The Foundations and Fundamentals of Quantitative Ethnography

Golnaz Arastoopour Irgens[1[000-0002-5308-3925]] and Brendan Eagan[2]

[1]Clemson University
garasto@clemson.edu

[1]University of Wisconsin-Madison
beagan@wisc.edu

Abstract. As the Quantitative Ethnography (QE) community becomes more inter-disciplinary, it will need multiple theoretical accounts to fit with the multiple epistemologies of researchers. Thus, in this paper, we provide one theoretical account. We argue that ethnography is foundational to QE, quantification augments ethnographic accounts. and that *critical reflexivity* is necessary in QE. Then, we outline ten iterative steps of QE analyses, explained through two examples, and articulate five main practices. Our goals for this paper are to 1) distill fundamental aspects of QE for new adopters, 2) offer a summarized account for established QE practitioners, 3) clarify underlying values and practices that drive the methodology, and 4) highlight which practices are essential to QE and which are flexible. This paper provides one accessible summarization of QE for an inter-disciplinary field.

Keywords: Modeling, Philosophy, Processes, Critical Reflexivity

1 Introduction

Quantitative Ethnography (QE) is a methodology that integrates qualitative and quantitative analysis methods. With roots in educational research, QE has been used by researchers in a variety of fields. The QE methodology became formalized with the release of Shaffer's *Quantitative Ethnography* book in 2017. However, as the community grows and become more interdisciplinary, it will need multiple accounts of the methodology from different perspectives. The purpose of this article is to provide an accessible, distilled description of QE for the broader developing community with new emphases on critical reflexivity practices. Blending our own ideas with Shaffer's, we begin by describing how ethnography is foundational to QE, meaning that ethnographic techniques provide the essential grounding. We continue by describing how quantification is fundamental, meaning that mathematical techniques are necessary to augment the power of ethnographic methods. Then, we outline ten steps in the iterative QE modeling process and provide two examples of studies that used these processes. Finally, we end by articulating five main practices for QE.

2 Ethnography is Foundational

Although QE has two facets, ethnography is at the heart. The science of ethnography is about making sense of how and why people do the things that they do. More specifically, an ethnographer interprets a culture by going through formal and systematic procedures. For example, in Hutchins' [1] landmark study of officers in the U.S. navy, he observed sailors tracking the position of their ship. The sailors needed to work together to succeed in the complex task of safely navigating their ship through sea and to land. Hutchins argued that the key system to facilitate communication and correct each other's mistakes was through a radio system in which sailors could hear all communications, even if they were not directly involved. This form of widely overheard conversation and communication is what Hutchins referred to as a wide horizon of observation and led to useful error detection and correction that kept the sailors efficient and safe in their work.

The purpose of ethnographic studies, such as the one conducted by Hutchins, is to make claims about one specific culture or community of people and weave these claims into an interpretation of the culture for a broader audience. Such claims are not necessarily meant to be generalizable beyond the community that was studied. Rather, the stories and conceptualizations that emerge from such ethnographic studies can be used as effective tools in other contexts and inspire related scientific explorations. For example, although Hutchins' study made claims about one particular group of sailors at one point in time, his work advanced the concept of how what an individual sees or hears influences cognition and teamwork. Years later, this general idea of horizon of observation inspired Blandford and Furniss's (2005) development of a methodological framework to analyze collaboration in small teams. The researchers tested their framework with telephone dispatchers at emergency ambulance centers in London and were able to draw similar conclusions to Hutchins.

To draw conclusions in ethnography, researchers collect data that are rich enough to yield a "thick description" [2] of a particular culture. Such data are collected through observational recordings, participant-observational recordings, interviews, and other artifacts of cultural meaning. These data contain evidence of the culturally specific ways of how people talk, listen, interact, and use tools in the community that is being studied. These forms of big-D Discourse [3] are the "overt manifestations of culture: what it actually looks like when someone is expressing meaning within some community" [4]. One of the goals of ethnographers and quantitative ethnographers is to shift from observable actions to interpreted meanings to provide an evidence-based account of why and how people from a particular community do what they do. To cross this bridge, ethnographers categorize data by tagging it with big-C Codes, which are categories of meaning derived by the ethnographer to make sense of collected data. This process is often, unsurprisingly, referred to as coding.

Once the data are associated with the appropriate Codes, the ethnographer analyzes how the Codes relate to one another in order to communicate a thick description of a culture. The thick description is created by linking multiple emic observations and interpretations (perspectives of the people in the culture) to etic observations and interpretations (perspectives of the researcher). As Shaffer [4] notes, Hutchins' study is a strong example of developing the etic description of horizon of observation from the emic contexts of U.S. sailors but can be applied across multiple emic contexts, such as with dispatchers at emergency ambulance centers in London.

When ethnographers systematically develop etic interpretations of emic observations, they construct knowledge through their own lens. Thus, the researcher themself plays a key role in the methodological process and "is the primary 'instrument' of data collection and analysis" [5] and knowledge creation. This knowledge is inherently situated within the researcher's intersecting identities of race, gender, class, and their ethical, theoretical, and epistemic commitments. To maintain scientific rigor, the researcher's "view from somewhere" [6] must be addressed through *critical reflexivity* practices [7]. Critical reflexivity involves interrogating the ethical decisions ethnographers make during the research process and making the encountered tensions visible to the scientific community [8]. This form of reflexivity becomes especially important when working with vulnerable or marginalized populations who are at risk of being exploited or harmed when participating in research [9].

3 Quantification is Fundamental

For quantitative ethnographers, critical reflexivity extends beyond human interactions to computational tools. In addition to the researcher being a primary instrument of data analysis in quantitative ethnography, the computational tools play a significant role. All tools were developed by people with particular worldviews that influence decision-making during development. Those people's assumptions, opinions, and biases are embedded into the tool [10]. Thus, computational tools are not neutral or objective artifacts for meaning-making. When researchers exercise critical reflexivity, they commit to uncovering the tools' limitations and ethical obligations in order to better understand the ways in which power and privilege amplify a particular point of view and obscure marginalized voices [11]. To exercise critical reflexivity, quantitative ethnographers should ask questions such as: Who designed this tool and for what purposes? What are the assumptions made in this tool about discourses and cultures? Which social, cultural, and political values are amplified and which are minimized? How does the tool (re)enforce inequities and oppression when creating samples, making statistical calculations, and visualizing data?

Through a critical reflexivity lens, the computational tools in QE are quite powerful. The unique combination of human researcher and digital tools in QE allows for an exploration of data and reveal of stories that likely would not have been uncovered without this analytical process [12]. For large datasets that emerge from modern digital spaces, such as social media, computational QE tools allow for researchers to access a "thick description" of such growing and changing digital cultures [4]. Even for comparatively smaller datasets, such as interview transcriptions, computational QE tools augment traditional qualitative analyses by exposing the researcher to new visualizations and quantifications that reveal and inspire stories grounded in the data [13].

Moreover, statistics provide forms of validity for qualitative analyses beyond accounts of "trustworthiness" or "authenticity" criterion [14]. Statistical tools in QE are used to generalize within the sample of data collected and not necessarily to generalize beyond the sample of data to a larger population. As mentioned before, although the broader ideas uncovered in ethnography may be useful in other contexts, limited generalizability is a fundamental aspect of ethnography. Thus, sampling and statistical tools in QE are not used in the same way that they are used in traditional quantitative studies. The goal is not to generalize findings to a broader population but rather, to provide some level of confidence that the interpretations and stories told about one particular culture are persistent throughout the group of people that were studied. In other words, statistical analyses in QE provide additional warrants or evidence for reaching a point in which analyzing additional data does not provide new insights. This point is called reaching *theoretical saturation* in qualitative analysis and is an important amalgamation of qualitative and quantitative in QE. As Shaffer [4] aptly puts it, "By reframing the role of sampling and statistical significance… the distinct logic of quantitative inquiry and the distinct logic become compatible. We find a point of contact between these two very different epistemological stances toward research."

4 Ten Steps in the Iterative QE Modeling Pathway

The interpretation techniques that are typically associated with qualitative analyses occur throughout the QE process. Likewise, the quantitative mindset is consistently active, as ethnographic data is organized, sorted, and coded in preparation for quantitative models. Here, we offer ten steps that incorporate both mindsets for developing quantitative models from ethnographic data. The ten steps are split into five categories: collection, segmentation, codification, accumulation, and measurement (Figure 1).

4.1 Collection

Because ethnography is the foundation of QE, data collection (Step 1) aligns with traditional qualitative procedures. Observations are central to ethnographic data collection and include researcher notes on how people talk and act in a particular setting [15]. Researcher-created data may be in the form of field notes in paper or digital forms containing observations and interpretations, photographs or videos and corresponding transcriptions, and researcher reflective diary entries. Data may also be in the form of participant-created documents or artifacts that have particular meaning to people within a culture. These include letters, multi-media art, emails, digital log data, articles, discussion boards, and social media postings. In some studies, data may be created by researchers and participants collaboratively.

4.2 Segmentation

After data are collected in a form that is likely to enable thick descriptions, the researcher prepares this data for statistical analyses. First, the researcher identifies and segments the data into *lines* (Step 2), which are defined as "the smallest unit of continuous action that is of interest in the data" [4]. In discourse data, lines can be defined in various ways including turns of talk, responses to a question, or moves made in a digital game. Lines are then grouped into *stanzas* (Step 3), which are "a set of lines that are within the same relevant context, and therefore related to one another" [4]. Stanzas are often compared to chapters in a book, verses in a song, or, quite literally, stanzas in a poem. Stanzas can be classified in various ways including as activities, interview sessions, or conversational topics on a discussion board. If a moving stanza window model is used, additional segmentation is completed to designate the start and end boundaries of the *conversation* as the moving stanza window slides through the discourse [16]. Researchers justify their segmentation choices theoretically by operationalizing existing conceptualization of theoretical discourse structures, if available. Alternatively, researchers justify their segmentation choices empirically by experimenting and choosing segmentation that is most aligned with a story grounded in the data. For example, Zörgő and colleagues [17] developed an approach for testing various segmentation choices in one dataset and discovered that segmentation choices change the interpretations of the findings in their particular dataset.

4.3 Codification

After segmentation, each line is codified numerically. Because of the ethnographic ideologies in QE, the coding process in QE is often grounded and derived from the data itself. In actual practice, steps three and four are interchangeable and iterative. Researchers may segment data into lines, code the data, and then segment the coded data into stanzas. To progress through the model-making process, codes must be represented as numbers. Many QE studies have used binary coding: displaying a 1 if the code appears in the line and 0 if the code does not appear in the line. Other studies have explored weighted code values. For example, Frey and colleagues [18] analyzed adolescents' emotions during peer victimization and used weighted coding in their analysis to model the strength of the emotions at the coding level. As of now, most QE researchers rely on "hand-coding" in which a human identified the codes in each line of data [19]. However, researchers also develop automated classifiers using regular expressions, topic modeling, and nCoder. Other tools, such as the Rho R Package, use statistics to validate that the automated classifiers are coding the data consistently in ways that align with human interpretations. The Reproducible Open Coding Kit (ROCK) also available on R, provides QE researchers with tools to segment and code data and transfer their codes to a model [20].

4.4 Accumulation

After codes are represented quantitatively, the data are accumulated in steps five through seven. A value is computed for each stanza based on the codes, represented as S1 in the figure. For example, when using ENA, S1 is represented as a vector that captures the number of co-occurrences among codes. Then, stanza values are accumulated for each unit of analysis. Units of analysis can be defined in many ways including each person in the community, teams of people, or documents. In practice, steps five and six may happen concurrently, in which a stanza value is determined for each unit of analysis. As of now, ENA is the most widely used tool for accumulating data [19]. In ENA, the observation of the unit is visualized in two forms: 1) as a weighted network representation of the accumulated, and often scaled, co-occurrences of codes, and 2) as a point in space that roughly corresponds to the centroid of the weighted network. Other tools that have been developed to accumulate and visualize coded data include Social Epistemic Network Signature (SENS) [21], Socio-Semantic Network Analysis (SSNA) [22], and using R to merge sentiment analysis with domain-specific discourse [23]. Some QE analyses have stopped at step seven of computing and visualizing observations. In these studies, the main purpose is to provide a description of the data and tell a visual story but not to make claims about statistical differences within dataset.

4.5 Measurement

If researchers are interested in measuring differences between or within groups in the data, then they can continue to steps 8, 9, and 10 in which statistical analysis are conducted. A parameter is computed for each sample and a statistic is computed to determine whether there is a difference between samples that is statistically significant. Examples of parameters could be calculating a mean or a median. Examples of statistics could be a t statistic from a t-test or a u statistic from a non-parametric Mann-Whitney test. Although two-sample inferential statistics are commonly used, other forms of statistical tests could be conducted that align with study's goals. For example, one study used k-means clustering as an exploratory way to group and measure samples [24]. In a more recent study, one-sample inferential statistics were used to determine whether observed ENA models were statistically different from models created by chance and thus, making statistical claims for theoretical saturation within the dataset [25].

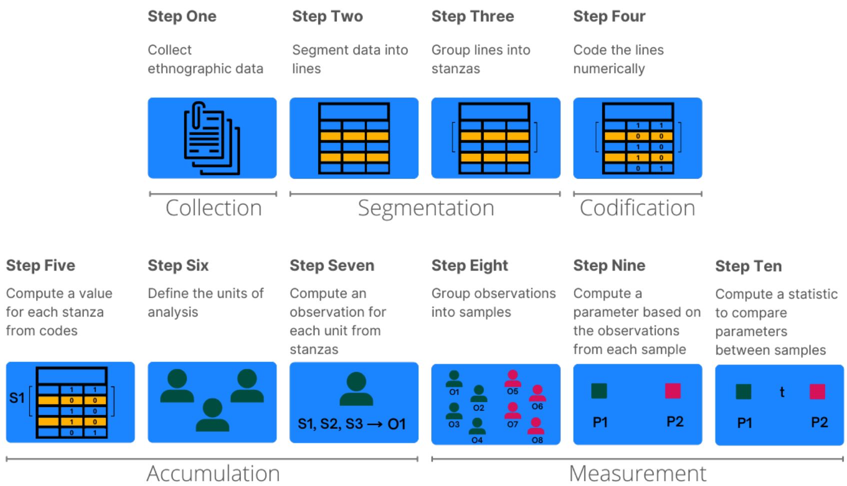

Fig. 1. Ten iterative steps for creating a QE model modified from Shaffer [10].

5 Examples of QE Processes

5.1 Augmenting Descriptive Analyses

In this first example, we describe a QE analysis by Vega and colleagues [26] in which they explored the identity process of four Costa Rican pre-service teachers in training to become English as a foreign language teachers. Given that pre-service education is a critical period for teacher identity formation, the purpose of the study was to illuminate the ways in which dominant discourses can contribute to tensions in identity development. One researcher conducted semi-structured, open ended interviews via video calls. Interviews, which ranged from 45-75 minutes, were recorded and transcribed. Researchers were interested in characterizing tensions that pre-service teachers felt in their training programs when negotiating and managing their identity development and wanted to provide a powerful descriptive analysis. Thus, the research team employed steps one through six of the iterative QE modeling pathway and did not engage with the measurement aspects.

The research team segmented the data into 143 lines by turns of talk (Step 2) and grouped the lines into stanzas that represented a sliding window of two turns of talk (Step 3) bounded by each interview session. To code the dataset, the team engaged in three iterative rounds of coding. In the first round, they created deductive codes derived from a theoretical framework, and then followed an inductive process to extend and add codes. In the third round, for reduction of the codes and reaching theoretical saturation, categories generated in the previous round were collapsed. Using the refined coding scheme, two of the researchers coded the data separately and met to discuss the inconsistencies until they reach mutual agreement. The coding was coded numerically through binary coding (Step 4).

To accumulate and visualize the dataset, Vega and colleagues relied on ENA. They chose pre-service teachers as the unit of analysis (Step 5). Then, they computed a value for each stanza and accumulated these stanzas for each unit to create an observation (Steps 6 and 7). The observations were represented as weighted networks for each teacher. Analysis of the networks and discourse revealed that all four participants positioned themselves as inferior non-native English speakers and negatively compared themselves to native English speakers from the U.S. or U.K. In the paper, the researchers tell the stories of the four pre-service teachers who made strong connections between the Native speaker as a standard and Tensions, indicating concerns around their linguistic practices as language learners.

At the conclusion of the paper, the researchers argue for the affordances of ENA and QE that strengthen qualitative analyses of language teacher identity. First, they were able to test the deductive and inductive coding categories and refine based on their persistence and connectivity on the visualizations. They claim that “the iterative process of interaction between the qualitative data and ENA network visualizations provided grounded evidence of salient codes in our data set and clarity for the story that the data were telling.” Second, ENA allowed for descriptive storytelling at multiple levels. There were aspects of the teachers' developing identities that were prevalent between two teachers but even within a particular pair of teachers there were nuanced experiences that the networks could explain and visualize that were not visible in previous qualitative studies of this data. These distinctions between and within groups revealed undiscovered tensions at personal, group, and aggregated levels that participants experienced. The authors conclude that the use of ENA on previously analyzed qualitative data revealed new descriptive interactions in the data that were not seen before, thus augmenting the original qualitative analyses.

5.2 Employing Statistical Models to Make Claims

In this second example, we describe a QE analysis by Sweicki and colleagues [27] in which they examined a dataset of transcripts from 16 teams comprised of 94 naval officers during air defense warfare team training. The dataset was segmented into 12,027 lines of talk and split into an experimental condition, teams with access to a decision-support system, and a control condition, teams that did not have access to this system. The researchers were interested in determining if there was a statistical difference between the commanding officers in the experimental and control group in terms of collaborative problem-solving approaches. To make this comparison they segmented the transcripts of the Navy air defense warfare team training scenarios in lines determined by turns of talk and grouped the lines of data by team and condition (Steps 2 & 3). Then, the research team qualitatively analyzed the data starting with a grounded analysis and triangulation with related previous related studies to develop a coding scheme. They then applied this coding scheme using nCoder to develop and validate automated qualitative coding of the data (Step 4).

After segmentation and codification, Swiecki and colleagues used ENA to achieve accumulation and measurement. Through the ENA webtool, they identified units of analysis as the commanders in each condition (Step 6) and used a sliding stanza model with a window size of five to capture the recent temporal context of the conversation (Step 3). The ENA webtool then created a value for each stanza (Step 5) for each commander in each condition and accumulated these values to create one observation for each officer (Step 7). The observations were represented as weighted networks in which one network was created for each officer that revealed the connections they made between codes. To address their original research question, Swiecki and colleagues compared the patterns of connections made by commanders in the experimental and control conditions. They found that the discourse patterns of commanders in the control condition spent more time making connections related to Seeking Information while the commanders in the experimental condition we able to make more connections that Contributed Information and Linking Information about the tactical situation and tactical actions. Taking the analysis further, the research team used an alternative observational representation in ENA which represents each observation as a numeric value that roughly corresponds to the center of mass of the weighted network. Because each observation is a number, a distribution can be created for each group of interest and in turn, inferential statistics can be employed. In this study, the research team conducted a two-sample Mann-Whitney U test between distributions of the projected points in

ENA space for commanders in the two conditions (Steps 8, 9, and 10). The results revealed that the discourse patterns in the control group ($Mdn = -0.21$, $N = 13$) were significantly different from the experimental group ($Mdn = 0.25$, $N = 16$; $U = 206$, $p < 0.01$, $d = 2.98$, $power = 1.00$). This quantitative result supported the claim that commanders in the control condition made stronger connections to Seeking Information, while commanders in the experimental condition made stronger connections to codes related to tactical decision-making.

Finally, prior to reporting their quantitative results, the research team identified qualitative examples of all of the aspects of their quantitative results and provided interpretive explanations of the qualitative data. Moreover, after providing their qualitative results they used both the qualitative examples and linked quantitative models in ENA to frame further examples of individual commander performance to reinforce the linkages between their qualitative understanding and the computational model they were using.

6 Five Main Practices in QE

In addition to a set of ten iterative steps, we offer five main practices in QE. These practices differ from the iterative steps in that they are ways of thinking and doing that the researcher engages in throughout QE.

6.1 Practice the 3 C's of Data Hygiene: Clean, Complete, and Consistent

Because QE researchers integrate qualitative and quantitative ways of thinking, they must use data organization approaches that align with an integrated mindset. One way to frame such integrative practices is by referring to having proper "data hygiene." In everyday terms, hygiene is defined as the set of cleanliness practices conducive to maintaining health and preventing disease. Extending this metaphor to QE, having proper data hygiene means making qualitative and quantitative techniques compatible such that ethnographic data can be analyzed statistically without compromising validity and thick descriptions [4]. In actual practice, QE researchers practice proper data hygiene by organizing ethnographic data into a single qualitative data table that is clean, complete, and consistent. As stated above, the data table is comprised of lines, which are the smallest unit of action that is of interest in the data, and stanzas, which are groups of lines that are topically related. These lines and stanzas must be represented in the data table such that the data table is machine-readable but not necessarily human-readable. Bold borders, shading, and merged columns may help readers understand information presented in a data table [28]. However, these aesthetics are not usable for a computational program.

Researchers can ensure their data table is machine-readable by following the three C's of proper data hygiene. First, the same notation should be used throughout the table to indicate the same information. For example, if the location of the study is referred to as "Los Angeles" in some parts of the data table and "L.A." in other parts of the table, the quantitative analysis will assume these are two different locations although a human reader will assume they are the same location. The data table should be *clean* in terms of notation. Second, each row in the table should have all the information related to that unit of analysis. For example, if a row represents one line of discourse from a particular participant, then that row should also have all the data collected that is related to that participant, such as age, location, or dietary preferences. Again, completeness is not aesthetically pleasing to a human reader because the same information will be repeated in every row, but it is necessary for computation. Thus, each row in the data table should be *complete* and include all the data collected for that particular participant or unit of analysis. Last, not only should every row be complete but it should also only contain information related to that particular unit and line of discourse. The same idea applies to the columns in the data table–every column must contain one type of information. For example, if a column is labeled as "Location," then it should only have data about location. The "Location" column could have different data entries such as, "Los Angeles," "Lagos," or "Buenos Aires," but there is no other information except for location. Thus, each column and row in the data table should be *consistent* in terms of the information provided.

6.2 Get a Grip

After data collection and organization, ethnographers develop stories about a particular culture. The analysis process can be thought of as putting together pieces of a socio-cultural puzzle [29] but one with multiple possible solutions. Just like most research processes, quantitative ethnography, is a messy, ambiguous process. To navigate through making sense of the collected data, QE researchers must "get a grip" on the story that they would like to tell from the data. Shaffer [4] describes this process as getting a mechanical grip on a Discourse. He argues that the term emphasizes how the mechanical tools of research are used to "grab hold" of the complex phenomena in the world that we are trying to explain. In practice, QE researchers are consistently getting and refining a grip on the Discourse throughout the analytical process. The first move towards getting a grip is by familiarizing oneself with the data and developing initial Codes. As indicated by the example from Vega and colleagues in section 5.1, researchers may refine their Codes during the coding process but also after modeling their data and interpreting their models. In QE, this iterative process of Codes informing models and models informing Codes continues until the grip has tightened sufficiently and a fair thick description has been developed. Getting a grip is a metaphor for how QE researchers consistently use the mechanisms of ethnography and statistics to reveal the underlying Discourse in a culture.

6.3 Have a Conversation with the Tools

One thing that makes QE a unique methodology is that researchers get a grip on their data by using ethnographic and statistical tools together. Even in the initial stages of data organization, a seasoned QE researcher will rely on digital and computational tools to ensure that data tables are clean, complete, and consistent. In later stages of QE analysis, the digital tools play a more significant role. As stated in section 4.2, a researcher may rely on natural language processing tools to help identify phrases or lines that should be coded in the data. Such tools are skilled at finding additional Codes (or lines that should have been coded) that a researcher may have overlooked. However, such tools do not understand the discourse data in the way that human researchers understand and interpret discourse—at least, not yet anyway. Each party—computational tool and human researcher—have certain strengths and weaknesses. And what makes QE such a powerful approach is the way it leverages the strengths of the automation power of computational tools with the interpretative power of human researchers. But to ensure rigor and quality, the researcher must check the outputs of the tool to see if the results make sense, are fair, and grounded in the data. What often happens is that the researcher and the tools engage in iterative cycles of feedback in which the research inputs information into the tools, the tools process the information, the tools output new information, the researcher interprets this information and decides their next move. This is called *having a conversation* with the tool. Similar to a conversation between two people, the researcher may decide to "ask" the tool a clarifying question by exploring the output further or move the conversation along by "asking" a new question and running a new analysis. The computational tool essentially becomes a member of a well-functioning research team that has shared responsibilities and goals, utilizes the talents of its team members, and promotes the exchange of feedback.

6.4 Close the Interpretative Loop

While conversations with the tools happen throughout QE, there is one particular class of conversations that occurs during modeling that is of critical importance. In this conversation, the researcher inputs coded data into a modeling tool, the tool produces a model, the researcher interprets the model, and then goes back to the original data to see if the interpretation is supported by the data. If the interpretation does not align with the collected data, then the researcher must reevaluate the Codes and the assumptions made in the model. This unique conversation is called *closing the interpretative loop*. It is a central mindset in QE because it is the fundamental pathway of validating a model and provides the qualitative evidence that created the quantitative result. Although a central aspect of QE, closing the interpretative loop is susceptible to being forgotten by researchers. One reason this process is ignored is because the conversations between tools and researchers can be lengthy and intense. After many long hours and dead ends, a researcher may be excited to find a model that has a significant result and seems to make sense at face value. However, by not grounding the interpretation of the model in the data, the researcher violates a central tenet in the science of ethnography. Thus, closing the interpretative loop is an important way of establishing validity in QE and should be a consistent mindset.

6.5 Embrace Multiple Forms of Validity

In quantitative research, validity is defined as the accuracy of measurement and the extent to which the tools are measuring what is intended to be measured. However, in QE, validity is a much broader concept that takes up multiple forms. Closing the interpretive loop is the fundamental validity philosophy that supports QE, but there are other validity checks that occur throughout. As Winter [30] argues, "validity is not a single, fixed or universal concept, but rather a contingent construct, inescapably grounded in the processes and intentions of particular research methodologies and projects." For example, when coding data, QE researchers establish construct validity by building a codebook with definitions of Codes and examples of discourse that are categorized by a Code [31]. The definitions provide defined constructs that can be traced back to theory and examples of discourse provide evidence grounded in the data. A researcher may also establish validity by doing member-checking or other forms of participatory QE research in which researchers discuss the Codes with the participants from the study [32]. Researchers may strengthen emic-etic connections for building thick descriptions and co-construct meaning with participants. However, member-checking may also result in tensions between researcher and participants and highlight unequal power dynamics [33]. Thus, when engaging in participatory QE research, researchers should engage in their critical reflexivity practices by interrogating power dynamics, reflecting on how participants will be harmed or put at risk by member-checking, and the roles of the "researcher" and the "researched" [34].

In some studies, one researcher will code all the data alone. In other studies, two or more researchers will code the data and inter-rater reliability metrics will be used to determine if the interpretations are relatively the same across two or more people and therefore, offer some confidence in the conceptual validity of the interpretations. In some cases, researchers will engage in social moderation and code all the data and discuss until mutual consensus is reached. When the dataset is too large, two researchers select a sample of the data, code the sample individually, calculate inter-rater reliability using a statistic, determine if the statistic has met a pre-determined cut-off or threshold of performance, and if it has, then split the dataset and code the remaining data individually. These steps confirm reliability for the sample that was coded but it is unclear if the sample is representative of the remaining dataset and thus, it is unclear whether the remaining dataset will have the same level of validity. Eagan and colleagues [35] recommend using rho as a statistical technique to take representative samples of the dataset and to control for Type I error (false positives) when coding data with two or more raters. In addition to two human raters, QE researchers may also train and use an automated classifier to code large datasets. In these cases, inter-rater reliability is measured between two human raters to determine conceptual validity. It is also measured between each human rater and the automated classifier to determine computational validity and whether the automated classifier is capable of consistently coding the data in ways that align with human interpretations. Whether during coding or modeling, validity checks such as exercising critical reflexivity, closing the interpretive loop, and inter-rater reliability facilitate the rigor that establishes QE as a science.

7 Conclusion

In this paper, we argue that ethnography is foundational, and quantification is fundamental to the science of QE. We provided ten iterative steps for creating QE models and two examples of how these steps are visible. In parallel to the steps, we provided five main practices that we have observed and experienced as seasoned QE researchers. We drew from Shaffer's visionary book [4] but have reconceptualized and summarized key ideas. The goals of this article were to 1) provide a distilled version of the fundamental tenets of QE to provide an access point for new scholars, 2) provide a summarized reference guide for those who are established users of the methodology, 3) bring clarity to the potentially hidden values and practices that drive the methodology, and 4) bring clarity to processes that are essential in QE and processes that are flexible and context-dependent. Overall, this work provides one form of an accessible description of QE for the broader inter-disciplinary community.

References

1. Hutchins, E.: Cognition in the Wild. MIT press (1995)
2. Geertz, C.: The interpretation of cultures. Basic books (1973)
3. Gee, J.P.: An introduction to discourse analysis: theory and method. Routledge, Milton Park, Abingdon ; New York (2011)
4. Shaffer, D.W.: Quantitative Ethnography. Cathcart Press, Madison, WI (2017)
5. Watt, D.: On Becoming a Qualitative Researcher: The Value of Reflexivity. The Qualitative Report. 12, 82–101 (2007)
6. Haraway, D.: Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. Feminist Studies. 14, 575–599 (1988). https://doi.org/10.2307/3178066
7. Corple, D.J., Linabary, J.R.: From data points to people: feminist situated ethics in online big data research. International Journal of Social Research Methodology. 23, 155–168 (2020). https://doi.org/10.1080/13645579.2019.1649832
8. Arastoopour Irgens, G.: Using Knowledgeable Agents of the Digital and data feminism to uncover social identities in the #blackgirlmagic Twitter community. Learning, Media and Technology. 47, 79–94 (2022). https://doi.org/10.1080/17439884.2021.2018608
9. Pacheco-Vega, R., Parizeau, K.: Doubly Engaged Ethnography: Opportunities and Challenges When Working With Vulnerable Communities. International Journal of Qualitative Methods. 17, 160940691879065 (2018). https://doi.org/10.1177/1609406918790653
10. Pea, R.: Practices of distributed intelligence and designs for education. In: Salomaon, G. (ed.) Distributed cognitions: Psychological and educational considerations. Cambridge University Press, New York (1993)
11. D'Ignazio, C., Klein, L.F.: Data Feminism. The MIT Press, Cambridge, Massachusetts (2020)
12. Arastoopour Irgens, G.: Quantitative Ethnography Across Domains: Where we are and where we are going. , Madison, WI (2019)
13. Hod, Y., Katz, S., Eagan, B.: Refining qualitative ethnographies using Epistemic Network Analysis: A study of socioemotional learning dimensions in a Humanistic Knowledge Building Community. Computers & Education. 156, 103943 (2020). https://doi.org/10.1016/j.compedu.2020.103943

14. Guba, E.G., Lincoln, Y.S.: Fourth generation evaluation. Sage (1989)
15. Glesne, C.: Becoming Qualitative Researchers: An Introduction. Pearson, Boston (2014)
16. Siebert-Evenstone, A.L., Arastoopour Irgens, G., Collier, W., Swiecki, Z., Ruis, A.R., Shaffer, D.W.: In search of conversational grain size: Modeling semantic structure using moving stanza windows. Journal of Learning Analytics. 4, 123–139 (2017)
17. Zörgő, S., Swiecki, Z., Ruis, A.R.: Exploring the Effects of Segmentation on Semi-structured Interview Data with Epistemic Network Analysis. In: Ruis, A.R. and Lee, S.B. (eds.) Advances in Quantitative Ethnography. pp. 78–90. Springer International Publishing, Cham (2021)
18. Frey, K.S., McDonald, K.L., Onyewuenyi, A.C., Germinaro, K., Eagan, B.R.: "I Felt Like a Hero:" Adolescents' Understanding of Resolution-Promoting and Vengeful Actions on Behalf of Their Peers. J Youth Adolescence. 50, 521–535 (2021). https://doi.org/10.1007/s10964-020-01346-3
19. Kaliisa, R., Misiejuk, K., Irgens, G.A., Misfeldt, M.: Scoping the Emerging Field of Quantitative Ethnography: Opportunities, Challenges and Future Directions. In: Ruis, A.R. and Lee, S.B. (eds.) Advances in Quantitative Ethnography. pp. 3–17. Springer International Publishing, Cham (2021)
20. Peters, G.-J., Zörgő, S.: Introduction to the Reproducible Open Coding Kit (ROCK), https://psyarxiv.com/stcx9/, (2019)
21. Gašević, D., Joksimović, S., Eagan, B.R., Shaffer, D.W.: SENS: Network analytics to combine social and cognitive perspectives of collaborative learning. Computers in Human Behavior. 92, 562–577 (2019). https://doi.org/10.1016/j.chb.2018.07.003
22. Oshima, J., Oshima, R., Fujita, W.: A Mixed-Methods Approach to Analyze Shared Epistemic Agency in Jigsaw Instruction at Multiple Scales of Temporality. Journal of Learning Analytics. 5, 10–24 (2018). https://doi.org/10.18608/jla.2018.51.2
23. Arastoopour Irgens, G.: Facilitating a Sense of Belonging for Women of Color in Engineering. In: Roscoe, R.D., Chiou, E.K., and Wooldrige, A.R. (eds.) Advancing Diversity, Inclusion, and Social Justice Through Human Systems Engineering. pp. 221–239. CRC Press, Boca Raton (2019)
24. Arastoopour Irgens, G., Chandra, S., Dabholkar, S., Horn, M., Wilensky, U.: Classifying Emergent Student Learning in a High School Computational Chemistry Unit. In: American Education Research Association (AERA) Conference. Toronto, CA (2019)
25. Swiecki, Z.: The expected value test: A new statistical warrant for theoretical saturation. In: International Conference on Quantitative Ethnography. pp. 49–65. Springer (2021)
26. Vega, H., Arastoopour Irgens, G., Bailey, C.: Negotiating Tensions: A Study of Pre-service English as Foreign Language Teachers' Sense of Identity Within Their Community of Practice. In: Ruis, A.R. and Lee, S.B. (eds.) Advances in Quantitative Ethnography. pp. 277–291. Springer International Publishing, Cham (2021)

27. Swiecki, Z., Ruis, A.R., Farrell, C., Shaffer, D.W.: Assessing individual contributions to Collaborative Problem Solving: A network analysis approach. Computers in Human Behavior. 104, 105876 (2020). https://doi.org/10.1016/j.chb.2019.01.009
28. Wong, D.M.: The Wall Street Journal Guide to Information Graphics: The Dos and Don'ts of Presenting Data, Facts, and Figures. W. W. Norton & Company, New York; London (2013)
29. Burns, A.: Analyzing Ethnographic Data. In: Tyner-Mullings, A., Gatta, M., and Coughlan, R. (eds.) Ethnography Made Easy. CUNY Open Educational Resource (2019)
30. Winter, G.: A Comparative Discussion of the Notion of "Validity" in Qualitative and Quantitative Research. TQR. (2000). https://doi.org/10.46743/2160-3715/2000.2078
31. Shaffer, D.W., Ruis, A.R.: How We Code. In: Ruis, A.R. and Lee, S.B. (eds.) Advances in Quantitative Ethnography. pp. 62–77. Springer International Publishing, Cham (2021)
32. Vega, H., Arastoopour Irgens, G.A.: Constructing Interpretations with Participants Through Epistemic Network Analysis: Towards Participatory Approaches in Quantitative Ethnography. In: Wasson, B. and Zörgő, S. (eds.) Advances in Quantitative Ethnography. pp. 3–16. Springer International Publishing, Cham (2022)
33. Candela, A.: Exploring the Function of Member Checking. TQR. (2019). https://doi.org/10.46743/2160-3715/2019.3726
34. Shum, S.B., Arastoopour Irgens, G., Moots, H., Phillips, M., Shah, M., Vega, H., Wooldridge, A.: Participatory Quantitative Ethnography. In: Advances in Quantitative Ethnography. pp. 126–138. , Malibu, CA
35. Eagan, B., Rogers, B., Serlin, R., Ruis, A.R., Arastoopour Irgens, G., Shaffer, D.W.: Can we rely on irr? testing the assumptions of inter-rater reliability. In: International Conference on Computer Supported Collaborative Learning (2017)