The Journal of Extension

Volume 62 | Number 1 Article 15

6-19-2024

The Key to Coding Qualitative Data: Examples from the 4-H Youth Retention Study

Sarah Hensley
University of Florida IFAS, sarahzt@ufl.edu

Janet Fox
University of Tennessee Institute of Agriculture, jfox35@ukt.edu

Missy Cummins
Washington State University Extension, missy.cummins@wsu.edu

Meggan Franks
Louisiana State University Ag Center, msfranks@agcenter.lsu.edu

Marianne Bird
University of California Agriculture and Natural Resources, mbird@ucanr.edu

See next page for additional authors

Recommended Citation

Hensley, S., Fox, J., Cummins, M., Franks, M., Bird, M., Wells, C., & Miller, J. (2024). The Key to Coding Qualitative Data: Examples from the 4-H Youth Retention Study. *The Journal of Extension, 62*(1), Article 15. https://doi.org/10.34068/joe.62.01.15

This Tools of the Trade is brought to you for free and open access by the Conferences at TigerPrints. It has been accepted for inclusion in The Journal of Extension by an authorized editor of TigerPrints. For more information, please contact kokeefe@clemson.edu.

The Key to Coding Qualitative Data: Examples from the 4-H Youth Retention Study

Cover Page Footnote

The 4-H Youth Retention Study team members would like to dedicate this article to the memory of Jeannette Rae-Keywood. Jeannette was with Rutgers University at the time of submission and has since passed.

Authors

Sarah Hensley, Janet Fox, Missy Cummins, Meggan Franks, Marianne Bird, Cindy Wells, and JoLynn Miller

JOURNAL OF Extension Tools of the Trade Volume 62, Issue 1, 2024

The Key to Coding Qualitative Data: Examples from the 4-H Youth Retention Study

Sarah Hensley[1], Janet Fox[2], Missy Cummins[3], Meggan Franks[4], Marianna Bird[5], Cindy Wells[6], and JoLynn Miller[5]

AUTHORS: [1]University of Florida. [2]University of Tennessee. [3]Washington State University. [4]Louisiana State University. [5]University of California Agriculture and Natural Resources. [6]University of Missouri.

Abstract. Cooperative Extension professionals utilize proven qualitative techniques to collect and analyze information to make data-driven decisions that guide program direction and determine impact. While the process may not always look the same, it is indeed essential to ensure findings are credible and reflective of the data. A codebook is a valuable tool that promotes interrater reliability among teams and enhances the reliability of findings. This article provides an overview of the process used to analyze qualitative data and the steps followed to create a codebook. Examples of the analysis process and codebook development are provided from the 4-H multi-state Youth Retention Study.

INTRODUCTION

Collecting data that is reflective of the true nature of a program participant's experience is vital to both formative and summative evaluation results. Qualitative data allows survey participants to provide rich and detailed answers about their program experience. Using data in program development, organizational development, and external communications has become increasingly imperative for Extension institutions (Schmieder et al., 2018). Extension professionals must ensure that the true perception of their clientele is captured. Qualitative data provides opportunities to collect data that will provide rich and multiple levels of understanding.

The diverse subject matter backgrounds of Extension professionals lead to varying levels of evaluation and research competence. A basic grasp of the processes for effective qualitative research is challenging for some; however, qualitative research is crucial for fully evaluating Extension programs. A recent study conducted by Diaz et al. (2019) found that "determining program impacts and how to measure them" is a top challenge faced by "newer Extension Agents" (p. 2). Additional challenges include "development of an accurate evaluation instrument for a given situation, evaluating newly developed programs, management and analysis of data collected, and evaluating long-term impacts of Extension programming" (p. 2).

This article will address the element of qualitative data organization commonly referred to as "coding" and provide steps to develop a codebook. Examples of this process will be provided from the multistate 4-H Youth Retention Study (YRS). The YRS explores first-year participation in 4-H from both the member and adult perspective to better understand their program experiences. The online survey administered via Qualtrics asked questions related to family expectations, experiences, and intent to re-enroll (Lewis et al., 2018). Over the past five years, youth and their parents from nine states completed surveys. The first year of data collection is considered a pilot year. This article will highlight the process used to code the qualitative data.

The research team used a coding process to analyze qualitative data in the form of open-ended questions. In the first year of the study the team analyzed qualitative responses collaboratively through open coding and consensus. A codebook was developed based on themes derived from the responses of both youth and adults. As qualitative results were coded, the findings began to reveal information about the experience of first-year families in the

4-H program, and what influenced their 4-H experience and intentions to re-enroll in 4-H for a second year. The in-depth nature of the qualitative findings has provided valuable insight into factors impacting retention of youth.

QUALITATIVE CODING AS A PROCESS

Qualitative research is a broad term describing several research methods and paradigms that rely on the collection, analysis, and interpretation of nonmathematical data (Whitely & Crawford, 2005) and attempts to "translate life into text" (Glesne, 2016, p. iv). Researchers use coding as an approach to organize copious amounts of data into organized, understandable, informed, and useful texts (Belotto, 2018; Liamputtong, 2009).

This article provides an overview of the coding process with examples from the multistate youth retention study. The steps to code qualitative data are an orderly process requiring organization, yet many times one phase will overlap with another and require moving back and forth between the various phases of data coding (White & Scanga, 2019).

PHASE ONE: ORGANIZING DATA

Analysis of qualitative data requires the data to be organized efficiently and effectively, enabling the researcher to accomplish the intent of the research project (Sandelowski, 1995). The process of organizing data to prepare for coding involves streamlining all the collected data files in one comprehensive document (Akinyode & Khan, 2018). In this phase, the YRS study researchers organized all the open-ended survey responses into one Excel file, labeling each tab of the file with the corresponding question, and cleaning out any missing entries. This made the process of working through all the responses to identify codes efficient.

PHASE TWO: IDENTIFY CODES

Qualitative coding is just one component of the broader process of qualitative data analysis and occurs early in the analysis process (Childs & Demers, 2018). Coding is defined as "taking transcribed text data and making sense of them" (Cresswell, 2015, p. 159). In this step, words and sentences that have a similar meaning are identified and labeled with the same code (Belotto, 2018). Miles and Huberman (1994) define codes as labels that are used to describe the information within a statement. Codes can be assigned to pieces of data including single words, phrases, sentences, or paragraphs. In addition, multiple codes may be used to explain the meaning of words or phrases found in the data.

According to White and Scanga (2019), the process of assigning codes to qualitative data can be done through either a deductive or an inductive process. Deductive codes are specific codes based on a preset research framework, while codes assigned through an inductive process emerge from the data without a preconceived notion of what themes will emerge. Similarly, a researcher may also use *a priori* knowledge by reflecting on their practical experience as they review data. When coding, researchers may find a primary code, multiple codes, or subcodes associated with a key statement. Corbin and Strauss (2008) recognized two stages of coding: open and axial. Open coding is aimed at exploring the ideas and concepts that emerge in the raw data. In axial coding, the researcher examines relatedness between and among codes. Boyatzis (1998) recognizes two different approaches when developing codebooks based on the approach taken: theory-based (deductive) and data-driven (inductive) codes. A theory-based coding approach has three steps: 1) generate code; 2) review and revise codes; and 3) establish reliability of the code and coders. Theory-based codes are derived from theoretical frameworks that guide the study.

The YRS study utilized an inductive, data-driven approach in which codes emerged directly from the data. Open coding was initially conducted in which all research team members explored a subset of the data to generate codes. As codes were derived from the data, they were examined through consensus discussion among the researchers to group and refine them. Themes were then generated from the codes. For example, a response to a question would state the following: "The teacher is very disorganized." The team would discuss this response and decide to assign the following code: Organization/Structure (13). Other examples of data quotes and the assigned codes are provided in Table 1.

Oftentimes data will provide more than one meaning requiring the researcher to assign multiple codes to a response. Table 2 provides examples of quotes that were identified as having multiple codes.

Table 1. 4-H Youth Retention Study Data Sample Quote and Code

Quote		Code
"My photos going to state fair"	1	Fair
"To learn how to build and run a robot"	7	Learning
"To learn valuable skills, understand responsibility, recordkeeping, and competition"	10	Personal or Character Development

Table 2. Data Assigned Multiple Codes

Example		Codes Assigned
"Leadership skills and more knowledge about our animals and of course social experiences"	10	Personal or Character Development
	7	Learning
	2	Topic/Content Specific
	9	Social
"We expected someone to orientate us to the group and how things are done. We ended up having to figure things out for ourselves which took a while and we missed several things because of it"	5	Mentoring, guidance, and instruction
	13	Communication

PHASE THREE: CODEBOOK DEVELOPMENT

A codebook facilitates the concentrated and efficient study of raw data. "A codebook is a set of codes, descriptions, and examples used to examine data" (DeCuir-Gunby et al., 2011, p. 138). One benefit of a solid codebook development strategy is that it will support coding by multiple team members (MacQueen et al., 2009). It allows for multiple coders to analyze data with consistency. A codebook consists of three major components including: code, code definition (which includes standards to include or exclude statements), and an example statement (DeCuir-Gunby et al., 2011).

Boyatzis (1998) recognizes two different approaches when developing codebooks based on the approach taken: theory-based (deductive) and data-driven (inductive) codes. During the third phase, to increase reliability, researchers engage in rival thinking and challenge each other's perceptions of theoretical frameworks (Yin, 2016). This process enhances the integrity and validity of findings. When generating codes, it is imperative that researchers reach a consensus with the codes and their meaning. Researchers will review the suitability of codes and determine if they are reflective of the true meaning of data.

Data-driven codes involve the following steps for inductively establishing a codebook: 1) reduce raw data information; 2) determine subsample themes; 3) compare themes across subsamples; 4) establish codes; and 5) ascertain code reliability (Boyatzis, 1998). The first step to develop data-driven codes is to reduce raw data into smaller units by comparing, contrasting, sorting, and ordering data. During this step, researchers remove irrelevant data and look for observed patterns in data to construct and assess codes. When reading statements, it is important to focus on the meaning of the statement instead of coding line by line. If a statement has multiple codes, splitting the message is a possibility but the statement should stand on its own in conveying the code. Researchers should constantly compare codes within the context of multiple passages through examining patterns. Codes are established when observed patterns have been identified and a point of saturation is reached. The point of saturation is when all possible explanations of data have been observed and identified in order to accurately capture respondents' perceptions (Boyatzis, 1998).

As the codes are identified, a code book or listing of codes can be organized along with definitions of each code. Assigned codes should be meaningful, succinct, and accurately represent the data, and the definitions should be precise in defining the concept. A sample of codes from the codebook for the YRS are shown in Table 3. The development of the codebook has allowed the research team to divide the large data set among members for individual coding and then to check for interrater reliability across the findings. After the initial establishment of the codebook, the research team used it to code a subset of data and they engaged in discussion to ensure the useability and reliability of the codebook.

Table 3. Excerpt from 4-H Youth Retention Study Codebook

Code Title	Description	Data Example
Learning	- Learning new things - Gaining knowledge on specific topics - Lack of learning; uninformative	"Her accomplishments in learning, preparing, and meeting her goals"
Fun/enjoyment	- Having fun - Enjoying activities - Love doing something	"4-H provides a good opportunity for my kids to help with community events and have fun doing it"
Social	- Meeting people - Making friends - Networking and connecting with others	"Making friends and personal growth"
Organization/Structure	- Expectations for program/activities to be organized - Coordination of meetings, events, projects, etc. - Clarity of activities and events	"Monthly meetings were chaotic"

PHASE FOUR: RELIABILITY CHECKS

A critical stage of data coding is that of ensuring reliability, which is guaranteeing that codes are assigned consistently across the data set. When multiple people are involved in coding, there should be consistency between individuals when applying codes. All coders should understand what each code represents and interpret and code the data uniformly. A process called intercoder agreement ensures that those coding separately all assign similar codes to data (Cresswell, 2015). Interrater reliability is a term used to determine the consistency of interpretation among independent coders (Lombard et al., 2002). Having agreement among coders is an important part of coding, yet one must not assume this will ensure validity (Lombard et al., 2002). For example, even when a group of coders agree on the same meaning of a statement, they may not be accurately interpreting the intended meaning of that statement. Researchers must not assume meaning that is not clearly stated in the data.

Paying special attention to this step allowed the YRS team to divide the qualitative data among researchers. The YRS research team members were each assigned a percentage of the data to code and a partner. After each individual finished coding, they reviewed their codes with their assigned partners. Any difference in coding was discussed until consensus was achieved. Practicing interrater reliability at the beginning ensures consistency among coders throughout the process. As part of the process to establish interrater reliability and ensure that the coding remained consistent, team members were also assigned another subset of data to review the codes for consistency. This process of checking for reliability also allowed the team to refine the codebook as needed.

PHASE FIVE: CODEBOOK MODIFICATION

Analysis of qualitative data requires special attention to credibility. A "credible study" will show the efforts taken by the researcher to collect and interpret data properly to ensure findings "accurately reflect and represent" those who were studied (Yin, 2016, p. 85). Throughout the process of analysis, a researcher should engage in "rival thinking" (Yin, 2016, p. 90) to ensure that appropriate meaning is assigned to data. Rival thinking is the intentional search for differing patterns or themes in the data. It is a quest to look at the data in new ways that might support alternate explanations or understandings (Patton, 2001). Researchers will want to consider as many rival explanations as possible, including their own original assumptions.

The YRS team reviewed a sample of the data coded each year to engage in efforts to ensure reliability while coding. Through this process team members engaged in rival thinking which led to further code clarification and expanded understanding of what each code represented. The peer-debriefing process and the time spent reflecting on codes assigned resulted in modification and improvement to the codebook. At the end of the study the team reviewed all previously coded data based on codebook modifications to ensure all four years of data were coded consistently and correctly. This process further enhanced the credibility of the study's findings.

PHASE SIX: REPORTING FINDINGS

The final phase of qualitative data coding is that of reporting findings. This phase involves interpreting the data, seeking patterns in codes and themes, and sharing the results objectively. Data must be interpreted in a way that is complete, accurate, and credible (Yin, 2016). Once coded, researchers can explore the statements within codes and investigate connections between codes. As researchers begin to organize subcodes into clusters, themes or concepts will begin to emerge (Saldana, 2014, p. 14). Themes may be presented in order of prevalence but should not be quantified. According to Yin, 2016, it is inadvisable to make statistical generalizations with qualitative data that imply findings apply to a larger population (Yin, 2016). Qualitative data can be viewed with a more holistic lens, allowing researchers to understand the gestalt of that which is being studied. When reporting results, the researcher must ensure they provide an objective report of the findings without inserting their own interpretation. If the earlier phases are conducted appropriately, findings should provide a rich understanding of the phenomena being studied.

The completion of a comprehensive and credible codebook enabled the YRS team to review data from all four years of the study to gain deeper insight into the first-year 4-H experience for members and their parents. The team identified and coded notions that were frequently reported as influencing youth and parent decisions to re-enroll. For example, when youth were asked about "other reasons you might not come back to 4-H," their responses reflected the most prominent themes of communication, adult interactions, and organization.

CONCLUSION

Qualitative data provides rich information that helps researchers gain insight into data that quantitative measures cannot provide. The open-ended nature of qualitative data allows researchers to capture participants' thoughts and ideas to better comprehend their experiences and perspectives. For Extension professionals, this is important to our work in understanding and serving our clientele.

It is essential to ensure findings in qualitative research are credible and reflective of the data, and coding is a process that helps achieve this. Creating a codebook is complex and time-consuming, yet it is a valuable tool that promotes consistent codes that interpret data, leading to a greater understanding of a contextual phenomenon. A codebook is especially helpful when working with multiple people in data analysis. The process of building it leads to greater clarity of the data as well as greater consistency between coders. Divergent viewpoints and rival thinking ensure that researchers are exposed to multiple interpretations.

Developing a codebook with the YRS team unearthed the richness of 4-H participants' experience in the program. The process of developing the codebook promoted open conversations, which helped reduce bias in the coding process. The team's discussion around findings brought theoretical foundations to life, from the abstract to concrete, and facilitated recommendations for retention in the 4-H Youth Development Program. In the same way, other Extension programs may benefit from collecting and analyzing qualitative data utilizing a codebook.

REFERENCES

Akinyode, B. F., & Khan, T. H. (2018). Step by step approach for qualitative data analysis. *International Journal of Built Environment and Sustainability*, *5*(3).

Belotto, M. J. (2018). Data analysis methods for qualitative research: Managing the challenges of coding, inter-rater reliability, and thematic analysis. *Qualitative Report*, *23*(11), 2622–2633.

Boyatzis, R. (1998). *Transforming qualitative information: Thematic analysis and code development.* SAGE.

Childs, E., & Demers, L. B. (2018). Qualitative coding boot camp: An intensive training and overview for clinicians, educators, and administrators. *MedEdPORTAL: The Journal of Teaching and Learning Resources*, *14*, 10769. https://doi.org/10.15766/mep_2374–8265.10769

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). SAGE.

Cresswell, J. (2015). *Thirty essential skills for the qualitative researcher*. SAGE.

DeCuir-Gunby, J. T., Marshall, P. L., & McCulloch, A. W. (2011). Developing and using a codebook for the analysis of interview data: An example from a professional development research project. *Field Methods*, *23*(3), 136–155.

Diaz, J., Chaudhary, A. K., Jayaratne, K., & Warner, L. A. (2019). Program evaluation challenges and obstacles faced by new extension agents: Implications for capacity building. *Journal of Extension*, *57*(4). https://tigerprints.clemson.edu/joe/vol57/iss4/26

Glesne, C. (2016). *Becoming qualitative researchers: An introduction (*5th ed.*)*. Pearson/Allyn & Bacon.

Lewis, K., Ewers, T., Miller, J.C., Bird, M., Borba, J., Hill, R.D, Rea-Keywood, J. Shelstad, N. & Trzesniewski, K. (2018). Addressing retention in youth programs: A survey for understanding families' experiences. *Journal of Extension*, *56*(3). https://tigerprints.clemson.edu/joe/vol56/iss3/5/

Lewis, K.M., Hensley, S., Bird, M., Rea-Keywood, J., Miller, J. C., Kok, C. M., & Shelstad, N. (2022). Why youth leave 4-H after the first year: A multistate study. *Journal of Human Sciences and Extension*.

Liamputtong, P. (2009). Qualitative data analysis: conceptual and practical considerations. Health promotion journal of Australia, 20(2), 133–139.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human communication research*, *28*(4), 587–604.

MacQueen, McLellan, E., Kay, K., & Milstein, B. (1998). Codebook development for team-based qualitative analysis. *CAM Journal*, *10*(2), 31–36. https://doi.org.lp.hscl.ufl.edu/10.1177/1525822X980100020301

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). SAGE.

Patton, M. Q. (2001). *Qualitative research and evaluation methods* (2nd ed.). SAGE.

Saldaña, J. (2014). *The coding manual for qualitative research* (3rd ed.). SAGE. https://us.sagepub.com/en-us/nam/book/coding-manual-qualitative-researchers-1

Sandelowski, M. (1995). Qualitative analysis: What it is and how to begin. *Research in Nursing & Health*, *18*(4), 371–375.

Schmieder, C., Caldwell, K. E. H., & Bechtol, E. (2018). Readying Extension for the systematic analysis of large qualitative data sets. *Journal of Extension*, *56*(6). https://tigerprints.clemson.edu/joe/vol56/iss6/26/

White, A. J. & Scanga, L. H. (2019). Volunteer research and knowledge competency codebook: A tool for identifying volunteer needs. *Journal of Extension*, *57*(1). https://tigerprints.clemson.edu/joe/vol57/iss1/2/

Whitley, R., & Crawford, M. (2005). Qualitative research in psychiatry. *Canadian Journal of Psychiatry*, *50*(2), 108–114. https://doi.org/10.1177/070674370505000206

Yin, Robert K. (2016). *Qualitative research from start to finish* (2nd ed). The Guilford Press.