The Art of Coding and Thematic Exploration in Qualitative Research

Michael Williams
School of Business and Management,
Thomas Edison State University, Trenton, NJ, USA

Tami Moser
Department of Pharmaceutical Sciences, College of
Pharmacy, Southwestern Oklahoma State University, Weatherford, OK, USA

[Abstract] Coding in qualitative research is comprised of processes that enable collected data to be assembled, categorized, and thematically sorted, providing an organized platform for the construction of meaning. While qualitative research orientations differ theoretically and operationally relative to managing collected data, each employs a method for organizing it through coding data. Coding methods employ processes that reveal themes embedded in the data, in turn suggesting thematic directionality toward categorizing data through which meaning can be negotiated, codified, and presented. Coding is a key structural operation in qualitative research, enabling data analysis and successive steps to serve the purpose of the study.

This paper focuses on identifying, defining, and describing the coding techniques available to researchers, the function of each stage in the coding method, the iterative review process associated within the coding cycle, and the transition from codes to themes toward constructing meaning from the data. In addition, it references/provides examples of manual coding practices and identifies qualitative research software available for coding.

[Keywords] Coding; thematic exploration; qualitative research

> *Not everything that counts can be counted, and not everything that can be counted counts.*
>
> **----Albert Einstein**

Introduction

Qualitative research provides opportunities to locate the genesis of a phenomenon, explore possible reasons for its occurrence, codify what the experience of the phenomenon meant to those involved, and determine if the experience created a theoretical frame or conceptual understanding associated with the phenomenon. While quantitative research methods seek to count and provide statistical relevance related to how often a phenomenon occurs and then generalize the findings, qualitative research methods provide opportunities to delve into the phenomenon and determine its meaning while and after it occurs. Regardless of the research approach, the methodology employed for data collection and organization must be clear and repeatable, leading to and enabling data analysis. As in any research design, if its data collection and organization methods lack rigor, analysis can be impeded, in turn minimizing the value of outcomes. This approach supports the evolution of constructing meaning from the data, in turn enabling contributions to the related literature and enhancing our understanding of the world.

Context

Authors writing about qualitative research methods1–3 indicate that the evolution of qualitative research has migrated through decades of "methodological consolidation complemented by a concentration on procedural questions in a growing research practice" (Flick, 2009, p.20). Philosophical and methodological yields from this migration are the different orientations and procedures associated with conducting qualitative research. Given the plethora of qualitative research formats available, researchers need to decide which methodological approach will most effectively

enable their study. "Qualitative research is not based on a unified theoretical and methodological concept. The variety of approaches results from different developmental lines in the history of qualitative research, which evolved partly in parallel and partly in sequence" (Flick, 2009, p. 306).

As qualitative research has evolved and methodologies for collecting and organizing data have matured, specific strategies and structures for managing data in these areas have emerged and become common practice. A key data organizing structure in qualitative research is coding. "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" (Saldafia 2009, p.3). Initially, coding began as a progressive three-part schema; First, Second and Third level coding guided by the formula "from codes and categories to theory" (Saldafia 2009, p. 4). From these early coding strategies, additional coding strategies evolved associated with the emergent types of qualitative research methods (e.g., Phenomenological, Grounded Theory). This paper reviews the coding strategies associated with qualitative research using Grounded Theory method as an example for demonstrating the uses of the three-step coding process; open, axial, and selective coding.

Brief Historical Perspectives

The Ground Theory Method (GTM) of qualitative research emerged from the work of Barney Glaser and Anselm Strauss, aligned with the Chicago School of symbolic interactionism, which rose to prominence in the early part of the twentieth century (Glaser, 1967). In their seminal work "The discovery of grounded theory: Strategies for qualitative research," the authors suggested a pluralist and flexible approach to data coding (Strauss, 1998). "Our principal aim is to stimulate other theorists to codify and publish their own methods for generating theory and join us in telling those who have not yet attempted to generate theory that it is not a residual chore in this age of verification'' (Glaser, 1967. p. 8).

While encouraging researchers towards a "pluralist and flexible" orientation to coding and to original methods for "generating theory," the reference to "not a residual chore in this age of verification" is significant, as it is a reference to the quantitative research as being the dominant research method of the time. This initial philosophic frame suggests that researchers employing qualitative research methods, in this case GTM, should not engage in data organizing strategies indiscriminately. Instead, they should apply guiding principles that intentionally enable them to "codify and publish their own methods for generating theory" (Strauss, 1998, p.189). The focus on articulating a clear methodological framework that is both rigorous and able to be replicated, suggests a researcher engaged in qualitative research is using a viable research method. "The pluralistic nature of GTM [does not] mean that researchers can do pretty much whatever they want…there are certain principles about which grounded theory, proponents concur and as long as these principles are kept in mind, the details of the procedure can be modified to suit a researcher's needs" (Larossa, 2005, p. 840).

In this context, the GTM was one of the first qualitative methods to have a systematic approach for codifying and categorizing data in order to generate theory. As a result, researchers were provided with the methodological means to construct meaning from research findings through a three-phase coding method. The coding method enabled a progressive and verifiable mechanism for establishing codes, their origins, relationships to each other, and integration resulting in themes used to construct meaning.

The construction of meaning from collected data is the result of the progressive data coding process. In order for researchers to generate theory, researchers need to evidence employing an analytic approach and rationale methodological decisions. "Through explicating their decisions, grounded theorists gain control over their subject matter and their next analytic or methodological move. The construction of the process, as well as the analytic product, is emergent theory" (Charmaz, 2008, p. 408).

Open, axial and selective coding of collected data results in the creation of theory, leading the researcher to construct deeper theoretical meaning. This method of coding provides researchers with nuanced access to study informants' thoughts, perspectives, and reactions to study topics. Coding enables informant data to be gathered and analyzed relative to "what they do, how they do it, and why they do it interacting in the research setting" (Charmaz, 2008, p. 408).

Coding

Qualitative research generally and Grounded Theory Methods specifically is an inductive, not a deductive, approach to qualitative research. While deductive research focuses on casualty and testing theory, inductive research focuses on generating theory from collected data. In the GTM approach, data collection activities (e.g. interview, observation, and artifact review) requires the researcher to be present and be aware of the dynamic nature of the data, its thematic connectivity, intersectionality, and emergence toward theory creation. "Data collection, analysis and resultant theory generation has a reciprocal relationship…it requires a constant interplay between the researcher and the data" (Charmaz, 2008, p. 47). Central to the coding process is ensuring that coding procedures are defined, rigorous, and consistently applied in order to conform with validity and reliability standards associated with qualitative research. Historically, this orientation of insisting on rigorous data coding procedures is traceable to seminal work in qualitative research, indicating that "joint collection, coding and analysis of data is the underlying operation [toward] the generation of theory" (Glaser, & Strauss, 1967, p. 43).

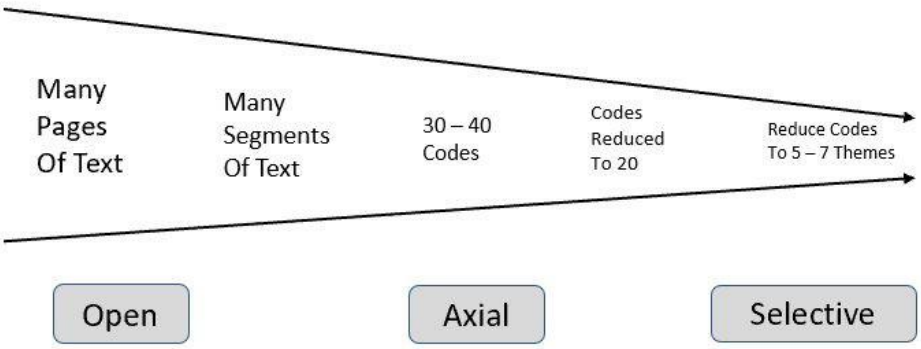

Figure 1. Overview of coding process: Open, Axial and Selective Coding

Recognizing the interdependent relationship among data organization, categorization, and theory development construction of meaning, coding plays a pivotal role in facilitating the researcher's ability to advance effectively the research process. "Coding is oriented around the central concept of [seeking] to represent the interplay of subjects' and researcher's perceptions of the nature and dimensions of phenomena under study" (Douglas, 2003, p. 48).

Importantly, the open, axial, and selective coding strategy enables a cyclical and evolving data loop in which the researcher interacts, is constantly comparing data and applying data reduction, and consolidation techniques. As the coding process progresses, its dynamic function and nonlinear directionality enables essential themes to be identified, codified, and interpreted in the service of a research study's focus and contributes to the associated literature. This cyclical process is both an art and science, requiring the researcher to understand intimately the data by continuously reading and rereading the collected data in order for theory to evolve.

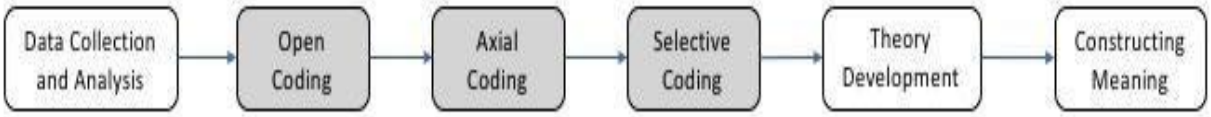

Figure 2. Linear Process for Qualitative Research

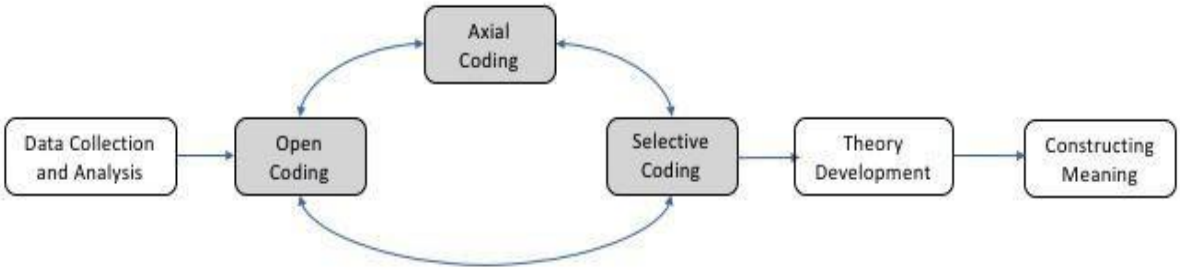

Figure 3. Non-Linear Process: Qualitative Research

Note. The process evolves into a cyclical process when moving between the three coding methods. The researcher must treat these as a non-linear.

Open Coding

Open coding is the first level of coding. In open coding, the researcher is identifying distinct concepts and themes for categorization. The first level of data is organized by creating initial broad thematic domains for data assemblage. "The first step [open coding] aims at expressing data and phenomena in the form of concepts. Units of meaning classifying expressions (single words, short sequences of words) in order to attach annotations and "concepts" (Flick, 2009, p. 307). In open coding, this process was termed the concept-indicator model. In brief, the concept-indicator model used constant comparison of [textual] indicators and focused on comparing regularly occurring textual material. This approach was accompanied by the ongoing coding of themes as an indicator of a concept, always comparing it to previous indicators that had been similarly coded (Saldafia, 2009).

In practice, the researcher needs to sift through informant's responses and organize similar words and phrases, concept-indicators, in broad initial thematic domains. For example, "indicators are symbols or conventional signs, thus a man and woman are concepts. So are love, mate selection, divorce, death, and depression" (Strauss, 1998. p.841).

Central to the efficacy of open coding is approaching the thematic fragments and coalescing concepts identified during data collection in an organized and systematic way. Prior to the use of qualitative research software programs, organizing data for open coding required a multifaceted research skill set. For example, researchers would read and re-read interview transcriptions, field notes, and associated data sources involved in the data collection searching for thematic connectivity leading to thematic patterns. Next, the researcher would color code aligned themes, cut the themes out (producing small paper fragments with the themes), and adhere the paper fragments on index cards in preparation for more precise assessment and axial coding. While this approach was subject to possible errors in overlooking or miscoding them, this rudimentary data organizing strategy could be a relatively effective process enabling open coding.

Today, researchers have the option of using qualitative software to enable the same process using complex data analysis tools. The advent of the use of qualitative software has expanded the ways that researchers can work through the coding cycles. In addition, the more advanced qualitative software packages provide opportunities for statistical analysis overlaying the coding process. The researcher still must move through each phase of coding; the software simply supports an easier capture of the researchers' coding and construction of meaning.

Table 1

Qualitative Software

Software	**Price: Education**	**Operating Systems**		**Statistics**	**Mixed Methods**
		IOS	Windows		
MAXQDA Plus 2018	Faculty: $565.00 Student: $99 (24 months)	√	√	Pro Version: Correlation, Descriptive Statistics, Anova	√
Nvivo Plus	Faculty: $600.00 Student: $85 (24 months)	√	√	Export to SPSS	√
Atlas.ti	Faculty: $670.00 Student: $99 (24 months)	√	√	No	√
QDA Miner	Faculty: $595.00 Student: No	√	√	No	√
HyperResearch	Faculty: $499.00 Students: $199.00	√	√	No	√
Quirkos	Faculty: $340.00 Student: $69.00	√	√	No	√
Dedoose	$10.95 per month $131.40 (12 months)	Web	Web	No	√
webQDA	$50.00 (90 days) $165.00 (12 months)	Web	Web	No	√

Once the researcher determines that a theme has emerged and is recognized, it would be provided with a code. The object is "to arrange things in a systematic order, to make something part of a system or classification…this permits data to be "segregated, grouped, regrouped and relinked in order to consolidate meaning and explanation" (Lincoln, 1985, p.21).

Determining a code for emergent themes from the data can be more art than science. For example, as themes or patterns coalesce, there may be a variety of codes that could effectively corral the themes. However, providing a code prematurely, prior to fully understanding a theme's content and directionality, could hinder its evolving associations with other themes. One approach to choosing a code is employing "classification reasoning plus tacit and intuitive senses to determine which data "look alike" and "feel alike" when grouping them together" (Lincoln, 1985, p.347). Identifying a sufficiently developed theme and determining an appropriate code requires attention to thematic association and a subjective sense of a code's accurate representation of the essence of a theme.

Open coding in qualitative research presents opportunities for sub-coding data. Determining what data to capture and how to display it is a critical aspect of the research design. Data presentation in open coding can be managed in numerous ways. Often, the form of presentation reflects the processes of its collection. For example, words, phrases, or sentence fragments of different emergent themes can be listed on different pages, and field notes counting the number of times a word was repeated in an interview could be graphed or relevant characteristics from photographs of an informant group could be referenced in a multi-photo archive. "The result of open coding should be a list characterizing codes and categories attached to the text and supported by code notes that were produced to explain the content

of codes. These notes could be striking observations and thoughts that are relevant to the development of theory" (Flick, 2009, p. 310).

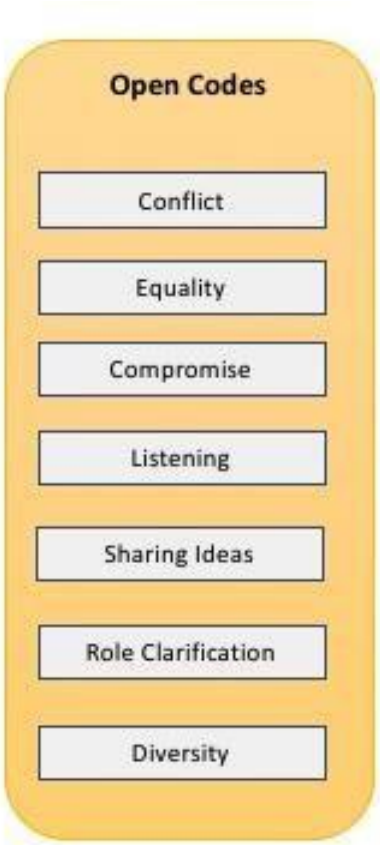

Figure 4. Open Coding: Thematic material identified in reviewing collected data that could serve as categories in axial coding.

A practical approach to determining codes is using the "*5W-1H"* (e.g. who, what, where, when, when, and how) questions as a foundational way for exploring and examining data in order to "list characterizing codes and categories attached to the text" (Flick, 2009, p. 311). This approach enables the researcher to parse and organize thematically similar data so that unique codes can be applied. Code selection can be used in open, axial, and selective coding in order to identify foundational thematic content and directionality (Flick, 2009). Operationally, there are various strategies for displaying data in open coding. Therefore, researchers can be creative and innovative in designing data open coding mechanisms that will best enable their research activities.

Axial Coding

Axial coding is the second level of coding. In contrast to open coding, which focuses on identifying emergent themes, axial coding further refines, aligns, and categorizes the themes. With the completion of open coding and transition to axial coding, collected data can be sifted, refined, and categorized with the goal of creating distinct thematic categories in preparation for selective coding. "Axial coding identifies relationships between open codes, for the purpose of developing core codes. Major (core) codes emerge as aggregates of the most closely interrelated (or overlapping) open codes for which supporting evidence is strong" (Strauss, 1998, p. 109). In order to achieve this organizing objective, researchers need to engage in continuous analysis, cross referencing, and refining theme categorization. There are three refinement activities associated with axial coding that enable and advance effective content categorization.

First is possessing a clear understanding of the analytic methods used in refining data and category

construction. As the researcher reviews the thematic material collected through open coding, the materials must be examined in the context of inductive and deductive analysis. As stated earlier in this paper, deductive reasoning tests theory by collecting and examining empirical data to determine if it is true, while inductive reasoning seeks to construct theory from data collected and analyzed with the goal of explaining research findings. The process of analyzing data is dynamic, requiring the researcher to consider a multivariate field of possible influencers relative to findings. Using inductive and deductive approaches to data analysis can maximize analytic acuity and enable precise thematic categorization. The categories that are finalized from axial coding serve as the axis point or hub in axial coding. For example, a wooden wheel metaphor is used to describe axial coding by locating key categories as the hub and subcategories to the spokes of the wheel (Glaser, & Strauss, 1967). Another description of axial coding is the "Six C's Model." This model encourages categorization using provided key perspectives for further organizing and categorizing data through ''causes, contexts, contingencies, consequences, covariance, and conditions'' (Larossa, 2005, p.98). While these coding activities are associated with qualitative research, considering relevant deductive approaches to testing theory remains an important data comparison strategy facilitating continuous review, reconsideration, and reflection.

Second is the constant comparison method. The constant comparison method is a data organizing and refining activity. While there are differing approaches to implementing the constant comparison method, its focus is to compare continually data collected, emergent themes, and their coding in order to continually create, refine, and newly create categories in preparation for selective coding. Thematic comparison and analysis are central to axial coding, as the critical focus is on organizing themes into cogent and comprehensive categories. In axial coding, as well as the other coding types, the researcher must understand the function of the coding and associated analytic activities in order to make informed research designs. ''Bringing process into the analysis is an important part of any grounded theory study" (Strauss, 1998, p. 163). This is an important understanding relative to the relationship between coding and analysis as analysis facilitates coding.

Third is "line-by-line" coding. In line-by-line coding, each textual line of an interview or document is scrutinized with the goal of maintaining the researcher's focus on the text. Through this approach, the researcher can deeply engage the text, and, in turn, recognize and codify nuances and discrete thematic connectivity with other themes. "Researchers do not want to impose a pre-existing framework onto the data, but rather to let new themes emerge from it. Through keeping 'close to the data' continuously sifting through themes, idea fragments and seemingly unrelated utterances, data categories can become thematically stabilized, defined and differentiated" (Charmaz, 2014, p. 80). Remaining "close to the data" requires the researcher to immerse herself in the text, explore its nuances and surrender biases. By progressing engaging data "line-by-line," a rhythm or cadence of analysis occurs, assisting the researcher in being methodical and, perhaps, pedantically focused on identifying textual subtleties fueling the construct of meaning. Through employing inductive and deductive reasoning, the constant comparison method, and line-by-line coding, the integrated essence of the thematic material can be identified and categorized.

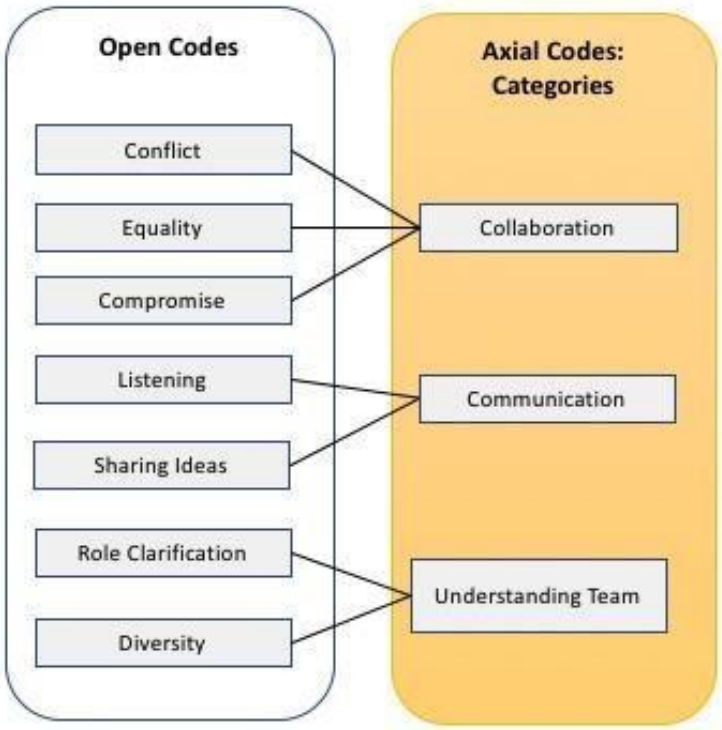

Figure 5. Axial Codes: Creating categories from open codes

In open coding, themes are being developed: an informant's musing, a hand-written note in the margin of a memo, or an elaborate monologue from a spontaneous utterance from an unscheduled actor. However, the interrelatedness of the thematic material remains unexplored and unanalyzed. In axial coding, the relationships among themes are explicitly stated, examined, and categorized. "If the development of theory rests heavily though not entirely on explanation and if explanation rests on how variables and their interrelatedness are empirically or logically established, then axial coding is the phase in which research begins to fulfill its theoretical promise" (Bengston, 2006, p.28).

Selective Coding

Selective coding is the third level of coding. It enables the researcher to *select* and integrate categories of organized data from axial coding in cohesive and meaning-filled expressions. "Selective coding continues the axial coding at a higher level of abstraction [through] actions that lead to an elaboration or formulation of the story of the case" (Flick, 2009, p. 310). Central to enabling the story or case to emerge from the data categories is the process of enabling further refinement of the data, selecting the main thematic category, and then in a systematic manner aligning the main theme to other categories that have been selectively coded. "The conceptualization of the yield from the selective coding as a 'case' or 'story' is significant as it provides researchers with flexible and multi-type vehicles for codifying and presenting study results" (Strauss, 1998, p. 158). This approach to data framing enables the researcher to work continually toward thematic specificity and, in turn, theory creation. In selective coding, degrees of causality or predictability can emerge from the thematic refining process, allowing the researcher to identify sets of circumstances in which certain responses will elicit responses that suggest certain circumstances receive unique and differentiated responses. With the work of selective coding done, the researcher can move toward developing theory and ultimately constructing meaning.

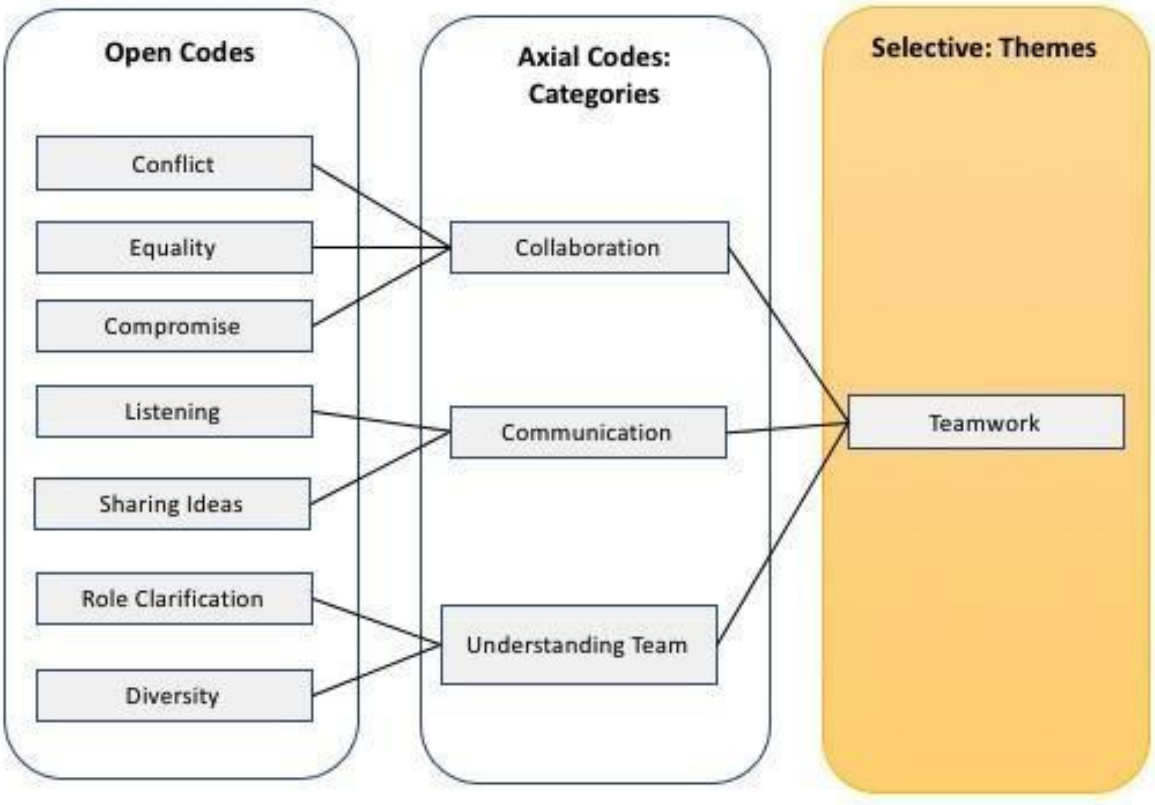

Figure 6. Open Codes to Selective Theme

Note: Typically, numerous thematic fragments in the collected data can be captured in open coding, enabling as if thematic materials to be integrated forming key categories using axial coding that are then further refined to one unique theme in selective coding. Frequently, research studies yield numerous selective codes; enabling researchers develop theory resulting in a theoretical framework and the ability to construct meaning.

Selective coding is a uniquely challenging phase of the data collection process of the research design in that it influences not only what theoretical constructs emerge, but also how meaning is created through presentation, impacting the reception of the findings. For example, if the meaning of the findings is expressed in the form of a case of story, the presentation can vary in form and style. Authors write about the format when finalizing the outcome of selective coding process. "Some researchers may prefer to tell idiographic stories, with anecdotal indicators fleshing out the particulars, while others may lean toward theoretical stories, accounts of how a complex of variables are interrelated" (Larossa, 2005, p. 201). The outcome of selective coding enables researchers to craft case stories that accurately and powerfully present the sum of the progressive coding process. Selective coding fuels expression and facilitates the construction of meaning.

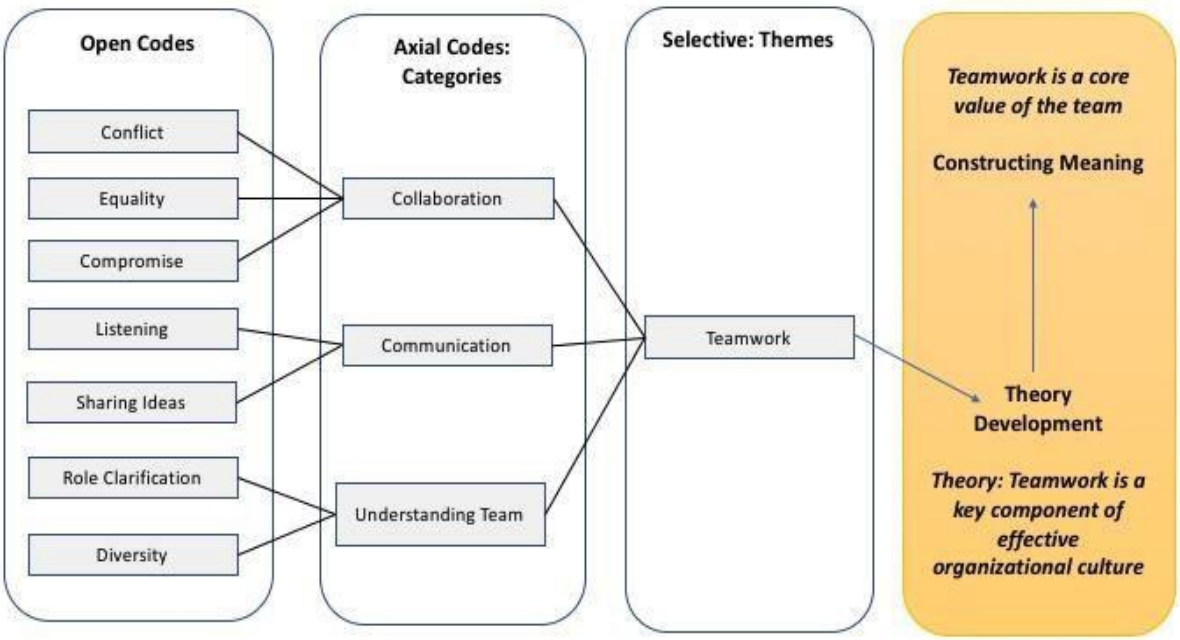

Figure 7. Creation of Theory and Meaning

Note. The coding process outcome can be theory development and the construction of meaning associated with the research purpose.

Coding in qualitative research enables researchers to identify, organize, and build theory. The roles of open, axial, and selective coding are critical to achieving the research goals of a study, as they provide opportunities for researchers to immerse themselves in the data. Coding promotes thematic integration and organizational strength, enabling researchers to be reflective and reflexive in joining the data in nuanced and intimate ways and employing the outcomes from the coding process to create meaning.

Conclusion

Coding in qualitative research enables researchers to identify, organize, and build theory. The roles of open, axial, and selective coding are critical to achieving the research goals of a study, as they provide opportunities for researchers to immerse themselves in the data. Each stage of the coding process progressively integrates the emergent themes acquired during data collection and continually refines the themes culminating in theory development and the creation of meaning.

References

Flick, O., (2009). *An Introduction to Qualitative Research:* Sage Publications.

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

Glaser, B.G. & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Saldafia, J. (2009). *The coding manual for qualitative researchers*. London: Sage Publications.

Strauss, A. (1998). *Qualitative analysis for social scientists*. New York: Cambridge University Press.

Larossa, R. (2005). Grounded theory methods and qualitative family research. *Journal of Marriage and Family, 67*(4).

Charmaz, K. (2008). Constructionism and the grounded theory. In: Holstein JA, Gubrium JF, eds. *Handbook of constructionist research*. New York: The Gilford Press, 397-412.

Douglas, D. (2003). Inductive theory generation: A grounded approach to business inquiry. *The*

Electronic Journal of Business Research Methods (EJBRM).
Lincoln. Y.S. & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage Publications.
Charmaz, K. (2014). *Constructing grounded theory: A practical guide through qualitative analysis.* 2nd ed. Thousand Oaks, California: Sage Publications.
Bengston, A., Allen. D., Anderson, P., & Klein, O. (2006). *Sourcebook of family theory and research.* 1st. London: Sage Publications.