

The Quality in Qualitative Methods¹

Manfred Max Bergman & Anthony P.M. Coxon

Key words:

quality
assessment,
validity, reliability,
secondary data
analysis, archiving

Abstract: Quality concerns play a central role throughout all steps of the research process in qualitative methods, from the inception of a research question and data collection, to the analysis and interpretation of research findings. For instance, the type of instrument or procedure to collect data may be evaluated in relation to quality criteria, and these may be different from those which are used to judge the data obtained from such instruments or procedures. All these may yet again be different from quality criteria that may apply to the qualitative analyses of data. A national resource center for qualitative methods can contribute to the establishment and maintenance of certain quality standards. In this article, we will explore some of these quality criteria and how they can be established and maintained by a national resource center for qualitative methods.

Table of Contents

- [1. Introduction](#)
- [2. Quality in Qualitative Methods](#)
 - [2.1 Conceptualization of a research question](#)
 - [2.2 Data collection](#)
 - [2.3 Data quality](#)
 - [2.4 Internal consistency](#)
 - [2.5 Additional remarks on probes and a caveat](#)
 - [2.6 Data analysis and interpretation](#)
 - [2.7 Quality of analysis and interpretation](#)
 - [2.8 Validity and sorting](#)
 - [2.9 Sorting](#)
 - [2.10 Sorting of what?](#)
- [3. The Contributions of a Resource Center](#)
- [4. Conclusions](#)
- [References](#)
- [Authors](#)
- [Citation](#)

1 Parts of this paper have appeared previously (BERGMAN 2000, 2002; BERGMAN & COXON 2002). This paper was financed in part by a grant from the Swiss National Science Foundation (grant number 3346-61710: Sexual Interaction and the Dynamics of Intimacy in the Context of HIV/AIDS) and by SIDOS. We would like to thank the students and staff of the Essex Summer School in Social Science Data Analysis and Collection for stimulating discussions about qualitative methods.

1. Introduction

In this article, we will explore various quality criteria for qualitative methods and how some of these can be maintained by a national resource center and data archive. To accomplish this, we will first establish a selection of quality criteria for qualitative methods. We will also examine which of these criteria can be more effectively pursued by a national resource center, while others ought to be left to the discretion of researchers. [1]

Two qualifying remarks are necessary here. First, we do not believe that the quality of qualitative research can be encapsulated *a priori* within a set of rigid rules. However, this does not mean that quality criteria do not apply to studies in the so-called qualitative tradition. In other words, certain guidelines may help improve either the quality or the credibility of research results, while their adoption does not automatically guarantee these. Second, good qualitative research design is not dependent on the existence of a national resource center. Instead, such a center should facilitate the appropriate application of data collection, documentation, and analysis. It should serve as a flexible and adaptable information and training center, but not attempt to define the research landscape by setting methodological agendas or suggesting universally valid procedures. As such, its credo should be that there are no bad methods, only inappropriate or misapplied methods. [2]

2. Quality in Qualitative Methods

What if interviews and focus groups are merely local accomplishments, ephemeral and context-bound elaborations of meaning between interactors? And what if interpretations of interview and textual data are incontestable due to intersubjectivity and hermeneutic considerations? What can be the significance of empirical social science research if observations are not generalizable beyond the moment of observation and, thus, not representative of anything other than of that moment of production? If we were to push some of these propositions further, empirical research would be rendered obsolete because any or no data could support any argument. [3]

In contrast, a perspective steeped in the positivistic² and post-positivistic tradition would suggest that the social sciences can uncover objective and universally valid facts by following clear procedures and rules, which include carefully controlled observations of empirical phenomena, impartial and logical argumentation, and objective analysis, i.e. the elimination of interpretation by the researcher. According to the so-called scientific method, facts should be reported objectively, rather than interpreted. [4]

2 We use this term rather loosely here, i.e. in a manner that is representative of current epistemological discussions in the literature on qualitative methods. Some would argue that what we "really" mean with positivism is indeed empiricism. However, this terminology would be confusing as well because even interpretive methods, the set of qualitative methods that proclaims greatest distance to positivism, deal with empirical evidence, i.e. data, and, as such, have to be considered empirical in their nature.

In our opinion, neither of these two starting positions is particularly helpful in qualitative research. An extreme subjectivistic position is untenable because it is not falsifiable in that all data and their interpretations are valid *a priori*, while a (post-) positivistic position is untenable because observation, analysis, and interpretation are bound to be subjective (see also the contributions of [FIELDING](#) and [MOTTIER](#) in this volume). As elegant and parsimonious as these two positions may be in theory, we must abandon them in order to elaborate quality criteria that, although lacking in universalism, may be more useful for qualitative research. [5]

Most fundamentally, any observed "fact" has already been interpreted at least in the sense that meaning has been assigned to an empirical observation. Description, explanation, prediction, and the assessment of causes and consequences of social phenomena cannot be achieved in the absence of evaluation and interpretation. To understand is to interpret. Most disciplines in the social sciences have long recognized the interplay between context, culture and tradition, our senses, and our understanding. Strangely, these factors are generally considered only in terms of the people whom we study. Sometimes we recognize the importance of context in the interpretation of empirical data when we discuss the research process in abstract terms, but we tend to fail to take into consideration contextual elements when we elaborate a research question or present our substantive research results. Yet it is not only those we study who are stuck in a subjectively constructed reality within which meaning is elaborated; we researchers, as we attribute meaning to empirical phenomena, are equally stuck in our own subjectivity (e.g. GERGEN 1973; WEBER 1949/1977). While we make sense of empirical observations by attributing meanings to them, two elements prevent us from grasping empirical observations objectively: first, social science data is produced within a particular political, historical, and socio-cultural context and, second, researchers themselves belong to various cultures and traditions (e.g. academic, political, historical, socio-cultural). These two interconnected elements—the contextualized research phenomenon and the contextualized researcher—contribute in different ways to the impossibility of sensing and describing anything without interpretation. Rather than considering these contexts as extrinsic and undesirable side-products or sources of bias, which must be eliminated in order to get at objective data and their unbiased interpretation, we have to concede that it is precisely these contexts that provide us with their significance in terms of "meaning frames," and which permit us to understand and communicate empirical observations. While there certainly exists a reality independent of our understanding thereof, in order to conceive and understand any phenomenon, we must capture "external" reality through observation, i.e. translate bits of information from an objectively unknowable reality into subjectively interpreted and context-bound, yet knowable, realities. [6]

This position implies a controversy: on the one hand, we are unable to access an objective and universal understanding of empirical phenomena but, on the other, we wish to evaluate the quality of incommensurable truth-claims. What quality guidelines can we adopt for the empirical social and political science research process that acknowledge the subjective and interpretative nature of our

endeavors while, concurrently, convey findings that are both empirically rigorous and credible? [7]

Quality considerations in empirical research tend to be addressed by the concepts "validity" and "reliability," especially in the areas of psychometrics and econometrics. We will critically examine some possibilities of these concepts, while concurrently realizing that we cannot simply transpose these two concepts from one theoretical basis to another.³ [8]

To make our task more manageable, we will separate conceptually the research process into four parts:

- elaboration of a research question;
- data collection;
- data analysis; and
- interpretation. [9]

We do this because quality concerns relate differently to these four areas and because we would like to emphasize that, contrary to many texts on research methods, data collection and data analysis are two interdependent, yet separate, research processes. For example, unobtrusive or participant observation, interviews, and focus groups are usually thought of as qualitative research methods. Upon closer examination, however, it becomes clear that these methods relate primarily to how data is collected, i.e. they are data collection methods. Data analysis methods, i.e. things we do with our observations after they are collected, may include some forms of content analysis, narrative analysis, discourse analysis, etc. In practice, of course, data collection and analysis are closely related. However, separating these processes conceptually into sub-processes—collection and analysis—exposes different aspects of quality considerations because things we can do to improve the quality of our research during the collection process differ from those we can do to improve the quality of analysis and interpretation. [10]

2.1 Conceptualization of a research question

Two issues relate to the quality of qualitative research methods that emerge during the conceptualization and elaboration of a research question: choice of a meta-theory and assumptions associated with a particular approach. By the time researchers find themselves in charge of a research project, they tend to be fully socialized within a particular research tradition. Departments dominated by a particular meta-theory, supervisors subscribing to specific ideologies, theories, and views, or funding bodies' explicit or implicit preferences for certain approaches tend to direct scholars into specific ways of conceiving, conducting,

³ It has often been suggested that these terms are inappropriate since they have emerged from a positivistic tradition. However, we argue that concerns about data quality transcend positivism; while we have nothing against coining new terms, particularly if this would avoid the conceptual baggage that may be attached to a certain terminology, we believe that we may want to examine existing tools before adding new terms to potentially similar concepts.

and presenting their research. For instance, while one political science department may be supportive of studies on how and why gender is important to the study of political behavior, another department may find such questions irrelevant to the subject and relegate them to the Center of Gender Studies. Thus, institutionalized traditions or preferences provide different colored and shaped lenses through which research questions are selected and conceptualized. Approaches marked by symbolic interactionism, DURKHEIMian structuralism, rational choice theory, Marxism, functionalism, psychoanalytic theory, feminist theory, historicism, etc. are likely to find different answers to similar questions, regardless of how carefully the studies are being conducted. More precisely, meta theoretical approaches make different assumptions about human thought and action and, thus, are likely to accumulate, code, analyze, and interpret data differently. Most researchers simply follow a particular tradition, whether or not they are conscious thereof. Only the most optimistic and zealous converts of a meta-theory would proclaim outright victory of one fundamental theory over another (e.g. rational choice theory vs. symbolic interactionism) based on empirical evidence. [11]

A national resource center should not partake in discussions about which meta-theory should be adopted for a research project. This is clearly a decision for researchers and their institutions. Usually, researchers already work within a particular framework, whether or not they are aware of it. However, a resource center can assist in examining the specific assumptions underlying a particular meta-theory and, accordingly, make suggestions as to what data and analytical methods may be suitable for such a study. In addition, this center could maintain a well-documented catalogue or inventory, made up of prior research studies in the qualitative tradition. [12]

Related to the selection of a fundamental theory are *a priori* knowledge and assumptions, which guide not only the choice of a research topic, but also the definition of the constructs and the scope of the study. Generally, we are unable to select all aspects which relate to a research topic. Only a few of the potential aspects relating to a research topic are recognized and, of these, only a very few are explicitly examined in the course of the research process. For example, the study of risky sexual behavior could include issues relating to age, gender, social position, sexual identity, national, regional and cultural considerations, morals, perceptions of agency, the Zeitgeist, economics, personal and collective histories, religiosity, motives, attitudes, behaviors, values, cognitions, affect, peer and reference groups, childhood experiences, relations with parents throughout the life span, relationship history, etc. This list represents only a subgroup of aspects that could be studied. Furthermore, while most of these aspects may be relevant to risky sexual behavior, only a small group of potential aspects related to our research topic can be studied. Frequently, the selection process about aspects which eventually find their way into a research project are based on not only their relevance to the topic but fulfill other criteria, such as the interest, ability, experience, and habits of the researcher, the requirements and restrictions by the funding body, the coherence between the aspects, research findings from related

fields or other countries, access to data, familiarity with specific analytic methods, etc. [13]

Also here, a national resource center should not interfere with the choices relating to the elements included in, or excluded from, a study. Instead, it could assist in examining the coherence of the elements with each other, as well as the line of argumentation which ties these elements together. [14]

Many detrimental research decisions have been made long before the "doing" of research in terms of data collection and analysis. Simply the choice of a research topic and aspects relating to it, as well as the meta-theory from which the researcher is examining the topic, preclude objective and unbiased examination in a positivistic sense. While unfocused research goes everywhere and, thus, nowhere, focused research often tells us as much about the researchers and their context as it does about the subject under investigation. [15]

Judging the reliability and validity of a research project based on the choice of meta-theory is futile. General discussions about whether Marxist theory is better or worse than rational choice theory, for instance, should probably be left to ideologues. It is far more fruitful to be aware of the specific ideological baggage as this may allow us at times to be self-critical in terms of the limitations embedded in all conceptualizations of the research questions and their relevant constructs, the interpretative tasks during the sorting and classification of data, and inferences that are drawn from the results of an analysis. [16]

2.2 Data collection⁴

The quality of the data collection process in qualitative methods can be divided conceptually into the quality of the instrument or other method of data collection, and the quality of the data obtained from the instrument. To keep this topic manageable, we will limit our discussion to quality concerns as they relate to interviews and focus groups. Nevertheless, similar arguments can be made toward other data collection methods, including documental research or observational methods. [17]

2.3 Data quality

A number of conditions are placed on observable phenomena which are selected to the status of data: we assign the status of data to that which we are able to observe, which we think is relevant to answering our research question, and which, for various reasons, we give priority over other empirical observations that may also be observable and relevant. The subjective nature of the choice and detection of data with respect to researchers and their context has been discussed above. Data collection is a process of selectively choosing empirical phenomena and attributing relevance to them with respect to the research question. Therefore, data are nothing but interpreted observations and our

4 We have omitted micro-analytic themes about data collection during the interview or focus group process, which have been covered in many how-to books on the individual techniques.

findings are strongly dependent upon what we accord the status of data (e.g. COOMBS 1964). All the baggage which shapes our selection criteria, whether or not we are aware of it, shapes also the body of raw data through, for example, question phrasing, probed responses, and a variety of non-verbal cues, which enter the interaction between researchers, their data collection methods and devices, and respondents. [18]

The following example will illustrate the subjective nature of the research process as it relates to data collection. Let us assume that we want to determine the similarities and differences between men and women with regard to how they conceptualize relationships. If we ask interviewees "What do you think are the differences between men and women with regard to relationships?" the respondents are likely to focus on differences. Data obtained from exploratory interviews on the differences between men and women differ indeed markedly from data obtained from interviews that explored the question "What do you think are the similarities between men and women with regard to relationships?" In addition, most interviewees assume that the interview is about romantic relationships between men and women. If we were interested in what our respondents believe are the differences and similarities between men and women with regard to relationships, either of the two interview questions (an exploration of differences and similarities of romantic relationships) is flawed because, first, the instrument did not address sufficiently the research question (i.e. similarities and differences) and, second, it may not accurately reflect the subject under investigation because, although not stated explicitly by the interviewer, the respondents assume that the topic under investigation is about romantic relations. If the interviewer is unaware of what was understood by the participant, an analysis of data from either of the two interview questions would lead to incorrect conclusions. [19]

In order to assure that the interviewee understands the questions in the way intended, we suggest, as one possible strategy, to conduct at least two types of pilot studies. First, we propose to conduct unstructured, exploratory interviews in which we ask interviewees to describe key concepts relating to our research question (e.g. types and boundaries of relationships, types and boundaries of intimacy). Second, using this material to help us construct our interview schedule (or questionnaire), we ask interviewees to paraphrase each of the questions and to tell us what they think we are trying to assess. The data we obtain from these pilot studies reveal not only various aspects of research methodology (e.g. question order, question threat) but also substantial information about the topic under investigation with regard to participants' assumptions about what the researcher knows or wants to know. In short, the pilot studies give us insight into whether we understand sufficiently what the respondents mean when they refer to certain key terms. During the process of constructing the research question and the associated instruments, the importance of pilot studies in close collaboration with members of the study population cannot be overemphasized. [20]

Much can be said about various aspects of data quality. Here, we merely use one of these, internal consistency, to demonstrate the issue. [21]

2.4 Internal consistency

Especially social psychologists have demonstrated that people are often inconsistent: their behaviors seem inconsistent with stated attitudes or values, behaviors seem inconsistent with other behaviors, attitudes and values seem inconsistent with each other, etc. When we talk about internal consistency in people's narratives during focus groups or interviews, we are not trying to create consistency where none exists. Instead, we are looking for apparent inconsistencies that indicate yet unexamined or misinterpreted aspects relating to our research theme. For example, during an interview, the following statement was made by a male interviewee:

"Men and women are pretty much exactly the same. Both women and men are looking for gentleness, erotic moments, closeness, relief from tension. At times in my life, I have looked for lifelong commitments, and I know of women, who are seeking erotic fantasy and an affair. But I often think that women are more moody. Of course, this may have something to do with the female biological cycle but, honestly, I think there are no differences between them." [22]

If we were to take this passage at face value, we could either conclude from it that the interviewee does not see any difference between men and women with regard to relationships or, alternatively, that the respondent believes in the differences between men and women since he stated that women are more moody based on biological difference. This passage reveals interesting inconsistencies or at least tension between the two interpretations. To explore the tension, i.e. to seek a clearer understanding, the interviewer had to probe. The following is the subsequent exchange:

- Interviewer: Does this mean that "being moody" would be one of the differences between men and women with regard to relationships?
- Respondent: No. Well, yes, but that does not mean that men cannot be moody.
- I: Are there any other things that come to your mind where men are different from women?
- R: No, not really.
- I: So if there are two people, one of them wants a long-term relationship and the other wants an erotic adventure, an affair—which one would you guess is the women and which one is the man?
- R: I don't know. They both can have affairs.
- I: But if you had to choose. If you had to put money on this.
- R: Well, that's obvious. The man would be the one who wants an affair, and the woman is the one who wants a long-term relationship.
- I: Why?
- R: That's what's in their nature. [23]

The interviewer detected the inconsistency in the initial statements, i.e. that there is no difference between men and women versus the idea that women are more moody, apparently due to their biology. The lack of internal consistency within one statement alerted the interviewer that he may have misunderstood the statement, or that there may be explanations beyond this apparent inconsistency, which would reveal further aspects about the interviewee's position. Through probes, the interviewer was able to reveal that the interviewee clearly differentiated between the possibility and the probability of characteristics and behavior patterns between men and women. In other words, when the interviewee stated that there are no differences between men and women, he meant that all characteristics or behavior patterns are possible for both men and women. In contrast, certain characteristics were rated as more or less probable, according to gender. In sum, the interviewer was able to find out a great deal about the position of the respondent, simply by probing the apparently inconsistent information. More generally, we advocate research designs in which the collection of meta-data allows for an empirical assessment of the meaning construction within the immediate context by the respondent, rather than imputing these based on scant information by the researcher. While this may be too costly for many projects, we nevertheless strongly urge researchers to consider explorations of this kind at least during the pilot stage of a study. The old adage "No data is better than bad data" comes to mind here. [24]

A national resource center can assist in exploring consistency issues in a number of ways, including the examination of the questionnaire schedule, review of the interview or focus group process, exploration of the degree of detail of the collected data, adequacy of audio or video data for subsequent analysis, sampling procedures, explicit incorporation of context, etc. Given that the interview or focus group is fundamentally an interaction between the topic, the researcher, and the respondents, however, a micro-analytic examination of the data would go beyond the capacity of a resource center, which could only help in the training of interviewers and facilitators of focus groups, as well as reviewing the overall research design, question order, phrasing, wording, and integrity of the instrument. [25]

2.5 Additional remarks on probes and a caveat

We have discussed quality concerns relating to focus groups and interviews only insofar as they deal with the conceptualization of a research question and data collection. But, as stated at the beginning, a conceptual distinction between different aspects of the research process does not imply that conceptualization of the research question, data collection, analysis, and interpretation can be separated in practice. For instance, in-depth interviewing usually includes an implicit or explicit analysis, at least in terms of an exploration of what respondents mean when they say or do something during the interview. We use probes extensively during interviews to find out what interviewees mean, or to obtain more detailed information, but this also implies a pre-analysis and construction of meaning during the interview process. While chronologically separate from subsequent analysis of interview data, probes constitute a form of (pre-)analysis

in that they are often based upon an on-the-spot, ad hoc analysis of the content of an interviewee's statement. For instance, when an interviewee states that he does not consider himself at risk of HIV infection despite being sexually active, we may use probes not only to find out why he holds this view, but also what, precisely, he understands by "risky" and "safe" behavior. By using probes in this way, we attempt to determine not only the respondents' positions on an issue, but also which aspects they connect with this issue and what relevance such aspects have to their positions. We may have specific *a priori* theories in our mind when we formulate probes, or we may want to confirm if we understand the answer in the way it was intended to be understood. While probes represent a formidable means of clarification and acquiring additional information, they also clearly present strong cues to an interviewee. Awareness of their effect may help assess the degree to which the response has been formed by the interviewees' preconceived notions about the interview or the interviewer. [26]

So far, we have outlined specific issues relating to the consistency and credibility, or reliability and validity, as they relate to data collection. We have separated somewhat artificially the conceptualization of a research question and data collection techniques. However, as we have argued throughout, most decisions taken during the research process reflect preconceptions, i.e. pre-analyses; so, in practice, conceptualization, data collection, data analysis, and interpretation are iterative and irreducibly linked. [27]

We will now turn to some examples of quality criteria related to the analysis and interpretation of data. Here again, we will comment on how these may or may not be pursued in conjunction with a national resource center for qualitative methods. [28]

2.6 Data analysis and interpretation

In order to analyze and interpret empirical phenomena in the course of conducting research in the qualitative tradition, researchers are engaged in sorting and labeling. The subjective element inherent in these processes is well known (cf. BAUMEISTER 1987; BERGMAN 2000, 2002; COOMBS 1964; CUSHMAN 1990; GEERTZ 1975, 1988; GERGEN 1973; GIDDENS 1976; JAHODA 1993; SHERIF & SHERIF 1967; VAN DIJK 1998; WEBER 1949/1977). Researchers tend to deal with the subjective element in empirical research in one of three ways: accept subjectivity as an unavoidable shortcoming; consider this a fault that can be partially eliminated through careful research design; or embrace this phenomenon as a natural part of research. Here, we will examine different aspects of subjectivity in the analytical and interpretative process and discuss the extent to which research design may or may not improve on the quality of interpretations. [29]

The identification of presences and absences, as well as similarities and differences between empirical phenomena is the foundation on which meaning and understanding is based. The process of (re)creating categories, their content,

their boundaries, and their relations to each other is central to thought and actions. As COXON states:

"Categorization and classification—putting a number of things into a smaller number of groups and being able to give the rule by which such allocation is made—are probably the most fundamental operations in thinking and language and are central to a wide variety of disciplines" (1999, p.1). [30]

As part of the research process, categorization⁵ needs to be understood in a wider sense: first, in terms of its socio-cultural and political precursors, which give rise to the particularities of categorization, while inhibiting or making impossible others. Second, categorization in itself is fundamental to human understanding. As stated by Jerome BRUNER and his colleagues: "perceiving or registering an object or an event in the environment involves an act of categorization" (1956, S.92), and through this act of categorization, the object or event is concurrently imbued with meaning. Third, salient features and context of an event or position qualify the type, content, and boundaries of categories. [31]

Central to the social and political sciences are typologies which are a function of sorting and categorizing. Knowledge and meaning of all kind is produced, understood, and communicated through categories. In connection with the association between categorization and knowledge, it is obvious that it is not things or events themselves that produce categories but, first, the interpretation of the salient features of the things or events; second, pre-existing interpretations that set conditions and limitations for subsequent interpretations of certain features of things and events; and, third, the ability and power to impose the meaning embedded in these interpretations over other, alternative interpretations.⁶ Categorizations are a function of how meaning is constructed and, thus, related to socio-cultural norms, values, customs, ideology, practices, etc. Therefore, they influence needs to be understood less as a bias (i.e. something that needs to be "controlled" or eliminated in order to get to the truth) in our effort to sort and categorize empirical phenomena but rather as a framework of knowledge and reference from which empirical observations can be classified and, thus, understood and communicated. [32]

2.7 Quality of analysis and interpretation

Until now, we emphasized the subjective, context-dependent, and socio-cultural nature of categorization during the analytical process. Accordingly, it may appear contradictory to introduce the concept of validity. Many researchers and methodologists believe that validity and subjectivity are irreconcilable because of their different epistemological positions: validity supposedly belongs to a positivistic

5 We will use the terms "categorization," "sorting," and "clustering" interchangeably in order to avoid monotony of terminology.

6 These fundamental ideas have been discussed in far greater detail in phenomenology and serve here merely to emphasize these processes of interpretative research and its links to validity concerns of research-based categories and knowledge production.

paradigm, while subjectivity belongs to research endeavors of "the other" traditions (e.g. postmodern, qualitative, interpret(at)ive, or ethnographic).⁷ [33]

2.8 Validity and sorting

First and foremost, it is the research question and the elements associated with this question that drive data collection and, thus, influence the type of categories and sorting rules. Frequently, however, one finds that the link between the research question and the data is not entirely convincing. In this sense, John TUKEY proposed a Type III error, i.e. the greatest threat to validity is asking the wrong questions of the data (cited in RAIFFA 1968). To explore this problem more closely, let us examine the actual process of sorting and how it relates to validity. [34]

2.9 Sorting

Sorting or categorizing is the process of ordering empirical observations with regard to their similarity (or dissimilarity). The function of categorization in the research process is to sort empirical phenomena according to rules that are believed to relate to a specific research question. Categories, however, are not only based on naturally occurring patterns (e.g. the classification of chemical elements according to their weight, structure, or affinity to other elements)⁸ but, more importantly for the social sciences, are active constructions, which researchers impose on data. Multiple sorting of the same data often vary based on the research focus and theoretical paradigm. Thus, empirical judgments about the degree of similarity are dependent on the categories and the boundaries of the categories that the researcher is applying to an available set of empirical observations. Such categories effect in both degree and kind processes of perception and interpretation. In other words, a set of empirical phenomena are sorted according to meaning structures that both pre-exist and emerge from the sorting process, which, in turn, confirm the categories that are applied to the sorting process of empirical phenomena. Yet, categorization is not a purely intraindividual or egocentric process; rather, the researcher's social and cultural environment pre- and proscribes to a large extent the categories, their content, and their boundaries. Socio-cultural influences on the categorization process operate in at least two ways: first, they are influenced by shared rules, i.e. values, norms, ideologies, etc.; second, they are influenced by the immediate context, which conditions cultural rules and classification itself.⁹ [35]

So far, the structuring of socio-cultural influences has been emphasized. However, to reduce categorization to collectively shared meaning structures

7 See SHARROCK and ANDERSON (1991) for an excellent discussion on subjectivity.

8 Strictly speaking, no "naturally occurring categories" exist for two reasons: first, because selection of units such as weight and atomic structure over other considerations are based on particular foci that cannot be considered exhaustive; second, because the importance of criteria such as weight or structure need to be established within a particular theory or paradigm. For example, the classification of elements in the periodic table, despite its order according to physical qualities, is by no means the only, some would say, best, way to classify elements.

9 Of course, salience and the current context are socio-culturally moderated.

would not do justice to human understanding and agency. Categorization, although always socially and culturally bound, is not only a product of a collective or culture. At first glance contradictory, this statement emphasizes a subtle difference between encultured and culture-bound human processes: no non-trivial thought or action can be conceived of that occurs independent of culture, yet culture does not always, and rarely entirely, pre- and proscribe a particular interpretation of empirical phenomena. Under certain conditions, for example novel situations or contradictions (which are particularly present in empirical research), socio-cultural interpretations may be either totally absent or at least not salient. But even when confronted with novel or contradictory phenomena, the researcher is neither likely to stop interpreting, nor capable of interpretation beyond pre-existing and culture-bound meaning structures. Instead, interpretation will take place based on both the unique context as experienced by the interpreter, as well as by the matching of the novel or contradictory phenomena to socialized and encultured explanations that render the novel or contradictory understandable and coherent. [36]

2.10 Sorting of what?

Categorization can be applied to any entity that can be empirically or theoretically captured, e.g. situations, individuals, families, groups, tribes, organizations, cultures, nations, actions, words, statements, themes, phrases, pictures, sounds, odors, interactions, networks, systems, statuses, concepts, time periods, practices, social arrangements, theories, artifacts, architectural styles, facial expressions, etc. Furthermore, classification may include comparative aspects such as gender groups, nationalities, age groups, social class, educational achievement, etc. [37]

However, analysis begins before data has been collected. As Clyde COOMBS (1964) appropriately observed, empirical phenomena, which are eventually raised to the status of data, are but a tiny subset of observable phenomena. As shown in Figure 1, only a fraction of potentially observable (but, for whatever reasons, unobserved) empirical phenomena relevant to a research question are actually considered by a researcher, and only a fraction of this subgroup eventually ends up as research data. As such, a pre-sorting of empirical phenomena takes place, where observable empirical phenomena are implicitly divided into data and non-data. Selection criteria in this process include a multitude of decisions such as the researchers' methodological and theoretical preferences, convenience of acquisition, cost, researcher and academic habits, preferences expressed by supervisors or research funding bodies, etc. As VAN MAANEN, MANNING and MILLER state: "Deciding what is to count as a unit of analysis is fundamentally an interpretive issue requiring judgment and choice" (1986, p.5) or as COXON observes more generally: "the answer to the question of 'What do I do with the information once I have collected it?' is ... going to depend in large part on decisions already made and on what assumptions the researcher [brings] to interpreting the observations" (1999, p.5).

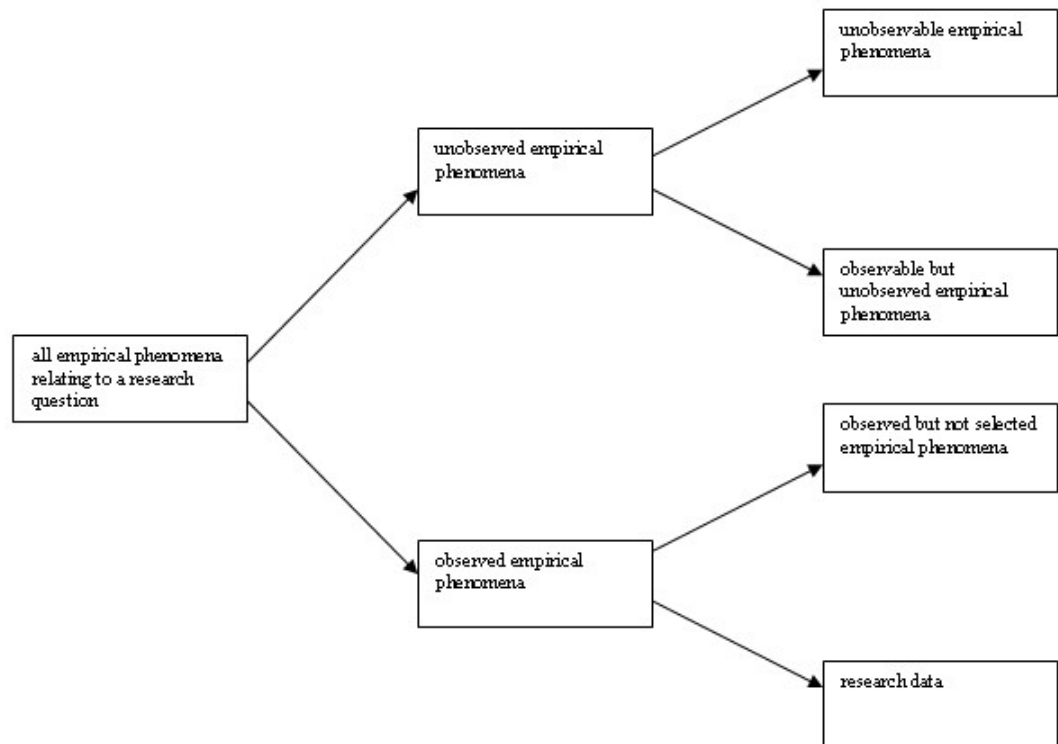


Figure 1: Types of Empirical Phenomena [38]

If data selection, i.e. identifying data from a larger set of empirical observations, is dependent on judgment and choice, and if these are strongly linked to the context within which researchers operate (BERGMAN 2002), it stands to reason that researchers may have to justify their choices. The appropriateness of data, as well as the appropriateness of selection of collection and analytical procedures, is demonstrated in at least four ways:

- Authority arguments: Claims about the appropriateness of a theory, statement, assumption, or procedure are often made through direct reference to what the authors consider reputable individuals, groups, or institutions. For example, citations of well-known academics (e.g. "According to Prof. Smith at Harvard University ..."), political figures, NGOs, and think tanks (e.g. "For years, UNICEF has shown that ...") are often used to argue for a particular choice of data, procedure, or categorization. Clearly, this is an effective way to justify research procedures and findings and can be thought of as cohort validity. Of course, skilled authors are able to selectively make authority arguments that bolster a particular line of argument, so this strategic verification of knowledge, although effective and widespread, is the most threatening to the integrity of a study and its results.
- Prior empirical studies: Obviously, it is nearly impossible to conduct research within an established area without comparing and contrasting it with the empirical work of prior studies. To insert one's study into the landscape of prior research in a particular area, authors cite empirical studies in order to either align or distance themselves with regard to paradigms, theories,

approaches, or procedures. Authors routinely employ phrases such as "What Smith and Brown failed to consider ..." or "As JOHNSON has convincingly demonstrated ...". Beyond embedding a research project within a discipline and research area, this strategy also serves as an authority argument, similar to that described above. A citation of a "scientific article," particularly if authored by an authoritative figure, is a far stronger discursive strategy to justify a particular approach, compared to simple authority arguments. It has to be admitted, however, that a careful selection and (re)interpretation of the "relevant literature" entails a significant amount of strategic positioning and, thus, relies heavily on cohort validity as well.

- Theory: As already discussed in the first part of this article, the theoretical framework pre- and proscribes to a great extent not only what will count as relevant data, but also how data is to be analyzed and interpreted. For example, Marxist approaches will be steeped in class conflict and analyze data accordingly, while rational choice theory will have different criteria not only for what constitutes relevant data, but how this data should be analyzed and interpreted. As before, while a theory imposes *a priori* limits to what can be studied and how, it also allows for a more coherent way to conceive of, conduct, and communicate research findings.
- Logical Argument: Selecting and strategically presenting authority arguments, prior empirical research, and theories are only one way in which logical argument justifies some research decisions over others. Comparing and contrasting, summarizing, synthesizing are also activities that help justify data selection and analysis. Sorting rules, i.e. based on what grounds certain categories have been selected, and empirical examples, i.e. how empirical evidence "fits" into the categories, also heavily rely on logical argument. Neither data nor categories speak for themselves but must be explained and justified. [39]

These four tools are obviously interdependent, e.g. authority arguments are often linked to prior empirical studies and, through crafty use of logic (and either omission or deconstruction of counter-arguments), can be combined to make analytical and interpretative choices appear consistent and coherent. [40]

Researchers impose conditions on their data and the categorization process that invariably lead to an oversimplification of findings. So-called "lumpers," for instance, may be able to devise a dichotomous categorization system for all world societies (e.g. individualistic vs. collectivistic societies; oriental vs. occidental cultures), while "splitters" will argue that, due to the complexity and uniqueness of societies, none can be compared with another.¹⁰ However, whatever categories we will propose, they are always, by definition, imperfect approximations that do not do justice to the within-category variation. Important to this idea for our purposes is that this relativistic insight does not free researchers from convincing

¹⁰ The degree of cultural consensus around meaning structures has been studied extensively (e.g. ARABIE, CAROLL & DESARBO 1987; ROMNEY, WELLER & BATCHELDER 1986) and is one of the main foci of the software package Anthropac (BORGATTI 1989). Applying these principles to the research activity itself, i.e. sorting and classifying data within the qualitative tradition, may shed light on why and how raters differ.

their audience of the coherence of their classification, their contents, and their boundaries. Once again, the four tools mentioned above—authority, previous studies, theory, and logic—will have to be used to achieve this goal. In this sense, it is less important to know how many cultures there may be but rather how cultural categories were established and by what empirical evidence this classification can be sustained. [41]

In an attempt to establish the validity of a classification (i.e. preventing selective reporting and overinterpretation of data), researchers often attempt to establish inter-rater reliability,¹¹ i.e. the convergence of sorting of the same material between two or more raters. This can be done formally, e.g. hiring research assistants as raters, or informally, e.g. trying to convince a supervisor or research partner. In more formal pursuits, a measure of association, most frequently a correlation coefficient, supposedly indicates the degree of inter-rater reliability and, by implication, convergence of a "correct reading" of the data. In other words, it is believed that if different raters produce very similar classifications of a set of stimuli, the classification is somehow correct and this correctness can be expressed by a statistic. [42]

However, a high convergence between two raters may imply other aspects: First, the simpler the sorting rules, the more likely raters will converge. For instance, the number of occurrences of the word freedom in a given set of presidential speeches is likely to lead to high convergence. But such simplicity is bought at a cost: frequency of a term tells us very little about the meaning that this frequency holds. The more complex the coding rule, the more raters are likely to diverge. However, this does not mean that social science should limit itself to simplistic coding schemes. Instead, the more complex the scheme, the more it has to be explicitly elaborated which, consequently, is likely to lead to greater convergence between raters. [43]

Second, and particularly applicable to more complex coding schemes, convergence of ratings often depend on a shared meaning space. This meaning space, as elaborated earlier, is culturally conditioned and contextually moderated. Detecting discursive strategies about sexually risky behavior, for instance, may strongly depend on the familiarity of the linguistic, social, historical, and cultural environment within which narratives are produced. The more both raters share the same conceptual and meaning frames, the more likely will their ratings converge. [44]

Third, and related to the previous point, training has a large influence of inter-rater reliability. Before raters are let loose on the data, they are often carefully selected (relationship with the researcher, field, approach, etc.) and trained (e.g. theoretical background of the research, research hypotheses, practice runs to see whether they understand and apply the coding scheme correctly, etc.). While it is unlikely that two raters will converge with their ratings if they do not agree on the research question and sorting rules, such agreement and the sharing of a

¹¹ Also known as inter-coder reliability or triangulation, although the latter is a far more inclusive term (see DENZIN 1970).

socio-cultural meaning space may create a fictitious sense of validity as alternative, also valid interpretations have been eliminated by the training process. [45]

Consequently, a convergence of ratings, either formally or informally, is a necessary but insufficient condition to validate a coding scheme or its content. In sum, we do not expect raters who apply different rules or a different logic to converge with their ratings, but some convergence should take place if raters follow similar rules and logic. [46]

3. The Contributions of a Resource Center

A national resource center for qualitative methods would have several strategies at its disposal to help establish and maintain quality standards. These strategies include teaching, consulting, maintaining an information base, and research. [47]

In our opinion, the teaching of research methods is the most effective long-term strategy to help establish and maintain quality standards in qualitative methods. Efforts in this area should be based not only on cooperation with universities by offering university-based courses (including summer school courses), but also by offering short-term courses for researchers and practitioners who are unable to attend university courses for a variety of reasons (cf. BERGMANN 1999, 2001). Effectiveness in the diffusion of information about such courses could take place in close collaboration with research funding institutions, such as the National Science Foundation, the Health Department, or other relevant organizations, such the Swiss Political Association, the Swiss Sociological Association, the Swiss Psychological Association, the Swiss Evaluation Society, etc. [48]

A professional consulting service should be part of a resource center. There are two constraints that need to be considered with regard to this service. Due to resource limitations, it will not be possible to offer such services to the entire research community. Instead, some rules should be elaborated that would make consultation requests more manageable. These could include, but are not limited to, giving preference to advanced graduate student projects or publicly funded research. Another limitation relates to a potential conflict of authority. More precisely, any consultation must take place with the explicit consent of the project or thesis supervisor in order to avoid a conflict of opinion, authority, or interest. [49]

The maintenance of an information base on qualitative methods would mainly consist of a regularly updated website. This site should include university courses relating to qualitative methods in Switzerland, information on national and international summer schools that offer courses in qualitative methods, and links to web-sites that specialize in qualitative methods, including discussion forums, software, and other relevant resources. [50]

Finally, the most effective resource centers of this kind are places where state-of-the-art research is not only preached, but also practiced. To make this option financially viable, one could envision project-specific collaborations with

researchers from universities and other institutions. In addition, non-stipendiary visiting fellowships have had success in other countries and could potentially find success in Switzerland also. [51]

4. Conclusions

We hope to have demonstrated in this article that quality concerns should be an integral and explicit part in qualitative research. While practically all research decisions in the research process, from the inception of a research question to the interpretation of findings, depend to some extent on subjective evaluation and judgment, it does not relieve researchers from making these elements as explicit as possible—less to make the research content but instead the research process more coherent and convincing. Whether or not researchers coin their own terminology because they reject constructs that may have emerged from another epistemological tradition, or whether they begin their quality considerations by adopting the existing terminology is not important at this point. Instead, it is the accountability of research practices through explicit description of research steps, which allow an audience to judge the plausibility of a particular study and its findings. In this sense, SWANBORN (1996) proposes to expose the reader to methodological decisions and interpretative judgments. SEALE (1999) goes further and suggests making the research community the arbiter of quality in the hope that, through a form of democratic negotiation within the research community, quality procedures and guidelines emerge. He states:

"All that we are usually left with, once individualistic, paranoid and egotistical tendencies have finally played themselves out, are some rather well-worn principles for encouraging cooperative human enterprises ... a research community exists, to which researchers in practice must relate and which possesses various mechanisms of reward and sanctions for encouraging good-quality research work ..." (1999, pp.30-31). [52]

But just as democracies cannot protect us from paranoid and egotistical leadership, so the agreement among the majority of experts (or their elected spokespersons) fails to guarantee the quality of research procedures and interpretations. Moreover, only in very abstract terms can we speak of the nearly always divided research community in the singular. Instead, research communities are usually ruled as fiefdoms and jealously guarded against invasion or revolution (e.g. KUHN 1962). [53]

Validity concerns, in whatever language they are translated, should always be present at each research decision. Although it will never be possible to prove that a procedure or result is valid in an objective sense, making quality concerns a "fertile obsession" (LATHER 1993) will render empirical research far more plausible and convincing. And this is very timely as qualitative research has left the defensive trenches and is now in many areas well integrated in the policy and governance process. [54]

References

- Arabie, Phipps; Carroll, J. Douglas & DeSarbo, Wayne S. (1987). *Three-Way Scaling and Clustering*. Thousand Oaks, CA: Sage.
- Baumeister, Roy F. (1987). How the Self Became a Problem: A Psychological Review of Historical Research. *Journal of Personality and Social Psychology*, 52(1), 163-176.
- Bergman, M. Max (2000). To Whom Will you Liken Me and Make Me Equal? Reliability and Validity in Interpretative Research Methods. Part I: Conception of Research Question and Data Collection. In Jörg Blasius, Joop Hox, Edith de Leeuw & Peter Schmidt (Eds.), *Social Science Methodology in the New Millennium*. Cologne: TT-Publikaties.
- Bergman, M. Max (2002). Reliability and Validity in Interpretative Research During the Conception of the Research Topic and Data Collection. *Sozialer Sinn*, 2, 317-331.
- Bergman, M. Max & Coxon, Anthony P.M. (2002). When Worlds Collide: Validity and the Sorting of Observations in Qualitative Research. Conference presentation. In [Thomas S. Eberle](#) (Ed.), *Quality issues in Qualitative Research* (RC33) at the XVth ISA World Congress of Sociology, Brisbane, Australia. 7-13 July.
- Borgatti, Stephen P. (1989). Using Anthropic to Investigate a Cultural Domain. *Cultural Anthropology Methods Newsletter*, 1(2), 11.
- Bruner, Jerome S.; Goodnow, Jacqueline J. & Austin, George A. (1956). *A Study of Thinking*. New York: Wiley.
- Coombs, Clyde H. (1964). *A Theory of Data*. New York: Wiley.
- Coxon, Anthony P.M. (1999). *Sorting Data: Collection and Analysis*. Thousand Oakes, CA: Sage.
- Cushman, Philip (1990). Why the Self is Empty: Toward a Historically Situated Psychology. *American Psychologist*, 45, 599-611.
- [Denzin, Norman K.](#) (1970). Strategies of Multiple Triangulation. In Norman K. Denzin (Ed.), *The Research Act in Sociology: A Theoretical Introduction to Sociological Method* (pp.297-313). New York: McGraw-Hill.
- [Fielding, Nigel](#) (2005). The Resurgence, Legitimation and Institutionalization of Qualitative Methods [23 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 6(2), Art. 32. Available at: <http://www.qualitative-research.net/fqs-texte/2-05/05-2-32-e.htm> [Date of Access: May 1, 2005].
- Geertz, Clifford (1975). On the Nature of Anthropological Understanding. *American Scientist*, 63, 47-53.
- Geertz, Clifford (1988). *Works and Lives: The Anthropologist as Author*. Stanford: Stanford University Press.
- [Gergen, Kenneth J.](#) (1973). Social Psychology as History. *Journal of Personality and Social Psychology*, 26, 309-320.
- Giddens, Anthony (1976). *New Rules of Sociological Method: A Positive Critique of Interpretative Sociologies*. New York: Basic.
- Jahoda, Gustav (1993). *Crossroads between Culture and Mind: Continuities and Change in Theories of Human Nature*. Cambridge, MA: Harvard University Press.
- Kuhn, Thomas S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Lather, Patti (1993). Fertile Obsession: Validity after Poststructuralism. *Sociological Quarterly*, 34(4), 673-693.
- Mottier, Véronique (2005). The Interpretive Turn: History, Memory, and Storage in Qualitative Research [21 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 6(2), Art. 33. Available at: <http://www.qualitative-research.net/fqs-texte/2-05/05-2-33-e.htm> [Date of Access: May 1, 2005].
- Raiffa, Howard (1968). *Decision Analysis*. Reading, MA: Addison-Wesley.
- Romney, A. Kimball; Weller, Susan C. & Batchelder, William H. (1986). Culture as Consensus: A Theory of Culture and Informant Accuracy. *American Anthropologist*, 88, 313-338.
- Seale, Clive (1999). *The Quality of Qualitative Research*. London: Sage.

Sharrock, Wes & Anderson, Bob (1991). Epistemology: Professional Skepticism. In Graham Button (Ed.), *Ethnomethodology and the Human Sciences* (pp.1-25). Cambridge: CUP.

Sherif, Carolyn W. & Sherif, Muzafer (1967). *Attitude, Ego-Involvement, and Change*. New York: Wiley.

Stone, Philip J.; Dunphy, Dexter C.; Smith, Marshall S. & Ogilvie, Daniel M. (1966). *The General Inquirer*. Cambridge, Ma: MIT Press.

Swanborn, Peter G. (first names missing) (1996). A Common Base for Quality Control Criteria in Quantitative and Qualitative Research. *Quality and Quantity*, 30, 19-35.

van Dijk, Teun (1998). *Ideology*. London: Sage.

van Maanen, John.; Manning, Peter K. & Miller, Marc L. (1986). *Qualitative Research Methods Series*. Newbury Park, CA: Sage.

Weber, Max (1949/1977). Objectivity in Social Science and Social Policy. In Fred R. Dallmayr & Thomas A. McCarthy, *Understanding and Social Inquiry* (pp.24-37). Notre Dame: University of Notre Dame Press.

Authors

Manfred Max BERGMAN

Present position: Professor of Sociology,
University of Basel, Switzerland.

Major research areas: quality assessment; data theory; mixed methods design; inequality; stratification; exclusion; discrimination.

Contact:

Manfred Max Bergman

Institute of Sociology
University of Basel
Petersgraben 27
4051 Basel

E-mail: max.bergman@unibas.ch

URL:

<http://pages.unibas.ch/soziologie/org/personal/index.html>

MTony Macmillan COXON

Present position: Honorary Professorial Fellow,
Graduate School of Social and Political Studies,
University of Edinburgh.

Major research areas: multidimensional scaling;
health studies (gay men and Aids).

Contact:

Tony Macmillan Coxon

Tigh Cargaman
Port Ellen
Isle of Islay
Argyll PA42 7BX
Scotland, UK

E-mail: apm.coxon@ed.ac.uk

URL: <http://www.tonycoxon.com/>

Citation

Bergman, Manfred Max & Coxon, Anthony P.M. (2005). The Quality in Qualitative Methods [54 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(2), Art. 34, <http://nbn-resolving.de/urn:nbn:de:0114-fqs0502344>.