Journal of Rural Social Sciences

Volume 34 | Issue 1 Article 2

8-9-2019

Benefits to Qualitative Data Quality with Multiple Coders: Two Case Studies in Multi-coder Data Analysis

Sarah P. Church
Montana State University, sarah.church@montana.edu

Michael Dunn
Centre for Ecosystems, Society and Biosecurity, Forest Research, Northern Research Station (UK), Michael.dunn@forestresearch.gov.uk

Linda S. Prokopy
Purdue University, lprokopy@purdue.edu

Benefits to Qualitative Data Quality with Multiple Coders: Two Case Studies in Multi-coder Data Analysis

Cover Page Footnote

Please address all correspondence to Dr. Sarah Church (sarah.church@montana.edu).

Benefits to Qualitative Data Quality with Multiple Coders: Two Case Studies in Multi-coder Data Analysis

Sarah P. Church
Montana State University

Michael Dunn
Centre for Ecosystems, Society and Biosecurity, Forest Research, Northern Research Station (UK)

Linda S. Prokopy
Purdue University

ABSTRACT
Qualitative research methods contend with debates surrounding subjectivity and bias. Researchers use a variety of techniques to help ensure data trustworthiness. One such technique is to involve multiple coders in data analysis. The deliberative nature of codebook development among multiple coders produces rich data analysis that may not otherwise be achieved with a single (or even two) researcher(s). In this manuscript, we make a plea for researchers and journals to include data analysis procedures and descriptions in published literature. In addition, we illustrate minimal reporting of qualitative data analysis processes through a synthesis of 21 years of agricultural best management practice adoption literature. We present two rural agricultural case studies on multi-coder team codebook development and intercoder reliability processes specific to interviews, focus groups, and content analysis. Overall, we argue that multi-coder teams can improve data quality, and reporting data analysis procedures can mitigate implications of subjectivity in qualitative methods.

KEYWORDS
Content analysis; inter-coder analysis; interviews; methods; natural resources

INTRODUCTION

There are many instances in rural agricultural lands management where researchers seek to answer questions such as *how* a phenomenon is perceived by individuals or reported through policy, plans, and media accounts, *why* people behave the way they do, and *what* contextual elements contribute to perceptions and behavior. In cases such as these – *how, why,* and *what* questions – qualitative methodologies such as interviews, focus groups, observations, and content analysis are appropriate approaches to data collection (Creswell 2013). Analysis of qualitative data requires the researcher to interpret the meaning of research participants' words and actions, as well as text found in reports and publications. There are extensive discussions surrounding validity, reliability, and trustworthiness of qualitative research (e.g. Prokopy 2011) – how can researchers assure readers that their conclusions are not subjective or biased in some way? Some qualitative researchers question whether data validity standards to indicate rigor can be incorporated into qualitative data analysis, while also allowing for the creativity and nuance of qualitative methodologies and the voices of the researched and researcher (Whittemore, Chase, and Mandle 2001). As a response to such debates, scholars have offered a variety of techniques qualitative researchers can use to help ensure trustworthiness of their data. Such techniques include data triangulation, checking for negative evidence or rival explanations, reporting potential researcher biases or preconceptions, member checking, intercoder reliability, inclusion of quotations, context descriptions, and more (Whittemore et al. 2001; Creswell 2013; Sin 2010; Prokopy 2011; Noble and Smith 2015).

In this article, we focus on the interconnected processes of codebook development and intercoder reliability particular to interviews, focus groups, and content analysis. We acknowledge that there are qualitative researchers and methodologies that place more value on acknowledging and addressing biases and/or preconceptions (a good practice in any case) than processes like intercoder reliability (e.g. Whittemore et al. 2001; Noble and Smith 2015). Here we work within a paradigm of rural agricultural lands management research in the United States, which in our experiences has been dominated by quantitative approaches (e.g. Prokopy et al. 2019). In our opinion, this thereby

necessitates clear documentation of data analysis procedures to assure perceptions of data quality. In this article, we argue two points: 1) the value of large coding teams in producing quality data; and 2) the necessity of reporting intercoder reliability processes in the published literature. In our experiences, there is a lack of guidance on intercoder processes in rural agricultural lands management literature. Moreover, we suggest that there is minimal reporting in the peer-reviewed literature on the processes researchers undertake to ensure qualitative data quality and reliability. These two issues are interconnected and perpetuate the myth that qualitative research is lesser than quantitative (e.g. Leavy 2014).

Qualitative analysis is difficult and complex, and should be acknowledged and reported as such. In qualitative analysis, text is read in context and then placed into meaningful categories (i.e. codes). Codes are the means through which data is interpreted and analyzed, and ultimately how researchers develop research outcomes and conclusions. Ensuring a project's codes and codebook (group of coding themes) are an accurate interpretation of any given text is therefore a crucial component of judging the validity, reliability, trustworthiness, and perceived quality of research conclusions. Including two or more coders in the coding process is a good step in ensuring coding reliability, whereby multiple coders achieve coding consistency or intercoder reliability (Kurasaki 2000). At the same time, incorporating more than two coders on the coding team inserts an additional level of scrutiny and rigor to the coding process through added perspectives of different researchers that may produce a more thorough analysis than with a smaller coding team (e.g. MacQueen et al. 1998; Olson et al. 2016). Rather than doing away with subjectivity, multiple coders achieve some level of inter-subjectivity within the team.

Much has been written to provide guidance on the coding and intercoder reliability process. This includes advice on calculating intercoder reliability (e.g. Rust and Cooil 1994; Sim and Wright 2005), building a codebook through theory or data (e.g. DeCuir-Gunby, Marshall, and McCulloch 2011), the utility of coding software (e.g. Lu and Shulman, 2008), and descriptions of coding and reliability processes (e.g. MacQueen et al. 1998). Coding and reliability are discussed within the context of health (e.g. Burla et al. 2008), education (e.g. Basit 2003), and communications research (e.g. Lombard, Snyder-Duch, and Bracken 2002). At least one article described the process of achieving reliability in plan quality analysis (Stevens, Lyles, and Berke 2014). Researchers should follow and document rigorous data analysis procedures to illustrate

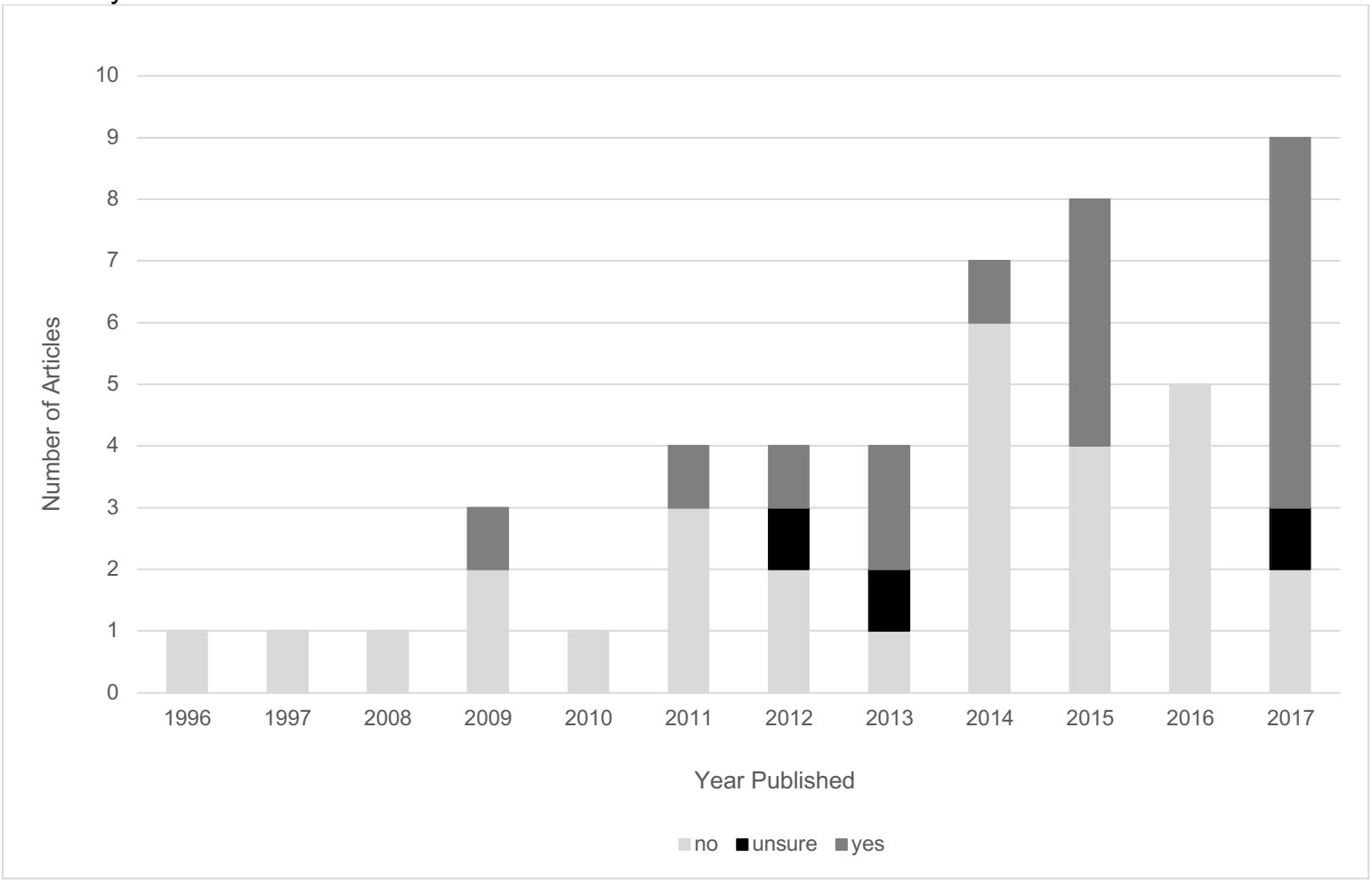
the complexity and difficulty of conducting qualitative research, while providing assurance of data quality.

Although it may seem obvious that intercoder reliability processes should be utilized in qualitative analysis, we contend that this semblance of status quo is not apparent in peer-reviewed manuscripts. Indeed, we synthesized 21 years of qualitative literature specific to rural agricultural land best management practice adoption, which included 48 qualitative articles published between 1996 and 2017. We found that 60 percent of articles (n=29) did not report intercoder reliability processes in this dataset, and six percent (n=3) gave some indication that intercoder procedures had taken place but did not give enough detail to understand exactly what that process was. In the 33 percent of articles (n=16) that reported intercoder reliability processes, half (n=8) used two researchers to verify coding, six articles reported three researchers involved in the inter-coder process, and two did not specify how many researchers participated in the process. Although the trend of this dataset over time indicates that intercoder reliability processes are reported in more frequency, reporting this information is not the dominant paradigm: between 1996 and 2012, 20 percent of articles (n=3) reported inter-coder processes, with 40 percent (n=13) reporting between 2013 and 2017 (Figure 1).

Although qualitative methods and analysis transcend discipline, our purpose in writing this article is to specifically engage rural social scientists and the larger natural resource and land use management community by documenting two rural agricultural land-based qualitative research studies' step-by-step intercoder reliability processes. We focus on projects that required large teams of coders working with large amounts of text, and argue that by using multi-coder teams, we achieved rich data outcomes that transcended subjectivity due to iterative processes of deliberation over code meanings and interpretations of text. Moreover, without reporting our processes in peer-reviewed literature, our rigorous analysis that resulted in quality data would be left in vacuum, unnoticed, thus perpetuating implications of subjectivity of qualitative methods.

In the following pages, we provide summaries of two research cases with a table outlining each process (Table 1), followed by a critique

Figure 1. Number of Agricultural Best Management Practices (BMP) Adoption Articles Reporting Intercoder Reliability Processes between 1996 and 2017



Note: “Unsure” indicates authors described reliability processes, but not in enough detail to understand the full process and whether more than one researcher was involved.

Table 1. Intercoder Reliability Process Summary of Two Research Studies

	Case Study 1 – Crop Advisor Interviews	Case Study 2 – Agricultural Trade Publications
# Documents coded	57 interview transcripts coded by lead coder after intercoder reliability achieved.	1,000 articles split evenly between 8 coders.
# Coders	9	8
Codebook development	• 3 researchers individually constructed coding themes to present to larger group. • Codebook refined through conference calls and individual meetings until final.	• Lead coder developed first codebook. • Codebook tested with 3 researchers and then refined. • Conference call held to discuss codebook. • Codebook further refined throughout 5 rounds of an intercoder reliability process.
# Broad code themes	12	9
# Subcode themes	0 (subcodes developed later)	86
Overall intercoder process	• 2 rounds of intercoder reliability kappa score testing. • Intercoder process began after final codebook. By this time, shared meaning had been developed through team phone conference calls. Thus, only 1 additional round of intercoder reliability tests was needed.	• 5 rounds of intercoder reliability kappa score testing. Although a codebook had been developed and discussed, the reliability coding rounds allowed the team to continue to refine the codebook until an acceptable kappa score was achieved.

	Case Study 1 – Crop Advisor Interviews	Case Study 2 – Agricultural Trade Publications
	• 9 transcripts coded. • 2 -3 coders per transcript. • Lead coder coded all 9 transcripts. • Repeated entire process to develop subcodes for specific research papers.	• 3 researchers coded 10 articles (not part of coding sample) – codebook revised. • 8 researchers coded same 10 articles – codebook revised. • 8 researchers coded 20 new articles (from coding sample) – codebook revised. • 8 researchers coded same 20 articles – codebook slightly revised. • Lead coder recoded each of the 8 coder's articles per the final codebook.
Detailed kappa score testing process	• 4 of 9 transcripts coded by 3 researchers; 5 remaining transcripts coded by 2 researchers. • Each researcher's coding compared with the lead coder to determine code-by-code kappa scores (through NVivo). • Ran separate queries for the lead coder against each coder, and 2 coders against each other for transcripts coded by 3 researchers. • Calculated mean average kappa score across all transcripts and coders.	• Compared lead coder's codes with each team member, one by one, through NVivo. Queries also run for other team members without the lead coder. • Determined average kappa scores for all nodes and sources, using Microsoft Excel*, for all 7 combinations of coders. • Averaged all 7 average kappa scores for final team kappa score for each coding round.
Intercoder coding method	• Transcripts coded individually through a variety of methods. • Lead coder manually entered each coder's work into a master NVivo file.	• Each coder coded articles in NVivo. • Lead coder combined each NVivo project into one master file.

	Case Study 1 – Crop Advisor Interviews	Case Study 2 – Agricultural Trade Publications
Discrepancy reconciliation process	• Kappa scores – low scores indicate problem codes. • Coding stripe analysis – visual representation of coding discrepancies. • One on one phone calls and meetings – negotiated consensus of code meanings.	• Kappa scores – low scores indicate problem codes. • Coding stripe analysis – visual representation of coding discrepancies. • One on one phone calls and meetings – negotiated consensus of code meanings. • Team conference calls – negotiated consensus of code meanings.
Timeframe	16 weeks	12 weeks
Average transcript/code kappa scores	We used Cohen's (1960) kappa coefficient, where a kappa score of 0 indicates agreement no better than chance and 1 signals perfect agreement.	
Target *Round 1* *Final round*	0.70 0.69 0.91	0.70 0.35 0.71
Challenges	• Length of text blocks coded: Reconciled through discussion and revision of NVivo file by lead researcher.	• Length of text blocks coded: Reconciled by coding entire articles. • Complex codebook with numerous subcodes made it difficult to come to shared agreement on code meanings with 8 coders: Reconciled by very detailed codebook with instructions and repeated conference calls and email discussions.

* The following spreadsheet was utilized for average kappa calculations: http://redirect.qsrinternational.com/examples-coding-comparison-nv10-en.htm.

of each intercoder reliability process. Finally, we present a brief conclusion in which we argue the value in utilizing a deliberative coding process with large coding teams to ensure quality of qualitative data interpretation.

THE CASES

Case Study 1 – Crop Advisor Interviews

This project entailed analyzing interviews of agricultural advisors in three US Midwestern states – Indiana, Iowa, and Nebraska. Past research suggests that agricultural advisors are influential components of producers' agricultural decision-making process (Arbuckle, Morton, and Hobbs 2015; Prokopy et al. 2015). Current and future risks to agriculture from climate change and the potential influence of agricultural advisors on producer behaviors, led us to conduct interviews with advisors to discover how they perceive and appraise risk from climate change and how their perceptions relate to risk management advice to their clients. The interview guide consisted of broad questions about agricultural risk generally and then moved to climate-specific topics such as advisors' climate change belief, concerns over climate related impacts, and climate change risk management strategies (Church et al. 2018). Fifty-seven interviews were recorded and transcribed. See Church et al. (2018) for detailed research design, data collection, and data analysis procedures. Table 1 outlines the intercoder reliability process.

Case Study 1 – Assessment of the analysis

This approach proved extremely successful in terms of reaching very high agreement between the lead coder and the other eight coders. Although it would have been relatively straightforward for a smaller coding team to analyze the nine sample transcripts, the inclusion of additional coders and their interpretations served to enhance the credibility of the framework that emerged. In this case, the impressive level of agreement was aided by the reliance on a relatively small number of broad codes. Through collaboration with the coding team, these broad codes were accompanied by clear cues to explain what each code should encompass. Early discussions about when cross coding should be permitted (i.e. coding one piece of text to two or more codes) and at what scale coding should occur (sentence, paragraph, etc.) also helped to ensure greater agreement. Owing to the unique constraints on the coders' availability and their individual interpretation of the initial framework, one-to-one meetings were preferred as a means of addressing instances of coding disagreement. In numerous cases, a coder was found to have simply missed an opportunity

to code rather than misinterpreting what the code itself encompassed. While a reliance on one-to-one meetings allowed for focused and pertinent discussions, it necessitated additional meetings when subsequent discussions led to new insights or ideas about the coding framework. In this case, we present the results of the development of broad coding themes. The same process was repeated to reach the level of detail required to answer specific research questions.

Case Study 2 – Content Analysis of Agricultural Trade Publications

This study entailed the analysis of agricultural trade publication articles published before, during, and after the 2012 Midwestern US drought. We investigated how these publications discussed drought over time, whether climate change was examined in relation to drought, and if linkages between drought and climate risk and variability were reported. Although we performed keyword counts of the overall article population (e.g. climate change), it was important to interpret the meaning of these words. Therefore, this project entailed qualitative coding of a sample of the article population. For example, although we found that 79 articles in the article population contained the words “climate change,” without reading those articles we could not say that, within the sample, climate change was generally not discussed as a cause of drought; or that when climate change was discussed, there was typically not a debate as to whether or not it was actually occurring (Church et al. 2017). Through this study, we sought to discover how climate change risk and adaptation communication strategies are conveyed to the agricultural community. We found 2,846 relevant articles and coded a sample of 1,000. See Church et al. (2017) for detailed research design, data collection, and data analysis procedures. Table 1 outlines the intercoder reliability process.

Case Study 2 – Assessment of the analysis

Overall, the coding process was successful for a large content analysis project. Indeed, without eight coders it would have been difficult to analyze such a large article sample. However, the benefit of having an eight-person coding team was also the primary difficulty of the project. While data management of each group of articles was a challenge (not to be taken lightly), achieving a suitable kappa coefficient was problematic. Low kappa coefficient scores did not occur because of wholesale disagreement with the codebook; rather, discrepancies occurred because of varied interpretations of just a few codes. These issues were resolved through team discussions about code meanings and article interpretations and,

importantly, details and coding “rules” being added to the codebook. These group discussions were an important aspect of coding agreement, as the deliberations fostered shared meaning of the codebook. There were other issues in this process, namely, there was some frustration among the coders as the codebook changed and, as mentioned previously, the project itself was a challenge to manage due to the large number of articles and large coding team. Despite these challenges, through the intercoder reliability process, we were able to achieve a suitable kappa score and analyze a large article sample with confidence in the entire team’s coding agreement. Notwithstanding, we recommend thinking through the challenges of undertaking a large content analysis project when laying out a project’s sampling strategy and subsequent approach to intercoder reliability.

CONCLUSIONS

Including multiple people in the qualitative research coding process is an important component of perceived reliability of research conclusions. Codebook development and the intercoder reliability process can occur with as few as two researchers. While there are times when larger coding teams may be warranted due to large amounts of text to be analyzed, we argue there is value in multi-coder teams, even with small collections of transcripts or articles. There are challenges to utilizing large, multi-coder teams, such as the long, iterative process of developing a shared meaning around coding themes and achieving an acceptable intercoder reliability score. We argue that this challenge is warranted. Not only are research outcomes perceived to be reliable, but we suggest that research quality also increases. Much time is spent discussing “good” reliability scores (e.g. Lombard et al. 2002), which is a goal we also sought to achieve. However, working with large teams entailed back and forth discussions of research ideas and meaning that added depth to each analysis that cannot be captured in a score. The kappa score (in our cases) put a number to this process; however, it was the process of working with a diversity of people that contributed to rich data analysis and outcomes. Could this be achieved with two people? Yes – but we argue not to the rich degree that occurred in the cases documented here. Finally, although joint processes of codebook development and intercoder reliability procedures can be utilized from groups of two coders to many more than two, this process is lost if not acknowledged in published literature. We thus put forth a plea to journals and researchers: report your qualitative

data analysis processes to address the hows, whys, and whats of rural agricultural lands management.

ACKNOWLEDGMENTS
Thanks go to our research teams, who worked through many rounds of coding and contributed to the original research discussed as case studies here. Thank you too, to anonymous reviewers whose comments helped improve this manuscript.

DISCLOSURE STATEMENT
No potential conflict of interest was reported by the authors.

FUNDING
Funding for this research was provided by the National Oceanic and Atmospheric Administration (NOAA) Sectoral Applications Research Program (SARP) grant number NA13OAR431012 and USDA National Institute of Food and Agriculture, Agriculture and Food Research Initiative Competitive Grant number 2011-68002-30220.

REFERENCES

Arbuckle Jr., J. Gordon, Lois Wright Morton, and Jon Hobbs. 2015. "Understanding Farmer Perspectives on Climate Change Adaptation and Mitigation: The Roles of Trust in Sources of Climate Information, Climate Change Beliefs, and Perceived Risk." *Environment and Behavior* 47(2):205-234. doi:10.1177/0013916513503832

Basit, Tehmina. 2003. "Manual or Electronic? The Role of Coding in Qualitative Data Analysis." *Educational Research* 45(2):143-154. doi:10.1080/0013188032000133548

Burla, Laila, Birte Knierim, Jurgen Barth, Katharina Liewald, Margreet Duetz, and Thomas Abel. 2008. "From Text to Codings: Intercoder Reliability Assessment in Qualitative Content Analysis." *Nursing Research* 57(2):113-117. doi:10.1097/01.NNR.0000313482.33917.7d

Church, Sarah P., Michael Dunn, Nicholas Babin, Amber Saylor Mase, Tonya Haigh, and L. S. Prokopy. 2018. "Do Advisors Perceive Climate Change as an Agricultural Risk? An In-depth Examination of Midwestern U.S. Ag Advisors' Views on Drought, Climate Change, and Risk Management." *Agriculture and Human Values* 35(2):349-365. https://doi.org/10.1007/s10460-017-9827-3

Church, Sarah P., Tonya Haigh, Melissa Widhalm, Silvestre Garcia de Jalon, Nicholas Babin, J. Stuart Carlton, Michael Dunn, Katie Fagan, Cody L. Knutson, and L. S. Prokopy. 2017. “Agricultural Trade Publications and the 2012 Midwestern U.S. Drought: A Missed Opportunity for Climate Risk Communication.” *Climate Risk Management* 15:45-60. https://doi.org/10.1016/j.crm.2016.10.006

Cohen, Jacob. 1960. “A Coefficient of Agreement for Nominal Scales.” *Educational and Psychological Measurement* 20(1):37-46. http://dx.doi.org/10.1177/001316446002000104

Creswell, John W. 2013. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, CA: Sage.

DeCuir-Gunby, Jessica T., Patricia L. Marshall, and Allison W. McCulloch. 2011. “Developing and Using a Codebook for the Analysis of Interview Data: An Example from a Professional Development Research Project.” *Field Methods* 23(2):136-155. doi:10.1177/1525822X10388468

Kurasaki, Karen S. 2000. “Intercoder Reliability for Validating Conclusions Drawn from Open-ended Interview Data.” *Field Methods* 12(3):179-194. https://doi.org/10.1177/1525822 X0001200301

Leavy, Patricia, ed. 2014. *The Oxford Handbook of Qualitative Research*. New York: Oxford University Press. doi:10.1093/oxfordhb/9780199811755.001.0001

Lombard, Matthew, Jennifer Snyder-Duch, and Cheryl Campanella Bracken. 2002. “Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability.” *Human Communication Research* 28(4):587-604. doi: 10.1111/j.1468-2958.2002.tb00826.x

Lu, Chi-Jung, and Stuart W. Shulman. 2008. “Rigor and Flexibility in Computer-based Qualitative Research: Introducing the Coding Analysis Toolkit.” *International Journal of Multiple Research Approaches* 2(1):105-117. doi:10.5172/mra.455.2.1.105

MacQueen, Kathleen M., Eleanor McLellan, Kelly Kay, and Bobby Milstein. 1998. “Codebook Development for Team-based Qualitative Analysis.” *Cultural Anthropology Methods* 10(2):31-36. https://doi.org/10.1177/1525822X980100020301

Noble, Helen, and Joanna Smith. 2015. “Issues of Validity and Reliability in Qualitative Research.” *Evidence-Based Nursing* 18(2):34-35. doi:10.1136/eb-2015-102054

Olson, Joel D., Chad McAllister, Lynn D. Grinnell, Kimberly Gehrke Walters, and Frank Appunn. 2016. “Applying Constant Comparative

Method with Multiple Investigators and Inter-Coder Reliability." *The Qualitative Report* 21(1):6-42. Retrieved from https://nsuworks.nova.edu/tqr/vol21/iss1/3/

Prokopy, Linda S. 2011. "Agricultural Human Dimensions Research: The Role of Qualitative Research Methods." *Journal of Soil and Water Conservation* 66(1):9A-12A. doi:10.2489/jswc.66.1.9A

Prokopy, Linda S., J. Stuart Carlton, J. Gordon Arbuckle, Tonya Haigh, Maria Carmen Lemos, Amber Saylor Mase, Nicholas Babin, Michael Dunn, Jeff Andresen, Jim Angel, Chad Hart, and Rebecca Power. 2015. "Extension′s Role in Disseminating Information about Climate Change to Agricultural Stakeholders in the United States." *Climatic Change* 130(2):261-272. https://doi.org/10.1007/s10584-015-1339-9

Prokopy, Linda S., Kristin Floress, J. Gordon Arbuckle, Sarah P. Church, Francis Eanes, Yuling Gao, Benjamin M. Gramig, Pranay Ranjan, and Ajay S. Singh. 2019. "Adoption of Agricultural Conservation Practices in the United States: Evidence from 35 Years of Quantitative Literature." *Journal of Soil and Water Conservation* 74(5).

Rust, Roland T., and Bruce Cooil. 1994. "Reliability Measures for Qualitative Data: Theory and Implications." *Journal of Marketing Research* 31(1):1-14. doi:10.2307/3151942

Sim, Julius, and Chris C. Wright. 2005. "The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements." *Physical Therapy* 85(3):257-268. doi:10.1093/ptj/85.3.257

Sin, Samantha. 2010. "Considerations of Quality in Phenomenographic Research." *International Journal of Qualitative Methods* 9(4):305-319. https://doi.org/10.1177/160940691000900401

Stevens, Mark R., Ward Lyles, and Philip R. Berke. 2014. "Measuring and Reporting Intercoder Reliability in Plan Quality Evaluation Research." *Journal of Planning Education and Research* 34(1):77-93. doi:10.1177/0739456X13513614

Whittemore, Robin, Susan K. Chase, and Carol Lynn Mandle. 2001. "Validity in Qualitative Research." *Qualitative Health Research* 11(4):522-537. https://doi.org/10.1177/104973201129119299