Forum Notes:

Coding Issues and Strategies

TONG SENG FAH[1]*, GEETHA SUBRAMANIAM[2] & ILYANA JANIS[3]

[1] Universiti Kebangsaan Malaysia
[2] Auckland Institute of Studies, New Zealand
[3] Universiti Tun Hussein Onn Malaysia

*Corresponding Author: Tsf@ppukm.Ukm.Edu.my

Introduction

Coding is often the first analytical strategy in qualitative studies, and it forms the "building block" for subsequent meaningful analysis. Acknowledging this, the Qualitative Research Association of Malaysia (QRAM) recently held an e-Forum with the theme "Coding Issues and Strategies". This article is based on the E-Forum by QRAM. Three panelists, Prof Tong Seng Fah, Prof. Steven Krauss and Dr Ng Lee Luan discussed and deliberated on three main issues namely:-

- Theory and Coding
- Subjectivity in Coding
- Computer Assisted Qualitative Data Analysis Software.

1.0 Theory and Coding

Two main issues related to theory and coding are: inevitable theory use in coding and use of theory in different stages of coding.

1.1 Theory Use in Coding is Inevitable

Prof. Tong Seng Fah started off by examining the coding process. Coding is basically labelling a meaningful piece of data with a phrase. In order to proceed with this process, firstly, the researchers will need to go back to the research questions, and this is how the researchers will cast actively - a perspective of how they would interpret a piece of data. Secondly, when putting a label to the piece of data, they will have to tap on their prior knowledge, experience and assumptions about what the data means. As Merriam & Tisdell (2018) said "our analysis and interpretation [hence coding] will reflect the constructs, concepts, language, models and theories that structure the study in the first place". Keza (2006) also highlighted that "rather than ignoring these influences, which is not possible, on our coding, we should acknowledge these assumptions explicitly, at least in our reflection, for us to reflect and critique our work accordingly".
The next issue that creates confusion is - what theory constitutes of and how they are defined.

- Theory can be defined loosely (Collins English Dictionary), as a system or rules, procedure, and assumptions used to produce a result. In this context, here the result refers to codes. On the other hand, theory can also be defined more "strictly" as a well substantiated explanation of some aspects of the natural world, with its aims for predictive power or explanatory force (Kazer, 2006, Tavallaei & Talib, 2010). This is more formal and often referred to in scientific writing. Nevertheless, while coding, our intuitive mind does not choose between loose and formal definition. They equally influence our mind in the process of coding.

- With such a wide range of definitions, theories can be classified at different levels in a spectrum (Fig 1). Macro-level theories are grand, general and abstract and loosely connected to what is being observed. They set a broad assumption of knowledge and perspective to the world. They set the proposition for knowledge, the beliefs and the truth about knowledge. Paradigm or worldviews represent this broad assumption at one extreme. The micro-level theories are local and specific. They carry substantive content to explain a specific phenomenon and are closely related to what is being observed. They are more practical to use. These represent the other extreme end within this spectrum. In between these extremes are grand theories and middle range theories, representing different levels of abstracts and reaching to a phenomenon between macro and micro theories.
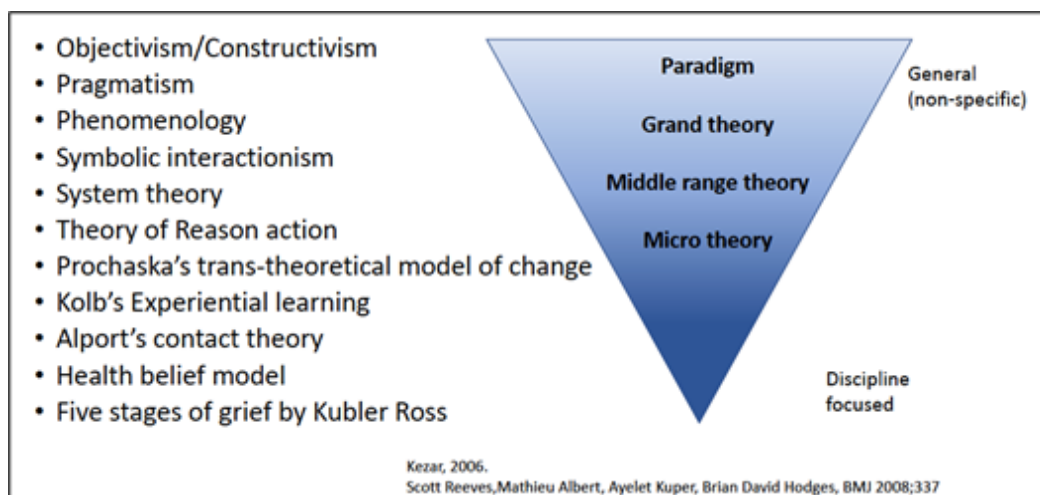

Figure 1: Different levels of theories and examples

- All of us carry with us assumptions about knowledge and to a certain level, depending on our experience, some micro theories. Whether we are aware of such is another debate. As stated earlier, they influence us in our interpretation and coding, albeit at different levels or extent.

1.2 Use of Theory in Different Stages of Coding

As argued above, theory shapes the perspective in analysis. Depending on the choice of methodology and stage of analysis, a good qualitative study needs to adopt different levels of theories.

- For studies with methodologies that slant towards inductive approach, such as grounded theory, narrative inquiry or phenomenology, adopting macro-level theories would be appropriate at the beginning. Whereas, for methodologies that slant towards deductive approach, such as case study and framework analysis, adopting micro-level theories would be of advantage to offer a more critical examination of data.
- Nevertheless, given the inductive nature of qualitative study at an early stage, engaging higher level theories is more appropriate rather than micro-level theories as the latter runs a risk of restricting researchers from openness in coding. As analysis progresses, engaging with micro-level theories would encourage sharper analysis and nuances in the findings. (Fig 2) At this later stage, analysis usually moves toward deductive and abductive approaches. Researchers may cast their own abstracts and theories at this stage, building on to what (theories and abstracts) is already known.
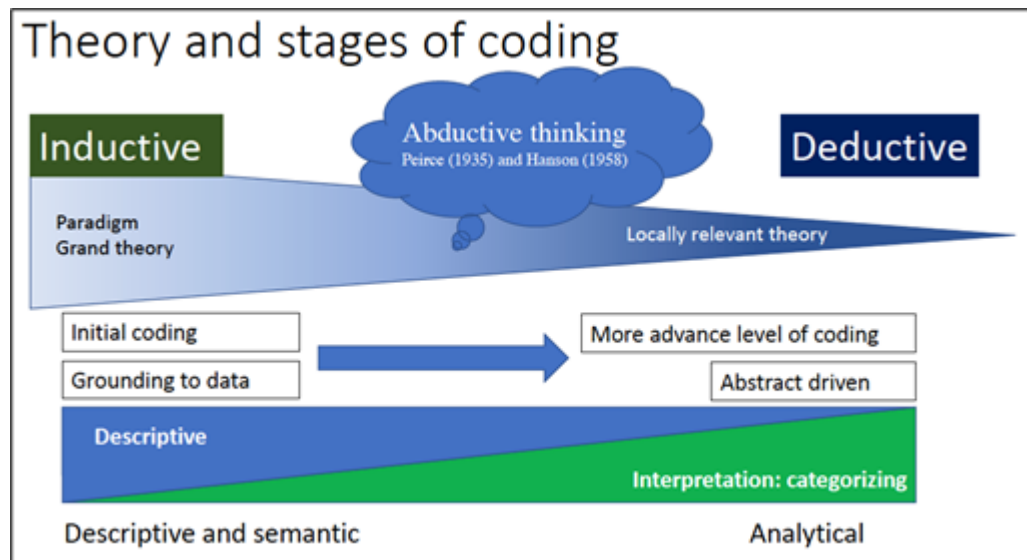
Figure 2: Progression of analysis and its relationship with theory

- According to Collins and Stockton (2018), a study with a low acknowledgment of theory is often simplistic, lacks empirical characteristics and lightly descriptive, or characterised by over emphasis of technique and lacks relevant insights into psychological/social organisation. They are likened to reporting with thin levels of interpretation.

1.3   Do you need a theory for coding?
Hence it can be concluded that qualitative research needs a theoretical lens in place before coding can be done. The role of theory will depend on three factors – that is the choice of methodology; the approach - whether deductive or inductive approach and finally the stage of analysis.

2. Subjectivity in Coding
Coding can be very subjective, and generally, researchers face challenges in terms of definition, interpretation and display.

2.1 Coding – Definition and purpose
Prof. Steven Krauss started his session by defining coding as "…examining a coherent portion of your empirical material - a word, a paragraph or a page - and labelling it with a short phrase that summaries its content" (Linneberg & Korsgaard, 2019, p.3); depending on the methodologies used. This act of interpretation as we code is much dictated by the role of theory.

In qualitative research, this scientific method requires considerable craft and the artfulness of researchers. Regardless of the reason for coding, whether it is discourse, social construction or lived experience, coding enables us to immerse ourselves in the data. This will allow us to pull the data apart where we will discover patterns at different levels within the data, such as putting together a jigsaw puzzle.

2.2 The Dilemma in Coding and how to get around it

Words have meaning, thus, the dilemma one faces is to decide what the words mean? Generally, the coding can be done by using various types of analysis and different types of evidence, which is essentially what qualitative study entails. Codes are like pieces of the jigsaw puzzle that form a larger coherent story later. Coding allows us to make a judgment about the meaning we assign to a piece of data.

Coding is the first phase of interpreting data. However, coding has become ubiquitous throughout the field. Some qualitative research, such as the discourse analysis and hermeneutic phenomenology, may not require coding and thus data is analysed differently. Hence the assumption that everybody needs to do coding in qualitative studies is incorrect. Coding and interpretation are not distinct phases, but they are interrelated processes (Linneberg and Korsgaard, 2019). Prof Krauss further presented two strategies to deal with this issue of subjectivity of interpretation in coding:

1. The first strategy is to be clear of the stance and background of the interpreter. Several scholars including Baptiste (2001) and Saldana (2013) discuss three elements of considerations we need when we begin our analysis of data: philosophical stance, design and context. (See Figure 3).
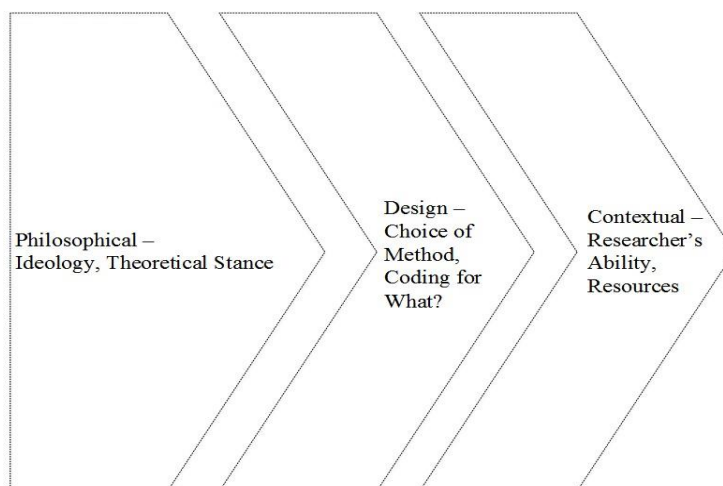

Figure 3: Considerations: Philosophical, Design, Contextual (Baptiste, 2001; Saldana, 2013)

How we see our data and code our data depend on the lenses of the analyst, which are very much coloured by his/her theoretical positions. Hence, the main point here is to be clear about these elements before we embark on coding. The philosophical stance includes personal values and beliefs of ontological, epistemological and ideological stance, and interests. In addition to philosophical and design considerations, research questions are important. Several points need to be considered such as the goal of the analysis; is it to describe or interpret? Is it to identify the structure? Is it to elicit values of our participants? Research questions function as a global positioning system (GPS) for our research, keeping us on track and guiding us throughout the research. They also dictate how we go about coding our data and the choice of method used for coding. Finally, the researchers need to be reflexive of his or her ability and feasibility of coding. This closely relates to their knowledge, skills and resources.

2. The second strategy is having a strong interpretation which can posit us on sound and past literature. Coding is an act of interpretation; it could be subjective, but we need to question whether it is effectively done (Gabriel, 2018). Prof. Krauss further shared several points one should take note of while making a strong interpretation:

i.   When interpreting text, new evidence may emerge which shifts its meaning – interpretations can change its meaning.
ii.   Not all interpretations are equally meaningful or valid - They may be original, clever, perceptive or incomplete, misleading or even plain wrong.
iii. While no interpretation can be 'proven', every interpretation can be corroborated through a variety of techniques.
iv.   Codes should work together to form an overall narrative – internal consistency of an interpretation is paramount, where the interpretation of parts have to be consistent with the interpretation of the whole.
v.   Use of negative cases - The researcher acts as a devil's advocate against their own interpretations, seeking to identify clues that would undermine or discredit their interpretation.
vi.   Can this be interpreted/coded differently? - Eliminating counter-interpretations is an effective way of corroborating an interpretation.

Prof Krauss ended his sharing by reminding us how to become transparent with interpretation. Taking an example of Bowen's coding process, where preliminary line-by-line codes at micro- level, becomes categories and themes, and how the codes fit together with other forms of data, field notes and documents. This demonstrates a triangulation of these sources, and also can portray the whole idea of how the codes and interpretation fit within the 'larger whole' of findings and thus, portray an effective way of being transparent in interpretation and establish a strong interpretation.

3. Computer Assisted Qualitative Data Analysis Software (CAQDAS)
As more tools are available for data analysis, one question researchers' consistently face about CAQDAS is - when should one use CAQDAS. The point highlighted by Dr. Ng is not whether one should use CAQDAS or not, rather how CAQDAS helps and what the possible pitfalls of CAQDAS are?

3.1 The role of CAQDAS in the qualitative analytical process
Dr. Ng discussed the role of CAQDAS in the qualitative analytical process:

- CAQDAS is a useful tool as it supports the coding process when dealing with a large amount of qualitative data. Manual coding can be time-consuming if there is a large amount of qualitative data. Thus, with a CAQDAS, the amount of time needed to deal with manual and clerical tasks associated with qualitative data can be reduced to allow the researchers spend more time in the "real" work of analysis.
- CAQDAS provides tools to undertake more rigorous analysis of data via increased flexibility and thoroughness in data handling.
- Data can be coded in multiple ways – even in the form of audio or video. Also, more than one code can be used for a particularly rich piece of datum. Codes can be developed, merged, deleted and moved as data are analysed.
- The researchers can link the textual data to pdf files, pictures, audios and videos.
- Data can be searched easily and the ideas about data can be linked directly to a particular datum.
- CAQDAS enables researchers to visualise their findings. The pattern can be shown in the form of matrices, charts, word clouds and data clustering. Codes can be cross-matched with demographic data, and frequencies of instances can be counted. The graphical representation of models and ideas also can be developed to facilitate readers' comprehension of data findings.
- CAQDAS provides a more visible audit trail in data analysis as it has its ways of storing memos and journals. Researchers can attach their analytical writing and information about the context of data as they develop ideas.

3.2 Eight possible pitfalls of CAQDAS

Dr. Ng highlighted eight possible pitfalls of CAQDAS:

i.   The focus of analysis on quantity rather than meaning will cause data to be analysed in more "breadth rather than depth". Dr. Ng described this pitfall to the scenario where insufficient and insubstantial analysis could occur when researchers need to deal with large volumes of data.

ii.   The researchers can almost "drown" in the number of codes created as generating codes can be so easily done in most qualitative data analysis software.

iii.   The researchers may face problems as they code and analyse data. They may not be able to see the conceptual "forest" for the coding "trees.

iv.   The "straight jacket" situation may occur, especially when using computers, data is structured before being entered into the computer, making analysis rigid and inflexible.

v.   The researchers may miss the essence of meanings in data as there is an overemphasis on coding and retrieval.

vi.   Code manipulation by researchers may cause the meaning of code to be separated from their contexts.

vii.   Activity and industriousness using CAQDAS cannot be mistaken for good analysis as sorting and filing data thoroughly does not always imply accurate and meaningful analysis. Diligently sorting and filing data does not necessarily equate with valid and meaningful analysis.

viii.   Researchers may face the possible negative aspect of commercialisation whenever a new version of the CAQDAS is introduced. This in turn will result in the researchers spending more time and energy to master the new, improved version of the CAQDAS.

3.3 How should we use CAQDAS wisely?

By being aware of the potentials and pitfalls of CAQDAS, we can choose to use it as a tool to suit our analysis rather than letting it limit our analysis potential. Dr Ng further reminded that the role of CAQDAS should not be confined to be used for data analysis only, but also before and after data analysis as highlighted below:

- **Before data analysis** – the CAQDAS can be used to conceptualise the main relevant ideas and could be used to manage reading and note-taking during the literature review process.
- **During data analysis** – the CAQDAS can be used to maintain collaboration among members via in-text discussion, and also could be used to determine the content for reporting.
- **After data analysis** – the CAQDAS can be used to manage the writing process and could be used to draft specific sections such as Introduction, Method and Discussion etc.

Acknowledgement

We would like to acknowledge the three panellists below for their input to this article.

- Dr Tong Seng Fah is a Professor of Family Medicine at the Faculty of Medicine, Universiti Kebangsaan Malaysia
- Dr. Steven Krauss is a Professor in the Department of Professional Development and Continuing Education at the Faculty of Education, Universiti Putra Malaysia.
- Dr Ng Lee Luan is a lecturer at the University of Malaya's Faculty of Languages and Linguistics and a certified professional NVIVO trainer.

References

Baptiste I (2001) Qualitative Data Analysis: Common Phases, Strategic Differences. Forum: Qualitative Social Research 2(3) pp.1-18 Available at https://www.qualitative-research.net/index.php/fqs/article/view/917/2003

Bowen, G. A. (2008). Naturalistic inquiry and the saturation concept: A research note. Qualitative Research, 8(0), 137-152.

Collins, C. S., & Stockton, C. M. (2018). The central role of theory in qualitative research. International Journal of Qualitative Methods, 17(1), 1-10

Gabriel, Y. (2018). Interpretation, reflexivity and imagination in qualitative research. In M. Ciesielska, & D. Jemielniak (Eds.), Qualitative Methodologies in Organization Studies (Vol. 1, pp. 137-157). Springer International Publishing

Kezar, A. (2006). To use or not to use theory: is that the question? In J. C., Smart (Eds.), Higher Education: Handbook of Theory and Research (pp. 283–344). Springer Netherlands.

Lee, R. M. & Esterhuizen, L. (2000) Computer software and qualitative analysis: Trends, issues and resources. International Journal of Social Research Methodology, 3(3), 231-243.

Linneberg, M. S., & Korsgaard, S. (2019). Coding qualitative data: a synthesis guiding the novice. Qualitative Research Journal, 19(3), 259-270.

Merriam, S. B., Tisdell, E. J. (2016). Qualitative research: A guide to design and implementation. San Francisco, CA: Jossey-Bass.

Saldaña, J. (2013). The coding manual for qualitative researchers. Los Angeles: SAGE Publications.

St John, W. & Johnson, P. (2000). The pros and cons of data analysis software for qualitative research. Journal of Nursing Scholarship, 32(4), 393-397.

Tavallaei, M., & Talib, M. A. (2010). A general perspective on the role of theory in qualitative research. Journal of International Social Research, 3(11), 570-577.