

Special Article

Validity and Reliability within Qualitative Research in the Caring Sciences

Phil Coleman, FSfET, M Ed, BSc (Hons), Dip N, Cert Ed, RMN

Staff Tutor/Senior Lecturer, School of Health, Wellbeing & Social Care, Faculty of Wellbeing, Education & Language Studies, The Open University, Walton Hall, Milton Keynes. MK7 6AA. United Kingdom.

Correspondence: Phil Coleman, FSfET, M Ed, BSc (Hons), Dip N, Cert Ed, RMN Staff Tutor/Senior Lecturer, School of Health, Wellbeing & Social Care, Faculty of Wellbeing, Education & Language Studies, The Open University, Walton Hall, Milton Keynes. MK7 6AA. United Kingdom.

Abstract

This paper identifies the considerable potential for well-designed, effectively implemented, and clearly described qualitative research to enhance understanding and service provision within the caring disciplines. Nevertheless, it also recognises the ever-present risk that bias presents when undertaking any investigation, not least a qualitative research study, and the need to ensure practitioners can have confidence in the results, conclusions, and recommendations of such work. Validity and reliability are defined from a qualitative research perspective and various techniques described which can be utilised to help ensure investigative rigour. As one of the most common qualitative data collection methods, the application of tools to promote validity and reliability in interviews is given particular emphasis. Finally, the article asserts that presenting a qualitative research study with a demonstrably robust design and implementation strategy contributes to the ongoing efforts to raise the status of qualitative research as a legitimate and important approach to developing the body of knowledge on which effective care practice should be founded.

Key words: Validity, Reliability, Qualitative Research, Interviews, Bias, Care

Introduction

Much research within the caring disciplines requires the use of qualitative methods, often semi-structured or unstructured interviews, to generate the rich data required to better understand care practices and experiences. Indeed, interviews are one of the most widely used data collection methods in qualitative research (Alvesson 2003, Silverman 2013, Bryman 2016). If such studies are to make a meaningful contribution to the body of knowledge which underpins practice, and ultimately improve care provision, they must demonstrate sufficient rigour to allow readers to have confidence in their conclusions. Noble and Smith (2015, p.35), however, comment that *'there is no universally accepted terminology and criteria used to evaluate qualitative research'* and Reed (2009) recognises that defining quality in interviews is particularly problematic.

Although validity and reliability have traditionally been associated with quantitative studies and, historically, were not used to

evaluate qualitative research (Tatano Beck 2009), they are now also being applied to such studies (Anderson 2010). There are, nevertheless, important differences in the operational definition of these concepts and their application within such investigations. From the perspective of qualitative research, both validity and reliability are broadly concerned with the issue of trustworthiness (Mischler 1990, Stiles 1993); validity referring to the *'correctness or credibility of a description, conclusion, explanation, interpretation, or other sort of account'* (Maxwell 2010, p.280) and reliability to the *'application and appropriateness of the methods undertaken and the integrity of the final conclusions'* (Noble & Smith 2015, p.34). Arksey and Knight (1999, p.55) comment that *'in a nutshell, the qualitative response to the issue of reliability and validity is to require researchers to demonstrate that what they do is fit for their research purpose'*. As in all social science investigations, bias inevitably exists in qualitative research (Smith & Noble 2014) since it is impossible to ever completely control or remove all social influences (Ryan 2019).

Indeed, within qualitative studies threats to validity are described as legion (Arksey & Knight 1999), including 'distortion by investigators', participants', and readers' expectations and values' (Stiles 1999, p.613). Since Maxwell (2010, p.279) claims that 'a lack of attention to validity threats is a common reason for the rejection of research proposals', it is of utmost importance that qualitative researchers clearly present and articulate a well-considered strategy for ensuring validity and reliability within such documents. This paper considers the issues of validity and reliability within qualitative research and in particular their application to interview-based studies; briefly describing a series of tools which can be utilised to help ensure rigour within these aspects of study design.

Validity

Mechanical recording and 'rich' data: Mears (2017, p.187) argues that 'the validity of interview research is related to its appropriateness for studying what it claims to inform and its veracity in reporting'. Although no method or procedure can guarantee validity, various tools can greatly assist in the reduction of validity threats and increase the credibility of the conclusions reached within a research study (Maxwell 2010) including mechanical recording, 'rich' data, use of contradictory evidence, member checking, respondent validation, quasi-statistics, neutrality, triangulation, and fair dealing (Arksey & Knight 1999, Mays & Pope 2000, Anderson 2010, Bisman 2010, Maxwell 2010, Birt et al 2016, Gray 2018). Use of audio or video recording devices, rather than researcher notes, allow raw data to be scrutinised (Gray 2018); whilst the production of verbatim interview transcripts instead of selective interviewer notes, termed 'rich' data, provides a deeper and more revealing picture (Arksey & Knight 1999, Maxwell 2010). Qualitative researchers intending to use interviews as a data collection method are therefore advised to capture interviews on a digital audio recording device (commonly achievable via use of a low-cost or free application on a smartphone) and either transcribe these recordings in full or, where financially possible, have this work carried out by an independent professional transcription service.

Use of contradictory evidence or deviant cases: Within a qualitative study, data can be inappropriately discounted (Gray 2018), so

'identifying and analysing discrepant data and negative cases is a key part of the logic of validity testing in qualitative research' (Maxwell 2010, p. 284). This approach, sometimes referred to as 'contradictory evidence' or 'deviant cases', requires the researcher to seek out, examine and account for all data which might otherwise be deemed to challenge their conclusions and in so doing reduce the risk that an investigator merely sets aside such findings to strengthen their argument (Anderson 2010, Smith & Noble 2014). Such contradictory data is a common occurrence within interviews and so researchers are urged to capture, present, and explain its occurrence within their study.

Member checking and respondent validation:

Validity may also be substantiated by 'member checking', which involves the researcher informally confirming the accuracy of their understanding with participants during the data collection process (Gray 2018). Researchers can implement member checking in interviews by echoing, paraphrasing, and seeking further clarification on respondent comments where these are ambiguous and, in so doing, allow the interviewees an opportunity to confirm or correct the interviewer's interpretation of their words. They should also be alert to the tone and emphasis within both respondents' speech and their own utterances (Gray 2018, Rutakumwa et al 2020) and, congruent with the advice of Bonello (2001), be continually aware of the extent to which the verbal and non-verbal communication of each respondent appears harmonious and, therefore, potentially demonstrates an authentic response. Beuving and de Vries (2015, p.44) suggest that an overall test of validity related to the findings from a qualitative research study can also be undertaken at the end of the process by sharing the report and providing an opportunity for participant feedback. Nevertheless, they stress that respondent agreement with the findings does not, in isolation, demonstrate validity and that, similarly, a rejection of the results by respondents may highlight an unpalatable truth rather than an inaccurate conclusion. Whilst recognising these limitations, it does indeed still appear both desirable and advisable for researchers to implement this technique upon completion of their study. A more intensive form of member checking, termed 'respondent validation', provides an opportunity for interviewees to later comment on and revise their transcribed interview record (Anderson 2010,

Birt et al 2016). Implementing respondent validation in a study can, however, greatly increase the activity burden on respondents; a potential problem in all instances where research time and resources are limited. Moreover, researchers should also be aware of arguments which may discourage use of this approach; not least that interviewees can have a partial and restricted view of a topic (Torrance 2012) and may, having examined a transcript, suggest their responses have been misunderstood and should be revised merely to present themselves or their organisation in a more favourable light (Alvesson 2003, Miliades 2008, Cohen et al 2011).

Quasi-statistics: Maxwell (2010, p.285) argues that *'many of the conclusions of qualitative studies have an implicit quantitative component'*, for example the prevalence of a given phenomenon within a setting or population. The use of simple descriptive numerical data, termed *'quasi-statistics'*, is therefore presented as a valuable supplementary form of evidence to promote validity in a predominantly qualitative investigation. Capturing the frequency of issues, preferences or other phenomena raised by respondents within a study may therefore be a further way in which the research could be made more robust.

Neutrality: Neutrality is described as *'a requirement that the researcher considers their own role in the research'* and the aim *'is not to try to standardize researchers, but to have them reflect on the ways in which their background (class, gender, race, special concerns), personality (which is critical to achieving rapport and trust), mind set (assumptions and preconceptions), and actions have contributed to their account'* (Arksey & Knight 1999, p.55). Whilst actual neutrality may be an unachievable goal (Diebel 2008), striving for neutrality and making such reflection explicit in a research report are deemed valuable activities which assist investigators to demonstrate rigour within their work (Bekhet & Zauszniewski 2012, Erlingsson & Brysiewicz 2013, Noble & Smith 2015). Qualitative researchers should therefore make a conscious effort to capture their thought processes and reflections associated with all aspects of the study in the report of their work.

Triangulation and fair dealing: Essentially, triangulation compares results from two or more different methods of data collection and/or two or more data sources; researchers looking for patterns of convergence which enable them to

formulate or corroborate an overall interpretation of the findings (Mays & Pope 2000, Beauving & de Vries 2015, Fusch et al 2018). This approach to increasing the rigour within a research study is based on the argument that *'no single method is likely to afford a comprehensive account of the phenomenon under investigation'* (Torrance 2012, p. 113) and that data from different sources may offer complementary perspectives on the same construct (Rolfe 2006, Scott 2007). Use of multiple methods, especially qualitative and quantitative approaches, may therefore provide an additional opportunity to demonstrate confirmation and completeness (McEvoy & Richards 2006, Bekhet & Zauszniewski 2012). Indeed, Barbour (2001, p.1117) argues that the *'heavy reliance on triangulation in grant applications testifies both to the respect accorded to this concept and to its perceived value in demonstrating rigour'*. Gathering qualitative data from different relevant stakeholder groups (for example, care recipients, their families and care providers) may not only provide triangulation but also ensure *'the research design explicitly incorporates a wide range of different perspectives so that the viewpoint of one group is never presented as if it represents the sole truth about any situation'*; an approach termed *'fair dealing'* (Mays & Pope 2000, p.51) or *'truth value'* (Arksey & Knight 1999).

Reliability

Within a qualitative investigation, reliability is sometimes referred to as *'dependability'* (Rolfe 2006, Erlingsson & Brysiewicz 2013), *'confirmability'* (Jensen 2008) or *'consistency'* (Arksey & Knight 1999). Demonstrating reliability within a qualitative study is challenging because, unlike quantitative research, there are no available statistical tests for this purpose (Sutton & Austin 2015). Usefully, triangulation may also provide a more compelling argument regarding the reliability of findings (Brannen 2005, McEvoy & Richards 2006, Hesse-Biber 2010, Gray 2018) since consistent findings from different data sources or collection methods are also likely to better evidence the integrity of the research conclusions.

Detail and transparency: Within interviews, reliability may be enhanced by greater control of, and uniformity within, the interview process but by imposing such structure, validity can be adversely affected as the interaction inevitably

becomes more stilted and less friendly; ultimately increasing the risk that participants feel inhibited and become less likely to share full, frank, and accurate accounts of their experiences (Alvesson 2003, Cohen et al 2011). A key element of reliability within such research is that *'the researcher shows how the research has been done and decisions have been made, so that the reader could conduct an 'audit trail', examining the good sense and plausibility of the researcher's thought and actions'* (Arksey & Knight 1999, p.54). Transparency and detailed description of the rationale for the research design and its implementation (Elo & Kyngas 2008, Fitzgerald & Dopson 2011) should therefore afford an opportunity for the reader of a study to better evaluate its reliability.

Multiple coding: Another widely advocated tool to promote reliability in qualitative research is the use of *'multiple coding'*; also referred to as *'peer review'*, *'consistency checks'* or *'intercoder reliability'* (Barbour 2001, Thomas 2006, Burnard et al 2008, Vaismoradi et al 2013, Smith & Noble 2014, Gray 2018). This approach, which Barbour (2001) describes as the equivalent of *'inter-rater reliability'* within a quantitative study, involves one or more additional qualitative researchers independently analysing study data (Ryan & Bernard 2003) and although there is *'a debate as to whether qualitative researchers should have their analyses verified or validated by a third party'*, it is also *'argued that this process can make the analysis more rigorous and reduce the element bias'* (Burnard et al 2008, p.431). It is therefore recommended that codes, or at least preliminary themes, are examined by a second disinterested party with some understanding of the topic of interest or field of practice. Where respondent comments are not deemed to be relevant or mutually exclusive to a particular code or theme by this third party, revisions should be made to the findings based on this feedback and such changes recorded within the report.

Replicability: Bisman (2010) asserts that one of the main criteria for judging the quality of a research study is replicability. Stiles (1993, p.602) refers to such replicability as *'procedural trustworthiness'* and notes that this *'concerns whether the observations are repeatable (after allowing for contextual differences) and whether the investigator's report conveys what you would have seen if you had been observing'*, so is deemed a key consideration in determining reliability within a qualitative research study.

Clearly, audio recordings and full transcriptions offer considerable opportunity to establish procedural trustworthiness when an interview has been used as a data collection method, providing a further argument for the use of such techniques to enhance reliability within this form of research study.

Conclusion: Although qualitative research can make a unique and critical contribution to better understanding organisation and practice within the care disciplines and in so doing improve both services and the service user experience, it is vital that such studies are designed and implemented in a way that enables practitioners to have confidence in their results. Historically, qualitative data has been regarded merely as supplementary to quantitative data (Hesse-Biber 2010) and wholly qualitative studies judged to be an inferior form of investigation (Hammersley 2001, Victor 2008, Edgley et al 2016, Corry et al 2018). Qualitative researchers must therefore demonstrate carefully considered application of appropriate tools to evidence the validity and reliability of their work. Arguably, the absence of such evidence in many instances has perpetuated the view that qualitative studies are *'less scientific'*. Whilst Rolfe (2006) argues that responsibility for appraising a qualitative research report must ultimately rest with the reader rather than the writer, he also stresses that researchers should strive to ensure the approach they adopt to their investigations is systematic, rigorous, clearly described, appropriately justified, and exhibits a robust design. Application of the techniques to promote validity and reliability described in this paper may greatly assist qualitative investigators to meet these criteria. Effective use of such techniques also makes a small, but important, contribution towards changing longstanding negative views in respect of the quality of qualitative research. Furthermore, every rigorous qualitative research study helps enhance recognition that such work makes an important contribution to the development of a more comprehensive body of empirical knowledge needed to underpin effective practice within the caring disciplines.

References

- Alvesson, M. (2003) Methodology for close up studies – struggling with closeness and closure, Higher Education, 46(2):167-193.
- Anderson, C. (2010) Presenting and Evaluating Qualitative Research, Am. J. Pharm. Educ. 74(8) Article 141:1-7.

- Arksey, H. & Knight, P. (1999) *Interviewing for Social Scientists: An Introductory Resource with Examples*. London, Sage.
- Barbour, R. (2001) Checklists for Improving Rigour in Qualitative Research: A Case of the Tail Wagging the Dog? *British Medical Journal*, 322(7294):1115-1117.
- Bekhet, A. & Zauszniewski, J. (2012) Methodological triangulation: an approach to understanding data, *Nurse Researcher*, 20(2), pp.40-43.
- Beuving, J. & de Vries, G. (2015) *Doing Qualitative Research. The Craft of Naturalistic Inquiry*. Amsterdam University Press.
- Birt, L., Scott, S., Cavers, D., Campbell, C. & Walter, F. (2016) Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation? *Qualitative Health Research*, 26(13):1802-1811.
- Bisman, J. (2010) Postpositivism and Accounting Research: A (Personal) Primer on Critical Realism, *Australasian Accounting Business and Finance Journal*, 4(4):3-25.
- Bonello, M. (2001) Perceptions of fieldwork education in Malta: Challenges and Opportunities, *Occupational Therapy International*, 8(1):17-33.
- Brannen, J. (2005) *Mixed Methods Research: A discussion paper*. Swindon, ESRC National Centre for Research Methods.
- Bryman, A. (2016) *Social Research Methods*. 5th Edition. Oxford University Press.
- Burnard, P., Gill, P., Stewart, K., Treasure, E. & Chadwick, B. (2008) Analysing and presenting qualitative data, *British Dental Journal*, 204(8):429-432.
- Cohen, L., Manion, L. & Morrison, K. (2011) *Research Methods in Education*. Abingdon, Routledge.
- Corry, M., Porter, S. & McKenna, H. (2018) The redundancy of positivism as a paradigm for nursing Research, *Nursing Philosophy*, 20, e12230:1-10.
- Diebel, A. (2008) Neutrality in Qualitative Research in: Given, L. (ed.) *The Sage Encyclopedia of Qualitative Research Methods*. Thousand Oaks, Sage.
- Edgley, A., Stickley, T., Timmons, S. & Meal, A. (2016) Critical realist review: exploring the real, beyond the empirical, *J. Furth. High. Educ.*, 40(3):316-330.
- Elo, S. & Kyngas, H. (2008) The qualitative content analysis process, *J Adv Nur*, 62(1):107-115.
- Erlingsson, C. & Brysiewicz, P. (2013) Orientation among multiple truths: An introduction to qualitative research, *AFEM*, 3:92-99.
- Fitzgerald, L. & Dopson, S. (2011) Comparative Case Designs: Their Utility and Development in Organisational Research in: Buchman, D. & Bryman, A. (eds.) *The Sage Handbook of Organizational Research Methods*. Paperback Edition. London, Sage.
- Fusch, P., Fusch, G. & Ness, L. (2018) Denzin's Paradigm Shift: Revisiting Triangulation in Qualitative Research, *Journal of Social Change*, 10(1):19-32.
- Gray, D. (2018) *Doing Research in the Real World*. 4th Edition. London, Sage.
- Hammersley, M. (2001) On 'Systematic' Reviews of Research Literatures: a 'narrative' response to Evans & Benefield, *Br Edu Res J*, 27(5):543-554.
- Hesse-Biber, S. (2010) Qualitative Approaches to Mixed Methods Practice, *Qualitative Inquiry*, 16(6):455-468.
- Maxwell, J. (2010) Chapter 17. Validity. How might you be wrong? in: Luttrell, W. (ed), *Qualitative Educational Research*. Abingdon, Routledge.
- Mays, N. & Pope, C. (2000) Assessing quality in qualitative research, *British Medical Journal*, 320:50-52.
- McEvoy, P., & Richards, D. (2006) A critical realist rationale for using a combination of quantitative and qualitative methods, *J Res Nurs*, 11(1):66-78.
- Mears C. (2017) Chapter 21: In-depth interviews in: Coe, R., Waring M., Hedges, L. and Arthur, J. (eds), *Research Methods & Methodologies in Education*. 2nd Edition. London, Sage.
- Miltiades, H. (2008) Interview as a Social Event: Cultural Influences Experienced While Interviewing Older Adults in India, *International Journal of Social Research Methodology*, 11(4):277-291.
- Mischler, E. (1990) Validation in Inquiry-Guided Research: The Role of Exemplars in Narrative Studies, *Harvard Educational Review*, 60(4):415-442.
- Noble, H. & Smith, J. (2015) Issues of validity and reliability in qualitative research, *Evid Based Nurs*, 18(2):34-35.
- Reed, M. (2009) Chapter 25. Critical Realism: Philosophy, Method, or Philosophy in Search of a Method? in: Buchanan, D. & Bryman, A. (eds) *The SAGE Handbook of Organizational Research Methods*, London, Sage.
- Rolfé, G. (2006) Validity, trustworthiness and rigour: quality and the idea of qualitative research, *J Adv Nur*, 53(3):304-310.
- Rutakumwa, R., Mugisha, O., Bernays, S., Kabunga, E., Tumwekwase, G., Mbonye, M. & Seeley, J. (2020) Conducting in-depth interviews with and without voice recorders: a comparative Analysis, *Qualitative Research*, 20(5):565-581.
- Ryan, G. (2019) Postpositivist critical realism: philosophy, methodology and method for nursing research, *Nurse Researcher*, 27(3):20-26.
- Ryan, G. & Bernard, H. (2003) Techniques to Identify Themes. *Field Methods*, 15(1):85-109.
- Scott, D. (2007) Resolving the quantitative-qualitative dilemma: a critical realist approach, *International Journal of Research & Method in Education*, 30(1):3-17.
- Silverman, D. (2013) What Counts as Qualitative Research? Some Cautionary Comments, *Qual. Sociol. Rev.*, 9(2):48-55.
- Smith, J. & Noble, H. (2014) Bias in research, *Evidence Based Nursing*, 17(4):100-101.
- Stiles, W. (1993) Quality Control in Quality Research, *Clinical Psychology Review*, 13:593-618.
- Sutton, J. & Austin, Z. (2015) *Qualitative Research: Data Collection, Analysis, and Management*, *Canadian Journal of Hospital Pharmacy*, 68(3):226-231.
- Tatano Beck, C. (2009) Critiquing Qualitative Research, *Archives of Research in Nursing Journal*, 90(4):543-554.
- Thomas, D. (2006) A General Inductive Approach for Analyzing Qualitative Evaluation Data, *American Journal of Evaluation*, 27(2):237-246.
- Torrance, H. (2012) Triangulation, Respondent Validation and Democratic Participation in Mixed Methods Research. *J. Mix. Methods Res.*, 6(2):111-123.
- Vaismoradi, M., Turunen, H. and Bondas, T. (2013) Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing and Health Sciences*, 15:398-405.
- Victor, L. (2008) *Systematic reviewing*. University of Surrey Social Research Update, Issue 54:1-4.