

Theory to Practice

What Can You Tell From An N of 1?: Issues of Validity and Reliability in Qualitative Research

Sharan B. Merriam

Abstract

This article deals with the issues of validity and reliability in qualitative research in education. Philosophical assumptions underlying the concepts of internal validity, reliability, and external validity or generalizability are discussed. In addition, strategies congruent with a qualitative research perspective for ensuring for rigor and trustworthiness of findings in qualitative research are presented.

At conference presentations, in reviews of journal articles, at thesis defenses, the trustworthiness of qualitative research continues to be challenged, and rightly so. Rigor is needed in all kinds of research to insure that findings are to be trusted and believed. In applied fields like education, social work, counseling, and administration, the question of the trustworthiness of research findings looms large; after all, much research is designed to understand and improve practice. We want to feel confident incorporating research findings into our practice, for what we do affects the lives of real people.

Questions most commonly posed to qualitative researchers reflect concerns with the validity and reliability of the research findings—questions such as the one in the title of this article, and others such as: “How can you generalize from a small, non-random sample?” “If somebody else did this study, would they get the same results?” “How do you know the researcher isn’t biased and just finding what he or she expects to find?” and “If the researcher is the primary instrument for data

Sharan Merriam is Professor of Adult Education at the University of Georgia.

collection and analysis, how can we be sure the researcher is a valid and reliable instrument?" These questions reflect legitimate concerns about the rigor of qualitative research; they also reflect philosophical assumptions underlying a quantitative or positivist worldview and are thus inappropriate for assessing the rigor of a qualitative study. The purpose of this paper is twofold: (1) to examine conceptions of validity and reliability from a qualitative or interpretive worldview and (2) to present strategies for insuring for validity and reliability that are consonant with assumptions underlying the qualitative paradigm.

Purposes of Qualitative Research

In assessing the trustworthiness of qualitative research, it is important to back up and ask what kinds of questions or problems qualitative research is designed to address. Qualitative research is ideal for the following: clarifying and understanding phenomena and situations when operative variables cannot be identified ahead of time; finding creative or fresh approaches to looking at over-familiar problems; understanding how participants perceive their roles or tasks in an organization; determining the history of a situation; and building theory, hypotheses, or generalizations. The question of trustworthiness becomes how well a particular study does what it is designed to do.

Notions of validity and reliability must be addressed from the perspective of the paradigm out of which the study has been conducted. That is, if I am trying to build hypotheses rather than test them, if I am trying to understand a phenomenon rather than "treat" it, if I am interested in participants' perspectives rather than my own, different questions will need to be asked about the conduct of the study.

Qualitative researchers have approached rigor from one of two angles. Some lay out the standard, positivist threats to validity and reliability made famous by Campbell and Stanley (1963) and Cook and Campbell (1979) and demonstrate how qualitative research addresses these threats. History, maturation, observer effects, selection and regression, mortality, spurious conclusions, and so on, *can* be addressed from a qualitative research perspective as demonstrated by Guba and Lincoln (1981) and Goetz and LeCompte (1984). More commonly, writers make the case that qualitative research is based on different assumptions regarding reality, thus demanding different conceptualizations of validity and reliability. Some have proposed using a different nomenclature. Agar (1986), for example, talks about credibility, accuracy of representation, and

authority of writer; Guba and Lincoln (1981) suggest credibility, dependability, and transferability.

The position expressed in this paper is that notions of validity and reliability need to be grounded in the worldview of qualitative research. Further, there are strategies that can be employed to ensure for trustworthiness that are highly compatible with this worldview. The discussion that follows is presented in terms of the three major aspects of rigor--internal validity, reliability, and external validity (generalizability).

Internal Validity

Internal validity asks the question: How congruent are one's findings with reality? In quantitative research the question is often more precisely stated: Are we observing or measuring what we think we are observing or measuring? Key to understanding internal validity is the notion of reality. Is reality fixed and stable as the positivists believe, or constructed and interpreted as qualitative researchers believe? These two views of reality are eloquently contrasted in Steinbeck's (1941) log of his scientific journey to the Sea of Cortez:

We knew that what we would see and record and construct would be warped, as all knowledge patterns are warped, first, by the collective pressure and stream of our time and race, second by the thrust of our individual personalities. But knowing this, we might not fall into too many holes—we might maintain some balance between our warp and the separate things, the external reality. The oneness of these two might take its contribution from both. For example, the Mexican sierra has 'XVII-15-IX' 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-IX.' There you have recorded a reality which cannot be assailed—probably the least important reality concerning either the fish or yourself. It is good to know what you are doing. 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)

Steinbeck goes on to say, "We were determined not to let a passion for unassailable little truths draw in the horizons and crowd the sky down on us. We knew that what seemed to us true could be only relatively true anyway. There is no other kind of observation. The man with his pickled fish has sacrificed a great observation about himself, the fish, and the focal point, which is his thought on both the sierra and himself" (pp. 2-3).

That reality is what we take to be true, and that it is only *relatively* true, is echoed by Ratcliffe (1983) who writes that rules purport to guarantee the truth of an inquiry have changed over time. Guarantors of validity "are merely rules that have been made by human beings (inquirers), and they thus necessarily serve human purposes and goals. Such rules are therefore inevitably normative: They 'prove' only that we are playing by the (current) rules" (p. 159).

Qualitative research assumes that reality is constructed, multidimensional, and ever-changing; there is no such thing as a single, immutable reality waiting to be observed and measured. Thus, there are interpretations of reality; in a sense the researcher offers his or her interpretation of someone else's interpretation of reality. Just as in quantitative research there are things you can do (such as control for extraneous variables) to ensure that findings are valid according to that paradigm's notion of reality, so too in qualitative research. The following strategies can be employed to strengthen the internal validity of a qualitative study. (These strategies and the ones discussed in the next two sections are drawn from experience and literature, in particular, Guba and Lincoln, 1981; Merriam, 1988; and Patton, 1991.)

1. Triangulation--the use of multiple investigators, multiple sources of data, or multiple methods to confirm the emerging findings (Denzin, 1970; Mathison, 1988). For example, if the researcher hears about the phenomenon in interviews, sees it taking place in observations, and reads about it in pertinent documents, he or she can be confident that the "reality" of the situation, as perceived by those in it, is being conveyed as "truthfully" as possible.

2. Member checks--taking data collected from study participants, and the tentative interpretations of these data, back to the people from whom they were derived and asking if the interpretations are plausible, if they "ring true."

3. Peer/colleague examination--asking peers or colleagues to

examine the data and to comment on the plausibility of the emerging findings.

4. Statement of researcher's experiences, assumptions, biases--presenting the orientation, biases, and so on, of the researcher at the outset of the study. This enables the reader to better understand how the data might have been interpreted in the manner in which they were.

5. Submersion/engagement in the research situation--collecting data over a long enough period of time to ensure an in-depth understanding of the phenomenon. Criteria for determining how long is long enough can be found in Guba and Lincoln (1981), Merriam (1988), and Patton (1991).

Most writers agree that internal validity is a strength of qualitative research. There are fewer "layers" between the researcher and the phenomenon under investigation. The strategies described above can help ensure that the interpretation of "reality" being presented is as "true" to the phenomenon as possible.

Reliability

Reliability is concerned with the question of the extent to which one's findings will be found again. That is, if the inquiry is replicated, would the findings be the same? Reliability in the "hard" sciences revolves around repeated measures of a phenomenon. Typically investigators disassociate themselves from the phenomenon being investigated by using "objective" measures. The more times the findings of a study can be replicated, the more stable or reliable the phenomenon is thought to be. This was precisely the problem with the cold fusion findings announced by the University of Utah scientists. Other scientists were unable to replicate their work; hence, the reliability of the findings of the original investigations were called into question.

In the social sciences the whole notion of reliability in and of itself is problematic. That is, studying people and human behavior is not the same as studying inanimate matter. Human behavior is never static. Classroom interaction is not the same, day after day, for example, nor are people's understanding of the world around them. Cronbach (1975) notes that "an actuarial table describing human affairs changes from science into history before it can be set in type" (p. 123). Further, the scientific notion of reliability assumes that repeated measures of a phenomenon (with the same results) establishes the truth of the results. However, measurements and observations can be repeatedly wrong, especially where human beings are involved.

Scriven (1972) makes the point that a lot of people experiencing the same thing does not necessarily mean that their accounts are more reliable than that of a single individual. Five hundred people reporting that they had seen a magician cut a person in half, for example, would not be as reliable as a report as that of the lone stagehand who had witnessed the event from behind the curtain.

Qualitative researchers are not seeking to establish “laws” in which reliability of observation and measurement are essential. Rather, qualitative researchers seek to understand the world from the perspectives of those in it. Since there are many perspectives, and many possible interpretations, “there is no benchmark by which one can take repeated measures and establish reliability in the traditional sense” (Merriam, 1988, p. 170). Clearly, replication of a qualitative investigation will not yield the same results (as it might in quantitative research). Rather, both sets of results stand as two interpretations of the phenomenon.

Instead of reliability, one can strive for what Lincoln and Guba (1985, p. 288) call “dependability” or “consistency.” The real question for qualitative researchers, they suggest, is not whether the results of one study are the same as the results of a second or third study, but *whether the results of a study are consistent with the data collected*. As with internal validity, there are strategies one can use to ensure for greater consistency. Three such strategies are listed below:

1. Triangulation--The use of multiple methods of data collection, in particular, as well as other forms of triangulation, can lead to dependability or consistency (as well as internal validity).
2. Peer examination--Again, this strategy provides a check that the investigator is plausibly interpreting the data; that is, someone else can be asked whether the emerging results appear to be consistent with the data collected.
3. Audit trail--This strategy, suggested by Guba and Lincoln (1981), operates on the same premise as when an auditor verifies the accounts of a business. “In order for an audit to take place, the investigator must describe in detail how data were collected, how categories were derived, and how decisions were made throughout the inquiry” (Merriam, 1988, p. 172). Goetz and LeCompte (1984) suggest that the audit trail should be so detailed “that other researchers can use the original report as an operating manual by which to replicate the study” (p. 216).

Reliability, then, cannot be thought of in qualitative research in the same way as it is in positivist research. The logic of reliability in quantitative research is based on philosophical assumptions and a worldview

different from that of qualitative research. What one strives for is consistency and dependability, a sort of internal reliability in which the findings of an investigation reflect, to the best of the researcher's ability, the data collected.

External Validity

The extent to which the findings of a study can be applied to other situations refers to the question of external validity, or generalizability. Indeed, this question seems to haunt qualitative research more than any other, probably because most people think of generalizability in the statistical sense of extrapolating from a sample to a population. Since qualitative researchers rarely select a random sample (which would then allow them to generalize to the population from which the sample was selected), it is thus concluded that one cannot generalize in qualitative research. However, even generalizability in the statistical sense has its limitations. First, although random sampling may have been used, generalizations are made within certain levels of confidence. Outlying scores are ignored (they become part of the variance). Second, when dealing with human beings, it is difficult, if not impossible, to apply statistically-based generalizations to individual persons. For example, a study might find that 70% of the surgeons in the Northeast voted Republican in the last election. But any individual surgeon, say a Dr. Jones, will have voted not 70% Republican, but either 100% Republican or 100% Democrat. Third, several writers have pointed out that very little social science research is based on true random sampling of a population; most populations from which samples are selected are confined to small geographic areas, homogenous groups, or highly defined subgroups (middle-school principals in a single state, for example).

While some qualitative researchers view generalizability as a limitation of the method, or just not appropriate for the social sciences, most prefer to think of generalizability as something different than going from a sample to a population. The goal of qualitative research, after all, is to understand the particular in depth, rather than finding out what is generally true of many. There are at least three alternative conceptions of generalizability that are congruent with the philosophical assumptions underlying qualitative research. Cronbach (1975) thinks that empirical generalizations are too lofty a goal for social science research; rather, we should think in terms of *working hypotheses*. He writes: "Instead of making generalization the ruling consideration in our research, I suggest

that we reverse our priorities. An observer collecting data in one particular situation is in a position to appraise a practice or proposition in that setting, observing effects in context...Generalization comes late...When we give proper weight to local conditions, any generalization is a working hypothesis, not a conclusion" (pp. 124-125). Working hypotheses reflect situation-specific conditions of a particular context. They can also be used to guide practice (Patton, 1991).

A second concept, called *concrete universals*, has been proposed by Erickson (1986). In attending to the particular, universals can be discovered. Concrete universals are based on the notion that particular situations convey insights that transcend the situation from which they emerge. The general lies in the particular. This is, in fact, how human beings make sense out of their world, how they cope with new situations. What is learned in a particular situation is applied to similar situations subsequently encountered. For example, a person who receives a speeding ticket on a particular highway will most likely "generalize" to subsequent instances of being on the same road and to other similar roads. People do not wait until they have a sample of experiences before they generalize to new situations.

A third way of viewing external validity is something becoming known as *reader or user generalizability*. In this view, the extent to which findings from an investigation can be applied to other situations is determined by the people in those situations. It is not up to the researcher to speculate how his or her findings can be applied to other settings; it is up to the consumer of the research. Wilson (1979) suggests the notion of "a continuum of usefulness" beginning on one end representing "the setting where the information was gathered and stretching to dissimilar settings" (p. 454).

Whether one thinks of generalizability in terms of working hypotheses, concrete universals, or user generalizability, there are strategies one can employ to strengthen this aspect of rigor in qualitative research. Four such strategies are discussed below:

1. Thick description--This involves providing enough information/description of the phenomenon under study so that readers will be able to determine how closely their situations match the research situation, and hence, whether findings can be transferred.

2. Multi-site designs--The use of several sites, cases, situations, especially those representing some variation (Glaser & Strauss, 1967), will allow the results to be applied to a greater range of other similar situations.

3. Modal comparison--This strategy involves describing how typical the program, event, or sample is compared with the majority of others in the same class. In Wolcott's (1973) case study of a school principal, for example, he tells the reader how representative his subject is compared to the typical school principal.

4. Sampling within--A phenomenon being studied may have numerous component parts (teachers, administrators, students in a school system, for example), each of which could be randomly sampled for inclusion in the study. This would allow one to "generalize" to the larger group within the unit of study.

In summary, external validity or generalizability seems to be most problematic for those not well acquainted with qualitative research. To consider generalizability a limitation of this kind of research is to be thinking in terms of statistical generalization based in the quantitative paradigm. By viewing external validity from the perspective of the assumptions underlying qualitative research, several reformulations of generalizability are possible, such as working hypotheses, concrete universals, and reader or user generalizability.

What Can You Tell From An N Of 1?

Viewed from a qualitative perspective, quite a bit can be learned from an N of 1. The trustworthiness of the findings of a study with a small N and no random sampling are dependent upon the internal validity, reliability, and external validity of the study. As was discussed in this article, there are ways to view each of these concerns that are congruent with the underlying assumptions and worldview of qualitative research. Likewise, there are strategies that investigators can employ that will ensure for the validity and reliability of the study. Rigor is as valid a concern in qualitative research as in any other kind of research. Qualitative researchers employ different means of "persuading" the reader that a study is trustworthy. This is what Firestone (1987) calls the "rhetoric" of this research. While "the quantitative study must convince the reader that procedures have been followed faithfully because very little concrete description of what anyone does is provided," qualitative research persuades through its "classical strengths" of "concrete depiction of detail, portrayal of process in an active mode, and attention to the perspectives of those studied" (pp. 19, 20).

References

- Agar, M. (1986). *Speaking of ethnography*. Beverly Hills, CA: Sage.
- Campbell, D. T., & Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally College Publishers.
- Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 30, 116-127.
- Denzin, N.K. (1970). *The research act: A theoretical introduction to sociological methods*. Chicago: Aldine.
- Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119-161). New York: Macmillan.
- Firestone, W. A. (1987). Meaning in method: The rhetoric of quantitative and qualitative research. *Educational Researcher*, 16(7), 16-21.
- Glaser, B. G., & Strauss, A. L. (1967). *Discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.
- Goetz, J. P., & LeCompte, M. D. (1984). *Ethnography and qualitative design in educational research*. Orlando, FL: Academic Press.
- Guba, E.G., & Lincoln, Y. S. (1981). *Effective evaluation*. San Francisco: Jossey-Bass.
- Mathison, S. (1988). Why triangulate? *Educational Researcher*, 17(2), 13-17.
- Merriam, S. B. (1988). *Case study research in education: A qualitative approach*. San Francisco: Jossey-Bass.
- Patton, M. Q. (1991). *Qualitative evaluation methods* (2nd ed.). Newbury Park, CA: Sage.
- Ratcliffe, J. W. (1983). Notions of validity in qualitative research methodology. *Knowledge; Creation; Diffusion; Utilization*, 5(2), 147-167.
- Scriven, M. (1972). Objectivity and subjectivity in educational research. In L. G. Thomas (Ed.), *Philosophical redirection of educational research: The seventy-first yearbook of the National Society for the Study of Education*. Chicago: University of Chicago Press.
- Steinbeck, J. (1941). *Sea of Cortez*. New York: The Viking Press.
- Wilson, S. (1979). Explorations of the usefulness of case study evaluations. *Evaluation Quarterly*, 3, 446-459.
- Wolcott, H. (1973). *The man in the principal's office*. New York: Holt, Rinehart and Winston.