*The Qualitative Report* 2018 Volume 23, Number 11, How To Article 5, 2850-2861

# Thinking about the Coding Process in Qualitative Data Analysis

Victoria Elliott
University of Oxford, United Kingdom

*Coding is a ubiquitous part of the qualitative research process, but it is often under-considered in research methods training and literature. This article explores a number of questions about the coding process which are often raised by beginning researchers, in the light of the recommendations of methods textbooks and the factors which contribute to an answer to these questions. I argue for a conceptualisation of coding as a decision-making process, in which decisions about aspects of coding such as density, frequency, size of data pieces to be coded, are all made by individual researchers in line with their methodological background, their research design and research questions, and the practicalities of their study. This has implications for the way that coding is carried out by researchers at all stages of their careers, as it requires that coding decisions should be made in the context of an individual study, not once and for all.* *Keywords: Coding, Qualitative Data Analysis, CAQDAS*

Coding is an almost universal process in qualitative research; it is a fundamental aspect of the analytical process and the ways in which researchers break down their data to make something new." Coding is the process of analyzing qualitative text data by taking them apart to see what they yield before putting the data back together in a meaningful way" (Creswell, 2015, p. 156).Yet coding is also a process which goes remarkably undocumented with some honourable exceptions (see Richards, 2015; Miles, Huberman, & Saldaña, 2014; or Bernard, Wutich, & Ryan, 2016, for a thorough exploration of the process). Grounded Theory also considers the coding process in great detail in that particular context (e.g., Corbin, Strauss, & Strauss, 2015; Holton, 2011; Saldaña, 2016, focuses exclusively on the coding process). However, Bryman and Burgess argue, "There is potential for considerable confusion regarding what coding actually is, so that it is doubtful whether writers who employ the term are referring to the same procedure" (1994, p. 218). For the beginning researcher, it can feel as if coding is a process which they must undergo to be initiated into the discipline but must feel their way through with little guidance. Teaching advanced qualitative methods, I am frequently faced with doctoral students who are well into their studies, having collected data, but are still left wondering just how they should approach the coding process. The answer is often: *it depends*. This article is an attempt to draw together some disparate principles for coding in qualitative research, but not to lay down hard and fast rules. Instead, I will conceptualise coding as a decision-making process, where the decisions must be made in the context of a particular piece of research. This may be frustrating to those looking for an easy to follow set of steps, but it is a necessary frustration: if there were a single set of rules, we would have printed them on a single side of A4 and handed them out in the first research methods class. I hope after seeing the contradictory accounts offered by various authors cited in this article, the reader will appreciate why any such easy instructions which can be found from any source, should be read with some scepticism.

Conceptualising coding as a decision-making process, I would argue, has the potential to raise problems given the way that most qualitative researchers have learned coding, via a process of trial and error with guidance from one or two supervisors during their thesis research. Since this is the way that many researchers learn to code, they are left with the impression that it is a natural process which they have discovered. It must be an instinctual thing that you learn by doing, or there would be literature on it like every other stage of the research process! This leaves us (myself included) with a tendency to replicate what we did in our first research in later projects, without considering the differences in contexts which might require different approaches to coding. As Richards argues, "Coding should always be for a purpose. It is never an end in itself" (2015, p. 105). But those purposes are not necessarily the same from project to project, or even from phase to phase, and it is essential that we understand the decisions we are making, consciously or not, when we do coding in a certain way. In order for this to happen, the areas where different roads may be taken must be highlighted and consciously considered when designing qualitative data analysis strategies for research projects. This article is based on several years of teaching a year-long Advanced Qualitative Research class in the Social Sciences Division at the University of Oxford, for doctoral students who had collected their data and were grappling with analysis. Students from Education, Criminology, Law, Refugee Studies, Anthropology and more gathered to consider the literature on qualitative research and to raise their own particular issues. Coding was raised repeatedly by each and every year group which I taught. Now as Director of Doctoral Research in the Department of Education at Oxford, I get similar questions from beginning researchers who are planning their studies and find it hard to conceive in advance what they need to think about in writing about coding for their research proposals. My own research takes in a range of qualitative methods, and focuses largely around issues of curriculum, policy and practice in the teaching of literature. These were the stimuli that led me to think about coding initially; this article considers the questions I explored and the answers I found.

## Why Do We Code?

The need for coding is simple: "Text data are dense data, and it takes a long time to go through them and make sense of them" (Creswell, 2015, p. 152).

Coding is a way of doing this, of essentially indexing or mapping data, to provide an overview of disparate data that allows the researcher to make sense of them in relation to their research questions. Most simply, it can be a way of tagging data that are relevant to a particular point; you may want to, for example, identify all the places in an interview where someone has said something relevant to question 1, rather than just looking at the answer they gave chronologically. Richards and Richards (1994) suggest that this kind of coding (or "code-and-retrieve"; p. 168) should be referred to instead as indexing, to prevent confusion with the more analytically important coding for theory and understanding. Rapley (2011) refers to "labels" and "labelling" instead of codes and coding.

This article deals with a number of questions which researchers might ask themselves about the coding process. These questions arise both from my experience teaching qualitative methods to doctoral students across the social sciences and from the advice given about coding in the literature, which I have problematized in places as part of the model of a decision-making process. This article will not deal with coding as part of grounded theory, because that is a process which is well documented. There are, however, some useful ideas in the grounded theory approach to coding which are more broadly applicable in qualitative research such as coding is the "conceptual abstraction of data" (Holton, 2011, p. 265).

Significantly, researchers code to get to grips with our data; to understand it, to spend time with it, and ultimately to render it into something we can report. Goodwin (1994) describes this as the development of a "professional vision" directed towards your data.

For an overview of the practical steps which coding may entail, see Creswell (2015) or Spencer, Ritchie, O'Connor, Morrell, and Ormston (2014).

## Questions about Coding

*What's the difference between codes and categories?*

Before engaging in decisions about the process, there are decisions to be made about the terminology to be used. In the search for precision and clarity, a semantic mire has developed in that different people use different terms to refer to the same thing and the same terms to refer to different things. However, there is general agreement that there are two levels of terms, and that whatever the labels used at those levels they represent a different order of concept.

> First level coding mainly uses these descriptive, low inference codes, which are very useful in summarising segments of data and which provide the basis for later higher order coding. Later codes may be more interpretive, requiring some degree of inference beyond the data. Thus second level coding tends to focus on pattern codes. A pattern code is more inferential, a sort of "meta-code." *Pattern* codes pull together material into a smaller number of more meaningful units.... a pattern code is a more abstract concept that brings together less abstract, more descriptive codes. (Punch, 2014, p. 174)

The broad distinction between codes and categories is drawn at the level of these two levels of coding. Also called themes in the literature, categories "are broad units of information that consist of several codes aggregated to form a common idea" (Creswell, 2013, p. 186). In qualitative software programmes (CAQDAS) the word "node" is frequently used to connote the same kind of grouping that we might call a category in other coding (Richards, 2015). Saldaña warns that a "theme can be an *outcome* of coding, categorization or analytic reflection, but it is not something that is, in itself, coded" (2016, p. 15). Thus, we have codes at a primary level and categories or themes at a secondary level, which are formed from analysis of codes rather than of data.

More broadly, the word "code" is used to designate the label of any kind which is attached to a piece of data, and it is in this sense that the word code is used throughout the rest of this article. "***Codes*** are labels that assign symbolic meaning to the descriptive or inferential information compiled during a study" (Miles, Huberman, & Saldaña 2014, p. 71). In this sense a category is a code, but of a higher order.

*How many codes?*

The number of codes is a question on which many scholars have a firm opinion, though the range suggests that these numbers may be rather arbitrarily chosen. Friese (2014) warns that the number of codes should not swell into the thousands, a rather frightening number but one which is due to the context of qualitative analysis software, which could enable such proliferation (see below). Other figures gathered by Saldaña (2016) range between 50-300 codes (Friese, 2014); 80-100 codes divided into 15-20 categories, eventually grouped into 5-7 major concepts (Lichtman, 2013); or 30-40 codes (MacQueen, McLellan, Kay, & Milstein, 2009).

Creswell (2015) has a more modest figure in mind:

> I try to code all of my text data (whether a small database of a few pages or a large one of thousands of pages) into about 30 to 50 codes. I then look for overlap and redundant codes and start to / reduce the number to, say, 20 codes.

> These 20 codes then collapse further into about five to seven themes that become the major headings in my findings section of my qualitative report. (pp. 155–156)

For purely descriptive codes, that enable analyses to be done later, such as male/female interviewee, job role, date, school type, it is tempting to code for as many attributes as possible, to facilitate later queries. The question, as Richards points out, becomes not "whether" but "how much" (2015, p. 109). She argues that "excess information makes it hard to see the data clearly" (2015, p. 110), so that there should not be too many descriptive codes. Richards also prudently reminds the researcher not to dispose of information about the data simply because it is not of use now because that data can always be stored elsewhere or coded for if it becomes relevant.

While most authors warn against too many codes, there are others who remind us that the repetition of just a few can be a sign of something. More codes might nominally exist, but the researcher may not see their relevance to the data. If it seems you are seeing the same thing over and over, it might be positive in that it is evidence of an emerging pattern and therefore a research finding, but in the early stages of analysis "It can also mean that your labels are just too large, that you are not thinking with your data at an adequate level of detail" (Rapley, 2011, p. 285). This suggests a need for refining or rethinking; Miles, Huberman and Saldaña are more generous, suggesting "This problem calls for breaking down codes into subcodes" (2014, p. 82).

Saldaña's own preference is that "the final number of major themes or concepts should be held to a minimum… but there is no standardized or magic number to achieve" (2016, p. 25). The final number of themes reported often seems to be governed by journal space, which rarely allows for more than five, but the number of codes necessary to attain any particular five is a matter for personal choice. The answer to the number of codes may also depend on the answer to the question of whether everything in the dataset should be coded, or whether the research is using a priori or emergent codes (see below).

*Can a piece of data be assigned to more than one code?*

The question of whether codes should overlap or be exclusive is one without a clear answer in the literature. Creswell tells us: "You can certainly code a text segment with multiple codes, but ask yourself, 'What is the main idea being conveyed?' and assign a single code" (2015, p. 160). Alternatively, Richards (2015) argues that almost all data will need coding at least three times: once with descriptive coding that gives information about that particular source (such as age of interviewee); once with topic codes, which she describes as the "hack work of the qualitative researcher, labelling text according to its subject" (p. 106); and finally with analytical coding, the higher order codes that do the thinking. But it may need coding far more: "one passage may now be coded 11 times, because it's about 11 different aspects of the project" (2015, p. 107).

The question may be decided on pragmatic grounds: if you want to count the incidence of certain ideas in the data, or to calculate the percentage of the data which deals with a particular idea or concept (but this also raises the question of whether everything should be coded or not), then keeping each piece of data assigned to one code is probably better. On the other hand, the co-incidence of different codes within the same piece of data may well be a useful finding: "single passages often contain a number of different themes each of which needs to be referenced; multiple indexing of this kind can begin to highlight patterns of association within the data" (Ritchie & Spencer 1994, p. 182). However, it might also indicate a problem with the coding system being used or the definitions of codes: "too much Simultaneous coding suggests an unclear or incomplete vision for a coding system and, thus, research design" (Miles, Huberman, & Saldaña, 2014, p. 86). A more practical difficulty will also arise: when coding by hand in particular, the multiple coding of the same piece of data can be difficult to record on a practical level, and possibly interfere with the very point of coding, to make your data more condensed and more understandable on a conceptual level. Richards (2015) notes that this is less problematic when working with qualitative data analysis software. The fact that you can assign data more than one code does not mean that you necessarily should. It is also useful to keep in mind the ultimate reporting of the data: a particularly juicy "chunk" may be a rich example of more than one code, but an empirical journal article will appear less trustworthy and reliable should the same quotation be used to illustrate more than one point; it is a sensible precaution for doctoral students to search their theses for their favourite quotations and see how many times they have actually used them.

If your project is designed to view data through more than one lens (testing the fit of different theories to the data, for example) then multiple coding is likely to be necessary. It may, however, be important to keep frameworks separate, so that a piece of data can be coded more than once, but only once within each framework. A hierarchical, nested coding schema worked out in advance is essential in such a case.

### *Should everything be coded?*

The coverage of data by codes is related to similar questions of the way in which you want to ultimately report data, and by what the research questions are that you are addressing. For most researchers the answer is that not everything should be coded; when the decision is made to be exhaustive there must be clear explanations given for any exceptions (Cunningham, 2004). Cunningham, for example, highlights that verbal fillers like "um" and purely procedural instructions like "tuck your chairs in" are excluded from her coding, but that everything else is included) and a justification for why these are not important in an exhaustively coded project.

Saldaña (2016) highlights that there are some that "feel every recorded fieldwork detail is worthy of consideration" while others suggest "only the most salient portions of the corpus related to the research questions merit examination" and that much else can be deleted (p. 17). It is a brave researcher who actually deletes data, and it is unnecessary in the age of digital storage. It is not uncommon for a researcher, particularly the doctoral researcher, to collect far more data than they can ever use. Setting aside some of the data requires careful thought to select the appropriate parts to sideline for the given project and will usually require specific reference to the research questions. In a larger group project, the coding is more likely to do the division of an over-large data set into discrete units that can be reported separately. Saldaña himself codes "only what rises to the surface" although he began his career by coding everything that had been collected, an approach he recommends to novices: "You, too, will eventually discover from experience what matters and what does not in the data corpus. Code smart, not hard" (2016, p. 18).

The consensus within the literature on data analysis seems to be that coding should not be exhaustive and is in fact a process for reducing the data. Creswell regards coding as an act of "winnowing" (2015, p. 160), while Miles, Huberman, and Saldaña consider it as "data condensation" (2014, p. 73). However, they also point out that coding is a method of discovery: "You determine the code for a chunk of data by careful reading and reflection on its core content or meaning. This gives you intimate, interpretative familiarity with every datum in the corpus" (2014, p. 73). Therefore, while not every line of data may end up with a code label attached to it, the researcher or researchers have to spend time familiarising themselves with that data in order to decide that it does not require a code, just as much as they do if it does require one.

*A priori/emergent?*

This is perhaps the most consuming question for the beginning coder: where do my codes come from? To answer it, we are again directed immediately back to the research question and to the epistemology of the design.

> At one end of the continuum we can have prespecified codes or more general coding frameworks. At the other end, we can start coding with no prespecified codes, and let the data suggest initial codes. *This decision is not independent of other such decisions concerning research questions, conceptual framework and the structuring of data generally*. (Punch, 2014, p. 174, my emphasis)

A design which tests theory against empirical data requires preset codes; grounded theory researchers insist upon, and others with a deeply held philosophical commitment to qualitative research are likely to prefer, emergent codes. Creswell notes that use of a priori codes "does serve to limit the analysis to the 'prefigured' codes rather than opening up the codes to reflect the view of participants in a traditional qualitative way. If a 'prefigured' coding scheme is used in analysis, I typically encourage the researchers to be open to additional codes emerging during the analysis" (2013, p. 185). The most pragmatic researchers will typically use both in the course of a single research project. If the project is a large one with multiple coders, however, a priori frameworks are common, and emergent codes are codified quickly into a framework that different people can use. In all cases emergent codes require some form of codification, so that the researcher can confidently say that they have considered all the data in the light of all their codes, going back over portions which were coded early on, and refining their analysis in the light of later code creation. Emergent codes may be specific words from participants' own voices, or they may be concepts which you as a researcher have been sensitized to in the process of reading the literature in preparation for your research.

Ritchie and Spencer refer to the use of a priori thematic frameworks as "indexing" (1994, p. 182). This is a useful reminder that these terms are not universal; we may also talk of inductive and deductive coding (emergent and a priori respectively).

*What do I call codes?*

> A code in qualitative inquiry is most often 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 2016, p. 4)

A good principle to start with when naming codes is that of length: the "word or short phrase" is ideal; a longer phrase or sentence is not. Miles, Huberman, and Saldaña (2014) differentiate a number of different types of code names, including "emotion," "values," and "evaluation"; these categories suggest that what you call codes depends on what you are using them for. A descriptive code, they say, is a basic label, and often a noun; a process code "uses gerunds ("-ing" words) exclusively to connote observable and conceptual action in the data" (2014, p. 75). The precise ways in which researchers name their codes are intimately related to both their research questions and the procedures they are adopting to generate their codes; a priori codes set beforehand can be categorised and made consistent within categories. Emergent codes are more likely to require editing after the fact to make them consistent with each other (such as being all processes, or all nouns); if this is the way you are working then it is also necessary to ensure that a change in the label does not mean that it no longer describes the data. The main principle that Miles, Huberman, and Saldaña espouse is that codes should "have some conceptual and structural unity. Codes should relate to one another in coherent, study-important ways; they should be part of a unified structure" (2014, p. 82).

The main debate in the literature is the question of whether or not to use in vivo codes that is using the exact words of a participant in your research study. Creswell calls these the "best" code labels because "you start to build codes and later themes that resonate with your participants" (2015, p. 160). He notes that:

> Other types of code labels would be a term you make up on the basis of your personal experiences (e.g., *stressed out*) or a good social science or health science label based on theory (e.g., *efficacy*). Still, "in vivo codes" are best because they move you towards the voices of participants, which you want to reflect in your realistic final report. (Creswell, 2015, p. 160)

However, Rapley argues against taking this too far and notes that altering the words or tenses does not mean you are being disrespectful to a participant's lived experience. He cautions that "it confuses the analytic phase with the phase of presentation of your argument to others. In notes to yourself and in publications, you will probably end up using verbatim quotes, and so give others access to these 'voices'" (2011, p. 282). Miles, Huberman, and Saldaña, on the other hand, argue that "phrases that are used repeatedly by participants are good leads; they often point to regularities or patterns in the setting" (2014, p. 74). The decision may ultimately rest in the researcher's underlying epistemological commitment, or on a more practical level. Colourful phrases from participants may be attractive as code labels for gaining the audience's interest; using words from the participants can also keep interpretation closer to the data and give it greater "face" validity.

The best principle for the naming of codes was that of Cunningham, who suggested that a priority was "precision of name: the codes should have names that capture the essence of their content" (2004, p. 67). Ideally, codes should also be clear enough to convey their meaning to a reader; familiarity with code labels breeds not contempt but understanding, and a blindness to whether they are conveying that "essence of their content" which does not endear a piece of research to its eventual user.

*How big a piece of data should I be coding?*

What counts as a "chunk" of data? Is it a paragraph, a line, a word? To a certain extent this question co-exists with the question of overlapping or exclusive codes. The larger the chunk of data, the more likely the researcher is to need to assign it more than one code. There is no simple answer to this question. Pieces of data "may be individual words, or small or large chunks of the data" (Punch, 2014, p. 173). Some researchers, such as Creswell (2013), suggest coding in rounds; coding paragraphs in a rough first draft of coding, before refining the labels to smaller pieces through further re-readings. Miles, Huberman, and Saldaña say it depends on the study and your aims within it, but suggest that "more typically, codes get applied to larger units—sentences, monothematic 'chunks' of sentences, or full paragraphs" (2014, p. 85). Richards (2015) does not explicitly comment on this but her example of a piece of text being coded as a chunk is a five-line paragraph of text.

Saldaña distinguishes between "lumpers" and "splitters" (a term he attributes to Bernard, 2011, p. 379); the former takes a "lump" or large excerpt and gives it one code; a "splitter" splits the data into smaller codable moments. "Lumping is an expedient coding method (with future detailed subcoding still possible), while splitting generates a more nuanced analysis from the start" (Saldaña, 2016, p. 24). He also suggests that the latter approach requires a line-by-line analysis.

*Should I count codes?*

"Counting is easy; thinking is hard work" (Saldaña, 2016, p. 41).Whether or not to count codes is rarely a question that beginning researchers ask, but is one that they should, and one which often emerges when it comes to reporting data. Some qualitative researchers are very much against the principle of counting in qualitative research, because "counting conveys a quantitative orientation of magnitude and frequency contrary to qualitative research" (Creswell, 2013, p. 185). Others have more pragmatic views or consider counting important for a systematic approach to qualitative research, which is often the view taken by reviewers for non-qualitative research journals.

Counting may also provide a useful indicator for the importance of a given code. Harding (2013), for example, suggests that a code shared by one quarter of the participants in a study is worth consideration in the final analysis. This raises an important point: it is not necessarily the number of times a code appears in the data, but how widespread it is among the data which might be significant. In an interview study we recently conducted, we interviewed three teachers within a school, with a sample of fifteen or so schools (Ingram, Elliott, Morin, Randhawa & Brown, 2018). It was important to the study to know across how many teachers *and* across how many schools certain responses came up—a code applied to six teachers in two schools was likely to be less worthy of detailed consideration than one applied to six teachers in six schools. As Saldaña points out, however, "frequency of occurrence is not necessarily an indicator of significance" (2016, p. 41). Creswell further notes that a count can suggest that "all codes should be given equal emphasis" as well as disregarding that "the passages coded may actually represent contradictory views" (2013, p. 185).

Counting also risks the possibility of overlooking the significant and interesting minority report that is seen just once in a dataset. Saldaña warns of the possibility that the "unique" code or indeed one that appears just two or three times in a dataset may be the key to unlocking the analysis, but with the caveat that:

> Unfortunately, that same number of just one, two, or three instances of a code may also suggest something unimportant, inconsequential, and unrelated to your research questions and purpose. *The analyst must reconcile which one of these possibilities is at work.* (Saldaña, 2016, p. 25, my emphasis)

If a researcher decides to count, it is essential that she also thinks carefully about how to use counts, and what they imply for the later stages of data analysis.

It is certainly true that any codes which you do not count will be the ones which a reviewer insists on having absolute numbers for given in your journal article before he or she will consent to its being publishable. This is not the place for a debate about the standards to which qualitative research is held by largely quantitative journals, but it is worth considering whether it will be more important to you to be pragmatic and be published in the journal you wanted to be in, or to hold to higher principles about your beliefs in what qualitative research should or should not do.

*Should I use coding software?*

The use of software, such as NVivo or MaxQDA, has become de rigueur in qualitative research, it sometimes seems. There is no doubt that the use of such software makes it much easier to count codes if the researcher decides to do so. It can also make it easier to develop complex stratified sets of codes, arranged around nodes, in different layers. It will also make the use of codes to identify data that is relevant to any given question (for example in interviews) more operationalisable: the software will automatically pull together all the data that has been coded with a particular code for you to review together. In this case, the question of how much data to code under a particular label becomes very important, so that the programme can bring up enough data that you can understand the context of a statement, not just see the statement itself. Richards and Richards (1994) argue that the ease of doing this kind of "code-and-retrieve" or indexing (p. 168) becomes problematic for the use of coding software; these kinds of codes are easy to make and identify, whereas the "more vulnerable and tentative ideas emerging from the data are harder to incorporate into ordered categories" (1994, p. 168) so that the indexing takes over from the more important analytical coding.

It is also argued that the use of software makes coding too easy: it is possible to proliferate codes beyond the level at which you will be able to remember them all or deal with them usefully (See Richards 2015, p. 118). It is also easy to be drawn into the data in a way which means you do not have an overview of what is going on. Taking a step back from the software and conceptualising your codes is an important step.

It has to be said that the default position for most people now is that qualitative analysis software is an essential part of qualitative research, just as the norm is to use referencing software for preparing texts. This is partly because it is part of research training to prepare for life in academia, to be able to use, for example, both SPSS and NVivo. This means that the question to be asked is not "should I" but "how should I," and careful thought needs to be given to how coding frameworks are to be developed and used, to avoid the proliferation warned about, but also to ensure that CAQDAS is being used in the service of your particular study, rather than the other way around.

### *How do I do reliability?*

Conceptions of reliability are largely taken from quantitative research. "Reliability is generally defined as the consistency of a measure, or the degree to which scores approximate each other across multiple assessments of an instrument or multiple ratings of the same event" (Syed & Nelson, 2015). To translate this into the language of coding suggests two types of reliability: consistency between researchers (or inter-rater reliability), and consistency over time with the same researcher. To check the latter, code a clean version of a document which you have previously coded, before comparing. For consistency between researchers, both code the same document independently and compare. With an agreed format (such as how big a chunk is) you can calculate statistically the level of agreement, but it might be more useful to do a qualitative comparison: which codes are least agreed on (is there a reason? is the definition unclear?); do coders both code the same density of codes?

Richards (2015) notes that qualitative researchers and methods base their claims for credibility on very different grounds and with different standards than those of quantitative research, where the inter-coder reliability is crucial (and reliability is a notable omission from Creswell (2015), presumably because of his principles about what qualitative research should be). However, as Richards says:

> …being reliable (to use the adjective) beats being unreliable. If a category is used in different ways, you will be unable to *rely* on it to bring you all the relevant data. Hence, you may wish to ensure that you yourself are reliably interpreting a code the same way across time, or that you can *rely* on your colleagues to use it the same way. (2015, p. 117, emphasis in the original)

Richards goes on to note that an element of reliability testing might be more desirable in larger projects, with a number of coders. Miles, Huberman, and Saldaña (2014) suggest an inter-coder agreement of between 85%-90% should be attainable over time. Luker suggests that the recoding by another person needs to be done by "an unknowing coder, that is, someone who has no idea what hypothesis (or hypotheses) you are generating" as a way of keeping ourselves "honest" (2008, pp. 202–203).

Richards, however, sounds a note of caution with the use of inter-coder reliability or consistency tests: consistency over time as you develop your understanding of the data is unlikely; consistency between two raters will not necessarily be desirable, when the two coders have been chosen precisely because of their different understanding of the data, or from two different disciplines. Coding is usually an iterative process for the doctoral project at least; as you develop your understanding of the data and your codes, you return to earlier data and re-code, or refine codes and combine them, requiring a revalidation of earlier coded material.

For some researchers, the idea that there can be a "right" answer to coding is antithetic to qualitative research; Spencer, Ritchie, Ormston, O'Connor, and Barnard argue that "the aim is not to produce a perfectly consistently coded set" because "labelling is done to manage data rather than to facilitate enumeration" (2014, p. 278). This implies a crossover with the question of counting.

> …the objective is to produce a meaningful account of the phenomenon that addresses key aspects of the research question, and to produce this account in a systematic and transparent way so that the reader can see how concepts, themes or categories were developed. Other researchers might well have devised alternative themes or developed different categories, but they should be able to see how the researcher(s) "got there" and be able to assess the value of the analysis. (Spencer, Ritchie, Ormston, et al., 2014, p. 278)

See also Ruth Wodak's retroductable' research, in which "analyses must be transparent, selections and interpretations justified, and value positions made explicit" (2014, p. 312). In the light of this "retroductability" I would endorse Richards' advice to keep impeccable records of changes in codes and coding method "as part of your continuing report on the story of your project" (2015, p. 121).

A related point is the question of defining codes in a framework or other format; "clear operational definitions are indispensable" (Miles, Huberman, & Saldaña 2014, p. 84), whether or not you are thinking of using them as a way of testing reliability. Creating these definitions might be another way to either increase reliability or to improve the validity and credibility of the research as a whole:

> Definitions become sharper when two researchers code the same data set and discuss their initial difficulties. A disagreement shows that a definition has to be expanded or otherwise amended. Time spent on this task is not hair-splitting but reaps real rewards by bringing you to an unequivocal, common vision of what the codes mean and which blocks of data best fit which code. (Miles, Huberman, & Saldaña 2014, p. 84)

While in journal articles coding frameworks and definitions are rarely articulated, it is common to do so in dissertations and theses, and the same warning applies here as to the naming of codes: it is all too easy to write a definition which you consider to be precise and clear, but which in practice is not so self-evident to another researcher (or examiner!).

## Conclusion

There is no clear and agreed upon answer to any of the questions raised above to be found in the literature. To answer any of them must be done in the light of a particular research project, a particular epistemology and/or methodology, specific research questions and with a particular purpose in mind. The answer to some questions about coding can depend pragmatically on the answer to others; they co-exist in a decision-making process that each researcher must undertake for themselves. I hope it may also prompt closer reflection on how the coding process was carried out and what decisions were made in the reporting of methods in future journal articles.

The aim of this article was to raise questions and to disturb assumptions, as well as to give a framework for thinking about some of the most common conundrums for the beginning coder. It is my hope that such disturbances will mean that the next time the reader is designing a qualitative data analysis procedure, if they are experienced, they will at least hesitate before doing as they have always done. For the beginning researcher I appreciate I have given no reassuring guidelines but hope that you now have an idea of what the real questions you need to ask are.

## References

Bernard, H. R. (2011) *Research methods in anthropology: Qualitative and quantitative approaches* (5th ed.). Walnut Creek, CA: Alta Mira Press.

Bernard, H. R., Wutich, A., & Ryan, G. W. (2016). *Analyzing qualitative data: Systematic approaches*. Los Angeles, CA: SAGE.

Bryman, A., & Burgess, R. G. (1994). Reflections on qualitative data analysis. In A. Bryman, & R. G. Burgess (Eds.), *Analyzing qualitative data* (pp. 216–226). London, UK: Routledge.

Corbin, J., & Strauss, A. L. (2015) *Basics of qualitative research techniques and procedures for developing Grounded Theory* (4th ed.). Los Angeles, CA: SAGE.

Creswell, J. (2013). *Qualitative inquiry and research design: Choosing among five approaches*. Los Angeles, CA: SAGE.

Creswell, J. (2015). *30 essential skills for the qualitative researcher*. Los Angeles, CA: SAGE.

Cunningham, D. (2004). *Professional practice and perspectives in the teaching of historical empathy* (Unpublished doctoral dissertation). University of Oxford, England.

Friese, S. (2014). *Qualitative data analysis with ATLAS.ti* (2nd ed.). London, UK: SAGE.

Goodwin, C. (1994). Professional vision. *American Anthropologist*, *96*(3), 606–633.

Harding, J. (2013). *Qualitative data analysis from start to finish*. London, UK: SAGE.

Holton, J. A. (2011). The coding process and its challenges. In A. Bryant, & K. Charmaz (Eds.), *The Sage handbook of grounded theory* (pp. 265-290). Los Angeles, CA: SAGE.

Ingram, J., Elliott, V., Morin, C., Randhawa, A., & Brown, C. (2018). Playing the system: Incentives to 'game' and educational ethics in school examination entry policies in England. *Oxford Review of Education*, *44*(5), 545-562.

Lichtman, M. (2013). *Qualitative research in education: A user's guide* (3rd ed.). Thousand Oaks, CA: SAGE.

Luker, K. (2008). *Salsa dancing into the social sciences*. Cambridge, MA: Harvard University Press.

MacQueen, K. M., McLellan, E., Kay, K., & Milstein, B. (2009). Codebook development for team-based qualitative analysis. In K. Krippendorff, & M. A. Bock (Eds.), *The content analysis reader* (pp. 211-219).Thousand Oaks, CA: SAGE.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods*

*sourcebook* (3rd ed.). London, UK: SAGE.
Punch, K. (2014). *Introduction to social research* (3rd ed.). London, UK: SAGE.
Rapley, T. (2011). Some pragmatics of qualitative data analysis. In D. Silverman (Ed.), *Qualitative research* (pp. 273-290). London, UK: SAGE.
Richards, L. (2015). *Handling qualitative data: A practical guide* (3rd ed.). London, UK: SAGE
Richards, L., & Richards, T. (1994). From filing cabinet to computer. In A. Bryman, & R. G. Burgess (Eds.), *Analyzing qualitative data* (pp. 146-172). London, UK: Routledge.
Ritchie, J. & Spencer, L. (1994). Qualitative data analysis for applied policy research. In A. Bryman, & R. G. Burgess (Eds.), *Analyzing qualitative data* (pp. 173-194). London, UK: Routledge.
Saldaña, J. (2016). *The coding manual for qualitative researchers*. London, UK: SAGE.
Spencer, L. Ritchie, J., O'Connor, W., Morrell, G., & Ormston, R. (2014). Analysis in practice. In J. Ritchie, J. Lewis, C. McNaughton Nicholls, & R. Ormston (Eds.), *Qualitative research practice* (2nd ed., pp. 295-345). London, UK: SAGE.
Spencer, L., Ritchie, J., Ormston, R., O'Connor, W., & Barnard, M. (2014). Analysis: Principles and processes. In J. Ritchie, J. Lewis, C. McNaughton Nicholls, & R. Ormston (Eds.), *Qualitative research practice* (2nd ed., pp. 269-293). London, UK: SAGE.
Syed, M., & Nelson, S. C. (2015). Guidelines for establishing reliability when coding narrative data. *Emerging Adulthood*, *3*(6), 375-387.
Wodak, R. (2014). Critical discourse analysis. In C. Leung, & B. Street (Eds.), *The Routledge companion to English studies* (pp. 302-316). London, UK: Routledge.

## Suggested Resource

Auerbach, C., & Silverstein, L. B. (2003). *Qualitative data: An introduction to coding and analysis*. New York, NY: New York University Press.

## Author Note

Victoria Elliott is Associate Professor of English and Literacy Education at the University of Oxford, UK. Correspondence regarding this article can be addressed directly to: velda.elliott@education.ox.ac.uk.

I am grateful to successive cohorts of Advanced Qualitative Research students in the Social Science Division for their prompting of many of the questions which stimulated the production of this article.



## Article Citation

Elliott, V. (2018).Thinking about the coding process in qualitative data analysis. *The Qualitative Report*, *23*(11), 2850-2861. Retrieved from Retrieved from https://nsuworks.nova.edu/tqr/vol23/iss11/14