Applying Constant Comparative Method with Multiple Investigators and Inter-Coder Reliability

Joel D. Olson
Kaplan University, jolson@kaplan.edu

Chad McAllister
Walden University, chad.mcallister@waldenu.edu

Lynn D. Grinnell
St. Petersburg College

Kimberly Gehrke Walters
Kaplan University

Frank Appunn
Northcentral University

Follow this and additional works at: https://nsuworks.nova.edu/tqr

Part of the Quantitative, Qualitative, Comparative, and Historical Methodologies Commons, and the Social Statistics Commons

Applying Constant Comparative Method with Multiple Investigators and Inter-Coder Reliability

Abstract

Building on practice, action research, and theory, the purpose of this paper is to present a 10-step method for applying the Constant Comparative Method (CCM) of grounded theory when multiple researchers perform data analysis and meaning making. CCM is a core qualitative analysis approach for grounded theory research. Literature suggests approaches for increasing the credibility of CCM using multiple researchers and inter-coder reliability (ICR), but documentation of methods for collaboration on CCM data analysis is sparse. The context for developing the10-step CCM approach was a qualitative study conducted to understand the impact of webcams on a virtual team. To develop a methodology for the study, the researchers reviewed grounded theory literature to synthesize an approach for conducting CCM with multiple researchers. Applying action research, an integration of literature and practical experience conducting the qualitative study resulted in a model for using CCM with multiple researchers performing data analysis. The method presented in this paper provides practical guidance for applying CCM collaboratively and shares the researchers' perspectives on the value of ICR.

Keywords

Grounded Theory, Constant Comparative Method, Inter-Coder Reliability, Researcher Collaboration, Action Research

Creative Commons License

Applying Constant Comparative Method with Multiple Investigators and Inter-Coder Reliability

Joel D. Olson
Kaplan University, Denver, Colorado, USA
Chad McAllister
Walden University, Minneapolis, Minnesota, USA
Lynn D. Grinnell
St. Petersburg College, St. Petersburg, Florida, USA
Kimberly Gehrke Walters
Kaplan University, St. Charles, Illinois, USA
Frank Appunn
Northcentral University, Prescott Valley, Arizona, USA

Building on practice, action research, and theory, the purpose of this paper is to present a 10-step method for applying the Constant Comparative Method (CCM) of grounded theory when multiple researchers perform data analysis and meaning making. CCM is a core qualitative analysis approach for grounded theory research. Literature suggests approaches for increasing the credibility of CCM using multiple researchers and inter-coder reliability (ICR), but documentation of methods for collaboration on CCM data analysis is sparse. The context for developing the10-step CCM approach was a qualitative study conducted to understand the impact of webcams on a virtual team. To develop a methodology for the study, the researchers reviewed grounded theory literature to synthesize an approach for conducting CCM with multiple researchers. Applying action research, an integration of literature and practical experience conducting the qualitative study resulted in a model for using CCM with multiple researchers performing data analysis. The method presented in this paper provides practical guidance for applying CCM collaboratively and shares the researchers' perspectives on the value of ICR. Keywords: Grounded Theory, Constant Comparative Method, Inter-Coder Reliability, Researcher Collaboration, Action Research

When qualitative research methods are used, data analysis may be completed by an individual or a group of two or more people. Researchers accustomed to completing independent data analysis may be surprised by the large amount of additional time and effort that working with a research group requires. Collaboration adds complexity to the work of data analysis and formulating findings, making a collaborative qualitative study more labor intensive (Miles & Huberman, 1994). Additional coordination and iteration are required for the qualitative coding process for creating themes, analyzing for meaning, and drawing conclusions. When members of a research team are geographically separated and working in a virtual environment, data analysis may be more challenging. However, the collaboration provides several benefits that derive from the additional perspectives provided by multiple researchers. In striving for consensus in the findings, the nuances in meaning brought by multiple researchers adds richness to the analysis by prompting deeper analysis. Inter-coder reliability (ICR) can be used to drive towards consensus but was found to be more suited for identifying nuance and significant meanings in the qualitative data. This paper explores the experiences of two geographically separated researchers who applied Constant Comparative

Method (CCM), based on grounded theory. The researchers applied action research to formulate a deliberate 10-step method for coding data, creating meaning, and structuring an exploratory model that represents findings. Collaboration was facilitated through synchronous online video discussions and email exchanges to work through analysis activities between the two researchers.

Literature Review

Literature on qualitative research, and specifically on the CCM methodology used by the researchers performing this study, reveals a diversity of positions that reflect the richness of qualitative research (Strauss & Corbin, 1998). There are supporters and opponents to qualitative research in general and CCM in particular. This review begins with a basic explanation of the approach that differentiates qualitative research from quantitative; then explores the methods used in qualitative research to address issues common to quantitative researchers involving validity and reliability. Finally, the review will focus on the literature concerning advantages, disadvantages, and potential roles of ICR measures in CCM.

Inductive Approach

The original purpose of qualitative methods was to design a structured approach for generating new theory that purports to explain an experience or phenomenon for which current understanding is inadequate. Qualitative research uses inductive reasoning (i.e., developing explanations from information) rather than the deductive (i.e., using theory to predict outcomes based on information) to draw conclusions from data. It explores a deliberately selected set of data, such as interviews, observations, or video/audio logs, to identify patterns that can be linked causally in a model or theory (Thomas, 2003). Models generated by qualitative theory can be tested using quantitative methods to provide further support for the theory. Quantitative research uses existing theory to generate a question or hypothesis that can be tested empirically (Curry, Nembhard, & Bradley, 2009).

Grounded Theory

Grounded theory is a qualitative research method developed to facilitate discovering patterns in data (Glaser & Strauss, 1967). It uses a systematic approach to review participant views collected from an experience in order to allow patterns and themes to emerge over multiple passes through the data. Strauss (1987) further elaborated on the data analysis methodology, creating CCM, in which the researcher developed codes while reviewing transcripts or other verbatim data to identify constructs, and iteratively compared texts identified with the same code to ensure they were representative of the same construct. Connections observed between constructs were described as patterns, and generalizations drawn from patterns observed in case studies were described as themes. A synthesis of the information results in an exploratory model.

Challenges exist when generating codes. Initially codes may be fairly shallow, leading to an overwhelming number of codes, but deepen throughout the process as multiple instances of concepts occurring in close proximity to each other highlight connections between codes (Scott, 2009). These patterns identify more substantive codes, ultimately providing a theoretical structure that can be tested through further analysis of existing data. A line-by-line (or unit-by-unit) examination of text to identify codes is critical to reaching the level of saturation needed to mitigate the risk of missing concepts important to the analysis (Holton, 2010). Key to the process is a researcher's intuitive sense to discern repeating patterns in very

different contexts. Patterns or themes extending across an entire phenomenon being observed could be constructed into a theoretical model, validated through further empirical studies.

Balancing process and purpose. More recent research has considered the messy aspects of qualitative research, challenging a too-methodical approach to qualitative analysis because it does not reflect either the intent or the actual development of qualitative research results (Sinkovics & Alfoldi, 2012). Conducting qualitative analysis too methodically, such as in a checklist manner, could risk leading to a more deductive approach rather than the desired inductive approach, searching for patterns in the data rather than imposing them (Barbour, 2001). Sinkovics and Alfoldi suggested use of computer-aided qualitative data analysis could capitalize on the more fluid nature of theme development while still preserving the rigorous nature of a qualitative approach such as CCM. By exploiting the fluid and spontaneous connections between data and theory, researchers inevitably engage in a *progressive focusing* to concentrate their efforts on emerging themes. Progressive focusing often begins with a thorough review of current literature, but it is important to note that most researchers will have extensive expertise in some aspect of the literature that will cause certain patterns to capture their attention. Exploring this messier aspect of qualitative research has been called "intellectual bricolage" and "the creative exploitation of existing resources or materials" (Lambotte & Meunier, 2013, p. 86). Although CCM is described as a linear iterative process, the personal perspectives of researchers' impact making relevant connections in the data, and creates a "thickness" that enhances the value of qualitative analysis. Recognition and embrace of the iterative learning process is a necessary step toward identifying new theory.

Case studies. Single and multiple case studies have provided effective venues for observing and collecting data (Gibbert, Ruigrok, & Wicki, 2008). When applying grounded theory analysis, case studies provided a deeper understanding of the social aspects of adoption of innovation (Lawrence & Tar, 2013). Diaries and reflections are recognized in qualitative research as rich and multi-layered sources of data, giving participants the opportunity to capture their recollections of events and the accompanying emotions in the relatively immediate aftermath of those events (Radcliffe, 2013). Structured reflection questions may have a restrictive effect on the range of issues discussed by participants, but can provide more substantive replies grounded in the common background of the participants (Grinnell, 2003).

CCM Research: Validity and Reliability

Research validity and reliability are common concepts in quantitative research but also applicable in qualitative research as both must establish credibility. This topic was investigated by Golafshani, who concluded "… when quantitative researchers speak of research validity and reliability, they are usually referring to a research that is credible while the credibility of a qualitative research depends on the ability and effort of the researcher" (2003, p. 600). Lincoln and Guba describe reliability, when applied to qualitative research, as the dependability of the research (1985). Validity in the qualitative context is viewed as producing findings that are trustworthy and defensible (Golfashani, 2003). These are the contexts used in this paper for the validity and reliability.

Gibbert et al. (2008) reported the questionable nature of the validity and reliability of qualitative research. The primary data of qualitative research is based on observations subject to participant bias and analysis methods subject to researcher bias. However, opposing views in critical realism (Madill et al., 2000) contend it is the nature of the human brain to filter new data through its past experiences, beliefs, and expectations. Hence, it is better to acknowledge the subjective nature of all observations and analysis (including quantitative), and adapt collection and analysis methods to address those concerns (Altheide & Johnson, 1998; Leininger, 1994).

Burton-Jones (2009) identified methods bias, defined as "the difference between the measured score of a trait and the trait score that stems from the rater, instrument, and/or procedure used to obtain the score" (p. 448), as a key issue affecting reliability and validity in qualitative studies. This is considered a strength and weakness of qualitative research (Strauss & Corbin, 1998) – a strength because advocates believe that all research is biased, and qualitative research is more honest in acknowledging the sources of probable bias in the researchers' background and experience; and a weakness in that qualitative studies must be more tentative in their conclusions. Burton-Jones suggested that knowledge bias (i.e., the ability of analysts to accurately identify intrinsic traits from self-reports) will "significantly affect relationships in the theoretical model" (p. 465). If researchers are participants, their analysis could be less biased because they know their own mind, but could be more biased because of social desirability bias (Crown & Marlowe, 1964) or a lack of self-awareness (Luft & Ingham, 1955). If researchers are independent of the participants, their analysis could be less biased because they are not affected by others' perceptions of their answers; however, they might project their own beliefs on the participants.

Triangulation has been one of the most widely used methods for increasing validity of qualitative research (Gibbert et al., 2008; Johnson, 1997; Jonsen & Jehn, 2009; Kirk & Miller, 1986; LeCompte & Preissle, 1993; Lincoln & Guba, 1985). Triangulation uses multiple data points (events, times, locations, participants, etc.), researchers, and analysis methods to generate findings. Where findings from different sources converge, triangulation identifies this as evidence of stronger support for the findings. Independent coding by outside researchers can enhance the trustworthiness of CCM data analysis to provide a powerful boost to construct validity (Thomas, 2003). External validity can come from cross-case analysis (Gibbert et al., 2008). Four to ten case studies that are within one organization (nested) or across different organizations can provide rich data for thoughtful theory development. Finally, internal validity of any theories or models generated from the data is strengthened by triangulating those with existing theories. Overall validity of research design is gauged by its use of triangulation to contribute to trustworthiness and defensibility (Jonsen & Jehn, 2009).

Inter-coder Reliability

Using multiple coders to analyze data can increase validity through triangulating but may reduce reliability because of inconsistency in coding (Harris & Burke, 2011). Reliability is improved with the transparency provided through a clear protocol, codebook, and database (Gibbert et al., 2008). While this provides information to aid in replicating studies, it does not provide a measure of internal consistency, a reliability measure common to quantitative studies that improves defensibility. Inter-coder reliability (ICR) checks have been proposed to further improve reliability by measuring agreement between multiple coders (Harris & Burke, 2011). Four methods of ICR have been proposed to measure consistency between coders: (a) percent agreement, (b) Chi-Square, which calculated the association, but not agreement between two coders; (c) Intraclass Correlation Coefficient (ICC), which compares variability on individual items to variability for all items; and (d) Cohen's Kappa, which measures the percent agreement, correcting for chance.

Inter-coder reliability has been applied in content analysis, providing a quasi-quantitative analysis of qualitative data, in which multiple researchers code, clarify, and re-code data until a specific level of agreement is achieved (Neuendorf, 2002). Cohen's Kappa is commonly used for calculating this agreement, since it corrects for chance agreement (Foster, Urquhart, & Turner, 2008). Content analysis is based on exploring the presence of key constructs from existing theories in a qualitative data set. ICR is a verification strategy used during this process to help researchers clarify their understanding of what constitutes the

presence of a particular construct in the data. Content analyses typically report their ICR as a measure of reliability of their analysis (Grinnell, 2003; Olson et al., 2012). However, ICR checks' usefulness in CCM analysis has been questioned (Marques & McCall, 2005).

Inter-coder reliability may have a different role in CCM (Marques & McCall, 2005). In CCM, codes representing patterns and themes are allowed to emerge from the data rather than being selected from the literature prior to the study (Strauss & Corbin, 1998). Studies using ICRs in CCM or other grounded theory analysis have much lower percentage of agreement between researchers than what is expected in content analysis (Marques & McCall, 2005). Explanations from the studies suggest that the codes generated by the researchers are inevitably influenced by the different background and experiences of the researchers. Despite the low agreement, however, ICR can play an important part in the analysis process. In Foster et al. (2008), ICRs were initially low, but increased to satisfactory levels after discussion and clarification of the codes being generated. Granularity in coding (i.e., level of detail) made it more difficult for each researcher to select the same codes to describe the same unit of analysis. Marques and McCall (2005) suggested ICR can be used as a "solidification" strategy. Diversity in researcher coding, they purported, is a strength rather than a weakness. Insights gained through the discussions of disagreements can be a powerful aid to understanding the patterns emerging from the data (Curry et al., 2009). Researcher analytic style, for example, can reveal personal internal models of reality that influence discernment and interpretation of patterns (Foster et al., 2008). This can play a dual role of recognizing and ameliorating bias on one hand (Madill et al., 2000), and highlighting differences in viewpoints that contribute to the richness of data interpretation on the other hand (Foster et al., 2008).

Qualitative Pilots

In quantitative research, pilot studies have been used to help researchers test questionnaire design or sampling (Pritchard & Whiting, 2012), thus improving validity and reliability of their study. More rarely used in qualitative studies, pilots have helped researchers understand the context and the process of data gathering and analysis. This can be particularly helpful for training coders, as well as developing "mutual exchange and interaction through the establishment of research relationships" (Caine et al., 2009, p. 491). Piloting the data analysis phase of qualitative research can allow for exploration of issues that may affect the overall analysis approach, a process Pritchard and Whiting call "forward reflexivity" (2012, p. 350). Pilots can therefore serve as test beds of innovation and conflict resolution, leading ultimately to more valid conclusions.

Literature Review Summary

After reviewing the development of qualitative research related to grounded theory since Glaser and Strauss wrote their 1967 paper, it is clear that its subjective nature has caused significant concerns, yet benefits remain (Gibbert et al., 2008). Qualitative research has filled an important gap in knowledge creation. Quantitative research depended on the existence of valid and reliable exploratory models to predict outcomes in new studies. Qualitative research provided structure and method to the process of developing new theory. However, the subjective nature of qualitative data and methods of analysis has lent credence to the criticism of CCM. As a result, there has been considerable effort in developing strategies to increase validity and reliability of CCM. Triangulation has helped validity, but there is still much disagreement on methods to increase reliability (Foster et al., 2008). The usefulness of ICR, which has proved its worth in content analysis, for improving CCM's reliability has been

questioned. Given the variation in conclusions drawn from the research, this study adds to the body of literature by determining one dimension of value in utilizing ICR in CCM.

Role of Researchers

The research team has studied the performance impact of incorporating video into team communications. The original research was qualitative in nature and involved multiple researchers collaboratively analyzing logs recorded by virtual team participants. CCM was chosen as a credible methodology, but guidance for using it in a collaborative setting was minimal, resulting in the need to develop a protocol for applying CCM with multiple researchers. The findings of the research team regarding the impact of video on virtual team performance have been previously published (Olson et al., 2012).

Study Context and Method Review

The context for the application of CCM was an analysis of the impact of adding webcams to synchronous virtual team interactions. The findings have been published by the authors (Olson et al., 2012). The study employed the CCM method, which is based on grounded theory (Glaser, 1992; Glaser & Strauss, 1967; Strauss, 1987). The study data consisted of transcribed logs by five team members (study participants) who participated in a virtual team. We added webcams to the virtual team meetings and participants recorded their impressions of the changes. We recorded data weekly for seven weeks, resulting in seven sets of logs, Week 1 to Week 7. Using sentences as the unit of analysis, this resulted in a rich sample of 1271 sentences across five participants and seven logs.

Boeije (2002) provided an application of CCM by a single researcher, illustrated by example. We adapted this model for use by multiple researchers as well as applying the comparative model to previously collected data. Both of these factors introduced additional issues to manage:

1. Finding ways to synthesize the analysis created by individuals, addressing the inevitable differences in perspectives that arise (Miles, Huberman, & Saldana, 2014).
2. Identifying and agreeing upon the significant themes found in the data to pursue for further analysis.
3. Applying CCM retrospectively to previously collected data without having the opportunity to refine interview questions as the data is collected.

We employed an action research approach to create a CCM research method that resolved these problems, integrating practice and theory. Action research provides a framework for professional learning and "enables practitioners everywhere to investigate and evaluate their work" (McNiff & Whitehead, 2006, p. 7). Action research uses a systematic learning loop to resolve problems or concerns: observe – reflect – act – evaluate – modify (McNiff & Whitehead, 2006). We began by applying Boeije's application of CCM individually, reflecting on how it would need to be modified for use with multiple researchers. This was put into action and the process evaluated. Weaknesses were discussed and the process further modified. Several iterations of the action research learning loop were conducted, resulting in the 10-step method and the effective use of ICR presented below.

Although we used action research throughout the application of CCM to the webcam study to create the 10-step method, we first used the action research learning loop on a pilot study. This resulted in a well-reasoned approach to applying CCM with multiple researchers

before being exposed to the actual data of the webcam study. The pilot study used data previously collected for a different purpose. This was the Cinderella Study, which has been used by researchers who are learning qualitative methods (Schutt, 2012). The pilot study provided the opportunity to reframe the research method outlined by Boeije (2002) and gain experience working together through iterations of the learning loop. It proved to be a valuable activity, building on each researcher's prior experience analyzing qualitative data individually (Pritchard & Whiting, 2012).

The three problems listed above and resolved with action research are described below and further addressed later in the paper as the details of the method are presented.

Synthesizing Individual Analysis

CCM applies the art of comparison, the foundation of grounded theory. As a qualitative research method, the focus is theory building and specifically grounding the theory to the data. Strauss and Corbin (1998) describe it as the interaction between researchers and data using repeatable processes.

Miles and Huberman (1994) provided guidelines for two researchers working together on qualitative analysis. This included independently coding the same data and discussing difficulties and disagreements. Further, inter-coder reliability (ICR) could be used to identify where disagreement exists. They shared:

> Each coder will have preferences – sociologists see organization-level codes for the same blocks of data that are intrapersonal for psychologists and interpersonal for social psychologists – and each vision is usually legitimate, especially for inferential codes. Clarifying these differences is also useful; each analyst tends to be more ecumenical during later coding for having assimilated a colleague's rival vision of data that initially looked codable in only one way. (p. 64)

Therefore, we adopted ICR as a tool to foster focused collaboration. Initially, we used ICR to quantify agreement between application of codes. The Coding Analysis Toolkit ("CAT," 2010) was employed to calculate Fleiss' Kappa as the measure of agreement. This identified differences in understanding or perspectives of codes, as well as providing insights into coding variation. As an example, Table 1 shows a subset of the first round of codes related to identifying communication and connection/authenticity when webcams are added to virtual team meetings and the Kappa value for each. A Kappa equal to 1 indicates we applied the code to identical sentences in the data – perfect agreement. Kappa values approaching 0 indicate we did not apply the codes in the same manner, nor to the same sentences – a high degree of disagreement. We met via online video conference to discuss codes with low Kappa values, identifying differences in code definitions and application of the code. This led to refining code definitions and coding guidelines for the next round of coding.

Table 1. Example ICR values for Codes

Code	Kappa
Communication--Added Richness	0.78
Communication--Decreased	0.77
Communication--Increased	1.00
Communication--Unchanged	0.67
Connection/Authenticity--Decreased	1.00

Connection/Authenticity--Increased   0.91

Connection/Authenticity--Unchanged 0.20

Initially we collaborated to increase the Kappa values, striving to reach a minimum agreement of 70%. However, based on our previous qualitative research experience, we felt overly constrained by using ICR in this manner. We found ourselves thinking less about the meaning of the data for ourselves and more about how the other researcher was likely to apply codes. The objective became the Kappa value and we were concerned some of the richness derived from the data could be lost. Video collaboration helped us transition to new uses for ICR. As a result of observing discomfort with the initial process, the use of ICR was changed from seeking a specific minimum Kappa value to using ICR to guide collaboration and identify nuances in the data brought to light by our prior experience, knowledge, and perspectives. This moved the emphasis on ICR from a tool for quantitatively assessing solidification to a tool for focusing collaboration (Marques & McCall, 2005).

Identifying Significant Themes

Coding data is an iterative process, with researchers individually identifying codes and creating definitions, synthesizing their individual codes to create a unified master set of codes, and reapplying the unified codes to the data. Codes self-emerge as researchers analyze the data for themes. In the process of individually identifying codes, unifying codes, and coding additional data per the CCM method, the number of codes can shrink and expand during the process. For the study, we settled on 47 codes organized in 10 categories. This provided many possible dimensions for further analysis. Dimensions that appeared significant to one researcher may not be the same for the other researcher. A consistent method was sought to identify dimensions that would receive additional analysis. We strived to maintain an inductive approach by searching for patterns in the data rather than forcing preconceived ideas on the data.

After much researcher collaboration, it emerged in discussions that the themes in the data, identified by the coding process, changed as a function of time. Study participants were emphasizing or deemphasizing their focus regarding the use of Webcams over the seven-week period. For example, there was an initial hesitation and stress related to the use of Webcams when they were first introduced but this decreased in a few weeks. Therefore, we chose to identify the themes for further analysis by determining which ones varied the most as a function of time.

Retrospective Comparisons

In the CCM steps described by Boeije (2002), data are collected from a study participant and then analyzed using open coding. Based on the coding analysis, the data collection protocol may be enhanced before collecting data from the next study participant. This sequence can continue throughout the study. This is the constant comparative nature of CCM -- using data analysis to refine future interactions. However, our task in this study was to analyze data that had already been collected. This presented the issue of how to adhere to the spirit of CCM without having the possibility of asking modified or additional questions. We took the approach of analyzing the data in time sequence.

The time-dependent analysis took the form of open coding the oldest transcript from the first participant, moving to the second participant, and so on. After analyzing each week's set of transcripts we collaborated on their analysis, synthesizing the codes to create a unified codebook. They then reapplied the codes to the same set of transcripts. Then we progressed to

the second oldest set of transcripts from the participants, conducting the same analysis, collaboration, synthesis, and reapplication. This time ordered analysis continued until all transcripts were analyzed, resulting in a master coding structure that was then once again reapplied to all transcripts in order from oldest to most recent. This resulted in a mechanism for observing changes in data as a function of time and allowing comparisons from any point in time to be made.

Results: A CCM Model for Collaboration

The CCM method we employed is described and illustrated by example.

We used the following 10 step CCM method:

1. Each researcher performed open-coding of Week 1 logs.
2. Collaborated to unify codes.
3. Each researcher re-coded Week 1 logs using unified codes.
4. Calculated ICR.
5. Collaborated to discuss each code and identify areas lacking agreement.
6. repeated the above process for each week of logs, producing a unified codebook applicable to all logs.
7. re-coded all logs, producing themes.
8. selected themes for further analysis.
9. conducted co-occurrence analysis.
10. Constructed an exploratory model – the findings of the study.

The tools employed in the method included:

1. ATLAS.ti Qualitative Data Analysis & Research Software (see http://www.atlasti.com)
2. Coding Analysis Toolkit (see http://cat.ucsur.pitt.edu)
3. Microsoft Excel

Step 1: Open-Coding

ATLAS.ti software was used to manage open-coding. Transcripts of the Week 1 logs were loaded into ATLAS.ti, with one document for each participant. Creating separate documents for each log and each participant allowed for the possibility of comparing data by participant and week. This was a critical decision to make before coding began to simplify data analysis later. We collaborated on the unit for assigning codes, such as phrase, sentence, or paragraph. In previous qualitative research, we were accustomed to free coding at any level of unit appropriate for the data, but the same level needed to be used by both researchers to make the ICR calculation meaningful. We chose a sentence as the unit (Straus & Corbin, 1998). Consequently, in ATLAS.ti, a complete sentence would be selected and codes applied to it. Further, we discussed how many codes to apply per sentence and chose to strive for the one or two most important codes and not apply more than four codes (Miles, Huberman, & Saldana, 2014). As qualitative research tends to generate more data than can be managed, it necessitates a selection process based on the researchers' best judgment. This coding process helped to make the ICR calcAlion more meaningful and focused analysis on more significant codes per the researchers' judgment.

We then independently open-coded the Week 1 logs, creating codes as they read the logs. Definitions for each code were written and added to the Code Manager in ATLAS.ti. For example, the definition for Communications—Added Richness was, *Subject had additional means to express themselves via the video, allowing for more or different forms of creativity, use of humor, or facial expressions. It provides new ways, compared to a phone conference, to interact.*

We then individually refined the codes produced to identify redundancies and delete insignificant codes or those not directly related to the phenomena being studied. This was accomplished using the Merge Codes capability, renaming codes, and deleting codes. This decreased the initial number of codes and produced codes with more clear definitions.

Step 2: Code Unification

We exchanged codebooks by email so they could be reviewed before talking. Then we met via phone or Skype to discuss each code and its definition. The preferred means of communication was Skype, as video could also be used, increasing opportunities to make connections through body language and context from one's surroundings, which helped build a feeling of community between us (Olson et al., 2012). During the discussion, we merged codes with similar definitions and refined definitions. One of the researchers then created a new codebook in ATLAS.ti that reflected the decisions and sent it to the other researcher. The codebook exchange via email could be avoided by using a cloud-based research tool that supports collaboration of multiple researchers on a project. The advantage of such tools is that all researchers have access to the most current version of the project, simplifying version control.

Step 3: Re-Coding

We independently re-coded the Week 1 logs using the unified codebook. Each researcher avoided creating new codes during this process, but there were times that a researcher believed a significant theme had been missed previously. If so, a code was defined and added to the unified codebook. We discussed the addition during the next round of coding with the logs for the following week. The result of this step was two sets of Week 1 logs, each with codes from the same codebook applied at the sentence level.

Step 4: ICR Calculation

We merged the coding files from each researcher in ATLAS.ti. Quotes (sentences with codes applied by each researcher) were exported from ATLAS.ti and imported into Coding Analysis Toolkit (CAT). CAT calculated Fleiss' Kappa as a measure of ICR. An example is shown in Table 2 for the Communication and Connection/Authenticity coding that resulted from the unified codebook. CAT identifies the number of times the code was applied by each researcher as well as the amount of agreement, expressed as Kappa.

Table 2. Example CAT Output for ICR

Code	Researcher 1	Researcher 2	Kappa
Communication--Added Richness	26	22	0.78
Communication--Decreased	11	12	0.77
Communication--Increased	7	7	1.00

Communication--Unchanged	2	3	0.67
Connection/Authenticity--Decreased	3	3	1.00
Connection/Authenticity--Increased	34	31	0.91
Connection/Authenticity--Unchanged	5	1	0.20

Step 5: Researcher Collaboration

We met again via phone or Skype to review the ICR analysis. The information quickly identified areas of agreement and disagreement. As seen in Table 2, the code counts for Researcher 1 and 2 indicate whether each researcher is applying a code with similar frequency. For example, "Connection/Authenticity-Unchanged" was used much more by Researcher 1 than Researcher 2, resulting in a low Kappa value. It is also possible to have similar frequency but a small Kappa because each researcher applied the code to different sentences. Such occurrences were the focus of discussion and we reviewed code definitions and shared examples of how they were applying the code. The definition of the code was refined if needed. This step resulted in an enhanced codebook with a clearer understanding of the codes.

Step 6: Repeat and Unify

We repeated the previous five steps for each set of weekly logs. The analysis began with creating a codebook for Week 1. We used the same process to extend the codebook with new codes that were applicable to Week 2 as well as to apply the previous codes created in Week 1 as needed. This continued until logs of all seven weeks were coded. As discussed previously in the Retrospective Comparison section, the logs were analyzed in time order so codes could emerge and develop in the same order as the participants' experience emerged. This allowed for the nature of comparisons in the CCM method. Step 6, while rather tedious and lengthy (occurring over several weeks) resulted in a unified codebook that represented all codes we identified for the logs.

Step 7: Re-Code All Logs

We once again coded the logs in time order using the comprehensive codebook. We were careful to not add new codes – only apply those previously agreed to. The key concern in this step was coder fatigue (Miles, Huberman, & Saldana, 2014). Each researcher had already coded all of the logs twice and Step 7 was the third pass with the largest set of codes yet. To prepare for coding, we once again collaborated via Skype to review the codebook. They found opportunities to group some codes into themes, creating a manageable number of 10 categories, while the number of individual codes was fairly high at 47. Also, we found it useful to read the code definitions each time before they started coding. This provided grounding to their previous work and helped to add consistency to the coding process and combat fatigue. The result of this step was two sets of all logs, each with codes from the master codebook applied at the sentence level. As before, the coding files were merged into one in ATLAS.ti for further analysis.

Step 8: Analyze Trends

At this point a great deal of analysis had been conducted to identify the notable themes in the logs. The next step was to begin synthesizing the analysis and make meaning of it. We could explore many dimensions of the data: (a) comparing the responses of particular

participants, (b) considering what was not shared that may have been expected, (c) comparing changes in prevalence of themes over time, or (d) identifying the most important themes, among others. For the research findings to be useful, they needed to separate the wheat from the chaff -- the important from the not important. However, determining what is important is challenging for qualitative researchers. While the number of times a particular code is used may help to identify it as important, it is not a true indication of weight (Patton, 1990). A code only applied once may relate to a profound contribution by a study participant.

However, both researchers had identified interesting patterns in the narrative as we read through the logs in time order. Having reviewed the logs in sequence three separate times over the course of several weeks, they were very aware of participants' reactions varying over time. Consequently, we chose to explore those trends represented in the data. We sought a visual way for doing so as trends are often best visualized. To identify trends, relying on the number of times a code was applied was appropriate as this was not a measure of weight but an indicator of frequency in the perceptions of the participants.

ATLAS.ti was used to determine the code occurrence value for each week of logs. Given the existence of seven logs, seven values for each code were totaled, which could be plotted in a histogram. Microsoft Excel was used to create the histogram views based on data exported from ATLAS.ti. To remove magnitude as a consideration when viewing code histograms together (since magnitude is not necessarily an indication of importance), each histogram was normalized so they all would have similar peak values. To focus the analysis, we chose to further examine those themes that appeared to change the most throughout the seven weeks. Themes with little change were excluded from the analysis. The reasoning was that this would indicate what the participants found important and relevant to the phenomena as they chose to talk about it more or less over time – the change made it important. This resulted in the trend analysis shown in Figure 1, accounting for 18 themes of the 47 coded.
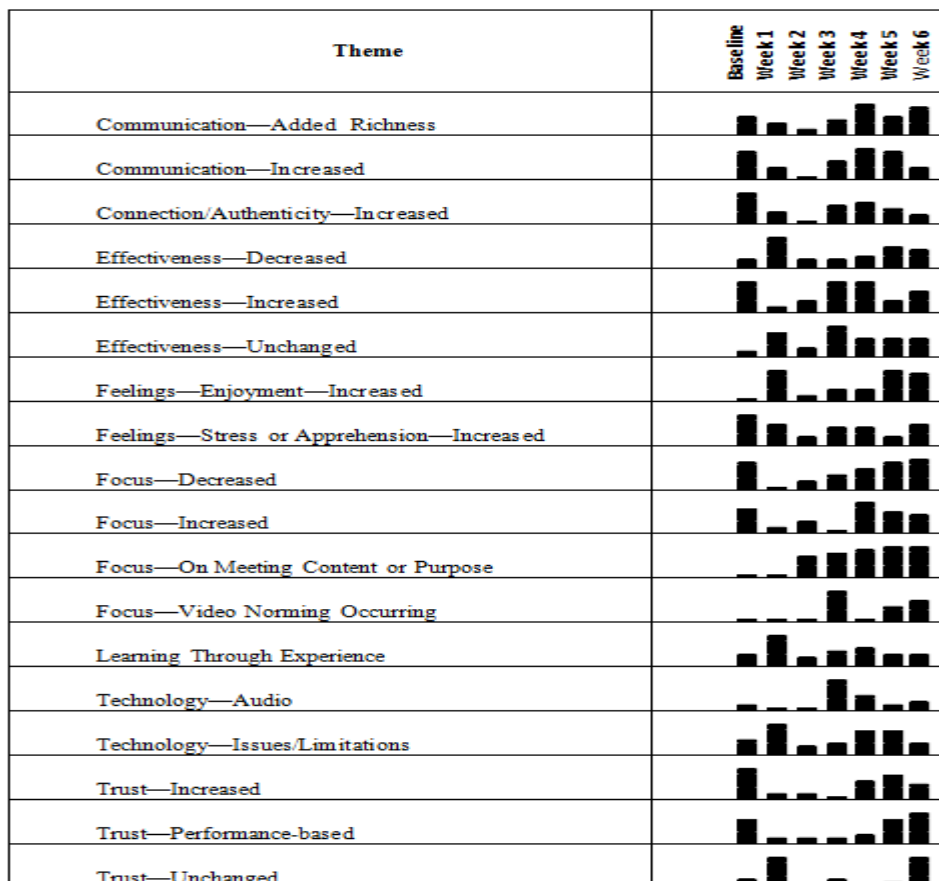
Theme	Baseline / Week 1 / Week 2 / Week 3 / Week 4 / Week 5 / Week 6
Communication—Added Richness	
Communication—Increased	
Connection/Authenticity—Increased	
Effectiveness—Decreased	
Effectiveness—Increased	
Effectiveness—Unchanged	
Feelings—Enjoyment—Increased	
Feelings—Stress or Apprehension—Increased	
Focus—Decreased	
Focus—Increased	
Focus—On Meeting Content or Purpose	
Focus—Video Norming Occurring	
Learning Through Experience	
Technology—Audio	
Technology—Issues/Limitations	
Trust—Increased	
Trust—Performance-based	
Trust—Unchanged	

Figure 1. Trends in themes based on normalized coding occurrence as a function of time. Adapted from "Towards a Theoretical Model of the Impacts of Incorporating Webcams in Virtual Teams," by J. Olson, L. Grinnell, and C. McAllister, F. Appunn, K. Walters, 2012, Review of Business Information Systems, 16(2), p. 79. Reprinted with permission.

Step 9: Co-Occurrence Analysis

Due to the familiarity we had with the narrative contained in the logs, a frequent topic of discussion was how one theme appeared to be related to a different theme. We used ATLAS.ti to conduct a co-occurrence analysis, identifying codes that were frequently used together, to further examine these observations. For example, it showed high co-occurrence between an increase in focus during a meeting and an increase in communication richness. As a result of Steps 8 and 9, we had a basis for making meaning – trends and dominant relationships in the trends.

Step 10: Constructing an Exploratory Model

The analysis of trends and relationships reduced the number of potential variables to explore from 47 to the four key themes we determined were the most significant for the studied phenomena. These themes included authenticity, focus, stress and learning technology, and effectiveness. A diagram was created to show the relationships, without an inference of absolute magnitude but only relative relationship, representing the exploratory model created from CCM. This is shown in Figure 2.
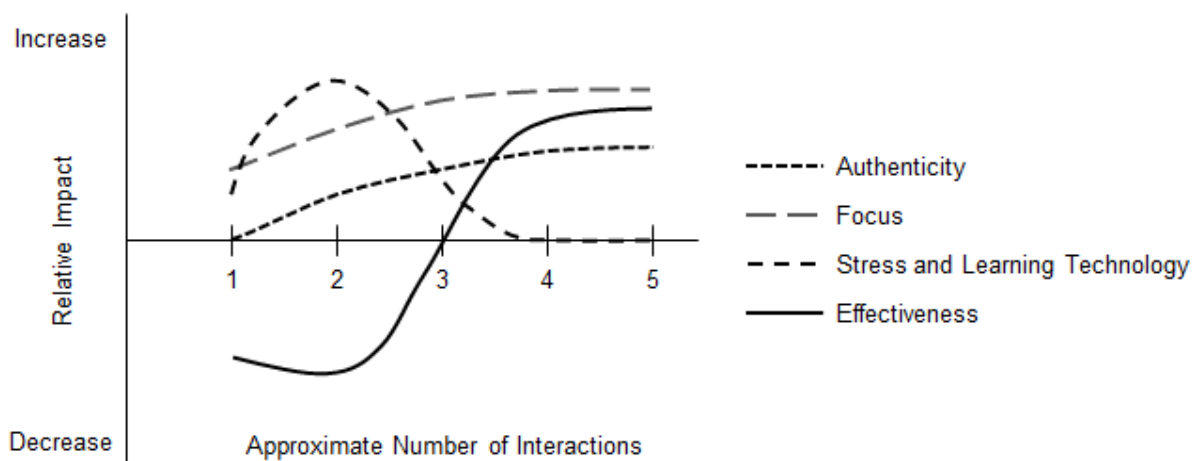

Figure 2. Proposed theoretical model for effect of video on virtual team development. Adapted from "Towards a Theoretical Model of the Impacts of Incorporating Webcams in Virtual Teams," by J. Olson, L. Grinnell, and C. McAllister, F. Appunn, K. Walters, 2012, Review of Business Information Systems, 16(2), p. 79. Reprinted with permission.

Discussion

Data analysis of qualitative data is a time consuming and tedious process. It often involves reading and re-reading transcripts, coding to identify themes, analyzing connections between themes, and recognizing trends. The process is further complicated when multiple researchers work together. Parsimony in generating codes is somewhat preferable to

proliferation of codes because more codes lead to lower reliability, even with a detailed codebook. Collaboration is required to unify codebooks, which are then reapplied to the transcripts. Initial discomfort is to be expected when using ICR to help solidify understanding of coding schemes, as differences between researchers are highlighted. Further discussion is required for all additional analysis ultimately leading to findings and recommendations. Two researchers collaborating on data analysis can create twice the work for each other that would have been required for a single researcher. The results of this collaboration must be worth the extra investment in time and effort.

The investment of the collaboration does have the benefits of providing additional perspectives on the data that would not be possible with a single researcher. Although qualitative researchers take steps to minimize bias in their analysis and findings, they are conducting their work through the lens of experiences, education, skills, and perspectives (Miles, Huberman, & Saldaña, 2014). A theme may appear insignificant (or be ignored) by one researcher and be very meaningful to another. Collaboration allows each researcher to share their position and decide together on the importance of the theme based on evidence.

Another benefit of the collaboration is increased validity through triangulation (Miles, Huberman, & Saldaña, 2014). Threats to validity of qualitative research occur on the dimensions of (a) description, (b) interpretation, and (c) theory (Maxwell, 1992). Description captures the need to accurately account for what was heard or seen. Interpretation represents the ability to reach defensible conclusions. Theory considers if there are alternative explanations for the findings. Collaboration between researchers can provide an additional defense for each of these dimensions. Researchers agree on the content of the data (description) through the coding process. They discuss support for conclusions (interpretation) when analyzing for meaning and trends. Further, alternative explanations (theory) are discussed when moving from analysis to an exploratory model. The 10-step method shared in this paper enhanced the research process and resulted in more robust data analysis because of the researchers' collaboration.

Further, when researchers are geographically separated, collaboration is greatly facilitated through the use of online video communication tools, such as Skype. A community of researchers can be built not only through the shared experiences and common language of periodic collaboration, but also through increases in communication richness through body language, visual setting, and the natural sharing of personal information external to the study. This can reduce potential conflict through increased visibility of feelings and reactions, thus speeding development of consensus within the research community.

In summary, this study confirmed that validity in qualitative research can be increased and analysis enriched through the use of multiple researchers in the data analysis phase. ICR can be used as an additional tool to aid discussions to clarify and solidify codebooks, increasing the reliability and replicability of analysis; however, it should be used with caution as it can be a constraint that violates the tenets of qualitative research if used to force agreement. As the data analysis phase of qualitative research can be daunting, it is important to build community among the researchers that will survive the lengthy and sometimes disconcerting analysis process. Frequent collaboration through direct, synchronous communication and use of a structured process, such as the ten-step process described here, can provide a roadmap for building consensus in a community of geographically separated researchers.

References

Altheide, D., & Johnson, J. M. C. (1998). Criteria for assessing interpretive validity in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Collecting and*

interpreting qualitative materials (pp. 283-312). Thousand Oaks, CA: Sage Publications.

Boeije, H. (2002). A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Quality and Quantity*, *36*(4), 391-409.

Barbour, R. (2001). Checklists for improving rigor in qualitative research: A case of the tail wagging the dog? *British Medical Journal*, *322*(7294), 1115-1117.

CAT. (2010). *Coding analysis toolkit.* Retrieved from http://cat.ucsur.pitt.edu/default.aspx

Caine, R. N., Caine, G., McClintic, C., & Klimek, K. J. (2009). *12 brain/mind learning principles in action: Developing executive functions of the human brain* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Curry, L., Nembhard, I., & Bradley, E. (2009). Qualitative and mixed methods provide unique contribution to outcomes research. *Circulation, 119*, 1442-1452.

Foster, A., Urquhart, C., & Turner, J. (2008). Validating coding for a theoretical model of information behaviour. *Information Research*, *13*(4). Retrieved from http://InformationR.net/ir/13-4/paper358.html

Gibbert, M., Ruigrok, W., & Wicki, B. (2008). What passes as a rigorous case study? *Strategic Management Journal, 29,* 1465-1474.

Glaser, B. G. (1992). *Emergence vs. forcing. Basics of grounded theory analysis.* Mill Valley, CA: Sociology Press.

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

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report, 8*(4), 597-607. Retrieved from http://nsuworks.nova.edu/tqr/vol8/iss4/6/

Harris, J. K., & Burke, R. C. (2011). Do you see what I see? An application of inter-coder reliability in qualitative research. *Center for Tobacco Policy at Saint Louis University School of Public Health.*

Holton, J. (2010). The coding process and its challenges. *The Grounded Theory Review*, *9*, 21-40.

Johnson, R. B. (1997). Examining the validity structure of qualitative research. *Education*, *118*(2), 282-292.

Jonsen, K., & Jehn, K. A. (2009). Using triangulation to validate themes in qualitative studies. *Qualitative Research in Organizations and Management: An International Journal*, *4*(2), 123-150.

Kirk, J., & Miller, M. (1986). *Reliability and validity in qualitative research*. Newbury, CA: Sage.

Lambotte, F., & Meunier, D. (2013). From bricolage to thickness: Making the most of the messiness of research narratives. *Qualitative Research in Organizations and Management*, *8*, 85-100.

Lawrence, J., & Tar, U. (2013). The use of grounded theory technique as a practical tool for qualitative data collection and analysis. *The Electronic Journal of Business Research Methods*, *11*, 29-40.

LeCompte, M., & Preissle, J. (1993). *Ethnography and qualitative design in educational research*. London, UK: Academic Press.

Leininger, M. (1994). Evaluation criteria and critique of qualitative research studies. In J. M. Morse (Ed.), *Critical issues in qualitative research methods* (pp. 95-115). London, UK: Sage.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Luft, J., & Ingham, H. (1961). The Johari window: A graphic model of awareness in interpersonal relations. *Human Relations Training News*, *5*(9), 6-7.

Madill, A., Jordan, A., & Shirley, C. (2000). Objectivity and reliability in qualitative analysis: Realist, contextualist, and radical constructionist epistemologies. *British Journal of Psychology*, *91*, 1-20.

Marques, J., & McCall, C. (2005). The application of interrater reliability as a solidification instrument in a phenomenological study. *The Qualitative Report*, *10,* 439-462. Retrieved from http://nsuworks.nova.edu/tqr/vol10/iss3/3/

Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review*, *62*, 279-300.

McNiff, J., & Whitehead, J. (2006). *All you need to know about action research*. London, UK: Sage Publications.

Miles, M., & Huberman, A. (1994). *Qualitative data analysis*. Thousand Oaks, CA: Sage.

Miles, M., Huberman, A., & Saldana, J. (2014). *Qualitative data analysis* (3rd ed.). Thousand Oaks, CA: Sage.

Neuendorf, K. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Olson, J., Walters, K., Appunn, F., Grinnell, L., & McAllister, C. (2012). The value of webcams for virtual teams. *International Journal of Management & Information Systems, 16*(2). Retrieved from http://journals.cluteonline.com/index.php/IJMIS

Patton, M. (1990). *Qualitative evaluation and research methods*. Thousand Oaks, CA: Sage.

Pritchard, K., & Whiting, R. (2012) Autopilot? A reflexive review of the piloting process in qualitative e-research. *Qualitative Research in Organizations and Management*, *7*, 338-353.

Radcliff, L. (2013). Qualitative diaries: Uncovering the complexities of work-life decision-making. *Qualitative Research in Organizations and Management*, *8*, 163-180.

Schutt, R. (2012). Investigating the social world. Thousand Oaks, CA: Sage.

Scott, H. (2009). Data analysis: Getting conceptual. *Grounded Theory Review*, *8*, 89-108.

Sinkovics, R., & Alfoldi, E. (2012). Progressing focusing and trustworthiness in qualitative research: The enabling role of computer-assisted qualitative data analysis software. *Management International Review, 52*(6), 817-854.

Strauss, A. L. (1987). *Qualitative analysis for social scientists.* Cambridge, UK: Cambridge University Press.

Strauss, A. L., & Corbin, J. (1998). *Basics of qualitative research*. Thousand Oaks, CA: Sage.

Thomas, D. (2003). A general inductive approach for qualitative data analysis. Auckland, New Zealand: University of Auckland.

Author Note

Joel D. Olson, PhD., has extensive experience in nonprofit leadership and consultation, education, and instructional design. He has published papers on virtual team performance, webcams, trust and technology acceptance. His current research continues in the area of technology and virtual teams, as well as online leadership education. Most recently, he has served the Reformed Church in America and the Evangelical Presbyterian Church as a consultant for churches in crisis. Currently he serves as the Leadership and Management Academic Department Chair for the School of Business and Management in Kaplan University. Correspondence regarding this article can be addressed directly to: Joel D. Olson at, jolson@kaplan.edu.

Chad McAllister, PhD., is lead faculty at Walden University in the DBA program. His research interests involve issues in new product development and innovation. He serves as VP of Education for the Rocky Mountain Product Development and Management Association and advices the Association of International Product Marketers and Managers on innovation topics

and professional certification. Correspondence regarding this article can also be addressed directly to: Chad McAllister at, chad.mcallister@waldenu.edu.

Lynn Grinnell, PhD., has published on sustainability management, educational measurement, and teams. Her research examines the ethical influence of individuals and teams and measurement of learning.

Kimberly Gehrke Walters, PhD., is a Professor of Human Resources at Kaplan University. She has extensive experience in global organizational development, assessment and measurement, and human-computer interaction. Her research examines trust and effectiveness in virtual teams, and the influence of technology in teams.

Frank Appunn, PhD., has published papers on technology, information security, and teams. His research considers the confluence of technology, people, and organizations, while information assurance forms another interest area. He teaches at multiple institutions, heads security, IT and project management content areas. He is a faculty evaluator and dissertation chair mentor.

Article Citation

Olson, J. D., McAllister, C., Grinnell, L. D., Walters, K. G., & Appunn, F. (2016). Applying constant comparative method with multiple investigators and inter-coder reliability. *The Qualitative Report, 21*(1), 26-42. Retrieved from http://nsuworks.nova.edu/tqr/vol21/iss1/3