Journal of Ethnographic & Qualitative Research
2023, Vol. 17, 208–220 ISSN: 1935-3308

Demystifying Qualitative Data Analysis: Unveiling Strategies and Insights from NSF Proposals

Roli Varma
University of New Mexico

Frequently, I have found myself assigned to review papers grounded in qualitative data that have been submitted for publication. Sometimes, these papers have appeared to originate from authors who have recently obtained their Ph.D. or from students who are nearing the completion of their doctoral studies. In many instances, I have provided feedback to support the presentation of the qualitative data, as well as guidance on how to properly clean and organize it. Analyzing qualitative data poses a considerable challenge in qualitative research. Books on research methods often remain detached from real hands-on activities. In this article, my objective has been to show how qualitative data analysis is executed in practical scenarios. Rather than creating fictitious examples, I have provided a step-by-step account of my own research process from diverse projects that have received support from the National Science Foundation (NSF). This approach has offered the advantage of real-life examples. Publishing in peer-reviewed journals has subjected these findings to external validation and review. By sharing my experiences and techniques, I have hoped to offer valuable insights to emerging researchers in effectively analyzing qualitative data.

The formalization and recognition of qualitative methodology as a distinct research approach began to emerge in the early 20th century (O'Reilly, 2009). In the field of anthropology, the work of scholars such as Franz Boas, Bronisław Malinowski, and Margaret Mead in the late 19th and early 20th centuries laid the foundation for ethnographic research and participant observation, which are key qualitative methods. In sociology, qualitative research gained prominence during the mid-20th century with the Chicago School of Sociology and the development of methods such as in-depth interviews, participant observation, and content analysis (Erickson, 2018; Vidich & Lyman, 2000). In the latter half of the 20th century, qualitative research in communication, education, political science, psychology, public administration, and other social sciences gained momentum, with researchers employing qualitative approaches to study human development, learning, and subjective experiences (see, Denzin & Lincoln, 2018). Since then, qualitative methodology has continued to evolve and expand, incorporating various theoretical frameworks, methods, and

Roli Varma, Ph.D., is Full Professor of Public Administration at University of New Mexico in Albuquerque, New Mexico.

Correspondence regarding this article should be sent to varma@unm.edu.

This research was partially supported by the National Science Foundation (Grant 1655322).

This is an expanded version of the talk titled, *Qualitative Data Analysis: Some Tips for Early Career Academicians,* given to graduate students in the Department of Management, Indian Institute of Technology-Delhi, India, on January 11, 2023.

data collection modes. Since the 1990s, qualitative research has become an established and widely used methodology across disciplines, providing valuable insights into human behavior, social phenomena, and lived experiences.

Most contemporary books on research methods emphasize that qualitative and quantitative methodologies are distinct approaches to research, differing in their goals, data collection, and data analysis (Creswell & Creswell, 2018). Qualitative research focuses on collecting and analyzing non-numerical data, such as words, observations, interviews, and textual materials, with the aim of understanding the complexities, meanings, and interpretations of phenomena. In contrast, quantitative research deals with numerical data obtained through structured measurements, surveys, questionnaires, experiments, and statistical analysis, aiming to quantify relationships, patterns, and trends in a population or sample. Qualitative research often explores open-ended research questions, seeking an in-depth understanding of social phenomena, experiences, or perspectives, and investigating the "why" and "how" behind human behavior and social processes. On the other hand, quantitative research typically addresses research questions that can be answered through statistical analysis, with the goal of establishing relationships, predicting outcomes, or testing hypotheses using numerical data. Qualitative research frequently employs non-random sampling to ensure that participants are relevant to the research topic, while quantitative research commonly utilizes random sampling techniques to ensure the representativeness of the target population. In terms of data analysis, qualitative research involves interpreting and making sense of the collected data through processes such as thematic analysis, content analysis, or grounded theory, with a focus on identifying patterns, themes, and emergent insights. In contrast, quantitative data analysis employs statistical techniques to analyze numerical data, including descriptive statistics, inferential statistics, correlations, or regression analysis, with the aim of identifying patterns, trends, and relationships in a systematic and objective manner.

Despite the establishment of two distinct methodologies, there has been a polarization between qualitative and quantitative researchers (Onwuigbuzie & Leech, 2005). Quantitative research has traditionally been hailed as scientific and objective, while qualitative research has often been criticized as biased, anecdotal, and lacking rigor (Hagger & Chatzisarantis, 2011; Hammersley, 2008). One of the primary sources of this perception is the distinction between deductive and inductive reasoning. Quantitative research is closely associated with deductive reasoning, wherein hypotheses are tested based on existing theories, and data is collected to either confirm or reject these hypotheses. In contrast, qualitative researchers frequently employ inductive reasoning, where theories or concepts are developed from the data itself. This distinction has led some to perceive quantitative research as more scientific due to its deductive approach. For instance, Karl Popper (2001) stated: "Induction, i.e., inference based on many observations, is a myth. It is neither a psychological fact nor a scientific procedure" (p. 52). Consequently, some argue that qualitative methods lack the crucial criterion of falsifiability necessary for scientific inquiry, making them susceptible to subjective interpretations and unverifiable claims.

My goal for this article is not to conduct a comparative study of two methods or show the superiority of one method over the other. I write with the assumption that both qualitative and quantitative methodologies have their strengths and limitations, with the choice between them depending on the research objectives, the nature of the phenomenon under study, available resources, and the researcher's preferences and expertise. Since the 1980s and early 1990s, there has been a growing recognition of the value of mixed methods research, which combines elements of both qualitative and quantitative approaches (Creswell & Creswell, 2018; Maxwell, 2015). Ultimately, the choice between qualitative and quantitative research is guided by the research question and the specific needs of the study.

Since the mid-1990s, I have accumulated extensive information by employing qualitative data collection techniques on various topics, including the career paths of industrial scientists and engineers (Varma, 1996-1998), the experiences of minority women in computer science and computer engineering (CS/CE) education (Varma & Kapur, 2003-2007; Varma et al., 2001-2003), the challenges faced by women in CS/CE education in the United States and India (Varma, 2007-2008), the integration of Indian immigrant scientists and engineers into the U.S. workforce (Varma, 2002-2005), the dynamics of the tenure system in engineering (Varma, 2006), and the phenomenon of scientists and engineers returning from the United States to India (Sabharwal & Varma, 2012-2015; Varma & Sabharwal, 2017-2021). These research endeavors have been supported by funding from the National Science Foundation (NSF). In this article, my aim is to illustrate the techniques I have employed to analyze qualitative data from these studies. I believe that the insights shared in this article will be particularly beneficial to emerging researchers who are embarking on their own research journeys to employ qualitative research.

Nature of Qualitative Data

Qualitative data is typically collected through methods such as interviews, observations, focus groups, and open-ended survey questions. I have mostly conducted in-depth interviews which were audio recorded. Additionally, I have collected open-ended survey responses and conducted some focus group discussions. Although I have audio recorded, I have also taken detailed notes during fieldwork. I have hired university students to transcribe the interviews verbatim; with the financial support from the NSF, this practice has become feasible. When I have not have a support from the NSF, I have transcribed my own data (e.g., for my PhD dissertation). My qualitative data has consisted of verbatim records of interviews, focus group discussions, open-ended texts from survey responses, and detailed field notes. I have created electronic duplicates, keeping the original intact, and worked on the duplicates.

Cleaning Qualitative Data

The first step I have taken in analyzing qualitative data has been to clean the transcribed data. In qualitative research, data cleaning refers to the process of preparing qualitative data for analysis (Saldaña, 2011). I have ensured that all data elements have been present and that there have been no missing or incomplete sections. If any parts of the data have been missing, I have made every effort to retrieve them. In qualitative research, transcribed data often contain missing elements attributed to various factors, including transcriptionists not looking at their screens, challenges in accurately hearing audio, typographical errors due to keyboard misplacement, and difficulties in comprehending accents, jargon, or lengthy sentences. Then, I have formatted the data to have standard spellings, punctuation, and capitalization. Most importantly, I have removed identifying information to comply with Institutional Review Board (IRB) requirements and protect the privacy and confidentiality of participants. I have replaced names, locations, or any other identifying details with pseudonyms or generic terms. I have also assessed whether all portions of the transcript have been usable, and if not, I have deleted irrelevant and unusable sections. Additionally, I have maintained a detailed record of the data cleaning process, including any decisions I made or changes I implemented. I have found that cleaned data has made analyzing and deriving meaningful insights easier. Although hired university students have performed the transcription, I have not involved any student in the data cleaning process. Instead, I have handled data cleaning personally to ensure that important information has been preserved and that intended meanings of the sentences have not been altered.

As an example, in my tenure project (Varma, 2006), I posed the following straightforward question to participants: "What does tenure provide?" For me, passages of text in bold were not necessary, so I deleted them to clean the data. I put three dots to remind myself that some passages of text had been deleted and the sentence was no longer in the verbatim form. The following example represents a passage of data prior to cleaning:

> Pseudonym Mary: Tenure **means, I guess the big part, it will be my guess.** I haven't really had a great taste of it yet **because I was just told**. But to me it's really why you get promoted to associate. **Basically, you have**, one of my colleagues was saying, **you know**, once you get tenure you can tell them shove it. They can't **basically you know** fire you **or whatever. I guess that's mostly (unintelligible)**. So, **it seems** tenure provides secure permanent employment till you retire.

The following passage of text represents the same data after cleaning:

> Pseudonym Mary: Tenure...I haven't really had a great taste of it yet...But to me it's really why you get promoted to associate...One of my colleagues was saying...once you get tenure you can tell them shove it. They can't...fire you...[T]enure provides secure permanent employment till you retire.

In addition to open-ended questions during the interview process where participants can choose how to respond and the level of detail they provide, I have also included semi-closed questions where participants have a more limited range of responses. Even with such questions, it has been necessary to clean the data since individuals may express the same ideas in different ways. Therefore, I have standardized their responses to ensure uniform formatting across all data points. For example, in my return migration project (Varma & Sabharwal, 2017-2021), I asked participants how long they had been in the United States. Responses varied; some provided the number of years, while others were more precise, including years and months. Some respondents gave the exact dates of their arrival in the United States, while others provided the duration from their arrival year to the year of data collection, and so forth. Standardization in this case involved expressing all these responses in the same unit, that is, years spent in the United States.

Organizing Cleaned Qualitative Data

After completing the data cleaning process, the subsequent step is data organization. The decisions made in this phase depend on the researchers' objectives and the resources available. Researchers need to determine the formatting of the cleaned data, whether to present responses question by question or group them based on individual cases, groups, or sites (Saldaña, 2011). In essence, data organization empowers researchers to cut and arrange the data as required.

After the data cleaning process, I have hired university students to organize the data based on the questions presented to each participant, making it more manageable for analysis. Depending on the nature of my project, they ensure that there is a clear way to identify the source of all the data, such as by individual, location, gender, ethnicity, and other relevant factors. For each question, they create a dedicated file that is easily identifiable based on its content. Subsequently, they revisit each file and establish a data structure. When dealing with smaller amounts of text, they use spreadsheets, and for larger datasets, they opt for overview charts to visualize the data. Columns represent the subjects, while rows represent data points. I have found this approach useful as it facilitates data comparison.

As an example, I conducted interviews with 150 undergraduate students from diverse institutions, ensuring an equal distribution across gender and ethnicity/race (Varma & Kapur, 2003-2007). I inquired whether they had ever considered changing their major from

Figure 1

A Sample of Organizing Cleaned Qualitative Data

Site	White Male	White Female	Black Male	Black Female	Hispanic Male	Hispanic Female	Asian Male	Asian Female	Native American Male	Native American Female
Hispanic Serving Institution A	Clean data text	Clean data text			Clean data text	Clean data text	Clean data text	Clean data text	Clean data text	
Historically Black College			Clean data text	Clean data text			Clean data text			
Tribal College									Clean data text	Clean data text

Note. Clean data text refers to actual interview quotation.

computing to another field and, if so, the reasons behind their considerations. Subsequently, my hired university students organized the data into two distinct overview charts: one for respondents who answered "yes" and another for those who answered "no." I stored the organized data in two separate files, presented in a grid-like format as shown in Figure 1.

It is important to emphasize that organized data should be stored in a structured format that is both easily accessible and maintainable. This entails creating separate folders or directories for different categories or themes, adopting consistent file-naming conventions, and maintaining a record of the location of each data segment. The chosen data storage system should enable effortless retrieval and referencing during the analysis process.

Coding Qualitative Data

Coding qualitative data means naming segments of data with a label that instantaneously classifies and summarizes transcribed data. It is the process of categorizing and organizing textual or non-numeric information to identify patterns, themes, or concepts (Saldaña, 2011). Coding assists in making sense of a large amount of unstructured data. It brings structure to unstructured data, making it more manageable and enabling systematic analysis. Assigning codes is a fundamental step that allows researchers to capture the essence and content of each qualitative response, facilitating a deeper understanding of the data.

While I often start with some a priori concepts derived from literature review or theoretical frameworks, I remain open to the emergence of in vivo concepts that arise directly from the data itself. First, I immerse myself in the data, conducting a thorough review to develop a comprehensive understanding of its content. I begin with a detailed line-by-line analysis (see, Charmaz, 2014). I annotate sentences with question words such as what, when, where, who, why, how, and so on, to uncover their underlying meanings. With this process, I segment data into initial codes. Below is an example from the tenure project mentioned in the section on Cleaning Qualitative Data. I prefer to have as many initial codes as possible because they can be combined later to facilitate a higher level of inference. In response to the question, what tenure gives them, participants gave the following responses:

> Pseudonym Mary: Tenure...I haven't really had a great taste of it yet...But to me it's really why you get promoted to associate (initial code: promotion). ... One of my colleagues was saying...once you get tenure you can tell them shove it (initial code: freedom of expression). They can't...fire you (initial code: permanent job)...[T]enure provides secure permanent employment till you retire (initial code: job security).
>
> Pseudonym Alexi. Tenure to me means stability. It means security (initial code: job security). It means that you feel free to be able to express your thoughts (initial code: freedom of expression), show your research without fear of ramifications or punishment or retribution. It's freedom to do your job (initial code: freedom in research).
>
> Pseudonym Kate: It certainly gives you peace of mind (initial code: job security) but people take too much advantage of tenure (initial code: abuse of tenure). I am not sure that I believe in tenure anymore (initial code: questioning tenure). I've seen too many people sort of just relax & don't work hard after tenure (initial code: low productivity).

Above example shows that I have broken down the data into smaller, meaningful segments or labels. I have identified words, phrases, or sentences that conveyed important information which I have represented with initial codes.

For this process of initial coding, I enlist the help of two graduate students, made possible through financial support from the NSF. I provide these students with key articles that contain theories and concepts related to the project at hand and ask them to read and familiarize themselves with the material. For each question, these students separately generate initial codes, after which we compare codes and engage in discussion if there are any inconsistencies. We collectively develop what is referred to as a coding frame (Saldaña, 2011). This frame comprises a list of initial codes, each accompanied by a definition. Once we reach an agreement on the initial codes, the students take over the subsequent tasks, systematically applying the coding frame to the data and conducting the necessary counts. They separately use spreadsheets for this purpose. In the spreadsheet, they tally how many times each initial code occurred and who the participant was. Therefore, the spreadsheet includes the total occurrences of initial codes. Based on my research question(s), the spreadsheet will also include the occurrences of these initial codes categorized by characteristics such as gender, race/ethnicity, age, place of employment, and other relevant variables. It is important for me that students count the frequency of each initial code as doing so demonstrates their importance. If participants provided multiple responses, students make a note of the total occurrence of each initial code as well as the first or primary initial code mentioned by a participant. These two measures help illustrate the significance of initial codes, and the one mentioned first or is primary can later be used for statistical analysis.

To assess the consistency of initial coding between the two students, I typically employ percentage agreement to calculate the proportion of coding on which the coders agree. I simply divide the number of aggrements by the total number of agreements and disagreements. My goal is to achieve a high level of agreement, usually 80% or above, although the acceptable threshold may vary depending on the project. This process, known as inter-coder reliability, measures the agreement or consistency between two or more coders when independently coding the same data (Creswell & Poth, 2018). There is a general agreement that achieving inter-coder reliability is crucial because it provides an indication of the consistency in coding or rating procedures, which, in turn, contributes to the trustworthiness of a study's findings. It is important to recognize that achieving perfect inter-coder reliability is rare, and some level of discrepancy or variability is to be expected.

Why do I generate initial code and hire two students to do the same thing, and why, once the coding frame is developed, do I ask students to separately count the occurrence of initial codes? I do this to mitigate researcher's bias, which refers to the influence of the researcher's personal beliefs, values, attitudes, or expectations on the design, implementation, and interpretation of a research study (Saldaña, 2011). It occurs when researchers unconsciously or consciously may count data in a way that confirms their preconceived beliefs or desired outcomes, overlooking alternative explanations or contradictory evidence. This can lead to a one-sided understanding of the topic, subjective interpretations and biased conclusions. To me, counting initial codes demonstrates where the possibility of the researcher's bias can enter; thus, I allow students to do this task. Since a student can miss counting, I use two of them separately to do the counting. My simple logic is that when two students independently arrive at the same data count, it instills confidence in the analysis. It demonstrates the meticulous efforts taken to ensure the credibility of the analysis. However, I actively participate in generating initial codes alongside students to ensure that the nuances and key elements of the data are accurately captured. This way, I maintain control over the quality and comprehensiveness of the coding process, especially when dealing with students who may have a limited understanding of the subject matter. I believe this contributes to the rigor and accuracy of the qualitative research.

In cases where hiring independent coders is not feasible, which is often the case for graduate students working on their dissertations or researchers with limited financial resources, it becomes imperative for the researcher to actively work to minimize their bias or influence on the coding process. It is worth noting that a significant proportion of qualitative data is coded by a single coder (Cambell et al., 2013). Qualitative researchers have proposed various steps to substantiate the credibility of the coding process (Bauer & Gaskell, 2007). When I have not had support from the NSF, I have personally conducted the coding of my data. I have allocated some time, recoded the same data, and then compared the two sets of codes.

Initial codes serve as building blocks for the development of themes or concepts—abstract representations of events, objects, or actions that are deemed significant within the data (see Babchuk, 2019). For example, in my project on return migration of faculty to India (Sabharwal & Varma, 2012-2015), I directly asked returnees to identify the primary reasons for moving back to India. The initial coding process by graduate students resulted in the following 21 initial codes, which are noted under (i) total responses, i.e., multiple reasons given to return, and (ii) primary response, i.e., the main reason given to return:

1. Attractive job opportunities in India (total n = 16, primary n = 13)
2. Ample internal resources to support research in India than in the U.S. (total n = 9, primary n = 7)
3. High competition to obtain research grants in the U.S. than in India (total n = 12, primary n = 8)
4. Support for basic research in India than in the U.S. (total n = 9, primary n = 6)
5. Better job security in India than in the U.S. (total n = 7, primary n = 5)
6. No relationship between research productivity and time to obtain tenure in India (total n = 5, primary n = 0)
7. Access to better students in India than in the U.S. (total n = 3, primary n = 0)
8. Emphasis on being good teachers in India than in the U.S. (total n = 3, primary = 0)
9. Problems with getting permanent residency in the U.S. (total n = 17, primary = 13)
10. Delay in processing immigration paper in the U.S. (total n = 14, primary n = 6)
11. Spouse could not get a job due to dependent visa status in the U.S. (total n = 4, primary = 0)
12. Difficulty bringing immediate family members to the U.S. due to temporary visa (total n = 3, primary n = 0)
13. Always wanted to come back to India (total n = 7, primary n = 5)
14. Cultural/social association with India (total n = 6, primary n = 3)
15. Preference to raise children in India (total n = 4, primary n = 3)
16. Inability of spouse to adjust in the U.S. (total n = 2, primary - 0)
17. Preference for the Indian education system (total n = 2, primary n = 0)
18. Obligations to care for aging/ailing family members (total n = 13, primary n = 9)
19. To be with extended family (total n = 2, primary n = 0)
20. Desire to contribute back to India (total n = 6, primary n = 3)
21. Love for India (total n = 3, primary = 2)

After the initial coding, I move on to the next crucial step: identifying themes and concepts within my data. This stage involves a more refined and systematic approach to data analysis, with the aim of uncovering commonalities, similarities, and recurring patterns. While the initial coding is relatively loose and rapid, this subsequent step delves deeper into the analysis of the initially coded data to discern patterns and progress toward the development of final concepts and theories. At this stage, the objective is to streamline the coding process by reducing the number of codes generated during the initial phase. It also involves decisions on how to categorize these pre-existing codes effectively. I handle this task myself, grouping together initial codes that exhibit similarities or appear related to the same overarching themes and concepts. I do not involve students in this level of coding because they may not be familiar with my theoretical commitments to the project. As an example, the 21 codes mentioned above were conceptualized into the following five themes or concepts (Sabharwal & Varma, 2017):

1. Better Career Prospects in India: this category combined codes indicating that Indian institutions offer better opportunities for scientists and engineers to secure internal and external funding; faculty have more freedom to pursue curiosity-driven research; there is less pressure to achieve tenure quickly; there is a weaker relationship between research productivity and the time required to achieve tenure; excellence in teaching is valued; institutions provide robust support for students; there is flexibility in the timeline to establish a record for advancing from assistant professor to associate professor; and job security is guaranteed.
2. Immigration Problems in the U.S.: this category merged codes showing returnees feeling like second-class citizens in the United States due to lengthy immigration processing times; spouses without work visas face challenges in

joining the U.S. workforce; and difficulties in sponsoring immediate family members for visit to the country.

3. Indian Cultural Identity: this category joined codes that reflected an affinity for Indian culture and lifestyle; belief that India is a true representation of returnees' identities; the importance of raising their children within an Indian cultural framework; and spouses' preference for Indian over American society due to their heritage.
4. Family Reunification: this category put together codes related to returnees' concerns for aging or ailing parents in India; the desire of their family in India for them to return; and the importance of socializing with extended family members.
5. Indian patriotism: this category combined codes which expressed a strong sense of Indian nationalism among returnees; consideration of the stay in the United States as temporary; and altruistic intentions to give back to their homeland.

The above example shows that initial codes assisted in breaking the data into smaller, more manageable units. Concepts helped in simplifying complex information and identifying recurring themes and patterns. They made it easier to convey to others the essence of the data and the insights gained from the analysis.

Employing Tools for Qualitative Data Analysis

Since the 1980s, researchers have utilized various qualitative data analysis tools to organize, process, and analyze data. These tools have continuously improved and offer a range of functionalities. On one end, software programs enable users to tag and highlight important aspects of their research. On the other end, advanced software tools leverage artificial intelligence to expedite tasks such as tagging, analyzing, and visualizing data. The range of qualitative data analysis software tools provide features for coding, categorizing, and retrieving qualitative data. They enable researchers to establish hierarchies of codes, link codes to specific data segments, and generate reports or visual representations of the data organization.

Several popular qualitative data analysis software options include ATLAS.ti, Dedoose, MAXQDA, MonkeyLearn, NVivo, QDA Miner, Quirkos, and Taguette. Because each such tool has its own strengths and weaknesses, researchers should select the one that best suits their needs. Personally, I have used various versions of NVivo as it suits my work with smaller data sets. Despite utilizing NVivo, I continue to analyze data manually (see Creswell & Guetterman, 2019), as the elaboration on the coding scheme in the "Coding Qualitative Data" section exemplified. I believe manual coding provides in-depth understanding of the data. Often, I use a combination of both, manual and NVivo to make the most of qualitative data analysis.

Presenting Qualitative Data

Presenting qualitative data requires careful organization and clear communication to effectively convey the findings (Saldaña, 2011). I begin by clearly identifying the purpose of my presentation, whether to provide a comprehensive overview of the data or to focus on specific themes. To enhance understanding and engagement, I select appropriate visual aids that complement my qualitative data. Common options include charts, tables, diagrams, or images that visually represent the information I am presenting. For key findings, I highlight the most significant and interesting insights from my qualitative analysis by using direct quotes, anecdotes, or compelling examples to illustrate my points and ensure they represent the data. I enure that concepts or patterns are supported by multiple but diverse interview excerpts that demonstrate consistency and strengthen the overall argument. I additionally select passages from multiple participants to show they come from a diverse group of people. I incorporate both numbers and interview quotes to add confidence in the findings. By following these strategies, I strive to effectively present qualitative data and facilitate a comprehensive understanding of the research presentation.

In reporting qualitative data, I present findings in a ranked order by assigning a priority to different concepts based on their prevalence in the dataset. This method of organizing and prioritizing concepts identifies the most important themes or patterns within the data. I rank concepts based on the frequency of occurrence (i.e., how often a concept appears in the dataset). This process often results in a hierarchical structure whereby I present concepts from most to least prevalent. I elaborate concepts through interview excerpts. The examples in the following five subsections illustrate discussion of findings from my published papers.

Example 1: Defining Concepts

To establish a common understanding among readers and clarify the criteria used for categorizing responses, explicitly defining concepts is of utmost importance, even if those definitions seem obvious. In a study examining participants' engagement in international collaboration, my co-author and I classified the participants' responses into four distinct categories to provide readers with a framework for understanding how we counted and categorized responses, as the following excerpt from Varma and Sabharwal (2018) indicates:

> (i) yes, included statements that conveyed returnees were involved in at least one international collaborative project at the time of the interview; (ii) past, included declarations that showed current returnees were not involved in an international collaborative project, but had been in at least one prior to the interview; (iii) no, included sentences that suggested returnees were neither engaged in an international collaborative project at the time of the interview nor had any plan to do so in the near future; and (iv) future, included records that transmitted returnees were not engaged in an international collaborative project, but had the desire to do so in the near future. (p. 598)

The preceding example highlights the importance of clearly defining and categorizing data, even for seemingly simple categories like "yes" and "no." This practice ensures transparency and makes it easier for readers to understand how researchers have organized and classified findings. Clear and precise definitions and categorizations are essential for effectively reporting data analysis in qualitative research.

Example 2: Combining Numerical Evidence and Quoted Excerpts Within a Paragraph

To exemplify combining interview excerpts and numerical evidence in support of my findings, I draw upon my (Varma, 2021) study in which 40 Indian scientists and engineers shared their perspectives on the similarities and differences in work culture between themselves and their American counterparts:

> An overwhelming majority of interviewees (87.5%), however, noted significant cultural differences with their American colleagues in the workplace. As one interviewee generalized, "I view America as a nomadic place where people come from all over the world. They bring their own style, which makes things rather interesting, or I should say rich." Similarly, another declared, "Culture affects how you work, how you relate to people, what you think about, how you think about work, and the impact you have." This interviewee generalized, "Most of the work here is rather a high level in terms of technology, but it has some basic social ingredients." (p. 4)

In this example, I interwove qualitative data in the form of interview excerpts with a quantitative percentage. The interview excerpts revealed a particular trend which the numerical evidence supported, adding trustworthiness to the qualitative insights.

Example 3: Presenting Numerical Evidence and Interview Excerpts in Separate Paragraphs

I often adopt a structure by which I present numerical evidence and interview excerpts in separate paragraphs. This approach allows clear distinction between the quantitative and qualitative aspects of the findings. For instance, in Varma (2007), I had asked computer science and computer engineering (CS/CE) students to describe the typical culture within their program. First, I provided the numerical results pertaining to the students' responses, and then followed that with relevant interview excerpts to elaborate on their experiences and perceptions. In addition, I made sure to include diverse quotes from 150 participants (15 white male and 15 white female; 15 black male and 15 black female; 15 Hispanic male and 15 Hispanic female; 15 Asian male and 15 Asian female; and 15 Native American male and 15 Native American female):

> On the question of describing typical culture within their program, almost half of the interviewees (51% female and 45% male) believed there is a stereotypical computer culture mostly consisting of geeks, nerds, and/or hackers (which are substantially overlapping). In general, more whites (60%), blacks, (50%) and Hispanics (47%) than American Indians (43%) and Asians (40%) identified CS/CE as a geek culture. Among females, more whites (73%) and blacks (60%) identified computing as a geek culture than American Indians (47%), Hispanics (40%), and Asians (33%).
> A large majority of interviewees believed in the prevalence of geek culture. They pointed out that all CS/CE students know of geek values even if they do not possess them. An Asian female generalized typical CS/CE students as 'Someone with glasses, a geek, whatever'. 'Usually just a bunch of weirdoes', said a white female. A Hispanic female alleged that 'They are hacking on some sort of program until like three in the morning'. An American Indian female believed they are 'nerdy-type people...who teach themselves all computer stuff'. Several interviewees added lack of social relations or interpersonal communication skills such as they 'do not have a life other than school'; 'don't party that much'; 'don't have a girlfriend they complain about'; and 'buy the cheapest clothes so they can buy more computer stuff'. (pp. 365-366)

This example illustrates presentation of numerical evidence and interview excerpts in separate paragraphs to maintain clarity and separation between the quantitative and qualitative aspects of the findings. This clear distinction helps readers easily understand and navigate the different types of data and findings presented in the research.

Example 4: Incorporating Interview Excerpts in Tables With Descriptive Statistics in Text

Often in presenting qualitative data, I incorporate interview excerpts within the text, while presenting descriptive statistics in tables. However, in some instances, I have reversed this order to better suit the study's objectives. In the Varma (2010) case of focusing on early exposure to computers in India among female students in CS/CE, I inverted the approach. The article featured a table displaying the interview excerpts related to the participants' computer exposure, while I interwove descriptive statistics within the main body of the text, providing further insights and context to their experiences:

> Few students (5 out of 60) reported having a computer in their home as a child; if a computer was brought into the home, it was when they were in the secondary schools (high school, consisting of 9th to 10th grades, and intermediate college, consisting of 11th to 12th grades). A little over half of the students (32 out of 60) had personal access to a computer, either in their high school years (15 out of 60) or in their intermediate college years (17 out of 60). The remaining 23 students had no access until they went to a university. (Table 2; p. 260)

Reversing the order, as Example 4 has shown, is a flexible and strategic approach that demonstrates adapting the presentation format to best serve the specific needs of the study.

Example 5: Summarizing Findings in a Paragraph Followed by a Series of Interview Excerpts

In some of my earlier papers, I adopted a style in which I summarized the findings in a brief paragraph, followed by supporting it with a series of interview excerpts. However, I now advise against this approach as it may present several problems. First, quotes lack the necessary context for readers to understand their relevance to the research question and their place within the broader thematic framework. Second, absence of an analytic narrative may leave readers to interpret the quotes on their own, which may in turn lead to misinterpretation or

Table 2
Exposure to Computers at Home (Reproduced with Permission from Begell House, Varma, 2010, p. 260)

Subjects from	Comments
Top national institute	"When I was in the 9th or 10th grade, my parents bought a computer. But I did not work with it a lot because school kept me busy. If I used it, it was at a very minimal level."
	"My dad had a computer in his office, but not at home. There was no computer at home while I was growing up."
Regional well-known institute	"No ma'am. I did not have a computer at home. I used to go to the cybercafé to do assignments, projects, and search for things."
	"My father used a computer for his office work. But it was out of my reach. But it created more curiosity in me."
Historically Muslim university	"When I was in the 7th or 8th [grade], one of my close relatives bought a computer, and we used to go to his place to play on it or watch movies."
	"No one in my family knew about computers to have one at home."
Predominantly Sikh university	"No. I got a computer in the 2nd year of my B-tech. Since my mother and father are working, they have computers in their offices. So they did not need one at home. When I joined a course for computer languages, my teacher just advised me to get one. And we got a computer a year back."
	"My father bought an old computer when I was in the 9th [grade]. But I did not do anything complicated on a computer, just played games."

confusion. Third, analytic narrative allows researchers to connect the quotes to the research objectives, theories, or prior findings, which is vital for understanding the implications of the research. Fourth, the absence of an analytic narrative may reduce the credibility of the research. It may leave readers questioning the rigor of the analysis. As an example, consider the following presentation of early socialization bias from Varma (2002), which I no longer recommend (pp. 277-278):

> The following interview excerpts show that the under-representation of women in IT-related fields at the undergraduate level is at least in part inherited from the bias in early socialization both at home and in the school system.
> *Well, I guess machines, programming, things like that, came naturally to me. I have been mechanically inclined... I was always very interested in computers. But, in the last few years, I became intensely interested in them (white male student).*
> *In my junior high, we had a computer room. The only reason that you could go into that computer room was if you were taking a computer class. They had started some remedial training... When I asked my teacher for the computer class, he simply said that class will not be good for me because it was on programming (white female student).*
> *My counsellor had advised me to do accounting. When I asked him about engineering, he simply said that he was trying to place me on a career path where I would be successful (Native American female student).*
> *I got interested in computers because I made a bet with my dad. He told me that I could not do computer science, and I told him I could (Hispanic female student).*
> *My parents were keen on me studying computer science because it is so growing right now. If you don't know computers, that means you are way behind (Asian female student).*

The example above has shown that, while quotes valuably provide direct evidence from participants, they are most effective when presented within the context of a coherent and well-structured analytic narrative. Such a narrative would guide readers through the quotations, highlight key points, and ensure the credibility and impact of the research.

Conclusion

Textbooks on research methods often present methodologies in abstract forms, aiming to teach novice scholars about qualitative data analysis. While they define concepts and provide examples they often remain detached from hands-on activities. These books typically lack demonstrations of how to analyze data, even when they offer a roadmap for introducing emerging researchers to the analysis process (Lester et al., 2020). As a result, emerging researchers may develop only a superficial understanding of qualitative data analysis and struggle when applying these skills to their research projects. This limited exposure to real-world examples of qualitative data analysis may lead emerging researchers to prioritize memorization and theory over the development of critical evaluation skills. To address this issue, it is essential that teaching scholars provide emerging researchers with concrete and practical examples that illustrate the implementation of qualitative data analysis activities.

In this paper, my objective has been to aid emerging researchers in comprehending how qualitative data analysis may be executed in practical scenarios, thereby promoting active learning. Rather than creating fictitious examples, I have drawn upon instances from my diverse research projects. This approach has offered the advantage of real-life examples, exposing emerging researchers to genuine challenges in data analysis. Furthermore, the research projects I selected have received support from the NSF and their findings have been published in peer-reviewed journals, affirming their credibility through external validation and review.

In summary, qualitative analysis is a valuable phase of research that offers profound insights into human experiences and phenomena. However, it is not without limitations and criticisms (Aguinis & Solarino, 2019; Rocco, 2010). Challenges in qualitative data analysis may arise from various factors, impacting the trustworthiness and credibility of the findings. Addressing these issues requires researchers to adopt systematic approaches to qualitative data analysis. Without such measures, researchers employing qualitative research may inadvertently contribute to substantiating criticisms made in some quarters of quantitative researchers.

References

Aguinis, H., & Solarino, A. M. (2019). Transparency and replicability in qualitative research: The case of interviews with elite informants. *Strategic Management Journal*, *40*(8), 1291-1315. https://doi.org/10.1002/smj.3015

Babchuk, W. A. (2019). Fundamentals of qualitative analysis in family medicine. *Family Medicine and Community Health, 7*(2), e000040-. https://doi.org/10.1136/fmch-2018-000040

Bauer, M. W., & Gaskell, G. (2007). *Qualitative researching with text, image and sound: A practical handbook*. SAGE.

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, *42*(3), 294-320. https://doi.org/10.1177/0049124113500475

Charmaz, K. (2014). Grounded theory in global perspective: Reviews by international researchers. *Qualitative Inquiry*, *20* (9), 1074-1084. https://doi.org/10.1177/1077800414545235

Creswell, J. W., & Creswell, D. J. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design choosing among five approaches*. SAGE.

Creswell, J. W., & Guetternman, T. C. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research.* (6th ed.) Pearson.

Denzin, N. K., & Lincoln, Y. S. (2018). *The SAGE handbook of qualitative research*. SAGE.

Erickson, F. (2018). A history of qualitative inquiry in social and education research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (pp. 27-65). SAGE.

Hagger, M. S., & Chatzisarantis, N. L. D. (2011). Never the twain shall meet? Quantitative psychological researchers' perspectives on qualitative research. *Qualitative Research in Sport, Exercise and Health*, *3*(3), 266-277. https://doi.org/10.1080/2159676x.2011.607185

Hammersley, M. (2008). *Questioning qualitative inquiry*. SAGE.

Lester, J. N., Cho, Y., & Lochmiller, C. R. (2020). Learning to do qualitative data analysis: A starting point. *Human Resource Development Review*, *19*(1), 94-106. https://doi.org/10.1177/1534484320903890

Maxwell, J. A. (2015). Expanding the history and range of mixed methods research. *Journal of Mixed Methods Research, 10*(1), 12-27. https://doi.org/10.1177/1558689815571132

Onwuegbuzie, A., & Leech, N. (2005). Taking the "Q" out of research: Teaching research methodology courses without the divide between quantitative and qualitative paradigms. *Quantity and Quality: International Journal of Methodology,* 39, 267-296. https://doi.org/10.1007/s11135-004-1670-0

Popper, K. (2001). *The logic of scientific discovery.* Routledge.

Rocco T. S. (2010). Criteria for evaluating qualitative studies. *Human Resource Development International, 13*(4), 375-378. https://doi.org/10.1080/13678868.2010.501959

Sabharwal, M., & Varma, R. (2012-2015). Return migration of academic scientists and engineers from the United States to India. *National Science Foundation*, Grants 1229990 & 1230091, $207,179.

Sabharwal, M., & Varma, R. (2017). Grass is greener on the other side: Return migration of engineers and scientists in academia. *Bulletin of Science, Technology & Society, 37*(1), 34-44. https://doi.org/10.1177/0270467617738463

Saldaña, J. (2011). *Fundamentals of qualitative research.* Oxford University Press.

Varma, R. (1996-1998). Industrial scientists: New challenges with restructuring. *National Science Foundation*, Grant 9602200, $55,000.

Varma, R. (2002). Women in information technology: A case study of undergraduate students in a minority-serving institution. *Bulletin of Science, Technology & Society, 22*(4), 274-282. https://doi.org/10.1177/0270467602022004003

Varma, R. (2002-2005). New immigrants: Science and engineering in the United States. *National Science Foundation,* Grant 0136467, $79,766.

Varma, R. (2006). The tenure system in engineering. *National Science Foundation,* Grant 0610402, $10,247.

Varma, R. (2007). Women in computing: The role of geek culture. *Science as Culture, 16*(4), 359-376. https://doi.org/10.108%9505430701706707

Varma, R. (2007-2008). Cross-national differences in women's participation in computer science education in India and the United States. *National Science Foundation,* Grant 0650410, $61,766.

Varma, R. (2010). Computing self-efficacy among women in India. *Journal of Women and Minorities in Science and Engineering, 16*(3), 257-274. https://doi.org/10.1615/jwomenminorscieneng.v16.i3.40

Varma, R. (2021). Dissecting culture at work: Conversation with Indian immigrant scientists & engineers in the US technology sector. *Technology in Society, 66,* 1-8. https://doi.org/10.1016/j.techsoc.2021.101654

Varma, R., & Kapur, D. (2003-2007). Cross-ethnic differences in undergraduate women's preference for information technology. *National Science Foundation,* Grant 0305898, $350,392.

Varma, R., Kapur, D., & Perlman, B. (2001-2003). Why so few women in information technology? A comparative study. *National Science Foundation,* Grant 0120055, $102,426.

Varma, R., & Sabharwal, M. (2017-2021). Maintaining diversity in the US high-tech sector. *National Science Foundation*, Grants 1655322 & 1655366, $225,155.

Varma, R., & Sabharwal, M. (2018). International collaboration: Experiences of Indian academics after returning from the United States. *Perspectives on Global Development and Technology,17*(5-6), 593-613. https://doi.org/10.1163/15691497-12341498

Vidich, A. J., & Lyman, S. M. (2000). Qualitative methods: Their history in sociology and anthropology. In N. K. Denzin & Y. S. Lincoln (Eds). *The SAGE handbook of qualitative research* (pp. 37-82). SAGE.