if possible, clarify how this theme is related to other themes or to the overall topic under study
* collate segments of data, such as quotes, that epitomise or illustrate these themes
* name each theme; this name could resemble one of the quotes that epitomise this theme
* define each subtheme, also in a few sentences

According to Braun and Clarke, themes should not comprise too many subthemes. Themes that comprise too many subthemes may evolve into a superficial list of ideas rather than a coherent, meaningful, and nuanced account

|  |
| --- |
| **Phase 6: Generate the report** |

Like all research projects, the final phase of thematic analysis is to write the report. Proponents of thematic analysis, however, recommend some practices that are especially pertinent to this technique. Specifically

* while you write the report, you need to consider which order you should present the themes; you might rearrange the order several times to generate a cohesive narrative
* when you describe each facet or feature of a theme, insert relevant excerpts of data, such as quotes, to illustrate your arguments and to demonstrate that your conclusions are derived from data rather than preconceptions or biases

In contrast to some other approaches, proponents of thematic analysis, as delineated by Braun and Clarke, do not generally seek feedback from participants on whether the report is accurate—a method called member checking. That is, Braun and Clarke argue that researchers need to utilise their judgment and intuition to derive the themes and conclusions. Consequently, the account of researchers is unlikely to resonate entirely with the perspective of participants.

|  |
| --- |
| **Philosophical and logistical considerations** |

Thematic analysis is merely a method to analyse data rather than an entire methodology or philosophy on how to conduct research. Indeed, unlike some other qualitative techniques, thematic analysis permits significant discretion on the part of researchers. Researchers do not have to follow a particular set of procedures or uphold a specific philosophy. However

* researchers should specify the theoretical perspective they want to adopt, such as whether they espouse positivism, constructivism, pragmatism, critical theory, and so forth
* researchers can also embed thematic analysis within a broader methodology—such as case studies, narrative enquiry, or ethnography.

Regardless of the philosophical underpinnings that researchers adopt, Braun and Clarke (2006) enumerates some of the criteria that researchers can apply to evaluate thematic analysis. These criteria also provide some insights into how to undertake thematic analysis. Specifically, they argue

* the transcripts should be checked to ensure they align to the original recordings
* each item of data, such as a specific remark, should be devoted the same level of attention
* themes should be derived from a pattern of codes instead of merely a couple of vivid anecdotes, because these anecdotes might have attracted undue attention
* for each theme, all the relevant extracts should be collated
* themes should be contrasted to prevent undue overlap but integrated into a cohesive framework
* themes should help understand the phenomenon of interest
* the data should have been reread after the themes were developed to ensure these themes are comprehensive and applicable
* when the themes are reported, the researcher should include a balance between analytical descriptions and illustrative examples
* any underlying assumptions should be stated explicitly

**Sample size**

The amount of data you need to collect, such as the number of participants to interview, depends on many considerations, such as the duration of interviews, the variability across participants, and the scope of research. Nevertheless, in general

* the number of participants should exceed five; otherwise, researchers cannot readily identify patterns—such as similarities or differences—across the individuals
* the more prestigious journals often prefer the number of participants to exceed 20 or so

|  |
| --- |
| **Differences to other approaches** |

To appreciate thematic analysis, as delineated by Braun and Clarke (2006), you should understand how this technique diverges from other approaches. This section delineates some of the differences between thematic analysis and alternative methods or methodologies

**Thematic analysis versus grounded theory**

Grounded theory is one of the most common methodologies that qualitative researchers apply. In practice, several distinct variants of grounded theory have evolved over time—and most of these variants share features in common with thematic analysis. For example, the categories that proponents of grounded theory propose are similar to the themes that proponents of thematic analysis propose. Nevertheless, as outlined in the following table, grounded theory in general diverges from thematic analysis on several key attributes.

|  |  |  |
| --- | --- | --- |
|  | Grounded theory | Thematic analysis |
| Methodology versus method | Grounded theory is regarded as a methodology—a theory about how to conduct research; for example, each kind of grounded theory adopts a particular theoretical perspective. The variant of grounded theory that Charmaz advocated is predicated on constructivism—in which researchers are more interested in how individuals derive meaning from their world | Thematic analysis is a method—an approach that can be applied to analyse data |
| Construction of a theory | Proponents of grounded theory strive to develop a theory—a comprehensive account that explains how various categories are associated with each other and often depicted in a figure | Proponents of thematic analysis might discuss how the themes are associated with each other but seldom present a diagram to represent these associations. Likewise, these researchers do not centre their discussion on how these themes are related to each other |

**Thematic analysis versus content analysis**

Some researchers utilise the terms thematic analysis and content analysis interchangeably. Indeed, the differences between thematic analysis and content analysis are hard to characterise because

* thematic analysis refers to a range of practices, although this term is often used to describe the practices that Braun and Clarke (2006) recommend
* content analysis also refers to a range of practices
* the meaning of these terms varies across disciplines

Despite these complications, many researchers share a similar perspective on the difference between thematic analysis and content analysis. For example, some but not all scholars assume that content analysis implies the approach will tend to be somewhat methodical rather than intuitive. To illustrate, proponents of content analysis might

* develop a set of categories before collecting data and then ascertain the number of times these categories appear in the data
* use various algorithms, such as latent semantic analysis, to extract sets of related words from the data

Furthermore, most proponents of thematic analysis will stipulate their philosophical assumptions—such as constructivism or critical realism. In contrast, some but not all scholars assume that content analysis implies the approach is not predicated on a specific philosophy or theoretical perspective. That is, few proponents of content analysis stipulate their philosophical assumptions.

**Thematic analysis versus interpretative phenomenological analysis**

The practices of researchers, and the results they generate, often do not differ appreciably between interpretative phenomenological analysis and thematic analysis. This document has already broached a few differences, however. The following table outlines some other differences between interpretative phenomenological analysis and thematic analysis

|  |  |  |
| --- | --- | --- |
|  | Grounded theory | Thematic analysis |
| Methodology versus method | Interpretative phenomenological analysis is usually regarded as a methodology—a theory about how to conduct research; for example, interpretative phenomenological analysis is predicated on phenomenology, in which researchers strive to convey the lived experience of relevant individuals | Thematic analysis is a method—an approach that can be applied to analyse data—rather than a methodology; for instance, proponents of many research philosophies could apply thematic analysis |
| Sources of data | Interpretative phenomenological analysis is usually applied to analyse data that were collected during interviews—consistent with the notion that phenomenological studies are conducted to convey the lived experience of relevant individuals | Thematic analysis can be applied to many sources of data, including qualitative surveys, vignettes, story completion tasks, diaries, focus groups, as well as interviews |
| Applicability | Researchers should apply interpretative phenomenological analysis only when the number of participants is limited, perhaps fewer than 10, the data emanate from interviews that explore the personal or lived experience of participants, and the researcher is interested in the unique experience of each person | Thematic analysis can be applied in a broader range of circumstances |

|  |
| --- |
| **References** |

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3, 77-101.

Braun, V. & Clarke, V (2016). (Mis)conceptualising themes, thematic analysis, and other problems with Fugard and Potts' (2015) sample-size tool for thematic analysis". International Journal of Social Research Methodology, 19(6), 739–743.

Braun, V., & Clarke, V. (2018). Using thematic analysis in counselling and psychotherapy research: A critical reflection. Counselling & Psychotherapy Research, 18(2), 107–110.

Braun, V., & Clarke, V. (2019a). Thematic analysis. In Handbook of research methods in health social sciences (pp. 843–860). Hoboken, New Jersey: Springer

Braun, V., & Clarke, V. (2019b). Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health, 11(4), 589–597.

Charmaz, K. (2006). Constructing grounded theory: A practice guide through qualitative analysis. Thousand Oaks, CA: Sage.

Guest, G., MacQueen, K. M. & Namey, E. E. (2012). Applied thematic analysis. Thousand Oaks, CA: Sage.

Joffe, H. (2011). Thematic analysis. In D. Harper & A. R. Thompson (Eds.), Qualitative methods in mental health and psychotherapy: A guide for students and practitioners (pp. 209-223). Chichester: Wiley.

King, N. & Brooks, J. M. (2017). Template analysis for business and management students. London: Sage.

Ritchie, J. & Spencer, L. (1994). Qualitative data analysis for applied policy research. In A. Bryman & R. G. Burgess (Ed.), Analysing qualitative data (pp. 173-194). London: Taylor & Francis.