

Evaluation of qualitative research

DOROTHY HORSBURGH BA, MEd, PhD, RCNT, RGN, RNT

Lecturer, Napier University, School of Acute and Continuing Care Nursing, Faculty of Health and Life Sciences, Edinburgh, UK

Accepted for publication 1 May 2002

Summary

- It is necessary to subject published research to critical scrutiny, in order to evaluate the robustness of the findings.
- The criteria used in evaluation require to be appropriate for the research paradigm, i.e. quantitative or qualitative. Quantitative conceptualizations of reliability and validity are unsuitable for evaluation of qualitative research as they were not devised for this purpose.
- The use of quantitative criteria to evaluate qualitative research may create the impression that the latter is not academically rigorous.
- Evaluation criteria which are specific to qualitative research require identification and application, in order to provide a formalized and rigorous approach to critical appraisal.
- A formalized framework for evaluation will help to ensure that the contribution of qualitative studies, with specific reference to health services research, receives optimum recognition.
- The work of a number of writers is used in this paper to examine the features which distinguish qualitative research and the following are discussed:
 - the need for researcher reflexivity;
 - the use of the 'first person' in academic work;
 - the context in which research takes place;
 - the selection of research participants;
 - the interpretation of participants' accounts;
 - the active acknowledgement of 'lay' knowledge;
 - researcher flexibility within the research process;
 - the generalizability of findings.

Correspondence to: Dorothy Horsburgh, Napier University, School of Acute and Continuing Care Nursing, Faculty of Health and Life Sciences, Canaan Lane Campus, 74 Canaan Lane, Edinburgh, EH9 2TB, UK (tel.: +44 0131 536 5600; e-mail: d.horsburgh@napier.ac.uk).

- It is concluded that academically rigorous criteria, which are appropriate for evaluation of qualitative research, exist and are available for use by practitioners and researchers.

Keywords: criteria for evaluation, critical appraisal, qualitative research.

Introduction

One criticism which has been levelled at qualitative approaches to research is that they allegedly lack the 'scientific' rigour and credibility associated with traditionally accepted quantitative methods, in which inquiry is assumed to occur within a value-free framework and which rely on the measurement and analysis of causal relationships between variables. The notion that quantitative approaches are objective, impartial and value-neutral has been questioned by a number of writers (for example, Mishler, 1990; Denzin & Lincoln, 1994; Guba & Lincoln, 1995; Coffey, 1999) and the belief that the criteria for evaluation of quantitative research (i.e. reliability and validity) are appropriate in an identical format for evaluation of qualitative research is similarly debatable (Strauss & Corbin, 1990; Webb, 1992; Sandelowski, 1993; Koch, 1994; Popay *et al.*, 1998; Morse, 1999a).

It is arguable, however, that, whilst the precise means used to evaluate research should differ according to the question which the study has posed and the approach employed to address it, the fundamental principles of any evaluation process are the same. The plausibility and trustworthiness of the researcher's account are assessed, as is the study's potential and/or actual relevance to current and future theory and practice. The purpose of this paper is to discuss some of the hallmarks of qualitative research and identify criteria which may aid evaluation of qualitative research studies.

REFLEXIVITY

Qualitative research usually operates from the premise that total detachment on the part of the researcher is unattainable (even if deemed desirable) and that the individual who carries out research comprises an integral component of the entire process and product, as opposed to being a disembodied bystander with the capacity to provide an 'uncontaminated' account. Reflexivity refers to active acknowledgement by the researcher that her/his own actions and decisions will inevitably impact upon the meaning and context of the experience under investigation. By means of reflexivity the researcher realizes that (s)he is an integral part of the world that (s)he studies and that

neutrality and detachment in relation to data collection, analysis and interpretation are impossible (Henwood & Pidgeon, 1993; Porter, 1993; Mason, 1996).

Reflexivity may be demonstrated by use of the 'first person' when the researcher describes the aspects of the research in which (s)he has had personal involvement. Use of the first person, and consequent acceptance of the self as integral to the research enterprise, does not suggest that the research, or conclusions drawn, lack rigour (Webb, 1992, 1996), which should be evidenced by clear explication of the research process and product. This should provide sufficient information for the reader to identify the foundations upon which the findings of the study have been grounded.

Daly (1997) points out that all facts are *interpreted* facts and that the constructs developed by researchers are second degree, i.e. they are interpretations made by social scientists of the constructs of the participants as delivered to the researcher (p. 349). Daly argues (with specific reference to grounded theory) that, '...The challenge is one of preserving participants' definitions of reality' (p. 350), whilst concomitantly developing a theory which transcends these. Theory is, by definition, imposed upon the interpreted data, the role of the researcher being, '...to organize, select and construct explanation' (p. 350). Whilst this may appear obvious, Daly points out that:

Although there is greater acknowledgement of the role that the self plays in the research process, the self is usually left out of the final theoretical product of the research endeavor...For all intents and purposes, the theory...is presented as the product of a disembodied intellect. (Daly, 1997, p. 351).

The reality, however, is that theories are a product of a specific time, place and person and will also be interpreted by those who are time and context-bound. Daly suggests that theory development is both an individual and a co-operative enterprise, the aim being to construct an account which remains true to the data from participants, but is also subject to analysis and interpretation by the researcher.

Guba & Lincoln (1995) similarly argue that findings are not facts *per se*, but are created via the interaction between the participant, the data, the researcher and the evaluator. As such, they are dependent upon the value systems of each party and the context within which they operate. This

point is also made by Sandelowski (1993), who states that two researchers faced with the same qualitative task will produce different accounts due to their individual philosophies and theoretical commitments.

Coffey (1999) argues that, whilst many qualitative researchers and the authors of methods textbooks identify the role of the researcher in the construction of the research process, they nonetheless tend to:

...confine the discussion of the personal and the emotional to particular aspects of the research process, rather than establishing them as pervasive to the whole enterprise. (Coffey, 1999, p. 3)

Coffey (1999) uses the works of Lofland & Lofland (1995), Hammersley & Atkinson (1995) and LeCompte & Preissle (1993) to illustrate her argument, on the grounds that these are widely accepted within the UK as being 'gold standard' methods texts. Coffey (1999) proposes that the idea conveyed within such texts is that the researcher's identity requires to be manipulated in order to facilitate the research process and product, but she argues that this approach omits identification of the effects which undertaking research has upon the identities, relationships and emotions of *all* participants (i.e. 'researcher' and 'researched'). Coffey's (1999) insistence on an explicit recognition that researchers, accompanied by their epistemological and conceptual baggage and concomitant life experiences, permeate and suffuse the entire research enterprise, represents a view which has gained increasing momentum in recent years.

Similarly, Sword (1999) writes that:

Although some would criticize the subjectivity that is inherent in interpretive work, no research is free of the biases, assumptions, and personality of the researcher. We cannot separate self from those activities in which we are intimately involved. (Sword, 1999, p. 277)

Given that the researcher is, in this view, intimately involved in both the process and product of the research enterprise, it is necessary for the reader to evaluate the extent to which an author identifies and explicates their involvement and its potential or actual effect upon the findings.

Popay *et al.* (1998) identify that one criterion of good qualitative research is the provision of sufficient detail to enable the reader, '...to interpret the meaning and context of what is being researched... and exposes the experience as a process' (p. 347). They also argue that, 'Given the involvement of the researcher in the research process, the question is not whether the data are biased, but to what extent has the researcher rendered transparent the processes by which data have been collected, analysed and presented' (p. 348).

Koch (1994) contends that, whilst readers may not share the author's interpretation, they should nonetheless be able to discern the means by which it has been reached. She suggests that, in order to achieve this, a decision (or 'audit') trail should be provided by researchers. This entails clear explication of theoretical, methodological and analytic decisions made throughout the study. Indeed, the study in its entirety should form an auditable document.

In order to assess the plausibility and trustworthiness of a study, then, it is necessary to audit the process by which the end product has been achieved. This enables the reader to determine whether the analytical comments, or claims, made by the researcher appear to be justifiable. One problem, however, is that submission of a paper for publication limits wordage and this necessitates summarization. In addition, whilst some aspects of a study may, in theory, be relatively transparent for auditing purposes, e.g. audiotapes of interviews, in practice, access would prove problematic, both for logistical reasons and those of confidentiality/anonymity.

The aspects of the research process which are even less available to the reader are the theoretical speculations which are posited and tested, both during and subsequent to, data collection. The rationale for allocation of data to particular codes and categories may also prove difficult to discern in a summarized account, although some authors may provide an example as an illustration of this process. It is also important that the researcher acknowledges their responsibility for editing and selection of material. Omissions may be of equal, or greater, significance to inclusions, although by their nature are unlikely to be identifiable in publications!

THE ROLE OF CATEGORIES

Whilst the findings from a study are usually presented under the headings which describe the categories or themes which emerged, the aim being to facilitate the process and product of analysis and to provide clarity for the reader, there is a concern that this may produce a rather misleadingly clear-cut account. Daly (1997) points out that the production of logically consistent accounts of human experience and behaviour operates upon the assumption that logical consistency is inherent in the social processes that are described (p. 357). The reality, however, may be that complexity and contradiction are elemental to human existence and thus unavoidable. Imposition of a neat structure upon data has the potential to create order at the expense of accuracy and it may be misleading to view categories as discrete, self-sufficient entities, if appreciation of their

interdependence is a prerequisite to understanding of the overall context.

Criteria for evaluation of qualitative research

SUBJECTIVE MEANING

Popay *et al.* (1998, p. 341) identify three interrelated criteria, fulfilment of which comprise the hallmark of good qualitative health research. (Although they specify 'health' research, their comments are nonetheless relevant to qualitative work in general.) The first criterion is that there should be an *interpretation of subjective meaning* and within a study this would entail that the participants' accounts were used as the data base upon which all subsequent analysis and interpretation were firmly grounded.

Having acknowledged the importance of subjective meaning, it is nonetheless important to evaluate the reliance which may be placed upon specific sources of data, or on the approaches used in their collection. In recent years the use of narrative accounts as an appropriate data source for theory development has increased. Atkinson (1997) attributes this in part to the longstanding cultural emphasis within westernized societies upon the individual. Atkinson argues that narratives are of no higher standing than other forms of data and that their use within a research study requires to be integrated within a systematic approach. In some accounts of research studies which use narratives as data Atkinson (1997) points out that there is:

...an extraordinary absence of social context, social action, and social interaction [and]...remarkably little sense of how narratives are forged in face-to-face interaction or how they are elicited in given social contexts. (Atkinson, 1997, p. 339)

Atkinson argues that, if devoid of acknowledgement of their social context and with no attention accorded to their formal structure, such narratives provide no basis for social analysis. Researchers require to subject to scrutiny the means by which, within narratives, socially shared cultural resources are used in the production of comprehensible and convincing accounts of experience.

Similarly, Darbyshire (in Emden, 1997) says that:

...there is nothing sacred or sacrosanct about people's voices. They are data that is [sic] material for interpretation...we need...to bring our interpretive energies and creativity and insight to bear on this thing called the data. (Darbyshire, in Emden, 1997, p. 139)

Morse (1999b), writing in her role as editor of *Qualitative Health Research*, comments that some of the submissions

received by the journal provide minimal synthesis of data, let alone analysis, in the belief that the participants' voices will speak for themselves and that any interference by the researcher will, in some way, diminish their authenticity. In Morse's view, however, qualitative research must:

...add something more to the participants' words for it to be considered a research contribution, whether it be synthesis, interpretation, or development of a concept, model, or theory. (Morse, 1999b, p. 163)

PARTICIPANT 'VALIDATION'

In some qualitative research studies, the data and/or findings are returned to participants in order to obtain their validation. Such an approach is, however, problematic because the participants and the researcher will, to a greater or lesser degree, have different agendas and perspectives. A completed research study will, to some extent, conflate the accounts of all participants in order to obtain saturation of categories and should provide a shift from *substantive* situations to a more generalized and *theoretical* discussion of their underpinning concepts. Sandelowski (1993), for example, comments that, within her own work, she found that participants may have possessed a greater interest in their own substantive situation than in any abstract synthesis of their own accounts with those of others. Morse (1998) argues that, as theory is developed from a synthesis of the perspectives of a number of participants, it is inappropriate to expect that individual participants will have the ability to 'validate' the findings of the research study as a whole.

For the foregoing reasons, it may be inappropriate to return the findings from a study to participants for comment, in order to obtain participant 'validation'. This is not, of course, to suggest that, were the findings to be presented to the participants, they would prove unrecognizable, or incomprehensible. It would be expected that the account would elicit recognition by the participants as being applicable and relevant to their situations, experiences and perceptions. Within some interview settings participant validation may, indeed, be carried out, in that the interactive nature of an informal interview facilitates the asking of questions by the interviewer, in order to ensure that the individual participant's perception is identified and clarified.

DESCRIPTION OF CONTEXT

The foregoing comments about the need for *interpretation* of subjective meaning link with Popay *et al.*'s (1998)

second criterion for evaluation of qualitative health research, which is that there should be a description of the social context, for example, background information about the overall structures, settings and frameworks within which participants were situated. Active acknowledgement of the effect which these may have in facilitating, or inhibiting participants' actions is required, in order to place the data which is obtained from them within a wider context.

LAY KNOWLEDGE

Popay *et al.*'s (1998) third criterion for evaluation of qualitative research is that there is attention to *lay* knowledge and that, within a study, the participants, own perspectives are accorded an equivalent importance to those of 'experts'. For example, in a study in which patients were participants, their accounts would be considered equal in value to those of medical and nursing staff, who might be considered to possess 'expert' knowledge in the area under investigation.

Popay *et al.* (1998) argue that one key question is whether the research, as reported, illuminates, '...the subjective meaning, actions, and context of those being researched' (p. 345). If the lay knowledge possessed by the participants is fully acknowledged, then this *emic*, as opposed to *etic*, information has the potential to, '...address key policy issues that have eluded the efforts of many policy makers' (p. 345). For example, health service users are in a good position from which to identify the strengths and shortcomings of health service provision. Individuals who smoke are well-placed to provide insights into the lack of success of health promotion initiatives and may help to identify more useful approaches. Recent policy within the National Health Service has placed an increasing emphasis upon ensuring that the perceptions of the users of services are obtained and actively used to shape future service provision.

FLEXIBILITY

Rather than rigid adherence to a specific research design, Popay *et al.* (1998) argue that, '...the hallmark of good qualitative methodology is its variability, rather than its standardization. Thus, there should be some evidence of adaption and redesign in the writing up of research' (p. 346). In qualitative research, the collection and analysis of data proceed concomitantly throughout a study. The analysis of one set of data thus drives the research agenda in the selection of subsequent areas for investigation and the selection of participants who are best placed to provide

relevant data. Presentation of research should provide the rationale for all decisions made and a discussion of the effect that these may have had upon the enterprise as a whole.

SAMPLING

In relation to sampling, Popay *et al.* (1998) state that, in qualitative work, '...randomness and representativeness are of less concern than relevance...Does the sample produce the type of knowledge necessary to understand the structure and processes within which the individuals or situations are located?' (p. 346). Thus, initial sampling decisions should be purposive, in that selection of participants is made on the basis of their ability to provide relevant data on the area under investigation. Analysis of the data thus obtained should indicate the future direction that sampling should take, in order to develop a theory from data analysis. This, then, entails 'theoretical sampling' (Glaser, 1978).

GENERALIZABILITY

Whilst it is often argued that generalizability is not the purpose of qualitative research, Morse (1999a) states that if qualitative research is not considered to be generalizable, then it is arguably of little use (and is unlikely to be funded). She compares quantitative and qualitative approaches, in order to differentiate between the concept of generalizability as applicable to each. In quantitative work generalizability is statistical, i.e. the study sample is matched to the study population at large to ensure comparability of demographic characteristics and, if this is done correctly, then it is assumed that the findings from the sample are generalizable. In qualitative work, however, participants are selected by means of theoretical sampling, i.e. for their ability to provide information (and consequent theory development) about the area under investigation. Situational, rather than demographic, representativeness is what is sought. It may be said, then, that generalizability in qualitative research refers to the extent to which *theory* developed within one study may be *exported* (K.M. Melia, personal communication) to provide explanatory theory for the experiences of other individuals who are in comparable *situations*. This position is supported by the comments of Popay *et al.* (1998) on the subject of generalizability, who emphasize that, '...the aim is to make logical generalizations to a theoretical understanding of a similar class of phenomena rather than probabilistic generalizations to a population' (p. 348).

Conclusion

Application of quantitative conceptualizations of reliability and validity is inappropriate in evaluation of qualitative research, as the purpose and focus of the paradigms are not directly comparable. Use of quantitative criteria for a purpose for which they were not devised, and for which they are unsuited, has the potential to create the impression that qualitative research does not comprise an academically rigorous approach, at least in comparison to quantitative methods. It is, however, equally inappropriate to assume that criteria which are suitable for evaluation of qualitative research are unachievable, or unavailable. It may be seen, from examination of the work of Popay *et al.* (1998), that academically rigorous criteria are available, and accessible, for evaluation of qualitative research.

References

- Atkinson P. (1997) Narrative turn or blind alley? *Qualitative Health Research* 7(3), 325–344.
- Coffey A. (1999) *The Ethnographic Self: Fieldwork and the Representation of Identity*. Sage, London.
- Daly K. (1997) Re-placing theory in ethnography: a postmodern view. *Qualitative Inquiry* 3(3), 343–365.
- Denzin N.K. & Lincoln Y.S. (1994) *Handbook of Qualitative Research*. Sage, London.
- Emden C. (1997) A conversation with Margarete Sandelowski and Philip Darbyshire: issues in qualitative inquiry. *Nursing Inquiry* 4, 138–141.
- Glaser B.G. (1978) *Theoretical Sensitivity*. The Sociology Press, Mill Valley, CA, USA.
- Guba E.G. & Lincoln Y.S. (1995) *Fourth Generation Evaluation*. Sage, Newbury Park, CA, USA.
- Hammersley M. & Atkinson P.A. (1995) *Ethnography: Principles in Practice*, 2nd edn. Routledge, London.
- Henwood K.L. & Pidgeon N.F. (1993) Qualitative research and psychological theorising. In *Social Research: Philosophy, Politics and Practice* (Hammersley M., ed.). Sage, London.
- Koch T. (1994) Establishing rigour in qualitative research: the decision trail. *Journal of Advanced Nursing* 19, 976–986.
- LeCompte M.D., Preissle J. (with Tesch R) (1993) *Ethnography and Qualitative Design in Educational Research*, 2nd edn. Academic Press, San Diego.
- Lofland J. & Lofland L.H. (1995) *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*, 3rd edn. Wadsworth, Belmont, CA, USA.
- Mason J. (1996) *Qualitative Researching*. Sage, London.
- Mishler E.G. (1990) Validation in inquiry-guided research: the role of exemplars in narrative studies. *Harvard Educational Review* 60(4), 415–426.
- Morse J.M. (1998) Validity by committee. *Qualitative Health Research* 8(4), 443–445.
- Morse J.M. (1999a) Qualitative generalizability. *Qualitative Health Research* 9(1), 5–6.
- Morse J.M. (1999b) Silent debates in qualitative inquiry. *Qualitative Health Research* 9(2), 163–165.
- 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(3), 341–351.
- Porter S. (1993) Nursing research conventions: objectivity or obfuscation? *Journal of Advanced Nursing* 18, 137–143.
- Sandelowski M. (1993) Rigor or rigor mortis: the problem of rigor in qualitative research revisited. *Advances in Nursing Science* 16(2), 1–8.
- Strauss A.L. & Corbin J. (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Sage, Newbury Park, CA, USA.
- Sword W. (1999) Accounting for presence of self: reflections on doing qualitative research. *Qualitative Health Research* 9(2), 270–278.
- Webb C. (1992) The use of the first person in academic writing: objectivity, language and gatekeeping. *Journal of Advanced Nursing* 17, 747–752.
- Webb C. (1996) Caring, curing, coping: towards an integrated model. *Journal of Advanced Nursing* 23, 960–968.