AARHUS UNIVERSITY

Coversheet

This is the accepted manuscript (post-print version) of the article.
Contentwise, the accepted manuscript version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication
Please cite the final published version:

Publication metadata

Coding qualitative data: a synthesis guiding the novice

2 authors:

Mai S. Linneberg
Business and Social Science - Aarhus University

21 PUBLICATIONS   36 CITATIONS

SEE PROFILE

Steffen Korsgaard
University of Southern Denmark

45 PUBLICATIONS   868 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Exploring employees' proactive behavior in shaping environmental CSR initiatives View project

Project   PACE - Promoting A Culture of Entrepreneurship View project

Coding qualitative data: a synthesis to guide the novice

Mai Skjøtt Linneberg
Department of Management, Aarhus University

Steffen Korsgaard
Department of Entrepreneurship and Relationship Management

Purpose

Qualitative research has gained in importance in the social sciences. General knowledge about qualitative data analysis, how to code qualitative data and decisions concerning related research design in the analytical process are all important for novice researchers. This article offers researchers who are new to qualitative research a thorough yet practical introduction to the vocabulary and craft of coding.

Design

Having pooled our experience in coding qualitative material and teaching students how to code, in this paper we synthesize the extensive literature on coding in the form of a hands-on review.

Findings

Our aim in this paper is to provide a thorough yet practical presentation of the vocabulary and craft of coding. We thus discuss the central choices that have to be made before, during and after coding, providing support for novices in practicing careful and enlightening coding work, and joining in the debate on practices and quality in qualitative research.

Originality/value

While much material on coding exists, it tends to be either too comprehensive or too superficial to be practically useful for the novice researcher. This paper thus focusses on the central decisions that need to be made when engaging in qualitative data coding in order to help researchers new to qualitative research engage in thorough coding in order to enhance the quality of their analyses and findings, as well as improve quantitative researchers' understanding of qualitative coding.

Keywords: coding, qualitative data analysis, qualitative interview, validity, transparency.

Introduction

The analysis of qualitative data is challenging and often frustrating, especially for novice qualitative researchers. Findings and results do not emerge from your transcripts and documents by themselves, but require deliberate work to identify the most important elements and write them up into a coherent and convincing "story" that answers the research questions and provides insights that are loyal to the data (Miles, Huberman, and Saldana 2013).

An important tool in the process of turning raw qualitative data into a communicative and trustworthy "story" is coding. The core operation of coding involves examining a coherent portion of your empirical material – a word, a paragraph, a page – and labelling it with a word or short phrase that summarizes its content. Central to qualitative analysis, coding reduces large amounts of empirical material and makes data readily accessible for analysis, while at the same time increasing the quality of the analysis and findings. In particular, coding in itself is an early form of analysis *"in such a way that the 'final' conclusions can be drawn and verified*" (Miles and Huberman 1994: 11).

The process of teasing findings out of qualitative data requires craft and artfulness on the part of the researcher. With time it therefore becomes less frustrating and more exciting, even though it remains a cumbersome process (cf. Bochner 2018). This paper presents a set of coding techniques that can help pave the way to the researcher's interpretive judgements and improve their quality. By using this paper, novice researchers will be able to reflect more carefully on the repertoire of choices related to qualitative coding in order to improve their analyses and draw conclusions

while remaining loyal to the data. Thus, this paper discusses how and why coding can help qualitative data analysis and explains the steps associated with engaging in coding.

We were motivated to write this paper by the state of the current literature on coding, which we found to be either too comprehensive and detailed (see e.g. the otherwise excellent Saldaña 2015, Miles, Huberman, and Saldana 2013) or too superficial in its treatment of the basic craft of coding (most methods books that we have encountered when teaching qualitative methods) to be really helpful to novices seeking to familiarize themselves with coding for the first time. In this paper, we provide a comprehensive yet focused synthesis of the practice of coding in organization and management studies. The synthesis is an educational and reflexive resource that also provides researchers with practical hands-on knowledge. Having read this paper, researchers new to qualitative research will become familiar with the different perspectives that are present in the literature on coding, enabling informed choices to be made.

Qualitative research comprises heterogeneous approaches encompassing a multitude of philosophies, paradigms and methodologies, and there are significant differences in how qualitative researchers attend to research designs, analytical techniques and quality issues (cf. Gehman et al. 2018, Creswell and Miller 2000, Bochner 2018, Guba and Lincoln 1994, Creswell and Poth 2017). Despite these differences, however, we suggest that researchers i) share a concern for immersion and careful attention to the empirical material, coupled with an insistence on having their findings grounded in or supported by the data; and ii) are able to benefit from transparency in how the findings were developed from the data and perhaps the unfolding of

theory both within a research team and in the communication of results. Thus, whether you are "coding for" mechanisms, discourses, social constructions, metaphors or lived experiences, coding offers a useful craft that enables deep immersion in the data, as well as transparency in the development and presentation of findings that will be useful in particular to many novice researchers who are testing the waters of qualitative inquiry for the first time. In some ways coding is similar to how the craftsmanship of the stroke, the mixing of colours and the preparation of a canvas are conducive to the artfulness of both abstract and figurative painting. As such, structuring your data and achieving an overview of it, carefully considering the relevance, meaning and importance of segments of data, and making the data easily accessible for subsequent data analysis are all likely to constitute good practice and to enable good analytical work in most of the approaches used in qualitative research. In the following, we therefore treat coding as a craft that is useful in most qualitative approaches, but note when, for example, underlying philosophies (as in grounded theory) have implications for the coding process.

We start the paper by discussing what coding is, its rationales and what to consider prior to coding. In the following section we move on to discuss the actual activity of coding – whether it is deductive or inductive in nature, and how to implement coding cycles. We then highlight the usefulness of analytical memos, examples of how and why to display data and thoughts on interpretation before finally presenting some common pitfalls to keep in mind when engaging in qualitative coding. The paper concludes by discussing the limitations of and problems encountered in coding.

What is coding?

Coding in its most basic form is the simple operation of identifying segments of meaning in your data and labelling them with a code, which can be defined as *"a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data*" (Saldaña 2015: 3).

The basic coding operation can be done in various ways. In smaller projects with a limited amount of data, simple colour coding with markers may suffice, with one colour for each code. The copy-and-paste function in software such as Word or Excel will also allow you to copy portions of text or images from your data documents into new documents (making sure the data origin is visible, e.g. "interview #3"). With larger amounts of data or research groups, dedicated software programs can be useful. Such software programs use a basic copy-and-paste operation and have become very easy to use over time. We encourage our students to use this software in their qualitative projects and to observe how the use of software allows them to do better data analysis, going more deeply into the data, providing better presentations of the data and findings, and achieving higher levels of quality regarding the findings. However, using software is not without its dangers. The ease with which new codes can be generated often results in the creation of too many codes. Also, as using software can easily become mechanical in nature, the coding can end up being too rapid and perhaps lacking in reflective analytical depth and focus. Hence, beneath the simple operations lie a great number of challenges. So, while coding itself is easy to do, it is difficult to do well. This is perhaps why some researchers prefer to stick to paper and markers while coding.

Some advantages of coding

While at first glance coding might seem like an unnecessary and laborious step between data collection and coming up with findings, there are several good reasons for spending a considerable amount of your time on coding. It represents the gritty craftmanship that enables artful and creative interpretation and analysis of the data. Basically, the coding process creates an inventory of your data, which enables you to do the following (cf. Saldaña 2015, Miles, Huberman, and Saldana 2013):

Acquire deep, comprehensive and thorough insights into your data

As you code your data, you have to look at each individual sentence and paragraph in your data and make a judgement about its meaning. Simply reading through your data as if it were a book will not prevent you from overlooking potentially new and surprising data. Thus, coding is an approach which makes you revisit all aspects of the data you have collected, including those you may not have noticed during the actual data collection. This will likely trigger analytical ideas that are not simply derived from your (unconsciously) selective impressions and recollections of the data collection activities (interviews, observations etc.).

Make the data easily accessible and retrievable

Qualitative empirical work can often continue over long periods, during which the memory of the data may fade, thus making access to the data an important condition for analysis. Coding sorts the data into labelled segments, much like folders on your computer. This enables quicker access to data and allows the researcher to retrieve it for another look. Moreover, it makes collaborative work easier when more than one researcher is using the same data.

Sorting and structuring your data

Having your research question in mind, the most essential codes can often be used as a skeleton on which to build your analysis. While coding, you make judgements about each individual element in your data in order to decide whether it is relevant or not. This reduces the amount of data you have to take into the final analysis and makes the analytical tasks easier. If you are doing a comparative study or have multiple interviews, coding provides you with a structure that allows comparison of specific dimensions of interest.

Ensuring transparency

As indicated in the introduction, there is an intense debate over the quality criteria that are appropriate in qualitative research (see e.g. Sinkovics, Penz, and Ghauri 2008, Creswell and Miller 2000, Kvale 1995, Bochner 2018). We believe that, despite these terminological and philosophical debates, most qualitative researchers are keen on ensuring that their findings are fundamentally credible and trustworthy. In empirical work, this can be enhanced by observing transparency in respect of how your conclusions are linked to your data (cf. Guba and Lincoln 1994, Elo et al. 2014) because this allows the reader to understand the role played by the researcher in the shaping and analysis of the data. Also, it justifies the choices made by the researcher and can be seen as an attempt to even out the knowledge and the access asymmetry between the researcher and the reader (in particular when the reader is a practitioner rather than another researcher). In turn this can also compel a focus on the applicability and meaningfulness

of one's work to practitioners, which can be understood as a quality aspiration in its own right in many approaches to qualitative research (Kvale 1995, Bochner 2018).

Coding allows you to pick out relevant parts to show the reader. Even though interpretation will always be part of the researcher's task, too much telling and too little showing lowers the quality of the work. Showing the data to the reader forces you to develop a chain of evidence depicting your arguments and showing how you have reached your conclusions (Pratt 2009).

Ensuring validity

Coding is an important step in moving from the raw data to the findings, as well as being a means to maintain coherence between the objective and the results. Coding is a way to ensure that the questions asked are the questions that have been answered. Even though the questions asked initially can change throughout the research process, especially in inductive approaches (Charmaz 2014), reconciling the chosen question(s) and answer(s) is ultimately necessary. Moreover, this provides an excellent opportunity to come to terms with confirmatory bias by searching for contradictory evidence in the empirical material.

Giving a voice to one's participants

Although coding confronts the researcher's work with the empirical material, the coding process can also be seen as interactive (Charmaz 2014). Codes are created as a means to understand the phenomenon and/or participants and their perspectives. As such the researcher interacts with the phenomenon and the participants time and time again by means of the empirical material. As a result, through the dynamic development of codes, we come to understand participants' views

and actions from their own perspectives (Charmaz 2014). This view is particularly prominent in the literature on grounded theory, where researchers are not restricted by preconceived codes but understand codes as emerging inductively from scrutinizing meanings in the data (see below for an overview of inductive and deductive forms of coding respectively). However, in deductively inclined coding as well it is essential that the researcher engages with the data continuously during the analysis.

Preparing for the coding process

While it is always good to have an overview of your research process and to prepare the different steps in advance, certain things need to be in place before coding is initiated.

First, a research question (or a set of research questions) and a research design are established. The objectives and the research question are important, as they help define what kind of data are needed in order to complete the project successfully by answering this question, as well as serving as an arbiter in respect of any questions that appear during the research process. The research design outlines the nature of the research and examines the overall elements of the research project to determine how they fit together. It also defines the unit of analysis, the context and the data that need to be collected.

Secondly, in most qualitative research projects, even those with a more grounded approach, some form of reading or review of the relevant research literature is necessary. According to Eisenhardt, "I believe in knowing the literature, and then looking for a problem or questions where there's truly no known answer. It's almost impossible to find those problems without

knowing the literature" (Eisenhardt in Gehman et al. 2018: 4). Also, knowing the existing research literature enables you to delimit the area of research. Moreover, the existing literature can help you develop the tools for collecting the data, for example, an interview guide (cf. Rowley 2012, Kvale 1997, Miles, Huberman, and Saldana 2013) or the coding framework, assuming the coding will being using a deductive approach.

Thirdly, you need to collect at least some of the data and document it in a form that allows systematic analysis. Collecting data requires the development of a research design and a sampling strategy. In organization and management research, for example, the typical data for which qualitative coding is relevant are textual, such as transcribed interviews, written field notes documenting the activities observed and/or other forms of text, such as newspaper articles, reports, excerpts from social media etc. Visual data in the form of images and video are increasingly being used, and the software tools available for qualitative data analysis are getting better at handling such data. Regardless of the type of data, it is important to document them in a format that makes them accessible to labelling so they are visible and retain their form over time. Consequently, researchers taking field notes while observing must make sure to rewrite the hastily written jottings into more elaborate and legible notes as soon as possible in order to retain thick data in a form that is fully retrievable even after the impressions of the moment have faded from their memory (Spradley 2016).

Coding qualitative data

Inductive coding

There is a strong tradition in qualitative research of developing codes "directly" from the data. Researchers develop codes from the data by using phrases or terms used by the participants themselves, rather than using the, often theoretical, vocabulary of the researcher. In this way the codes stay close to the data, mirroring what is actually in them, rather than the ideas and prior understandings of the researcher, who is working vigorously to remain open-minded. This approach is most often referred to as the inductive approach or sometimes as grounded theory. The concept of grounded theory originates from the strategies developed by Glaser and Strauss (1967) and Strauss and Corbin (1990) and later Charmaz (2014), but it has since developed into a "big tent", that is, a set of approaches to the task of building a theory from data:

> "It almost invariably involves collecting data, breaking it up … and then abstracting at a higher level… this process is at the heart of what most theory-building qualitative researchers are doing." (Gehman et al. 2018: 5)

The inductive approach is relevant when doing an exploratory study or when no theoretical concepts are immediately available to help you grasp the phenomenon being studied. Working systematically with coding allows the inductive researcher to observe transparency and thus offer credible interpretations of the empirical material (Gioia, Corley, and Hamilton 2013).

When doing inductive coding, you will often find yourself creating several codes, often very precise and narrow ones, which is good for capturing the complexity and diversity of the data.

From the grounded theory approach, we have the notion of line-by-line coding, which depicts this exactly (Charmaz 2014). Finding a balance between having a workable number of codes and capturing the complexity and diversity of your data is difficult. Ultimately you may want to end up with a list of 50-70 initial codes. These codes can be submitted to a second cycle of coding in which higher-level categories are created from the initial code list (Gioia, Corley, and Hamilton 2013). This process moves the researcher from having a higher number of codes to having a smaller number of themes or categories. As you move forward in the coding process towards higher-level categories, you may want to draw on some of the existing theory and concepts that are related to your phenomenon. This helps you anchor the study in the literature and provides further support for the findings.

Deductive coding

While inductive codes have the advantage of being completely loyal to the data, there is a risk of the whole process becoming too complicated and lacking in focus, especially for the novice qualitative researcher. Sometimes researchers will therefore adopt a more narrow and deductive approach to coding. Here, a pre-defined list of codes is created in a so-called coding frame before you start coding your data (cf. Miles, Huberman, and Saldana 2013). This approach helps focus the coding on those issues that are known to be important in the existing literature, and it is often related to theory testing or theory refinement. It is also a helpful approach if the aim of the study is to generalize analytically across cases (Rowley 2002, Eisenhardt 1989). Indeed, if the study is theory-driven, the theoretical framework may be converted into a coding framework. Even if you engage predominantly in deductive coding, the process can still remain flexible.

Generally, the codes in deductive coding are theoretical concepts or themes drawn from the existing literature. In a deductive coding approach the number of codes will typically be relatively limited, with maybe just five to ten codes derived from the theoretical framework. During coding, the coding frame can be adjusted if interesting differences emerge within a given code or if some new and interesting things come up that are not captured by the existing codes.

In practice, a combination of inductive and deductive coding is the most commonly used approach, what is sometimes referred to as a blended approach (Graebner, Martin, and Roundy 2012) or abduction (Alvesson and Kärreman 2007). In order to get started, it is nevertheless a good idea to concentrate on one of the two approaches, as they have different strengths and weaknesses. Starting inductively ensures closeness or "giving voice" to the data, with the possibility of unfolding theory later. Deductive approaches ensure structure and theoretical relevance from the start, while still enabling a closer inductive exploration of the deductive codes in later coding cycles.

The notion of abduction captures this combination of inductive and deductive elements and suggests a cycling back and forth between data and theory (Pierce 1978). By doing this, the researcher remains open to surprises in the data while at the same time staying attuned to existing theories. Inductive coding stays more loyal to the data but may also be less focused. Turning a deductive coding process inductive will thus move you closer to the data, yet also lose theoretical focus, and vice versa. The abductive approach is attuned to looking for surprises in the data in the way that one has an interest in rethinking the existing theories (Pierce 1978). The notion of abduction thus encourages a flexible theoretical framework as well as a flexible approach to the

empirical framework, which manifests itself in a combination of inductive and deductive coding in accordance with the needs of the emerging analysis.

Cycles of coding

It is helpful to see coding as occurring in two or more cycles. In an inductive approach, the first coding cycle uses informant-centric terms, whereas the second coding cycle becomes more researcher-centric in the sense that concepts, themes and dimensions from existing theories may be introduced to lift the analysis to a higher level of abstraction (Gioia, Corley, and Hamilton 2013). This can help you think of the code types that are used in the initial phases as more descriptive. Here the codes are used to create an overview of the data and enable subsequent exploration of patterns of similarities and differences in the later cycle. Thus, the code types used in the second cycle are more analytical in nature and focus more on creating patterns in the data. Hence, the act of coding is not to be seen as linear: rather, it enters feedback stages, which is why we can use the labels 'first-' and 'second-cycle coding'. Below is an overview of selected coding methods:

INSERT TABLE 1 HERE

Different code types are commonly used in these two coding cycles. The text-book literature is swamped with code types, which seems to confuse novices more than helping them. For this reason, here we will stick to just a few. In the first coding cycle two types of codes are often of particular importance: descriptive codes and attribute codes.

Descriptive codes are assigned to segments of data based on what the segment "is about". Segments of data are thus summarized using a label that indicates the meaning of the segment of data in relation to the overall research topic. Descriptive codes are assigned to smaller units of data, usually in the form of sentences or a coherent set of statements. If we consider an example where entrepreneurs have been interviewed about their financing, whenever the researcher comes across statements by interviewees about their use of networks, he or she will code the statements "networks". If descriptive coding is done properly, it leads to a categorized inventory of the data providing an overview of what is in them (Saldaña 2015).

Descriptive coding may be more or less structured depending on whether the approach is deductive or inductive, and you may wish to uncover certain types of content in the data. For some purposes it may be useful to code for events, which may later be analysed in a process-oriented study, or alternatively to code for emotions, such as emotional responses to certain events. Notably, in deductive coding a higher level of abstraction can be reached in the second cycle, but it may also be useful to explore differences within codes if the deductive codes contain interesting sub-themes and variation. This second option will take you closer to the details of the data rather than to a higher level of abstraction. This depends on where one feels the analysis needs to go and is something one can experiment with.

Attribute codes are basic information assigned to larger segments of data, typically to the units in which the data were originally collected such as interviews, sites of observation, data sets for a given example. In certain types of project, typically those where the findings are derived in part from comparing cases or eliciting differences in experiences and perspectives (e.g. in

multiple case studies, cf. Yin 1989, Eisenhardt 1989), attribute codes can help fix an early structure for the data set establishing a good overview and easy access to data. Later this enables sources of differences and similarities in the data to be explored more effectively. At the individual level it might thus be useful to assign attribute codes for age, gender, experience or other attributes that are of relevance in the given study. At an organizational level, there are similar generic or specific attributes such as industry, size etc. that can be considered potential sources of insights into the phenomenon. This corresponds to a replication logic (the constant comparative logic) invoked on the data level, which involves the process of continuously comparing new data to extant data. The ongoing effort will confirm, enhance or discount the theoretical propositions under scrutiny.

From the first cycle of coding to the second cycle, codes develop from the initial phases, being more straightforwardly developed in later stages, when a second cycle of coding entails "classifying, prioritizing, integrating, synthesizing, abstracting and conceptualizing, and theory building" (Saldaña 2015: 58). As the research process develops, so does the type of coding, which also allows the researcher to move from basic descriptive codes towards answering the research question posed. The trick is to look for an overarching structure or process that can be understood at a theoretical level (Gehman et al. 2018). Therefore, when codes are clustered together according to similarity and regularity, patterns are born, and you can begin to analyse the connections between them. This also implies identifying concepts inspired by the relevant theoretical framework. In this process, both theories and extant empirical work come to play yet another role in data management, as you can compare your own categories and their relations with relevant theories and/or concepts from previous studies in order to validate or generate

further ideas for the analysis. This is sometimes referred to as theoretical triangulation (Miles, Huberman, and Saldana 2013)

A coding structure consists of the first, second and perhaps subsequent cycle codes and shows the progression from data to theory, thereby making visible the grounds on which you are drawing your conclusions (Gioia, Corley, and Hamilton 2013). Depicting the data structure is not promoted equally by all qualitative researchers; for instance, you can argue that comparing multiple cases through replication (at the case level) makes it difficult to see anything from the data structure that does not point to the particular elements in the single case (Eisenhardt in Gehman et al. 2018). In this instance you can choose to present each case narrative as the depiction of the empirical material.

After coding

Combining coding with analytic memos

Analytical memos can be described as the researcher's ongoing reflections during coding concerning the codes, the phenomenon, the informants and their interrelations. Analytical memos help you think about these elements, thus eventually bridging the distinctions between coding, analysis and results. Memos are creatively developed little documents based on intuition, hunches and serendipitous occurrences (Saldaña 2015) related to the above-mentioned elements. We use memos as a tool as early as in data collection. Repeatedly we find that data collection and coding run in parallel, and the analytical memo is a great tool to help materialize ongoing reflections, much like a log that can both inform subsequent data collection and lead to richer explanations in the analysis later on.

From coding to analysis

Coding and interpretation are not two distinct phases but interrelated processes that co-evolve, yet have different characteristics. While in a sense the first cycle of coding in particular represents a more mechanical application of labels to segments of data, the associated task of interpreting the codes and deriving an overall interpretation from them is not mechanical. It requires insight to get patterns to emerge. While the coding process and analytical memorizing enable the emergence of patterns in the data, it can only enable, not determine it. Ultimately a process of immersion in the data is required, and this will take time and be as difficult to control as the creative process of artists. Nonetheless some notes on what the output of the final analysis is expected to look like may help guide the novice to the process of qualitative data analysis.

In most conceptualizations of qualitative research, theory-building (as opposed to theory-testing) is considered the overall purpose (Gioia, Corley, and Hamilton 2013). This means that qualitative research should lead to, for example, the identification and/or elaboration of new concepts, ideas about how concepts relate to each other, or accounts of processes. The output may thus come in the form of typologies outlining the different aspects of a phenomenon, a concept with rich descriptions of the phenomenon, individuals or organizations that are categorized under this concept, or theories describing conceptual relations.

Importantly, findings in qualitative research should always emphasise elaborate detailed descriptions, specific examples and the inclusion and discussion of outliers rather than count

instances of a phenomenon. Good qualitative studies obtain a kind of undeniability to their results, provided the findings come with detailed accounts of the phenomena.

Displaying your data and codes

Visual display eases the reading and increases the understanding of scientific texts. A data display is anything in which data can be illustrated more efficiently than in longer pieces of text. In some methodological approaches, such as grounded theory, the visual representation of emergent findings is an essential step (Verdinelli and Scagnoli 2013). Inductive coding in particular can be visualized nicely using a form of horizontal tree structure. This displays the initial codes and shows how they are developed subsequently into categories and perhaps concepts. In qualitative analysis, visual displays have a role to play at various stages of the analysis (Saldaña 2015: 58). There are multiple types of visual display (see Miles, Huberman, and Saldana 2013 for a comprehensive overview); here we concentrate on just a few.

Boxed displays are the simplest displays and are "literally, text framed within a box … used to highlight a specific narrative … (it) emphasises the authors' interest or points of relevance." (Verdinelli and Scagnoli 2013: 12). This type of display underscores the extensive narratives that are central to the analysis in any project.

Matrices basically include two or more categories, dimensions or concepts that are cross-classified. The matrix is probably one of the most commonly used types of display. As with boxed displays, they can be used from the early stages of an analysis, for instance, to present a

literature review, to illustrate demographics or, later in the process, to illustrate the results of an analysis.

Network displays are described as collections of nodes that are connected or linked (see e.g. Miles, Huberman, and Saldana 2013), allowing a focus on many variables simultaneously. They help illustrate the relationship between themes and categories, and are also suitable for illustrating second-cycle coding or even the results of the study.

Discussion: A critical view of coding

Despite its widespread use, coding has also been exposed to criticism. First, some argue that coding splits the data into disjointed elements, the holistic element so central to qualitative analysis is thus being lost. This criticism is warranted in so far as coding does split your data, and there is a risk that in your analysis you focus solely on the relations between the codes at the expense of holistic and comprehensive understandings of the examples and phenomena you are studying, or else you neglect the context and situational factors. It is important to bear constantly in mind the overall understanding of the phenomena in order to make sure that the findings emerging from the coding process are valid and relevant. There are several techniques that can be employed throughout the process of analysis: rereading the full transcript and initial analytical memos, looking for negative evidence in the data, keeping an updated logbook that includes situational factors of relevance, and cooperating with one's peers.

Secondly, some criticisms are directed at the subjective nature of the coding process on the basis that judgements about which codes to develop and where a given segment of data fits into the

coding scheme are subjective to the coder. On the one hand, this is simply a necessary consequence of doing qualitative research, the drawbacks of which are offset by the richness and detail we can obtain through any qualitative research. On the other hand, it is important to note and consider this subjectivity, especially through reflexivity regarding oneself and one's perspectives. It has been suggested that using language-processing technology (LPT) for coding can address this subjectivism and increase the trustworthiness of the result (Crowston, Allen, and Heckman 2012). However, while using LPT allows the researcher to process very large amounts of data, there are obvious consequences in using such a system. First, having a system requires a trained analyst to develop the rules applied in the system; second, the system requires substantial inputs of data. It is therefore suggested it only has a raison d'être in empirical material where manual coding is not feasible due to the volume of data (Crowston, Allen, and Heckman 2012).

Along the same lines, a recurrent theme in the literature is the increase in trustworthiness that comes from having several people involved in the coding process, or perhaps establishing inter-rater reliability by, for example, conducting inter-coder reliability tests or working collaboratively to develop inter-coder agreement (a softer form of inter-coder reliability) (Campbell et al. 2013). However, researchers' varied perspectives and previous experiences are critical elements in the data analysis. This means that, instead of using one researcher to confirm the reliability of the other, one can draw on "the positional reflexivity of two researchers, each with a distinct perspective, as a potential strength to cogenerate themes and theory" (Anderson, Guerreiro, and Smith 2016: 44). This suggests that researchers working independently with thematic analysis and coding cooperation in a team can subsequently produce joint interpretations allowing for greater dimensionality.

Thirdly, researchers should be careful not to mistake qualitative coding for a mechanical quick fix neatly and swiftly de-contextualizing the data. In particular, the current literature on qualitative research reminds us that the urgency and the believability of the findings should be prioritized over their detachability from their empirical context (Bochner 2018). Importantly, coding and a count of the instances of a given phenomenon should not replace elaborate detailed descriptions, elaboration of contexts, specific examples, or the inclusion and discussion of contradictory results. This is neither the mission nor the value of qualitative research. Good qualitative studies obtain a kind of undeniability of results provided the findings come with detailed accounts of the phenomena that allow readers to see the stories unfold and to relate to the experiences of the people involved. Thus, ultimately the achievement of coding that is well done comes with a strong notion of the qualities of qualitative inquiries, but it can be challenged by the misconception that coding streamlines and standardizes the data.

Conclusion

This article has explored the key aspects of the coding of qualitative data: the reasons for doing coding; the roles coding plays; some of the different types of code; the use and combination of inductive and deductive approaches to coding; the phases involved in engaging in cycles of coding; how to display data; and the relationship between coding and interpretation.

With experience, focussing on coding provides structure and depth to the analytical process. In particular, coding should not be "just" a mechanical process chasing specifics. It is important to remain attuned and sensitive to the data and its context, as well as to one's role as a researcher in

collecting the empirical material. Engaging in coding allows the researcher to provide transparency for others and oneself in relation to how existing concepts are reflected in the empirical material, but also where there might be novel insights and possibilities for theoretical development. Alversson and Kärreman (2007) suggest doing several rounds and types of coding with different framings and to take a particular interest in the contradictions in the data, rather than merely looking for similarities. To engage in coding could be seen as an adventurous trip *revisiting* venues and allowing ongoing reflection and learning – not merely confirmation, but rather an art skilfully conducted (Probst and Bucholtz 2015). One should not be tempted into considering coding as merely a pragmatic tool for organising and reducing empirical material, as this material also allows the coding to enhance the quality of one's research by, for instance, checking for inconsistencies or contradictory findings in the empirical material, as well as providing transparency for oneself through the process, and the reader in the final output.

References

Alvesson, Mats, and Dan Kärreman. 2007. "Constructing mystery: Empirical matters in theory development." *Academy of management review* 32 (4):1265-1281.

Anderson, Ross C, Meg Guerreiro, and Joanna Smith. 2016. "Are all biases bad? Collaborative grounded theory in developmental evaluation of education policy." *Journal of Multidisciplinary Evaluation* 12 (27):44-57.

Bochner, Arthur P. 2018. "Unfurling rigor: On continuity and change in qualitative inquiry." *Qualitative Inquiry* 24 (6):359-368.

Campbell, John L., Charles Quincy, Jordan Osserman, and Ove K. Pedersen. 2013. "Coding In-depth Semistructured Interviews: Problems of Unitization and Intercoder Reliability and Agreement." *Sociological Methods & Research*. doi: 10.1177/0049124113500475.

Charmaz, Kathy. 2014. *Constructing grounded theory*: Sage.

Creswell, John W, and Cheryl N Poth. 2017. *Qualitative inquiry and research design: Choosing among five approaches*: Sage publications.

Creswell, John W., and Dana L. Miller. 2000. "Determining Validity in Qualitative Inquiry." *Theory Into Practice* 39 (3):124.

Crowston, Kevin, Eileen E Allen, and Robert Heckman. 2012. "Using natural language processing technology for qualitative data analysis." *International Journal of Social Research Methodology* 15 (6):523-543.

Eisenhardt, Kathleen M. 1989. "Building Theories From Case Study Research." *The Academy of Management Review* 14 (4):532-551.

Elo, Satu, Maria Kääriäinen, Outi Kanste, Tarja Pölkki, Kati Utriainen, and Helvi Kyngäs. 2014. "Qualitative Content Analysis:A Focus on Trustworthiness." *SAGE Open* 4 (1):2158244014522633. doi: 10.1177/2158244014522633.

Gehman, Joel, Vern L. Glaser, Kathleen M. Eisenhardt, Denny Gioia, Ann Langley, and Kevin G. Corley. 2018. "Finding Theory–Method Fit: A Comparison of Three Qualitative Approaches to Theory Building." *Journal of Management Inquiry* 27 (3):284-300. doi: 10.1177/1056492617706029.

Gioia, Dennis A., Kevin G. Corley, and Aimee L. Hamilton. 2013. "Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology." *Organizational Research Methods* 16 (1):15-31. doi: 10.1177/1094428112452151.

Glaser, Barney G., and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory*. Chicago: Aldine Publishing Company.

Graebner, Melissa E., Jeffrey A. Martin, and Philip T. Roundy. 2012. "Qualitative data: Cooking without a recipe." *Strategic Organization* 10 (3):276-284. doi: 10.1177/1476127012452821.

Guba, Egon G, and Yvonna S Lincoln. 1994. "Competing paradigms in qualitative research." In *Handbook of qualitative research*, edited by Norman K. Denzin and Yvonna S Lincoln, 163-194. New York: Sage.

Kvale, Steinar. 1995. "The Social Construction of Validity." *Qualitative Inquiry* 1 (1):19-40. doi: 10.1177/107780049500100103.

Kvale, Steinar. 1997. *InterView*. Copenhagen: Hans Reitzels Forlag.

Miles, Matthew B, A Michael Huberman, and Johnny Saldana. 2013. *Qualitative data analysis: A methods sourcebook*. Thousand Oaks, CA: SAGE Publications, Incorporated.

Miles, Matthew B., and Michael A. Huberman. 1994. *Qualitative Data Analysis*. 2nd edition ed. Thousand Oaks, CA.: Sage.

Pierce, CS. 1978. "Pragmatism and abduction." In *Collected papers*, edited by C. Hartshorne and P. Weiss, 180-212. Cambridge, MA.: Harvard University Press.

Pratt, Michael G. 2009. "From the Editors: For the Lack of a Boilerplate: Tips on Writing Up (and Reviewing) Qualitative Research." *Academy of Management Journal* 52 (5):856-862. doi: 10.5465/amj.2009.44632557.

Probst, Barbara, and John Bucholtz. 2015. "Polyphonic coding in qualitative analysis: conversation as musical motet." *Qualitative Research Journal* 15 (3):339-350.

Rowley, Jennifer. 2002. "Using case studies in research." *Management research news* 25 (1):16-27.

Rowley, Jennifer. 2012. "Conducting research interviews." *Management Research Review* 35 (3/4):260-271.

Saldaña, Johnny. 2015. *The coding manual for qualitative researchers*. Thousand Oaks, CA: Sage.

Sinkovics, Rudolf R., Elfriede Penz, and Pervez N. Ghauri. 2008. "Enhancing the Trustworthiness of Qualitative Research in International Business." *Management International Review* 48 (6):689-714. doi: 10.1007/s11575-008-0103-z.

Spradley, James P. 2016. *Participant observation*: Waveland Press.

Strauss, Anselm, and Juliet Corbin. 1990. *The Basics of Qualitative Research*. London: Sage.

Verdinelli, Susana, and Norma I. Scagnoli. 2013. "Data Display in Qualitative Research." *International Journal of Qualitative Methods* 12 (1):359-381. doi: 10.1177/160940691301200117.

Yin, Robert K. 1989. *Case Study Research*. Newbury Park, CA: Sage.

Table 1: Coding Methods

FIRST CYCLE CODING
Descriptive Coding (for field notes, documents), inductive and/or deductive
Attribute Coding (for data base structure and overview, and particularly valuable in comparative research projects)
SECOND CYCLE CODING
Eclectic Coding (for refining first cycle choices)

Pattern Coding (exploring patterns across first cycle codes)
Categorization (combining first cycle codes into – perhaps theoretically informed categories
Exploring patterns in across codes (temporal or processual structure, similarities and differences across descriptive and attritube codes)

(Adapted from Saldaña, 2015: pp. 59, 64)