SHARAN B. MERRIAM

Qualitative Research

A Guide to Design and Implementation

Revised and Expanded from *Qualitative Research and Case Study Applications in Education*

Internal Validity or Credibility

Internal validity deals with the question of how research findings match reality. How congruent are the findings with reality? Do the findings capture what is really there? Are investigators observing or measuring what they think they are measuring? Internal validity in all research thus hinges on the meaning of reality. Becker (1993) humorously points out that "reality is what we choose not to question at the moment," and "the leading cause of stress amongst those in touch with it" (p. 220). On a more serious note, Ratcliffe (1983) offers an interesting perspective on assessing validity in every kind of research. It should be remembered, he suggests, that (1) "data do not speak for themselves; there is always an interpreter, or a translator" (p. 149); (2) that "one cannot observe or measure a phenomenon/event without changing it, even in physics where reality is no longer considered to be single-faceted"; and (3) that numbers, equations, and words "are all abstract, symbolic representations of reality, but not reality itself" (p. 150). Validity, then, must be assessed in terms of something other than reality itself (which can never be grasped). That "something other than reality itself" is Lincoln and Guba's (1985) notion of credibility; that is, are the findings *credible* given the data presented?

One of the assumptions underlying qualitative research is that reality is holistic, multidimensional, and ever-changing; it is not a single, fixed, objective phenomenon waiting to be discovered, observed, and measured as in quantitative research. Assessing the isomorphism between data collected and the "reality" from which they were derived is thus an inappropriate determinant of validity. In writing about his scientific journey to the Sea

of Cortez more than fifty years ago, Steinbeck (1941) eloquently contrasted the two views of reality:

> The Mexican sierra has 'XVII–15–1X' spines in the dorsal fin. These can easily be counted. But if the sierra strikes hard on the line so that our hands are burned, if the fish sounds and nearly escapes and finally comes in over the rail, his colors pulsing and his tail beating the air, a whole new relational externality has come into being—an entity which is more than the sum of the fish plus the fisherman. The only way to count the spines of the sierra unaffected by this second relational reality is to sit in a laboratory, open an evil smelling jar, remove a stiff colorless fish from formalin solution, count the spines, and write the truth 'D. XVII–15–1X.' There you have recorded a reality which cannot be assailed—probably the least important reality concerning either the fish or yourself. The man with his pickled fish has set down one truth and has recorded in his experience many lies. The fish is not that color, that texture, that dead, nor does he smell that way. (p. 2)

Maxwell (2005) concurs that one can never really capture reality. "Validity," then, "is a goal rather than a product: it is never something that can be proven or taken for granted. Validity is also relative: It has to be assessed in relationship to the purposes and circumstances of the research, rather than being a context-independent property of methods or conclusions" (p. 105).

Then what *is* being studied in qualitative research, and how does a researcher assess the validity of those observations? What is being investigated are people's constructions of reality—how they understand the world. And just as there will be multiple accounts of eyewitnesses to a crime, so too, there will be multiple constructions of how people have experienced a particular phenomenon, how they have made meaning of their lives, or how they have come to understand certain processes.

Because human beings are the primary instrument of data collection and analysis in qualitative research, interpretations of reality are accessed directly through their observations and interviews. We are thus "closer" to reality than if a data collection instrument had been interjected between us and the participants. Most agree that when rigor is viewed in this manner, internal validity is a definite strength of qualitative research. In this type

of research it is important to understand the perspectives of those involved in the phenomenon of interest, to uncover the complexity of human behavior in a contextual framework, and to present a holistic interpretation of what is happening.

LeCompte and Preissle (1993) list four factors that lend support to the claim of high internal validity of ethnographic research:

> First, the ethnographer's common practice of living among participants and collecting data for long periods provides opportunities for continual data analysis and comparison to refine constructs; it ensures a match between researcher categories and participant realities. Second, informant interviews, a major ethnographic data source, are phrased in the empirical categories of participants; they are less abstract than many instruments used in other research designs. Third, participant observation, the ethnographer's second key source of data—is conducted in natural settings reflecting the life experiences of participants more accurately than do more contrived or laboratory settings. Finally, ethnographic analysis incorporates researcher reflection, introspection, and self-monitoring that Erickson (1973) calls disciplined subjectivity, and these expose all phases of the research to continual questioning and reevaluation. (p. 342)

Though qualitative researchers can never capture an objective "truth" or "reality," there are a number of strategies that you as a qualitative researcher can use to increase the "credibility" of your findings, or as Wolcott (2005, p.160) writes, increase "the correspondence between research and the real world." Probably the most well known strategy to shore up the internal validity of a study is what is known as *triangulation.* Usually associated with navigation or land surveying wherein two or three measurement points enable convergence on a site, the best known discussion of triangulation is Denzin's (1978), in which he proposes four types of triangulation: the use of multiple methods, multiple sources of data, multiple investigators, or multiple theories to confirm emerging findings. The use of multiple theories such as approaching "data with several hypotheses in mind, to see how each fares in relation to the data" (Seale, 1999, p. 54) is less common in qualitative research than in the other three forms.

With regard to the use of multiple *methods* of data collection, for example, what someone tells you in an interview can be checked against what you observe on site or what you read about in documents relevant to the phenomenon of interest. You have thus employed triangulation by using three methods of data collection—interviews, observations, and documents.

Triangulation using multiple sources of *data* means comparing and cross-checking data collected through observations at different times or in different places, or interview data collected from people with different perspectives or from follow-up interviews with the same people. *Investigator* triangulation occurs when there are multiple investigators collecting and analyzing data. Patton (2002, p. 560) suggests a related strategy, that of "*triangulating analysts*—that is, having two or more persons independently analyze the same qualitative data and compare their findings" (emphasis in original). This notion of multiple researchers has also been discussed in other contexts as collaborative or team research. In participatory research, where the goal of the research is political empowerment, the participants along with the researcher collectively define the problem to be addressed, conduct the study, and engage in collective action to bring about change.

It might be noted that as with other strategies for ensuring for trustworthiness in qualitative research, triangulation is being revisited in the literature from a postmodern perspective. Richardson (2000; Richardson & St. Pierre, 2005) points out that triangulation assumes a "'fixed point' or 'object' that can be triangulated." But in postmodern research, "we do not triangulate; we *crystallize*. We recognize that there are far more than three sides from which to approach the world" (Richardson, 2000, p. 934). Crystals exhibit "an infinite variety of shapes, substances, transmutations, multidimensionalities, and angles of approach. Crystals are prisms that reflect externalities and refract within themselves, creating different colors, patterns, and arrays casting off in different directions. What we see depends on our angle of response—not triangulation but rather crystallization" (Richardson, in Richardson & St. Pierre, 2005, p. 963). However, from an interpretive-constructivist perspective, which is the basis of this book, triangulation remains a principal strategy to ensure for validity and reliability.

A second common strategy for ensuring for internal validity or credibility is *member checks.* Also called *respondent validation,* the idea here is that you solicit feedback on your emerging findings from some of the people that you interviewed. "This is the single most important way of ruling out the possibility of misinterpreting the meaning of what participants say and do and the perspective they have on what is going on, as well as being an important way of identifying your own biases and misunderstanding of what you observed" (Maxwell, 2005, p. 111). The process involved in member checks is to take your preliminary analysis back to some of the participants and ask whether your interpretation "rings true." Although you may have used different words (it is *your* interpretation, after all, but derived directly from their experience), participants should be able to recognize their experience in your interpretation or suggest some fine-tuning to better capture their perspectives. Some writers suggest doing member checks throughout the course of the study. Exhibit 9.2 is a sample of the results from a member check. In this study, Crosby (2004) was interested in how learning experiences foster commitment to a career in teaching English as a foreign language. He asked several of his

EXHIBIT 9.2. MEMBER CHECK COMMENTS.

Name	*Comments*	*Action Taken*
Holly	"I think your statements are an accurate reflection of what I said and what my experience has been."	Write back and explain about meaning of "disorientating dilemma"
	The Category you term "disorientating dilemma" puzzles me. That as a category doesn't quite ring true for me. Perhaps it came across that way, although I should also say that I'm not sure what you mean with that term and how it fits into learning experiences. Do you mean my challenges in teaching have encouraged/discouraged my commitment to teaching EFL?	No action needed to change research results

(*Continued*)

EXHIBIT 9.2. (CONTINUED)

Name	*Comments*	*Action Taken*
Kate	"It was kind of fun to see a bunch of my own thoughts already categorized into a graphic!"	Spelling corrected; phrases need not be adjusted
	Change spelling of Bombera to Bambara.	
	Clarification of two phrases used as coding: Getting a Masters in TESOL, and looking for more teaching experiences	
Grace	"I would agree with your categorization of comments."	No action needed
	"I'd definitely agree with your conclusions." Charts gave "me greater insight into my own thinking."	
Mary	"Everything is right on! I have reviewed attachments and agree with what is written. The themes are accurate."	No action needed
	"I really like the table; it was exciting to see my progression through your eyes."	
Ann	"I'd say it's pretty accurate. I can't think of anything I would add, change, etc."	No action needed
Shauna	"I do believe that the analysis rings true."	Note comment of commitment first to God then profession
	"It was definitely an enlightening read. . . . It reminded me of certain convictions the Lord had placed on my heart to enter the field in the first place, and I feel encouraged as I look ahead towards my next step in the profession."	
	"My commitment is first to God and His will for my life more so that [sic] my profession."	
Bob	"Both documents look great."	No action needed
Oliver	"When I left my interview with you I didn't feel like I expressed myself well, but after looking at your documents I think what you have is fine and rings true."	No action needed

Source: Crosby (n.d). Used with permission.

participants to comment on his findings regarding their experiences teaching English in a cross-cultural setting.

Adequate engagement in data collection is a third strategy that makes sense when you are trying to get as close as possible to participants' understanding of a phenomenon. How long one needs to observe or how many people need to be interviewed are always difficult questions to answer, since the answers are always dependent on the particular study itself. The best rule of thumb is that the data and emerging findings must feel saturated; that is, you begin to see or hear the same things over and over again, and no new information surfaces as you collect more data.

Adequate time spent collecting data should also be coupled with purposefully looking for variation in the understanding of the phenomenon. Patton (2002) argues that credibility hinges partially on the integrity of the researcher, and one approach to dealing with this issue is for the researcher to "*look for data that support alternative explanations.* Failure to find strong supporting evidence for alternative ways of presenting the data or contrary explanations helps increase confidence in the original, principal explanation you generated" (p. 553, emphasis in original). Some writers even suggest that you should purposefully seek data that might disconfirm or challenge your expectations or emerging finding. This strategy has been labeled *negative* or *discrepant case analysis.*

Related to the integrity of the qualitative researcher is a fourth strategy sometimes labeled *researcher's position,* or more recently, *reflexivity*—"the process of reflecting critically on the self as researcher, the 'human as instrument'" (Lincoln & Guba, 2000, p. 183).. Investigators need to explain their biases, dispositions, and assumptions regarding the research to be undertaken. Even in journal articles authors are being called upon to articulate and clarify their assumptions, experiences, worldview, and theoretical orientation to the study at hand. Such a clarification allows the reader to better understand how the individual researcher might have arrived at the particular interpretation of the data. As Maxwell (2005, p. 108) explains, the reason for making your perspective, biases, and assumptions clear to the reader is not to eliminate "*variance* between researchers in values and expectations they bring to the study, but with understanding how

a *particular* researcher's values and expectations influence the conduct and conclusions of the study" (emphasis in original).

Yet another strategy is called *peer examination* or *peer review*. Certainly there's a sense in which all graduate students have this process built into their thesis or dissertation committee, since each member of the committee reads and comments on the findings. A similar process takes place when an article is sent into a peer-reviewed journal for publication; "peers" knowledgeable about the topic and the methodology review the manuscript and recommend publication (or not). But such an examination or review can also be conducted by a colleague either familiar with the research or one new to the topic. There are advantages to both, but either way, a thorough peer examination would involve asking a colleague to scan some of the raw data and assess whether the findings are plausible based on the data.

RELIABILITY OR CONSISTENCY

Reliability refers to the extent to which research findings can be replicated. In other words, if the study is repeated, will it yield the same results? Reliability is problematic in the social sciences simply because human behavior is never static. Even those in the hard sciences are asking similar questions about the constancy of phenomena. Reliability in a research design is based on the assumption that there is a single reality and that studying it repeatedly will yield the same results. This is a central concept of traditional experimental research, which focuses on discovering causal relationships among variables and uncovering laws to explain phenomena.

Qualitative research, however, is not conducted so that the laws of human behavior can be isolated. Rather, researchers seek to describe and explain the world as those in the world experience it. Since there are many interpretations of what is happening, there is no benchmark by which to take repeated measures and establish reliability in the traditional sense. Wolcott (2005) underscores the inappropriateness of considering reliability in studying human behavior: "In order to achieve reliability in that technical sense, a researcher has to manipulate conditions so that replicability can be assessed. Ordinarily, fieldworkers do not

try to make things happen at all, but whatever the circumstances, we most certainly cannot make them happen twice. And if something does happen more than once, we never for a minute insist that the repetition be exact" (p. 159).

Traditionally reliability is the extent to which research findings can be replicated. In other words, if the study were repeated would it yield the same results? Reliability is problematic in the social sciences simply because human behavior is never static, nor is what many experience necessarily more reliable than what one person experiences. All reports of personal experience are not necessarily unreliable, any more than all reports of events witnessed by a large number of people are reliable. Consider the magician who can fool the audience of hundreds but not the stagehand watching from the wings. Replication of a qualitative study will not yield the same results, but this does not discredit the results of any particular study; there can be numerous interpretations of the same data. The more important question for qualitative research is *whether the results are consistent with the data collected.* Lincoln and Guba (1985) were the first to conceptualize reliability in qualitative research as "dependability" or "consistency." That is, rather than demanding that outsiders get the same results, a researcher wishes outsiders to concur that, given the data collected, the results make sense—they are consistent and dependable. The question then is not whether findings will be found again but whether the results are consistent with the data collected.

The connection between reliability and internal validity from a traditional perspective rests for some on the assumption that a study is more valid if repeated observations in the same study or replications of the entire study produce the same results. This logic relies on repetition for the establishment of truth, but as everyone knows, measurements, observation, and people can be repeatedly wrong. A thermometer may repeatedly record boiling water at 85 degrees Fahrenheit; it is very reliable, since the measurement is consistent, but not at all valid. And in the social sciences, simply because a number of people have experienced the same phenomenon does not make the observations more reliable.

It is interesting, however, that the notion of reliability with regard to instrumentation can be applied to qualitative research

in a sense similar to its meaning in traditional research (Lincoln and Guba, 1985). Just as a quantitative researcher refines instruments and uses statistical techniques to ensure reliability, so too the human instrument can become more reliable through training and practice. Furthermore, the reliability of documents and personal accounts can be assessed through various techniques of analysis and triangulation.

Because what is being studied in the social world is assumed to be in flux, multifaceted, and highly contextual, because information gathered is a function of who gives it and how skilled the researcher is at getting it, and because the emergent design of a qualitative study precludes a priori controls, achieving reliability in the traditional sense is not only fanciful but impossible. Wolcott (2005) in fact wonders whether we need "address reliability at all" other than to say why it is an inappropriate measure for assessing the rigor of a qualitative study. His objection is that "similarity of responses is taken to be the same as accuracy of responses," and we know that is a problematic assumption (p. 159).

Thus, for the reasons discussed, replication of a qualitative study will not yield the same results. That fact, however, does not discredit the results of the original or subsequent studies. Several interpretations of the same data can be made, and all stand until directly contradicted by new evidence. So if the findings of a study are consistent with the data presented, the study can be considered dependable.

Strategies that a qualitative researcher can use to ensure for consistency and dependability or reliability are triangulation, peer examination, investigator's position, and the audit trail. The first three have been discussed already under Internal Validity or Credibility. The use of multiple methods of collecting data (methods triangulation), for example, can be seen as a strategy for obtaining consistent and dependable data, as well as data that are most congruent with reality as understood by the participants. The *audit trail* is a method suggested by Lincoln and Guba (1985). Just as an auditor authenticates the accounts of a business, independent readers can authenticate the findings of a study by following the trail of the researcher. While "we cannot expect others to replicate our account," Dey (1993, p. 251) writes, "the best we can do is explain how we arrived at our

results." Calling the audit trail a "log" as in what a ship might keep in detailing its journey, Richards (2005) writes that "good qualitative research gets much of its claim to validity from the researcher's ability to show convincingly how they got there, and how they built confidence that this was the best account possible. This is why qualitative research has a special need for project history, in the form of a diary or log of processes" (p. 143).

An audit trail in a qualitative study describes in detail how data were collected, how categories were derived, and how decisions were made throughout the inquiry. In order to construct this trail, the researcher must keep a research journal or record memos on the process of conducting the research as it is being undertaken. What exactly do you write in your journal or your memos? You write your reflections, your questions, and the decisions you make with regard to problems, issues, or ideas you encounter in collecting data. A running record of your interaction with the data as you engage in analysis and interpretation is also recommended. In a book-length or thesis-length report of the research, the audit trail is found in the methodology chapter (often with supporting appendices). Essentially it is a detailed account of how the study was conducted and how the data were analyzed. Due to space limitations, journal articles tend to have a very abbreviated audit trail or methodology section.