9

Thematic Analysis: Making Values Emerge from Texts

Arild Wæraas

Introduction

When you have transcribed your qualitative interviews, completed your field notes, and you have collected and sorted supporting documents, you most likely have a very large amount of data. How do you proceed when you want to understand the values that are conveyed in the texts, what they mean, and how they relate to each other?

If these are your questions, then thematic analysis could provide the answers. Thematic analysis is a flexible and systematic way of making sense of qualitative data. It can be applied to any kind of written document such as interview transcripts, annual reports, strategy documents and marketing materials, blogs, observation field notes, employment advertisements, letters to shareholders, press releases, and even YouTube videos and photographs. More importantly, thematic analysis can serve to analyse any way of expressing values, explicitly as well as implicitly.

A. Wæraas (✉)
VID Specialized University, Oslo, Norway
e-mail: arild.waeraas@vid.no

Thematic analysis is not a research design or methodology in its own right, as it only deals with the analysis of existing data. It does not exist in one single version, and many aspects of it can be found in other methods for analysis such as qualitative content analysis (Schreier, 2012), grounded theory (Glaser & Strauss, 1999), narrative analysis (Esin et al., 2014) (see Chap. 11 by Espedal and Synnes in this volume), and text condensation analysis (Malterud, 2012). These methods employ different concepts to describe similar aspects and stages of qualitative analysis without necessarily referring to their approach as thematic, the result of which can be confusing (Braun & Clarke, 2020). In this chapter I do not attempt to bring clarity to this variety, nor do I propose a new way of analysing qualitative data. Rather I discuss the merits of applying some principles of thematic analysis to a specific empirical field; the research on values in organisational settings, and I offer examples of how this can be done. In doing so I draw mainly on a reflexive approach to thematic analysis (Braun & Clarke, 2006, 2012, 2020), in contrast to reliability- or codebook-based versions (Boyatzis, 1998; Guest et al., 2011; Hayes, 1997). My outline of thematic analysis of values is also inspired by Gioia et al. (2012), emphasizing inductive-based analysis grounded in data rather than deductive, theory-based analysis.

The chapter is structured as follows: First I describe the general aspects of thematic analysis and present the main concepts of thematic analysis such as codes and themes. I then review some common principles for performing thematic analysis of texts. Finally, I show how thematic analysis of values can be performed. I will not address the use of computer software programmes, although readers should note that these can be very useful for handling the technical aspects of thematic analysis (see e.g. Paulus and Lester (2016) or Saillard (2011)).

Thematic Analysis: A Brief Overview

Thematic analysis is a method for systematically describing and interpreting the meaning of qualitative data by assigning codes to the data and reducing the codes into themes, followed by an analysis and presentation of these themes. Thematic analysis thus combines a structured approach

with the researcher's subjective interpretation. This combination is a key characteristic and strength of thematic analysis, as it draws on the merits of systematically documenting all the steps in the process of analysing data at the same time as it allows the researcher considerable creativity in attaching meaning to the data. The researcher determines the themes, how many, and what they should be called. As such, thematic analysis does not presuppose the existence of one single "truth" in the data, waiting to be discovered once and for all, nor does it assume that coding is necessarily "accurate" or "objective" (Braun & Clarke, 2020). Rather, it requires a sort of deep immersion by the researcher into the data that eventually leads to themes being generated *from* the data rather than discovered *in* the data.

The structured aspects of thematic analysis revolve around the concepts of codes and coding. The process involves initial coding of the data, followed by a second round of coding whereby codes are grouped into themes and often organized in relation to each other. In the following I briefly explain these steps. A third round of coding can be added to identify aggregate dimensions, followed by visual representations of the codes and themes. I will illustrate these last steps towards the end of the chapter.

Assigning Codes to Data

Codes are the building blocks of thematic analysis. In the first round of coding, you use them to label text segments (coding units) that seem relevant to your research question. Briefly stated, a code is a label assigned to a coding unit, intended to capture the meaning of that unit.

The coding unit may vary from a single word to several paragraphs. The meaning conveyed by the unit determines the coding unit. As a rule of thumb, the coded text segment should always be sufficiently large to retain its meaning when taken out of context.

Where do the codes come from? You can determine (at least some of) the codes before you begin the analysis, in which case you develop them in a theory-driven or deductive way. You can also develop them during the analysis, in which case your approach is inductive and data-driven, similar to open coding used in grounded theory (Strauss & Corbin,

1998). Alternatively, you can use a combination of deductive and inductive approaches. In any case, predetermined codes are rarely sufficient alone in order to capture the breadth of the data. This chapter focuses on the inductive, data-driven approach only, although it should be recognized that thematic analysis cannot be entirely inductive since your pre-existing knowledge and theoretical concepts will always influence what you see in the data.

Should you rephrase the words in the text when developing codes or use the same words as those in the text? A distinction can be made between in vivo codes and descriptive codes (Saldaña, 2015; Strauss & Corbin, 1998). In vivo codes are taken directly from the text, meaning that the code assigned to a coding unit is exactly the same as the coding unit. Thus, if the word "seeking integrity" appears in the text and is important with respect to the research question, the in vivo code for that specific coding unit is also "seeking integrity". In vivo codes are informant-centric, and useful if it is important for you to ensure an as close relationship as possible between informant/textual expressions and codes.

Descriptive codes, by contrast, are researcher-centric codes that you create yourself to describe the meaning of a coding unit by developing another, shorter way to express what you think is conveyed by that unit. For example, "seeking integrity" could be the descriptive code if you determine that this is the meaning of a sentence or a paragraph, even if the words "seeking" and "integrity" are not used in the text. Descriptive codes are useful when in vivo codes do not sufficiently represent the nuances and the meaning of the text, and/or when the coding unit is large.

A final distinction can be made between semantic and latent codes (Braun & Clarke, 2006). Semantic codes are descriptive codes or in vivo codes; they describe the explicit or manifest meanings of the data. By contrast, latent codes are descriptive codes that you develop to identify what you think goes on beyond the data by identifying the underlying ideas, assumptions, or ideologies that have produced the patterns in the data. Both semantic and latent codes can involve making inferences about something that is not directly observable. The difference is that whereas semantic codes seek to show patterns in semantic content and

establish the meaning of what is expressed, latent codes seek to determine what produced those meanings.

From Codes to Themes

When coding the data, you will eventually notice that some codes convey similar meanings. If so, something important about the data in relation to the research question has been observed. In a second round of coding, you can then decide to group these codes together into themes. Themes are higher level theoretical constructs than codes because they encapsulate the meanings conveyed by many codes. They are "patterns of shared meaning cohering around a central concept" (Braun & Clarke, 2020, p. 4).

Your judgement as a researcher is critical in order to determine not only which themes are important for your research question but also *when* a set of codes forms a theme and *how many* themes should be generated. There is no rule for how many themes you should end up with, although at some point you will probably notice that adding an extra theme to the ones you already have no longer provides useful information. You may actually be more likely to merge some of the themes you have identified, especially if you have a large number of them.

There is also no rule for how many occurrences of a code or similar codes are needed in order to create a theme. Whereas one theme may be prevalent in every interview transcript or text and backed by thirty codes with similar meanings, other themes may be present in much fewer transcripts and texts and supported by only a handful of codes. The themes that are less prevalent may still be very important if they capture something new, essential, or revealing about the phenomenon of interest.

Organizing Themes

Once you have generated themes from codes, your analysis may stop here, in which case the next step is to report your themes as your findings. However, you could also undertake an additional analytical step by examining how the themes are connected. To figure out the connections, ask

yourself the following questions (cf. Saldaña, 2015, p. 247): Do the themes make more sense if they are arranged chronologically? Which theme seems to logically precede the other themes? Does one theme influence another? Is there a hierarchical relationship between them? Can some themes be understood as subthemes and others as aggregate themes?

When developing answers to these questions, you may be able to see a connection between the themes that becomes an important part of your findings. If this is the case, your analysis may end up proposing a grounded theory model (Gioia et al., 2012). However, regardless of whether it does so or not, keep in mind that your themes are your findings. When presenting your findings, it is important that you structure your presentation around the themes and back up your claims with relevant quotes that address the research question.

Thematic Analysis of Values

The coding process in thematic analysis of values varies depending on some features of the values to be studied and the goal of your research. Two important questions to address are:

- Are the values explicit or implicit in the text? This is to say, do the informants and the documents you have collected make direct references to values, or do you need to "read between the lines" to observe them?
- Is the goal of your research primarily to report the values as they are articulated explicitly or implicitly in the text, or do you want to go "deeper" in order to understand how the values relate to latent beliefs and assumptions?

Coding Explicitly Expressed Values

Let us first consider the simplest case, which is when the data consists of texts that make explicit references to values, and your primary aim is to

describe these values. In this case, the values are easily identifiable, and you will only have to deal with the question of what counts as a value rather than interpreting the text in order to establish them. Perhaps you asked your informants to talk about the values that are important to them, or perhaps you are studying official core values statements retrieved from strategy documents or web pages. In both of these cases, core values will be explicitly mentioned in the texts or transcripts. An example is provided in Table 9.1 below. It shows an excerpt from an analysis of the values found in a large university's core values statement (NTNU, 2018).

The table highlights the values in the text in the left column. In the right column, each value is now an in vivo code. In other words, the coding unit is one word (or sometimes several words, but rarely many), and the code is the same as the coding unit. If you are using a software for qualitative data analysis, the table looks quite similar to what you would see on your screen. On the left you identify and highlight the values; on the right you assign codes. The codes used in this example are in vivo codes only. The procedure for descriptive codes is basically the same, except that the coding units are likely to be larger because more than one word is needed to represent a value.

If your data material consists of explicit text segments such as this one, you should be able to produce a long list of data-driven codes that correspond exactly or at least very closely to the values in the text and then look for themes emerging from that list that could provide a better understanding of the values and your research questions.

Table 9.1 Excerpt from coding of core values statements using in vivo codes

Text	Codes
Every employee and student has a **responsibility** to **contribute** to a work and study environment characterized by **respect** and **consideration**. We facilitate **personal growth** and **professional development**. We contribute to **diversity** and **equal opportunity** in society and in our own activities. We promote **equality** and **tolerance**.	Responsibility, Contribute Respect Consideration, Personal growth Professional development Diversity, Equal opportunity Equality, Tolerance

Coding Implicit Values

In some cases, your data material is likely to speak about values in a more subtle way. This could be because the abstract nature of values makes it difficult to elicit information about values from informants, even when they are asked direct questions. Also, many written documents and other sources are not created specifically for the purpose of describing values. This does not mean that these texts do not contain values. What it means is that you will need to *look* for the values that are hidden in the language of the text and make a judgement about which values are implicitly invoked. Coding at the implicit level requires interpretation, meaning that you will have to infer from your observations something that is not directly said. For this type of coding, it will be necessary to rely more on descriptive codes rather than in vivo codes.

Consider the example in Table 9.2 where the researcher wants to find out which values are expressed in different leadership philosophies. A

Table 9.2 Excerpt from thematic coding of a qualitative interview using descriptive codes

Text	Codes
I was dreaming of a company where the worker would become the operator.	Empowerment
A place where operators would be able to organize themselves, adjust machines themselves and auto-control themselves.	Autonomy
At that time, employees were clocking in and out and received sanctions for any delay ... **I was dreaming of a place where instead of putting sanctions on being late, we would inquire the reason why somebody was late**. Because nobody is late on purpose. **And if needed, we would adjust the time schedules, for example for a young father whose baby had been crying all night**.	Caring
	Flexibility
At that time, we were awarding bonuses every month, up to 20% of the monthly salary. Most of the bonuses were awarded based on the mood of the direct supervisor. **I was dreaming of a place where the bonuses were integrated in the base salary**, and **I was dreaming of a system where we would share our results**.	Solidarity
	Sharing

French factory CEO describes his dreams for the ideal workplace in the following way (Minnaar, 2017):

In this case, the CEO was not asked to reflect on the values on which his leadership is built, nor on what the values should be. He was simply asked to describe his leadership philosophy, and he actually does not explicitly mention a single value. However, the texts still express many important values. Generally, you should look for phrases such as "It's important that", "I like", "I love", "I need", "I think", "I feel", and "I want" (Saldaña, 2015, p. 113), or, as in the case described above, "I was dreaming of".

Did you agree with the coding in the table above? Note that there could be multiple ways of delimiting the relevant coding units and assigning codes in this case. Two different researchers may not arrive at the same codes. For example, take the first sentence; "I was dreaming of a company where the worker would become the operator". Alternative codes to "empowerment" could be "emancipation", "liberation", "enablement", and other synonyms. Also note that if the research question was different, for example, if it involved examining the various components of leadership philosophies rather than identifying values, then the code could be "vision" or "worker-centric", depending on the preferences of the researcher.

So far, we have seen an example of a text that was very explicit about its values, and another that was not. Usually, texts are not either explicit or implicit in this respect—they are a combination. Your coding should reflect this reality. Alternating between in vivo codes, descriptive codes, explicitly derived codes, and implicitly derived codes is perfectly possible in thematic analysis of texts.

Coding at the Latent Level

With some research questions your primary interest may be to understand what lies behind the values you see in the texts. In these cases, you are less interested in identifying *which* values the texts are talking about, explicitly and/or implicitly, and more interested in understanding the attitudes, ideas, beliefs, or assumptions that seem to underpin the values

you observe in the text. Hence, latent thematic coding of values is based on the assumption that our beliefs shape the values we talk about and how we talk about them (similar to discourse analysis described in chapter 10 by Kivle and Espedal). As such, latent thematic coding could be especially relevant for highlighting and explaining differences between groups of informants. Questions you may ask yourself are: What do the values that you observe "really" mean in the context in which they are expressed? With what kind of characteristics, assumptions, or ideals do the texts associate the values? To which world views do the values "belong"? Does the text highlight some values as more important or essential than others?

As an example, consider in Table 9.3 again the example of the CEO who expressed his leadership principles:

Although this informant makes indirect references to values, these values are not the main focus. Rather, we create codes for the assumptions and beliefs that we think *produce* these values. Doing so requires a

Table 9.3 Excerpt from thematic coding of a qualitative interview using latent coding

Text	Codes
I was dreaming of a company where the worker would become the operator. **A place where operators would be able to organize themselves, adjust machines themselves and auto-control themselves**.	Employees work better when they enjoy professional autonomy
At that time, employees were clocking in and out and received sanctions for any delay … **I was dreaming of a place where instead of putting sanctions on being late, we would inquire the reason why somebody was late**. Because nobody is late on purpose. **And if needed, we would adjust the time schedules, for example for a young father whose baby had been crying all night**.	A caring work environment is a goal in itself You should treat your workers well
At that time, we were awarding bonuses every month, up to 20% of the monthly salary. Most of the bonuses were awarded based on the mood of the direct supervisor. **I was dreaming of a place where the bonuses were integrated in the base salary**, and **I was dreaming of a system where we would share our results**.	Employees are more motivated when profits are shared.

thorough analysis of the claims in order to get an idea of what lies behind them. Notice that the codes consist of multiple words because they need to capture a more complex logic compared to descriptive codes that seek to reflect explicit or implicit values. This makes coding at the latent level more complex than coding at the explicit level.

Latent coding is complex also for a different reason: When analysing underlying assumptions and beliefs, your private beliefs could be challenged. For example, consider the statement: "I definitely feel like I need to hire more people with a different cultural and ethnic background". Which latent belief or assumption lies behind this view? Without examining the rest of the text, at least two different interpretations are possible depending on your own views. One is "diversity is good for the workplace", another is "political correctness is a necessary evil". These beliefs are contradictory, yet both could arguably have produced the statement above. So, be careful: Before deciding on the latent belief, make sure you can justify your coding based on how the informants talk about *their* values, practices, and beliefs in the context in which they find themselves.

Generating Themes from Codes

Having developed codes, your task is now to identify themes. The process of doing so can occur in different ways. Three alternatives are as follows:

Grouping synonyms: You are likely to discover that many of the values you have coded are synonyms with similar meanings. For example, according to the Merriam-Webster dictionary (2020), sincerity, openness, frankness, candour, honesty, impartiality, and trustworthiness are synonyms. If these codes are part of your list, they can form a theme. Choose a name for the theme that matches its contents (e.g. "sincerity"). Similarly, other synonyms such as empathy, sympathy, clemency, altruism, benevolence, kindness, and compassion can also be grouped into a theme and given a name, if they exist in your data. This is a straightforward way of generating themes from codes, although it is not well suited for latent codes. You also risk the possibility that some of the codes on your list do not have synonyms and consequently do not have a "home". As a result, you may want to consider the other two alternatives:

Grouping codes of the same type: Many codes are likely to share features even if they are not synonyms. For example, when scholars classify values as belonging to the same type, they look for something that the values have in common. An example is Kernaghan's (2000) typology of public service values. It groups values such as integrity and fairness into *ethical* values, impartiality and rule of law into *democratic* values, and effectiveness and service into *professional* values. The logic of this process of generating themes is the following: You consider whether a group of codes have similar meanings in the sense that they refer to similar aspects of organisational activities, practices, identities, or states. If they do, then you group them into a theme, and find a name for this theme. This approach is also relatively straightforward. However, again, this approach is not well suited for coding at the latent level, and it does not fully consider the semantic content of the codes.

Grouping codes based on semantic content: Finally, and perhaps most importantly, you can group codes based on their semantic content. In this case, the approach involves figuring out what the codes are saying about something or someone, and then condensing that information into themes, regardless of whether the codes that constitute the themes are synonyms and/or of the same type or not. This is usually not a straightforward process. Each theme will have to be phrased as a short sentence, and this can be done in a number of ways. You may be experimenting with some themes initially, discarding some, and splitting others into separate themes. You may also be moving codes back and forth from one theme to another multiple times before you make up your mind about which codes belong where and how to name the themes. Moreover, you may discover new themes as you are working with your data. In the end, you will have to make a decision about which codes go where, how many themes are necessary to represent the data, and how the themes should be named.

If possible, you should consider whether the themes can be further reduced into aggregate dimensions. This would be an additional round of coding and the last step of the coding process in which you connect all the themes around a few core dimensions. The aggregate dimensions could clarify certain shared aspects of the values or highlight common

underlying assumptions, and they could form the basis for grounded theory development (Gioia et al., 2012).

Visual Representations

It is always useful to draw visual representations of your themes and their relationship with the coded values. By doing so, you keep track of all the codes and make sure they are grouped somewhere, and you can better demonstrate how you generated the themes. There are many ways of visually displaying codes and themes. Figures 9.1 and 9.2 show two examples of themes generated from the same set of initial codes. Note that the figures are not complete representations of the data set. In your own thematic analysis, the number of codes and themes is likely to be higher (for a more complete example, see Vaccaro and Palazzo [2015]) .

When comparing the two figures you will notice that although the initial codes are the same, the themes are different. These differences not
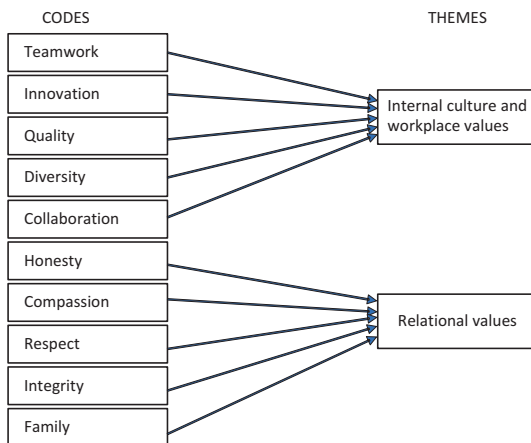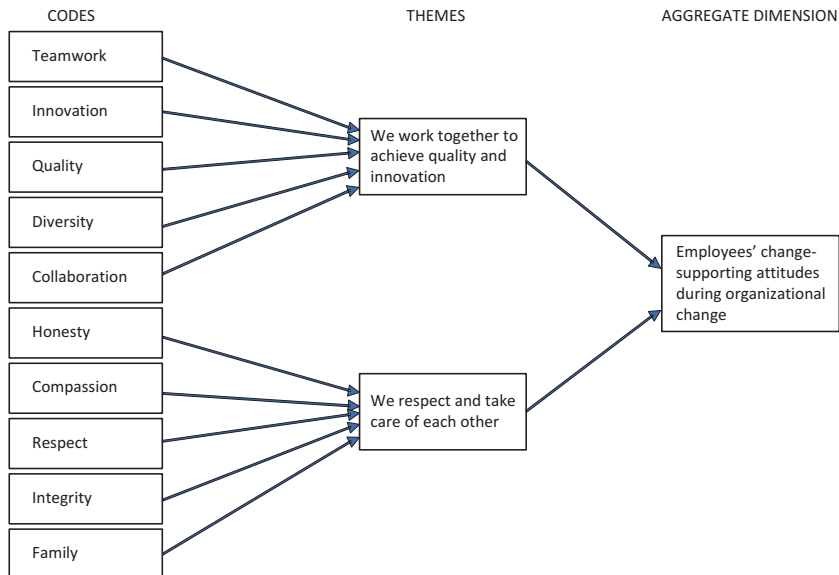
CODES   THEMES

| Teamwork |
| Innovation |
| Quality | → Internal culture and workplace values
| Diversity |
| Collaboration |

| Honesty |
| Compassion |
| Respect | → Relational values
| Integrity |
| Family |

Fig. 9.1 Codes grouped into themes based on type of code

Fig. 9.2 Codes grouped into themes based on semantic content of codes

only reflect different ways of generating themes (the first figure is based on type of code, the second on semantic content), but also different research questions or purposes. In the first case, the purpose may be to understand what characterises the values of a particular organisation or group as they are expressed by employees and top managers. In the second case, the figure could reflect the desire to understand the implications for successful organisational change of the values that employees attach to their own organisation or group. In this case it is possible to develop an aggregate dimension that highlights the overall pattern in the themes.

The themes can also be displayed quantitatively as frequencies. You could, for example, create charts that rank the different themes on the basis of how many codes they contain. This could be useful for summarizing your findings. Note, however, that frequency charts should not be used as the only basis for presenting codes and themes, as this would be
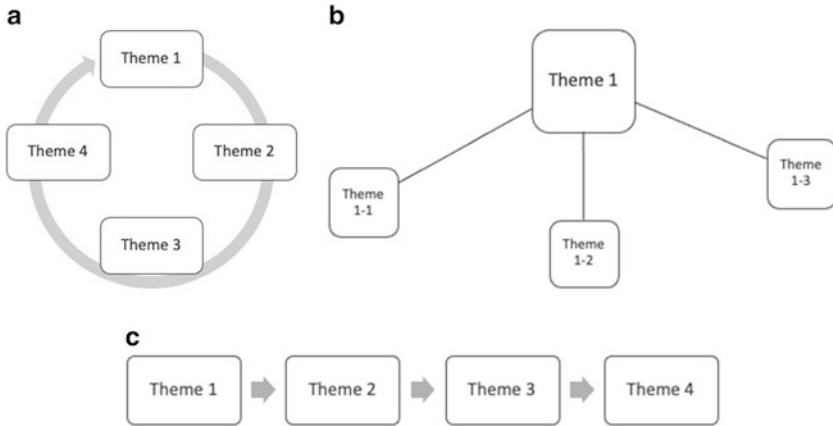
Fig. 9.3   (a) Themes organized in a cyclical model. (b) Themes organized as one central theme with three subthemes. (c) Themes organized in chronological order

more similar to quantitative content analysis (and in some cases, qualitative content analysis [Schreier, 2012]).

Finally, if your themes are developed on the basis of semantic content or latent codes, your analysis may benefit from showing visually how the themes are connected (see e.g. Braun & Clarke, 2006; Gioia et al., 2012). Figures 9.3a–c outline three possible models. The first model shows a cyclical relationship between the themes, the second shows one central theme and three subthemes, and the last shows four themes in chronological order. You may find that your themes fit one of the models, but if not, you could develop your own variation of one of them, or you could develop an entirely different one.

Conclusion

Values often manifest themselves in texts, and thematic analysis is one way of making them emerge from those texts. This chapter has suggested a few ways of doing so. As a stepwise approach, thematic analysis of values can be summarized in the following way: (1) Assign codes, (2) generate themes, and if possible, (3) organize themes, (4) create aggregate

dimensions from themes, and (5) make visual representations. The first two steps should be seen as essential to thematic analysis of values, the remaining ones can be added for further analysis and refinement.

The steps you take should be appropriate for your data and your research question, and you should never try to force fit your data to codes or themes or to a complex visual representation. If your research question only involves describing explicitly expressed values, steps 3 through 5 are probably redundant. If your goal is to understand how latent assumptions and ideas produce different value orientations in different types of settings, you may need all five steps. In any case, apply the principles outlined here with flexibility and creativity, and take your time to understand what kind of analysis your research questions require.

Values come to expression in different ways, and thematic analysis is one of many ways of understanding how. Its benefits lie in the reduction of information into a manageable and comprehensible body of data, which, in the end, is an important part of understanding abstract aspects of social life such as values.

References

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper (Ed.), *APA handbook of research methods in psychology. Research designs* (Vol. 2). APA Press.

Braun, V., & Clarke, V. (2020). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology, 18*(3), 328–352.

Esin, C., Fathi, M., & Squire, C. (2014). Narrative analysis: The constructionist approach. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 203–216). Sage.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2012). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods, 16*(1), 15–31.

Glaser, B. G., & Strauss, A. L. (1999). *The discovery of grounded theory: Strategies for qualitative research*. Aldine de Gruyter.

Guest, G., MacQueen, K. M., & Namey, E. E. (2011). *Applied thematic analysis*. Sage.

Hayes, N. (1997). Theory-led thematic analysis: Social identification in small companies. In N. Hayes (Ed.), *Doing qualitative analysis in psychology* (pp. 93–114). Psychology Press.

Kernaghan, K. (2000, March). The post-bureaucratic organization and public service values. *International Review of Administrative Sciences, 66* (1), 91–104. <Go to ISI>://000086531400009.

Malterud, K. (2012). Systematic text condensation: A strategy for qualitative analysis. *Scandinavian Journal of Public Health, 40*(8), 795–805.

Merriam-Webster Dictionaries. (2020). Thesaurus. https://www.merriam-webster.com/

Minnaar, J. (2017). FAVI: How Zobrist broke down FAVI's command-and-control structures. https://corporate-rebels.com/zobrist/

NTNU. (2018). *Kunnskap for en bedre verden. NTNU strategi 2018–2025*. NTNU. https://www.ntnu.no/ntnus-strategi/overordnet-mal#verdier

Paulus, T. M., & Lester, J. N. (2016). ATLAS. Ti for conversation and discourse analysis studies. *International Journal of Social Research Methodology, 19*(4), 405–428.

Saillard, E. K. (2011). Systematic versus interpretive analysis with two CAQDAS packages: NVivo and MAXQDA. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research, 12*(1), 1–21.

Saldaña, J. (2015). *The coding manual for qualitative researchers*. Sage.

Schreier, M. (2012). *Qualitative content analysis in practice*. Sage.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Sage.

Vaccaro, A., & Palazzo, G. (2015). Values against violence: Institutional change in societies dominated by organized crime. *Academy of Management Journal, 58*(4), 1075–1101.