
Article

Verification Strategies for Establishing Reliability and Validity in Qualitative Research

Janice M. Morse
University of Alberta
Edmonton, Alberta, Canada

Michael Barrett
University of Cambridge
Cambridge, England

Maria Mayan
International Institute for Qualitative Methodology
Edmonton, Alberta, Canada

Karin Olson
University of Alberta
Edmonton, Alberta, Canada

Jude Spiers
University of Alberta
Edmonton, Alberta, Canada

© 2002 Morse et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

The rejection of reliability and validity in qualitative inquiry in the 1980s has resulted in an interesting shift for "ensuring rigor" from the investigator's actions during the course of the research, to the reader or consumer of qualitative inquiry. The emphasis on strategies that are implemented during the research process has been replaced by strategies for evaluating trustworthiness and utility that are implemented once a study is completed. In this article, we argue that reliability and validity remain appropriate concepts for attaining rigor in qualitative research. We argue that qualitative researchers should reclaim responsibility for reliability and validity by implementing verification strategies integral and self-correcting during the conduct of inquiry itself. This ensures the attainment of rigor using strategies inherent within each qualitative design, and moves the responsibility for incorporating and maintaining reliability and validity from external reviewers' judgements to the investigators themselves. Finally, we make a plea for a return to terminology for ensuring rigor that is used by mainstream science.

Keywords: reliability; validity; verification; qualitative research

Acknowledgements: We acknowledge the support of the Council for Health Sciences, University of Alberta, and Alberta Heritage Foundation for Medical Research in preparation of this manuscript

Without rigor, research is worthless, becomes fiction, and loses its utility. Hence, a great deal of attention is applied to reliability and validity in all research methods. Challenges to rigor in qualitative inquiry interestingly paralleled the blossoming of statistical packages and the development of computing systems in quantitative research. Simultaneously, lacking the certainty of hard numbers and p values, qualitative inquiry expressed a crisis of confidence from both inside and outside the field. Rather than explicating how rigor was attained in qualitative inquiry, a number of leading qualitative researchers argued that reliability and validity were terms pertaining to the quantitative paradigm and were not pertinent to qualitative inquiry (Altheide & Johnson, 1998; Leininger, 1994). Some suggested adopting new criteria for determining reliability and validity, and hence ensuring rigor, in qualitative inquiry (Lincoln & Guba, 1985; Leininger, 1994; Rubin & Rubin, 1995).

In seminal work in the 1980s, Guba and Lincoln substituted reliability and validity with the parallel concept of "trustworthiness," containing four aspects: credibility, transferability, dependability, and confirmability. Within these were specific methodological strategies for demonstrating qualitative rigor, such as the audit trail, member checks when coding, categorizing, or confirming results with participants, peer debriefing, negative case analysis, structural corroboration, and referential material adequacy (Guba & Lincoln, 1981; Lincoln & Guba, 1985; Guba & Lincoln, 1982). Later, Guba and Lincoln developed authenticity criteria that were unique to the constructivist assumptions and that could be used to evaluate the quality of the research beyond the methodological dimensions (Guba & Lincoln, 1989). While Guba warned that their criteria were "primitive" (Guba, 1981, p. 90), and should be used as a set of guidelines rather than another orthodoxy (Guba & Lincoln, 1982), aspects of their criteria have, in fact, been fundamental to development of standards used to evaluate the quality of qualitative inquiry.

Thus, over the past two decades, reliability and validity have been subtly replaced by criteria and standards for evaluation of the overall significance, relevance, impact, and utility of completed research. Strategies to ensure rigor inherent in the research process itself were backstaged to these new criteria to the extent that, while they continue to be used, they are less likely to be valued or recognized as indices of rigor.

While researchers have continued to use the terminology of reliability and validity in qualitative inquiry in Great Britain and Europe, those who do so in North America are a minority voice. These few authors argue that the broad and abstract concepts of reliability and validity can be applied to all research because the goal of finding plausible and credible outcome explanations is central to all research (Hammersley, 1992; Kuzel & Engel, 2001; Yin, 1994). We are concerned, nonetheless, that the focus on evaluation strategies that lie outside core research procedures results in a deemphasis on strategies built into each phase of the research strategies that can act as a self-correcting mechanism to ensure the quality of the project.

This is an important issue and must be seen as more than just a paradigm debate. We suggest that by focusing on strategies to establish trustworthiness (Guba and Lincoln's 1981 term for rigor¹) at the end of the study, rather than focusing on processes of verification during the study, the investigator runs the risk of missing serious threats to the reliability and validity until it is too late to correct them.

This shift from constructive (during the process) to evaluative (post hoc) procedures occurred subtly and incrementally. Now, there is often no distinction between procedures that determine validity in the course of inquiry and those that provide research outcomes with such credentials. The literature on validity has become muddled to the point of making it unrecognizable, as Wolcott notes: "Whatever validity is, I apparently 'have' or 'get' or 'satisfy' or 'demonstrate' or 'establish' it. . ." (Wolcott, 1990, p. 121). We are also concerned that by refusing to acknowledge the centrality of reliability and validity in qualitative methods, qualitative methodologists have inadvertently fostered the default notion that qualitative research must therefore be unreliable and invalid, lacking in rigor, and unscientific (Morse, 1999). For the past two decades, qualitative researchers have complained of difficulty in getting funding and difficulty in getting published, and of being ignored by policy makers and practitioners. We suggest qualitative findings are still not regarded as solid empirical research. The purpose of this article is to reestablish reliability and validity as appropriate to qualitative inquiry; to identify the problems created by post hoc assessments of qualitative research; to review general verification strategies in relation to qualitative research, and to discuss the implications of returning the responsibility for the attainment of reliability and validity to the investigator.

Reliability and Validity

Guba and Lincoln (1981) stated that while all research must have "truth value", "applicability", "consistency", and "neutrality" in order to be considered worthwhile, the nature of knowledge within the rationalistic (or quantitative) paradigm is different from the knowledge in naturalistic (qualitative) paradigm. Consequently, each paradigm requires paradigm-specific criteria for addressing "rigor" (the term most often used in the rationalistic paradigm) or "trustworthiness", their parallel term for qualitative "rigor". They noted that, within the rationalistic paradigm, the criteria to reach the goal of rigor are internal validity, external validity, reliability, and objectivity. On the other hand, they proposed that the criteria in the qualitative paradigm to ensure "trustworthiness" are credibility, fittingness, auditability, and confirmability (Guba & Lincoln, 1981). These criteria were quickly refined to credibility, transferability, dependability, and confirmability (Lincoln & Guba, 1985). They recommended specific strategies be used to attain trustworthiness such as negative cases, peer debriefing, prolonged engagement and persistent observation, audit trails and member checks. Also important were characteristics of the investigator, who must be responsive and adaptable to changing circumstances, holistic, having professional immediacy, sensitivity, and ability for clarification and summarization (Guba & Lincoln, 1981).

These authors were rapidly followed by others either using Guba and Lincoln's criteria (e.g., Sandelowski, 1986) or suggesting different labels to meet similar goals or criteria (see Whittemore, Chase, & Mandle, 2001). This resulted in a plethora of terms and criteria introduced for minute variations and situations in which rigor could be applied. Presently, this situation is confusing and has resulted in a deteriorating ability to actually discern rigor. Perhaps as a result of this lack of clarity, standards were introduced in the 1980's for the post hoc evaluation of qualitative inquiry (see Creswell, 1997; Frankel, 1999; Hammersley, 1992; Howe & Eisenhardt, 1990; Lincoln, 1995; Popay, Rogers & Williams, 1998; Thorne, 1997).

Standards

While standards are a comprehensive approach to evaluating the research as a whole, they remain primarily reliant on procedures or checks by reviewers to be used following completion of the research. They represent either a minimally accepted level or an unobtainable gold standard for the researcher in the field. Subsequent clashes between the "ideal" and the "real" in the attainment of each standard are sometimes unavoidable. Those who evaluate completed research often forget that decisions that greatly influence the quality of the finished product may have, of necessity, been made quickly in the field

without the privilege of knowing the overall research outcome or without being able to see the ramifications of such a decision. Using standards, therefore, is a judgement of the relative worth of the research applied on completion of the project at a time when it is too late to correct problems that result in a poor rating.

Problems with post-hoc evaluation

Using standards for the purpose of post-hoc evaluation is to determine the extent to which the reviewers have confidence in the researcher's competence in conducting research following established norms. Rigor is supported by tangible evidence using audit trails, member checks, memos, and so forth. If the evaluation is positive, one assumes that the study was rigorous. We challenge this assumption and suggest that these processes have little to do with the actual attainment of reliability and validity. Contrary to current practices, rigor does not rely on special procedures external to the research process itself. For example, audit trails may be kept as proof of the decisions made throughout the project, but they do little to identify the quality of those decisions, the rationale behind those decisions, or the responsiveness and sensitivity of the investigator to data. Of importance, an audit trail is of little use for identifying or justifying actual shortcomings that have impaired reliability and validity. Thus, they can neither be used to guide the research process nor to ensure an excellent product, but only to document the course of development of the completed analysis.

Further, although Guba and Lincoln (1981) described member checks as a continuous process during data analysis (for example, by asking participants about hypothetical situations) this has largely been interpreted and used by researchers as verification of the overall results with participants. While it is an attractive idea to return the results to the original participants for verification, it is actually not a verification strategy. In fact, several methodologists (Hammersley, 1992; Morse, 1998), including Guba and Lincoln (1981), have warned against the tendency to define verification in terms of whether readers, participants, or potential users of the research judge the analysis to be correct, stating that it is actually more often a threat to validity.

The problem of member checks is that, with the exception of case study research and some narrative inquiry, study results have been synthesized, decontextualized, and abstracted from (and across) individual participants, so there is no reason for individuals to be able to recognize themselves or their particular experiences (Morse, 1998; Sandelowski, 1993). Investigators who want to be responsive to the particular concerns of their participants may be forced to restrain their results to a more descriptive level in order to address participants' individual concerns. Therefore, member checks may actually invalidate the work of the researcher and keep the level of analysis inappropriately close to the data. The result is that there is presently no distinction between procedures that determine validity during the course of inquiry, and those that provide the research with such credentials on completion of the project (Wolcott, 1994).

Moreover, we suggest that the terms reliability and validity remain pertinent in qualitative inquiry and should be maintained. We are concerned that introducing parallel terminology and criteria marginalizes qualitative inquiry from mainstream science and scientific legitimacy. Morse (1999) argues that, rather than clarifying, the development of alternative criteria actually undermines the issue of rigor.

Compounding the problem of duplicate terminology is the trend to treat standards, goals, and criteria synonymously, and the criterion adopted by one qualitative researcher may be stated as a goal by another scholar. For example, Yin (1994) describes trustworthiness as a criterion to test the quality of research design, while Guba and Lincoln (1989) refer to it as a goal of the research. Later, researchers followed Guba and Lincoln's 1989 shift toward post hoc evaluation, developing criteria as standards for evaluating

the worth of a project or as evidence that rigor had been attended to in the research process (see, for example, Popay, Rogers & Williams, 1998).

We are concerned that, in the time since Guba and Lincoln developed their criteria for trustworthiness, there has been a tendency for qualitative researchers to focus on the tangible outcomes of the research (which can be cited at the end of a study) rather than demonstrating how verification strategies were used to shape and direct the research during its development. While strategies of trustworthiness may be useful in attempting to *evaluate* rigor, they do not in themselves *ensure* rigor. While standards are useful for *evaluating* relevance and utility, they do not in themselves *ensure* that the research will be relevant and useful.

It is time to reconsider the importance of verification strategies used by the researcher in the process of inquiry so that reliability and validity are actively attained, rather than proclaimed by external reviewers on the completion of the project. We argue that strategies for ensuring rigor must be built into the qualitative research process *per se*. These strategies include investigator responsiveness, methodological coherence, theoretical sampling and sampling adequacy, an active analytic stance, and saturation. These strategies, when used appropriately, force the researcher to correct both the direction of the analysis and the development of the study as necessary, thus ensuring reliability and validity of the completed project.

The Nature of Verification in Qualitative Research

Verification is the process of checking, confirming, making sure, and being certain. In qualitative research, verification refers to the mechanisms used during the process of research to incrementally contribute to ensuring reliability and validity and, thus, the rigor of a study. These mechanisms are woven into every step of the inquiry to construct a solid product (Creswell, 1997; Kvale, 1989) by identifying and correcting errors before they are built in to the developing model and before they subvert the analysis. If the principles of qualitative inquiry are followed, the analysis is self-correcting. In other words, qualitative research is iterative rather than linear, so that a good qualitative researcher moves back and forth between design and implementation to ensure congruence among question formulation, literature, recruitment, data collection strategies, and analysis. Data are systematically checked, focus is maintained, and the fit of data and the conceptual work of analysis and interpretation are monitored and confirmed constantly. Verification strategies help the researcher identify when to continue, stop or modify the research process in order to achieve reliability and validity and ensure rigor.

While much has been written about the use of these strategies in various methods, the literature has focused on "how to do" rather than the contribution that these strategies make in optimizing the research outcome. In actual fact, it is the analytical work of the investigator that underlies these strategies that ensures their effectiveness. For example, many research decisions may underlie the sampling selection, which requires responsiveness to the needs of developing variation, verification, and the developing theory.

Investigator Responsiveness

Research is only as good as the investigator. It is the researcher's creativity, sensitivity, flexibility and skill in using the verification strategies that determines the reliability and validity of the evolving study. For example, ongoing analysis results in the dynamic formulation of conjectures and questions that force purposive sampling. The researcher analyses the data, which would then determine future participant recruitment. Within the notions of categorization and saturation lie sampling strategies to ensure replication and confirmation.

Responsiveness of the investigator to whether or not the categorization scheme actually holds (and is kept), or appears thin and muddled (and the scheme is changed), influences the outcome. In this way, it is essential that the investigator remain open, use sensitivity, creativity and insight, and be willing to relinquish any ideas that are poorly supported regardless of the excitement and the potential that they first appear to provide. It is these investigator qualities or actions that produce social inquiry and are crucial to the attainment of optimal reliability and validity.

The lack of responsiveness of the investigator at all stages of the research process is the greatest hidden threat to validity and one that is poorly detected using post hoc criteria of "trustworthiness." Lack of responsiveness of the investigator may be due to lack of knowledge, overly adhering to instructions rather than listening to data, the inability to abstract, synthesize or move beyond the technicalities of data coding, working deductively (implicitly or explicitly) from previously held assumptions or a theoretical framework, or following instructions in a rote fashion rather than using them strategically in decision making.

Verification Strategies

Within the conduct of inquiry itself, verification strategies that ensure both reliability and validity of data are activities such as ensuring methodological coherence, sampling sufficiency, developing a dynamic relationship between sampling, data collection and analysis, thinking theoretically, and theory development². Each of these will be discussed briefly.

First, the aim of *methodological coherence* is to ensure congruence between the research question and the components of the method. The interdependence of qualitative research demands that the question match the method, which matches the data and the analytic procedures. As the research unfolds, the process may not be linear. Data may demand to be treated differently so that the question may have to be changed or methods modified. Sampling plans may be expanded or change course altogether. The fit of these components with data to meet the analytic goals must be coherent, with each verifying the previous component and the methodological assumptions as a whole.³

Second, the *sample must be appropriate*, consisting of participants who best represent or have knowledge of the research topic. This ensures efficient and effective saturation of categories, with optimal quality data and minimum dross. Sampling adequacy, evidenced by saturation and replication (Morse, 1991), means that sufficient data to account for all aspects of the phenomenon have been obtained. Seeking negative cases is essential, ensuring validity by indicating aspects of the developing analysis that are initially less than obvious. By definition, saturating data ensures replication in categories; replication verifies, and ensures comprehension and completeness.⁴

Third, *collecting and analyzing data concurrently* forms a mutual interaction between what is known and what one needs to know. This pacing and the iterative interaction between data and analysis (as discussed earlier) is the essence of attaining reliability and validity.

The fourth aspect is *thinking theoretically*. Ideas emerging from data are reconfirmed in new data; this gives rise to new ideas that, in turn, must be verified in data already collected. Thinking theoretically requires macro-micro perspectives, inching forward without making cognitive leaps, constantly checking and rechecking, and building a solid foundation.

Lastly, the aspect of *theory development* is to move with deliberation between a micro perspective of the data and a macro conceptual/theoretical understanding. In this way, theory is developed through two mechanisms: (1) as an outcome of the research process, rather than being adopted as a framework to

move the analysis along; and (2) as a template for comparison and further development of the theory. Valid theories are well developed and informed, they are comprehensive, logical, parsimonious, and consistent (see Glaser 1978; Morse 1997).⁵

Together, all of these verification strategies incrementally and interactively contribute to and build reliability and validity, thus ensuring rigor. Thus, the rigor of qualitative inquiry should thus be beyond question, beyond challenge, and provide pragmatic scientific evidence that must be integrated into our developing knowledge base.

Discussion

We challenge the prevailing notion that the danger of using the generic term "validity" is that a particular method, for example ethnography, will be derailed from its philosophical underpinnings (Hammersley, 1992). Our argument is based on the premise that the concepts of reliability and validity as overarching constructs can be appropriately used in all scientific paradigms because, as Kvale (1989) states, to validate is to investigate, to check, to question, and to theorize. All of these activities are integral components of qualitative inquiry that insure rigor. Whether quantitative or qualitative methods are used, rigor is a desired goal that is met through specific verification strategies. While different strategies are used for each paradigm, the term validity is the most pertinent term for these processes. We advocate a return to Guba's (1981) early writings before Guba and Lincoln (1981) substituted trustworthiness for the qualitative paradigm. While this term bridges both reliability and validity concepts, the criteria they suggest still do not apply to all qualitative methods. For instance, Guba and Lincoln's confirmability is not pertinent to phenomenology, nor for postmodern philosophies such as feminism and critical theory in which the investigator's experience becomes part of data, and which perceive reality as dynamic and changing.

We argue for a return to validity as a means for obtaining rigor through using techniques of verification. Verification takes into account the varying philosophical perspectives inherent in qualitative inquiry, thus, the strategies used will be specific to, and inherent in, each methodological approach. At the same time, the terminology remains consistent with science.

Refocusing the qualitative research process to verification strategies is not without profound implications. It will, for example, enhance researcher's responsiveness to data and constantly remind researchers to be proactive, and take responsibility for rigor.⁶ Student projects, although necessarily smaller in scope, must also be responsive to rigor. We are concerned that in order for projects to be manageable within the constraints of student time-frames, abilities and budgets, rigor is seriously undermined by the narrow delimiting of the topics. We recommend that major concepts be verified and others left "hypothetical", rather than the student working with incomplete, thin data sets.⁷ Such strategies will enable students to assume projects small in scope but with the depth required by qualitative inquiry and thereby gain the grounding experience necessary to become an excellent researcher. Attending to rigor throughout the research process will have important ramifications for qualitative inquiry. Rather than relegating rigor to one section of a post hoc reflection on the finished work (such as stating that an audit trail was maintained, that member checks were done, or that the researcher was "reflective"), verification and attention to rigor will be evident in the quality of the text. Excellent inquiry is stunning: the arguments are sophisticated in that they are complex yet elegant, focused yet profound, surprising yet obvious.

In summary, we need to refocus our agenda for ensuring rigor and place responsibility with the investigator rather than external judges of the completed product. We need to return to recognizing and trusting the strategies within qualitative inquiry that ensure rigor. For too long, we have used the wrong tactics to defend qualitative inquiry. It is time to attend to our own research and work toward finding

consensus in broader criteria, appreciating how it is attained within the qualitative project itself, using criteria and terminology that is used in mainstream science.

Regardless of the standard or criteria used to evaluate the goal of rigor, our problem remains the same: they are applied after the research is completed, and therefore are used to judge of quality. Standards and criteria applied at the end of the study cannot direct the research as it is conducted, and thus cannot be used pro-actively to manage threats to reliability and validity.

Notes

1. We acknowledge Guba and Lincoln's evolution in regard to quality issues from trustworthiness (Lincoln & Guba, 1981), through authenticity criteria of fairness, knowledge sharing (ontological and educative authenticity), and social action (catalytic and tactical authenticity) (Guba & Lincoln, 1989) to Lincoln's (1995) social action commitments (Creswell, 1997). However, their work on "trustworthiness" is still regarded as seminal and pertinent.
2. See Meadows and Morse, 2001 for a detailed description of these strategies.
3. Methodological coherence does not exist when generic qualitative inquiry is conducted (i.e. no specific method is used, but rather the investigator limits analysis to themeing or categorizing), or when the researcher violates the method they purport to use.
4. One of the most common mistakes is that new investigators saturate their participants (that is, repeatedly interviewing the same participants until nothing new emerges) rather than saturating data (that is, continuing bringing new participants into the study until the data set is complete and data replicates). Returning to interview key participants for second or third time is oriented toward eliciting data to expand the depth or address gaps in the emerging analysis while interviewing additional participants is for the purpose of increasing the scope, adequacy and appropriateness of the data.
5. Premature closure of analysis or the completion of the study should not be limited by the researcher's timeline or budget. A strategy to reduce this threat to validity (by completing a shallow or prematurely closed study) is to plan research programs in components. In this, the researcher designs studies to investigate specific aspects of the phenomena and to complete segments that will contribute to a larger theoretical model.
6. Verification strategies may be problematic in pilot studies where data are thin. Recall, however, that the purpose of pilot studies, if used in qualitative inquiry, is to refine data collection strategies rather than to formulate an analytic scheme or develop theory.
7. Of course, adequate mentoring of graduate students and new researchers by experienced qualitative investigators is critical for learning to think qualitatively and for verifying emerging ideas.

References

- Altheide, D., & Johnson, J. M. C. (1998). Criteria for assessing interpretive validity in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Collecting and interpreting qualitative materials*. (pp. 283-312). Thousand Oaks, CA: Sage.

- Creswell, J. W. (1997). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.
- Frankel, R. (1999). Standards of qualitative research. In B. F. Crabtree & W. L. Miller (Eds.) *Doing qualitative research* (2nd ed). (pp. 333-346). Thousand Oaks, CA: Sage.
- Glaser, B. G. (1978). *Theoretical sensitivity*. Mill Valley, CA: Sociology Press.
- Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries, *Educational Communication and Technology Journal*, 29 (2), 75-91.
- Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco, CA: Jossey-Bass.
- Guba, E. G., & Lincoln, Y. S. (1982). Epistemological and methodological bases of naturalistic inquiry. *Educational Communication and Technology Journal* 30 (4), 233-252.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Hammersley, M. (1992). *What's wrong with ethnography?* London: Routledge.
- Howe, K., & Eisenhardt, M. (1990). Standards for qualitative (and quantitative) research: A prolegomenon. *Educational Researcher*, 19 (4), 2-9.
- Kuzel, A. & Engel, J. (2001). Some pragmatic thought on evaluating qualitative health research.. In J. Morse, J. Swanson, & A. Kuzel (Eds.), *The Nature of Qualitative Evidence* (pp. 114-138). Thousand Oaks, CA: Sage.
- Kvale, S. (1989). *Issues of validity in qualitative research*. Lund, Sweden: Chartwell Bratt.
- Leininger, M. (1994). Evaluation criteria and critique of qualitative research studies. In J. M. Morse (Ed.), *Critical Issues in Qualitative Research Methods*. Newbury Park, CA: Sage.
- Lincoln, Y. S. (1995). Emerging criteria for quality in qualitative and interpretive research. *Qualitative Inquiry*, 1, 275-289.
- Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Meadows, L. & Morse, J. M. (2001). Constructing evidence within the qualitative project. In J. M. Morse, J. Swanson & A. Kuzel, (Eds.). *The nature of evidence in qualitative inquiry*. (pp. 187 — 202), Thousand oaks, CA: Sage.
- Morse, J. M. (1991). Strategies for sampling. In J. Morse (Ed.), *Qualitative nursing research: A contemporary dialogue* (Rev. Ed.). (pp. 117-131). Newbury Park, CA: Sage.
- Morse, J. M. (1997). "Perfectly healthy, but dead": The myth of inter-rater reliability. [Editorial] *Qualitative Health Research*, 7, 445-447.
- Morse, J. M. (1998). Validity by committee. [Editorial] *Qualitative Health Research*, 8, 443-445.

- Morse, J. M. (1999). Myth #93: Reliability and validity are not relevant to qualitative inquiry. *Qualitative Health Research*, 9, 717.
- Popay, J., Rogers, A. & Williams, G. (1998). Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research*, 8, 341-351.
- Rubin, H. J. & Rubin, I. S. (1995). *Qualitative interviewing: The art of hearing data*. Thousand Oaks, CA: Sage.
- Sandelowski, M. (1986). The problem of rigor in qualitative research. *Advances in Nursing Science*, 8 (3), 27 — 37.
- Sandelowski, M. (1993). Rigor or rigor mortis: The problem of rigor in qualitative research revisited. *Advances in Nursing Science*, 16 (2), 1— 8.
- Seale, C. (1999). *The quality of qualitative research*. London: Sage.
- Thorne, S. (1997). The art (and science) of critiquing qualitative research. In J. Morse (Ed.), *Completing a qualitative project: Details and dialogue*, (p. 117 — 132). Thousand Oaks, CA: Sage.
- Wolcott, H. (1990). On seeking—and rejecting—validity in qualitative research. In E. W. Eisner & A. Peshkin, (Eds.), *Qualitative Inquiry in Education: The Continuing Debate*. (pp. 121-152. New York: Teachers College Press.
- Wolcott, H. (1994). *Transforming qualitative data: Description analysis and interpretation*. Thousand Oaks, CA: Sage.
- Whittemore, R., Chase, S. K., & Mandle, C. L. (2001). Validity in qualitative research. *Qualitative Health Research*, 11,.117-132.
- Yin, R. K. (1994). Discovering the future of the case study method in evaluation research. *Evaluation Practice*, 15, 283-290.