Quality in Qualitative Evaluation:
A framework for assessing research evidence

Liz Spencer, Jane Ritchie, Jane Lewis and Lucy Dillon
National Centre for Social Research

Other reports in the Occasional Papers series:

Morris, St., Greenberg, D., Riccio, J., Mittra, B., Green, H., Lissenburgh, S. and Blundell, R. (2003) *Designing a Demonstration Project – An Employment, Retention and Advancement Demonstration for Great Britain*, Government Chief Social Researcher's Office Occasional Papers Series No. 1., London: Cabinet Office.

CONTENTS

Preface

The Government's commitment to evidence-based policy is matched by its drive to develop excellence in government research and evaluation. To this end the Cabinet Office is working with Government departments to ensure that government research and evaluation is of the highest standard.

Qualitative techniques are used extensively in government social research so the Strategy Unit in the Cabinet Office commissioned the National Centre for Social Research to undertake a methodological review of quality standards in qualitative evaluation methods. The aim of this project was to produce guidance for undertaking research and evaluation that uses qualitative methods. The quality framework that has been produced by the National Centre for Social Research is an important step forward in ensuring that government policy is informed by robust, valid and appropriate research evidence.

We hope the framework will support work in Departments to deliver high-quality research and evaluation. The framework provides a useful and useable guide for assessing the credibility, rigour and relevance of individual research studies. We hope that over time a body of high-quality research evidence in qualitative evaluation will be assembled through the application of this framework.

This report and framework also contributes to the ongoing debate in the wider social research and academic community on the quality of all types of social research. The report aims to stimulate further debate and developmental work on these issues.

This project supports the wider activities of the Government Chief Social Researcher in promoting high-quality social research in Departments to support the development, implementation, review and evaluation of policy.

Government Chief Social Researcher
Prime Minister's Strategy Unit
Cabinet Office

Acknowledgements

The authors would like to express their sincere thanks to all those who contributed to this study:

Mr Richard Bartholomew, Department of Education and Skills
Professor Mildred Blaxter, University of East Anglia
Professor Alan Bryman, Loughborough University
Professor Martin Bulmer, University of Surrey
Mr George Clarke, Department for Work and Pensions
Ms Naomi Eistenstadt, Department of Education and Skills
Professor Ray Fitzpatrick, Nuffield College, Oxford
Ms Joan Garcia, Department for Work and Pensions
Professor Martyn Hammersley, Open University, Milton Keynes
Ms Angela Harden, Institute of Education
Dr Anthony Johnson, Department for Work and Pensions
Dr Janet Lewis, Joseph Rowntree Foundation
Professor Andrew Long, University of Salford
Dr Gordon Marshall, Economic and Social Research Council
Dr Jennifer Mason, University of Leeds
Professor Don Nutbeam, Department of Health
Professor Ann Oakley, Institute of Education
Dr Ray Pawson, University of London
Professor Roisin Pill, University of Wales College of Medicine
Professor Jennie Popay, University of Lancaster
Ms Shareen Sadiq, Home Office
Dr Roy Sainsbury, University of York
Professor Clive Seale, Goldsmiths College
Professor David Silverman, Goldsmiths College
Mr Elliott Stern, Tavistock Institute
Ms Jane Sweeting, Department for Work and Pensions
Dr Andrew Thomas, BMRB International
Dr Anthony Tomei, Nuffield Foundation
Professor Robert Walker, University of Nottingham
Ms Jan White, Office of the Deputy Prime Minister
Professor Paul Wiles, Home Office
Dr Sandra Williams, Department of Health
Dr Sharon Witherspoon, Nuffield Foundation
Mr David Woodward, Office of the Deputy Prime Minister

Sincere thanks also to Phil Davies, Rebecca Stanley and Annette King from the Strategy Unit in the Cabinet Office for their advice and support throughout the study.

EXECUTIVE SUMMARY

Background (Chapter 1)

This report presents the findings of a study carried out by a team of researchers based at the National Centre for Social Research, on behalf of the Strategy Unit in the Cabinet Office. The objective of the study was to develop a framework which would guide assessments of the quality of qualitative research evaluations. The study grew out the Cabinet Office's responsibility for ensuring excellence in government research and evaluation, and was a response to the fact that, despite their growing use, there are no explicitly agreed standards regarding what constitutes quality in qualitative research evaluations.

The study involved a number of elements:

- a comprehensive and systematic review of the research literature relating to standards in qualitative research
- a review of the qualitative research methods used in government-funded evaluation studies
- a review of existing frameworks for reviewing quality in qualitative research
- twenty-nine in-depth interviews with government-based commissioners and managers of research, government-based policy-makers, other funders of evaluation research (research councils and foundations), academics and practitioners involved in conducting qualitative research and writing about quality
- a workshop involving the above groups to refine the framework initially developed
- the trial application of the framework to a small number of studies to help to develop and refine it.

The framework (Chapters 2 and 7)

The framework is based around:

- **four guiding principles** – that research should be

    - **contributory** in advancing wider knowledge or understanding
    - **defensible in design** by providing a research strategy which can address the evaluation questions posed

- **rigorous in conduct** through the systematic and transparent collection, analysis and interpretation of qualitative data
- **credible in claim** through offering well-founded and plausible arguments about the significance of the data generated

- **eighteen appraisal questions:**

    1. How credible are the findings?
    2. How has knowledge or understanding been extended by the research?
    3. How well does the evaluation address its original aims and purpose?
    4. How well is the scope for drawing wider inference explained?
    5. How clear is the basis of evaluative appraisal?
    6. How defensible is the research design?
    7. How well defended are the sample design/target selection of cases/documents?
    8. How well is the eventual sample composition and coverage described?
    9. How well was the data collection carried out?
    10. How well has the approach to and formulation of analysis been conveyed?
    11. How well are the contexts of data sources retained and portrayed?
    12. How well has diversity of perspective and content been explored?
    13. How well has detail, depth and complexity (i.e. richness) of the data been conveyed?
    14. How clear are the links between data, interpretation and conclusions – i.e. how well can the route to any conclusions be seen?
    15. How clear and coherent is the reporting?
    16. How clear are the assumptions/theoretical perspectives/values that have shaped the form and output of the evaluation?
    17. What evidence is there of attention to ethical issues?
    18. How adequately has the research process been documented?

- **quality indicators:**

    - for each appraisal question, a series of possible features for consideration in the assessment of quality are proposed.

Scope and applicability of the framework (Chapter 4)

Key aspects of the scope and applicability of the framework:

- The framework has been designed to be applied to appraisals of the outputs of qualitative evaluations (reports, papers and journal articles), although it would also have some relevance to assessments of proposals or the conduct of a study.

- Although the four guiding principles would be relevant to any qualitative research method (and in many cases to quantitative research too), the appraisal

questions and quality indicators are designed to focus on four methods: in-depth interviews, focus groups, observation and documentary analysis.

- The framework is designed to aid the informed judgement of quality, but not to be prescriptive or to encourage the mechanistic following of rules. The questions are phrased as open-ended questions to reflect the fact that appraisals of quality must allow judgement, and that standards are inevitably shaped by the context and purpose of assessment.

- The framework is not intended to apply to the full range of traditions or paradigms in qualitative research, but equally it is not intended to be aligned with any individual specific schools or approaches. Instead, some key philosophical assumptions that are within or outside the scope of the framework are described.

- The framework is designed to be applied to a wide range of types of qualitative evaluation (including contextual research, which examines the world in which a policy or service operates; policy review and development; practice evaluation and appraisals of specific schemes or interventions). But the quality of the qualitative research on which an evaluation is based is seen as lying at the heart of assessments of quality.

- Most of the items included in the framework are heavily recurrent in the wider literature, in existing frameworks and among our interviews with research commissioners and managers, policy-makers, funders, academics and practitioners. They have also been selected to offer a series of readings on the guiding principles and to cover all the stages and processes involved in qualitative research.

Foundations of the framework (Chapters 3, 5 and 6)

The framework draws on and reflects a number of important and extensive debates relating to quality in qualitative research.

- There are different views about whether evaluation is an activity that is distinct from research and a number of differentiating features are proposed. There are also different definitions of qualitative research and different views about how it contributes to policy evaluation.

- There is debate about whether it is feasible or desirable to establish quality criteria for qualitative research and how far these are different to criteria for quantitative research. Positions range from a rejection of the notion of criteria altogether, to the identification of criteria or principles developed specifically to qualitative research, to the retention of concepts drawn from quantitative research.

- There are debates about what is meant by criteria, and about whether the particular philosophical assumptions and methods of qualitative research make criteria more problematic.

- Conceptions of quality are influenced by the various philosophical assumptions which underpin different approaches to qualitative research. These epistemological and ontological positions are diverse and span issues such as the nature of reality, the relationship between the researcher and the researched, the relationship between facts and values, the nature of knowledge, and appropriate methods of research.

- Within varying conceptions of quality, the notions of validity, reliability, objectivity and generalisation are often given key importance. The meanings attributed to these concepts, and how people view their applicability or otherwise to qualitative research, varies extensively.

- The research involved a review of 29 existing frameworks for assessing the quality of qualitative research. Many were developed in the fields of medical or health services research. They were developed for different purposes, including the assessment of written outputs, reviewing proposals or framing the conduct of research. Authors generally stress the development and permissive nature of their frameworks. There is much variety between the frameworks as to

  – how far their philosophical orientation is specified
  – how far their applicability to different qualitative methods is specified
  – the level of specification of criteria
  – the length, format and coverage of the frameworks.

- There are several common features, particularly around the need for clarity in aims and objectives; appropriate use of qualitative methods; appropriate sample design; clarity about the analytical process and clarity about how the evidence and conclusions are derived.

Concluding comments about the framework (Chapter 8)

- The framework draws heavily on existing frameworks, on the wider literature and on the contribution of those who participated in the study. It is, however, developmental, and it is recognised that there will be alternative views about its structure, focus and content. What is critical is that the framework is applied flexibly, and not rigidly or prescriptively. Judgement and discretion lie at the heart of quality appraisal, and assessments of quality will vary between different contexts and purposes.

- The study highlights two further points. First, assessing quality requires a degree of expertise in the conduct and use of research, and there is interest in more support, education and guidance for non-research experts in their use of qualitative research. Second, there is a need to consider a creative extension of the range of qualitative research methods used in government evaluations.

PART 1: THE FRAMEWORK

1   INTRODUCTION

1.1   Background and objectives

Qualitative research methods are widely used in policy evaluation. There is an increasingly rich understanding of how they contribute to policy formulation, evaluation and refinement, and of what they bring to policy analysis, both alone and in conjunction with quantitative methods. However, there are no explicitly agreed standards regarding what constitutes quality in qualitative policy evaluation methods, and no agreed formal criteria for judging the quality of qualitative evaluation research.

Within qualitative research generally, and particularly among those working in applied research, especially health, there has been a growing emphasis in the last decade or so on ways of formalising quality standards, and a large number of sets of guidelines, frameworks and checklists of criteria or considerations for assessing quality have been produced. (These are discussed in detail in Chapter 6 of this report.) Initiatives, such as the Campbell Collaboration, have formulated standards for distinguishing between high- and low-quality experimental and quasi-experimental studies. The growing use of systematic reviews has focused attention further on quality standards, and on how qualitative research can be incorporated alongside quantitative research in their formalisation. A number of recent initiatives have looked at quality standards in qualitative research, or in research more generally.[1] However, these developments have not focused specifically on how quality can be understood and judged in qualitative evaluation research.

The Strategy Unit in the Cabinet Office therefore commissioned a team of researchers based at the National Centre for Social Research to carry out a review of quality standards in qualitative evaluation methods, and to recommend a set of standards for assessing quality in that context. The Cabinet Office is responsible for encouraging excellence in government research and evaluation and for ensuring that government has access to up-to-date advice and information on evaluation methodology. The study grew out of a particular need to ensure that researchers and policy-makers in government have the tools they need to use qualitative research evidence with confidence. It was intended to address the need for agreed standards, both for judging *whether a particular qualitative study has been undertaken to the highest standards of inquiry*, and over the longer term for *appraising the existing research literature* so that qualitative studies that are of high quality can be distinguished from those that are not.

---

[1] These include initiatives being coordinated by the Health Development Agency, the Social Care Institute for Excellence (SCIE), the Evidence Network and the Campbell Collaboration.

Specifically, the objectives of the study, as set out by the Cabinet Office in the competitive tendering process, were to:

- identify the range of qualitative evaluation methods used in policy evaluation
- review relevant literature on qualitative methods and identify quality standards against which qualitative evaluation methods can be assessed
- produce a set of criteria that researchers and policy-makers can use to assess the extent to which a particular study demonstrates attention to key quality issues
- provide guidance on how standards can be used in appraising individual studies
- create a comprehensive bibliography of literature on quality standards.

1.2 Research strategy for developing the framework

The study was commissioned in February 2002. It was agreed between the research team and the Cabinet Office that the study would involve a number of stages. Each of these stages, and the activities involved, contributed to the development and refinement of the framework presented in this report. The first two stages of the project – a review of the literature and in-depth interviews – provided the information upon which an initial framework was developed. Subsequently, a workshop was held in which some of those who had taken part in the depth interviews, or their close colleagues, were asked to comment on and help us develop our draft framework. Having amended the framework based on the outcome of this workshop, it was then applied to a selection of research reports and articles and further refined. The following sections describe in more detail each of these stages.

1.2.1 Literature review

A comprehensive, systematic and broad-ranging review of the research literature relating to standards in qualitative research was carried out, running throughout most of the study. While the emphasis of the study was on evaluation and evaluation literature was of key importance, the scope of the review was broader and incorporated qualitative social research more generally. In addition, existing frameworks and checklists for quality criteria, and documents that specifically considered quality and quality standards were looked for and reviewed. In all, 29 frameworks were identified and our review examined their origins, purposes, form and content. Overall, the review of the literature sought to encapsulate views about the critical issues involved in assessing the quality in qualitative research, interpretations of key concepts, views about their relevance to qualitative research and to different traditions thereof, and views about the specific practices they require.

A number of different methods and sources were used in searching the literature. Electronic searches of The British Library Catalogue, PsychINFO, Web of Science, Caredata, BIDS and Medline were carried out using a number of keyword strings. This process produced over 5,000 hits. In addition, SIGLE, a grey-literature catalogue was

searched. The website publication lists for the Department for Work and Pensions (DWP), the Department for Education and Skills (DfES), the Department of Health (DoH), the Office of the Deputy Prime Minister (ODPM)[2] and the Home Office (HO) were also searched. However, the fact that discussion of methodological quality – beyond descriptions of methods used – was very limited in government research reports meant that, of around 600 reports checked, only five were considered relevant for inclusion in the review. Hand searches were also carried out of a range of journals and of the bibliographies from selected texts, including those of all the criteria and framework papers found. Using the inclusion and exclusion criteria which are shown in Appendix 1, the literature from all the different sources was sifted and an initial central bibliography compiled. This ran to around 800 references. The team then sifted this bibliography using a second, more evolved set of inclusion and exclusion criteria (see Appendix 1). This revised set of references ran to approximately 300 references.

Members of the research team then reviewed each document. This was done using an agreed template (see Appendix 2) to help prevent recording bias, to ensure a consistent approach across the team and across the course of the review and to facilitate synthesis of findings. Completed templates were stored centrally and were used directly to inform the subsequent stages of the study and framework development. While almost all the literature reviewed contributed to the development of the framework, not all is cited in this report. However, texts not cited are referenced in a separate bibliography at the end of this report.

An additional task carried out within the context of the literature review was an overview of government-funded evaluations using qualitative research methods, particularly identifying the specific methods they used. To ensure wide coverage of policy areas, publications of five government departments were included in this process: DWP, DoH, DfES, HO and ODPM. A search of the electronic publication listings on the relevant websites for all departments (other than ODPM, where hard copies of research summaries were provided to the team by the Strategy Unit in the Cabinet Office) was carried out in September/October 2002 (the search path followed is laid out in Appendix 3). A template was completed (see Appendix 3) for all evaluative reports purporting to have used qualitative methods. Depending on what information was available on the various websites, templates were completed based generally on abstracts or research summaries and, in a small number of cases, full reports. While attempts were made to ensure rigorous documentation of the methods used, this was limited by the lack of a detailed discussion or description of the methods in many of the research summaries or reports.

1.2.2   Depth interviews

Twenty-nine depth interviews were carried out in two waves (summer 2002 and autumn 2002), and all interviews were tape-recorded and fully transcribed for analysis. Four distinct groups of individuals were included in this sample: commissioners and managers of research within government departments; government-based policy-makers who use

---

[2] During the course of the study, the Department for Transport, Local Government and the Regions (DTLR) was restructured into the Office of the Deputy Prime Minister (ODPM) and the Department for Transport. In this report, the new title of ODPM is used.

qualitative research; funders of research from research councils and foundations; and finally a group of experts, academics and others involved in conducting qualitative research (for a detailed description of the interview process see Appendix 4). This final group included people who had themselves developed frameworks for assessing quality in qualitative research. Prior to the interviews, respondents received a letter of introduction (see Appendices 5 and 6) and a one page description of the study and its aims (see Appendix 7).

Topic guides were developed around the main aims of the study and were heavily informed by the information gathered in the first phase of the literature review (see Appendix 8). Interviews covered a range of topics related to the assessment of quality in qualitative research. Across the sample, respondents' views were explored in depth on the key contributions that qualitative research can make and what they saw as the main quality criteria for qualitative research. In particular, the team looked for respondents' views about how quality can be demonstrated and assessed. Interviews also investigated whether there were perceived to be any specific quality criteria for evaluations. Respondents' knowledge of and views about frameworks, their value and utility, and thoughts about their scope, structure, content and application were also explored. In addition, interviews with framework developers explored their experiences of developing the framework.

1.2.3 Developing the initial framework

Drawing on the review of the literature and the depth interviews, an initial draft of the framework was developed, in October 2002. At an early stage in its development, the team addressed a set of 13 questions based on the key issues around assessing quality arising from both the literature and the interviews. Under four broad themes, the questions addressed were:

Feasibility
1. Is it feasible to set quality criteria for qualitative research?
2. Is it feasible to frame these criteria in the form of a checklist, guidelines, quality framework, etc.?
3. Is it possible for the framework to cover all qualitative traditions?

The functions of the framework
1. What will the framework be used for?
2. Who will it be used by and in what circumstances?
3. What are the implications of our remit to examine quality standards for qualitative evaluations?

The philosophical base of the framework
1. What forms of criteria are appropriate for qualitative research?
2. What terminology/language should be used?

The content of the framework
1. What are the key areas that we need to cover in defining quality criteria?
2. Should we be deliberately selective about the criteria we include?
3. How are criteria to be assessed?
4. Should the framework have a hierarchical structure?
5. What specific criteria should we include?

Informed by both the literature consulted and interviews completed, the team discussed the above questions in depth. Stemming from these discussions formative decisions were taken on the scope, structure and content of a draft framework. The positions taken by the team in response to the questions above are presented throughout this report and demonstrated in the resulting framework.

1.2.4 Workshop

A workshop was held in November 2002 to explore reactions to the framework the team had developed to date and to identify aspects that needed further work. A total of 12 people – all people who had taken part in the earlier interviews, or their close colleagues – attended. At the workshop the research team presented the framework as it then stood, and participants discussed, in small groups and plenary sessions, the structure and content of the framework and issues which they felt required further development. Sessions were tape-recorded and detailed notes drawn up. The full programme is set out in Appendix 10. The discussion at the workshop contributed significantly to the further development of the framework (see further Chapter 7).

1.2.5 Applying the framework

Having revised the framework based on the outcome of the workshop, in early January 2003 the team applied the framework to several examples of evaluation reports. Testing, or applying the framework to 'real-life' research, was perceived by the team to be a crucial part of the process of developing the framework that would sharpen our design.

Eight documents were selected to which the framework would be applied. A number of criteria were developed that guided the selection of these documents:

- A broad definition of evaluation was adopted to include those evaluating specific policies and a small number of more contextual studies.
- A range of policy areas was covered.
- Studies using a range of qualitative methods (single- and multi-method studies) were included, as well as studies that also had a quantitative component.
- Research carried out by a range of authors was included, covering the fields of academia, commercial research and independent research institutes (including one report by NatCen).
- Both full reports and journal articles were included.

The focus of the exercise was on identifying areas where the content of the framework, or the phraseology used, needed to be modified. The process of applying the framework and the issues arising from this process led to further refinement of the framework (see further Chapter 7).

1.3 Structure of the report

This report provides a comprehensive account of the study. As well as presenting the framework, it aims to show how the framework was developed, to explain the decisions made about its scope, structure and content, and to show how it is based on the literature review, the review of existing frameworks, the interviews and the workshop.

The report is presented in two Parts. Part 1 focuses on the framework itself. Following this chapter's description of the study and its methods, Chapter 2 sets out the framework in full, with a brief summary of the key aspects of its focus, status and applicability, and guidance as to its purpose and use. This is the essential context of the framework. The aim of Part 2 is to display the material on which the framework is based, and to show how the framework was developed. We recognise that some readers will primarily be interested in the framework itself rather than in the material and debates upon which it is based. Chapter 2 therefore indicates where key elements of our approach are explored in detail in later chapters, but acts as a stand-alone description of the framework and our approach.

In Part 2, we begin our explanation of the foundations of the framework in Chapter 3 by exploring the broader debate about criteria and their application to qualitative policy evaluation. We look at the relationship between research and evaluation, describe the nature and contribution of qualitative research in evaluation, and explore different positions on the feasibility or desirability of establishing quality criteria in qualitative research. Chapter 4 shows how the literature review and interviews informed some early fundamental decisions about the scope of the framework – whether and how it could be applicable to different research paradigms and philosophical positions, to different types of evaluation, and to different qualitative research methods. Chapter 5 draws together the material from our literature review and interviews that shaped and informed the content of the framework. In particular it explores different concepts and components of quality which emerge from the literature, and comments on how they were discussed in our interviews with research managers, policy-makers, funders, academics and practitioners. It addresses the 'holy trinity' of validity, reliability and objectivity, and identifies other quality criteria and the practices and approaches which are seen as critical to them.

In Chapter 6 we turn to the existing frameworks for assessing the quality of qualitative research, which were also very influential on our decisions about the scope, structure and content of the framework shown in Chapter 2. The chapter describes key features of the approaches taken by other framework developers, looking at the origins and purposes of frameworks, how they relate to different philosophical positions and important distinctions in their structure and form. A detailed review of the content of frameworks is shown in Appendix 13.

Chapter 7 explains how our own framework was developed from these various resources and describes how its structure and content evolved. Finally, in Chapter 8 we briefly note some key issues relating to the application of the framework, and stress the need for thorough testing of it.

1.3.1 FRAMEWORK FOR ASSESSING QUALITATIVE EVALUATIONS

This chapter presents a framework for appraising the quality of qualitative evaluations. With minor modifications, the chapter will be reproduced as a 'stand-alone' document for those wishing to use the framework. It has therefore been written with this purpose in mind

2.1   Origins of the Framework

The framework was developed with particular reference to evaluations concerned with the development and implementation of social policy, programmes and practice. It was devised as part of a programme of research conducted on behalf of the Cabinet Office.

The research on which the framework was built involved:

- a review of literature on qualitative research methods
- a review of existing frameworks for assessing quality in qualitative research
- exploratory interviews with a range of people who have an interest in quality assessment of qualitative research and/or policy-related evaluations.

These three strands of activity have heavily influenced the content of the framework and the premises that surround its operation (*Chapters 3–7*).

The sections below describe the scope, application and coverage of the framework. Sections 2.2 and 2.3 provide important background information about the premises underpinning the framework and its intended usage. It is recommended that these sections are read by anyone using the framework for the first time. Those familiar with this background can move directly to Section 2.4 where the content of the framework is presented.

2.2 Scope of the framework

The framework has been built on certain premises surrounding the nature of qualitative enquiry, how it can be used for evaluative purposes and how its quality can be assessed. These premises derive from extensive debate in the literature on qualitative research and evaluation. A brief summary is given here but full documentation can be found in the referenced sections of this report.

For the purposes of this framework, the quality of the qualitative research that generates the evidence for an evaluation is seen as lying at the heart of any assessment. This is because of a primary interest in evaluations based on empirical enquiry, which form the majority of government-based evaluative investigations. Qualitative research and qualitative evaluation are therefore seen as broadly synonymous in terms of the principles surrounding quality assessment (*Chapter 3*). Because of this, the framework can also be used for qualitative research more generally, irrespective of whether it has an evaluative purpose.

2.2.1 Qualitative research

Qualitative research aims to provide an in depth understanding of people's experiences, perspectives and histories in the context of their personal circumstances or settings. Among many distinctive features, it is characterised by a concern with exploring phenomena from the perspective of those being studied; with the use of unstructured methods which are sensitive to the social context of the study; the capture of data which are detailed, rich and complex; a mainly inductive rather than deductive analytic process; developing explanations at the level of meaning or micro-social processes rather than context-free laws; and answering 'what is', 'how' and 'why' questions (*Chapter 3*). It employs a variety of methods, including: exploratory interviews; focus groups; observation; conversation, discourse and narrative analysis; and documentary and video analysis (*Chapters 3 and 4*).

Qualitative research is used in evaluation for a range of purposes (*Chapter 3*). For evaluations of programmes services or interventions, these include identifying the factors that contribute to successful or unsuccessful delivery; identifying outcomes (intended or unintended) and how they occur; examining the nature of requirements of different groups within the target population; exploring the contexts in which policies operate; and exploring organisational aspects of delivery. Qualitative research can also be used in advance of policy development or implementation, for example, to examine an issue or problem that is poorly understood or to inform the kind of intervention required. A further use is to help develop evaluative criteria where these are unclear or where alternative criteria are sought.

2.2.2   Quality standards in qualitative research

Despite many different approaches and traditions within qualitative and evaluative research, there are widespread concerns about quality. There is also shared interest in issues such as 'rigour'; the need for principles of practice to be made manifest; the importance of sound or 'robust' qualitative research evidence; and in the relevance and utility of research (*Chapter 5*).

2.2.3   The nature of quality assessment in qualitative research

There is debate in the literature about whether the concepts of quality used to assess qualitative research should be roughly the same as, parallel to, or quite different from those used to assess quantitative research (*Chapter 3*). This framework is based on the view that the concerns which lie behind customary conceptions of quality have relevance for qualitative enquiry but need to be *reformulated – and assessed quite differently –* within the domain of qualitative research. In other words, qualitative research should be assessed on its 'own terms' within premises that are central to its purpose, nature and conduct.

2.2.4   The formalisation of quality standards

There is some contention about the extent to which quality assessment of qualitative inquiry can be formalised (*Chapters 3 and 6*). Alongside this there have been increasing calls for guidance about quality assessment so that criteria appropriate to qualitative research (rather than those imported from other research traditions) are used. This, in turn, has led to the generation of a number of checklists, guidelines and lists of appraisal questions for assessing qualitative research (*Chapter 6*).

This framework draws heavily on previously developed quality criteria, both from the general methodological literature and from pre-existing guidelines. However, it also takes heed of the persistent concern that formalised criteria should avoid being rigidly procedural or over-prescriptive. It has therefore been devised to aid informed judgement, not mechanistic rule-following.

Because there has been an opportunity to build on other frameworks, it is comparatively comprehensive in its coverage. It is also distinctive in making explicit the philosophical assumptions within which it operates; and in the level of specificity of the quality 'indicators' on which to base an assessment.

2.2.5   Philosophical assumptions

There are numerous approaches, paradigms, schools and movements encompassed within what is broadly termed 'qualitative research'. They vary in terms of the ontological, epistemological and methodological assumptions on which they are based. The same issues arise in relation to different types of evaluation. A proliferation of approaches to evaluation exist, which, again, differ in terms of their philosophical,

ideological and methodological premises. In both cases, the various traditions are often categorised under labelled movements or schools although their identities are not always consistently described (*Chapter 4*).

It was not felt possible to design a single set of quality markers that would be appropriate for all the different qualitative and evaluative traditions that exist. The assumptions on which they are based are too diverse – in some cases contradictory – and the conception of what matters in terms of 'quality' can vary with these assumptions.

Instead, an 'elemental' approach has been used in which a range of premises about the nature of social reality and how the social world can be investigated have been identified and those which underpin the framework specified (*Chapter 4*). It is thought that these assumptions are broadly those within which most government-funded qualitative inquiry operates. Taking this approach means that the framework will be relevant for a range of types of qualitative evaluations including practice evaluation, policy development and appraisal as well as evaluations of particular interventions, schemes or programmes. It will be the assumptions within which an evaluation operates, rather than the model of qualitative research or evaluation used, that will determine whether the framework will be of value to assess its quality.

2.2.6   Choice of appraisal items

There are numerous appraisal items that could have been included in the framework (*Chapters 5, 6 and 7*). Those selected were chosen on the basis that:

- they operate within the boundaries of the philosophical assumptions prescribed
- they offer a series of readings on core principles concerning the contribution, defensibility, rigour and credibility of a qualitative study
- they cover different stages and processes within qualitative enquiry
- they are recurrently cited as markers of quality in the literature, in pre-existing frameworks and in the interviews conducted for this study.

2.3   Application of the framework

The framework has been designed primarily to assess the *outputs* of qualitative inquiry– that is reports, papers and journal articles. This has inevitably affected both the items covered and, most crucially, how they can be demonstrated for assessment. There is, for example, heavy emphasis on the quality of the evidence and the analysis that has informed it and rather less on aspects of conduct that cannot be well judged from written output.

Some translation of the criteria included can be made to assess proposals (see Section 2.4.5). It is also hoped that the framework will have a wider educational function in the

preparation of research protocols, the conduct and management of research and evaluation and in the training of social researchers, both within and outside government.

2.3.1　Relevance for different qualitative methods

The framework has been designed with a particular focus on the methods used most extensively in government-based evaluations, namely, interviews, focus groups, observation and documentary analysis (*Chapter 4*). Nevertheless, the principles that have guided the design of the framework, and many of the questions suggested for appraisal, will have application to a wider range of qualitative methods (e.g. linguistic analysis, historical and archival analysis, multimedia methods etc.). However, quality indicators that are particularly relevant to other methods will need to be added.

2.3.2　Use of multiple methods

It is common to find that evaluations use a mix of methods to collect the required information. Sometimes these are a combination of qualitative methods but in other cases a mix of qualitative and quantitative approaches. Although some of the appraisal questions in the framework are method generic (e.g. Qs 1–3, 5, 6 etc.) others are specific to qualitative methods (e.g. Qs 11–13). It is therefore recommended that to assess any quantitative components of an evaluation, quality criteria that are specific to the method concerned will need to be added.

2.3.3　Application to different policy domains

The framework has been designed to assess qualitative evaluation across the spectrum of substantive fields covered by government departments. However, it could well be that assessment within a particular substantive area might need to address some additional questions. If so, these can be added to the framework as necessary.

2.4　Content of the framework

The framework is built on a set of **principles** around which to frame and structure **appraisal questions** that might be asked of a piece of work in order to critically assess its quality. In each case, a set of **quality indicators** is listed – features that will help to form a judgement about how well the appraisal question has been addressed. Further questions might also be added depending on the purpose of the research and the approach it uses. It is then for the assessor to judge overall merit, based on the questions and indicators that are most relevant to the evaluation concerned.

2.4.1 Guiding principles

There are four[3] central principles that underpin the content of the framework. All of these are based on themes that are highly recurrent in the literature and in the interviews conducted (*Chapters 5, 6 and 7*). They advise that research should be:

- ***contributory*** in advancing wider knowledge or understanding about policy, practice, theory or a particular substantive field
- ***defensible in design*** by providing a research strategy that can address the evaluative questions posed
- ***rigorous in conduct*** through the systematic and transparent collection, analysis and interpretation of qualitative data
- ***credible in claim*** through offering well-founded and plausible arguments about the significance of the evidence generated.

These principles are presented at a sufficiently high level of abstraction that they would apply to a diversity of qualitative approaches. Indeed, most of them are simply emblems of sound and logical enquiry, whatever its form or purpose.

2.4.2 Appraisal questions

The guiding principles have been used to identify 18 appraisal questions to aid an assessment. The questions are listed in column a) of the framework. Between them, they cover all of the key features and processes involved in qualitative enquiry. They begin with assessment of the **findings**, move through different stages of the research process (**design**, **sampling**, **data collection**, **analysis** and **reporting**) and end with some general features of research conduct (**reflexivity and neutrality**, **ethics and auditability**).

It is suggested that the findings of the enquiry are given attention first, even though this is not a logical procedural order. This is because the nature of the evidence presented will help in assessing features of the research process (for example, the quality of the data collected, the visibility and logic of the analytic process). However, if readers prefer to look at research design and conduct before considering the evidence, they will need to return to Qs 6–18 before completing the assessment.

2.4.3 Quality indicators

Beside each question, there is a series of quality indictors (column b)) which will help in answering the appraisal question. These provide pointers to the kinds of information needed to judge whether or not the quality feature concerned has been secured. They are not intended to be comprehensive and other indicators might well be added for specific studies. Equally, they are not intended to suggest essential requirements – it is highly unlikely that all these indicators will be present, or even relevant, in any one study.

---

[3] Two of the four principles deal with the way in which a study has been devised and conducted. This is not surprising since, in qualitative research, there are no 'validated' instruments or standardised methods, which means that quality cannot be *assumed* on the basis that certain methods have been used, but must be *demonstrated* in each case.

Many of the quality indicators relate only to the methods specified in Section 2.3.1 (i.e. interviews, focus groups, observation and documentary analysis). For most appraisal questions, however, quality indicators that are relevant to other methods could be added.

2.4.4 The need for professional judgement

The assessment of a qualitative inquiry, using this framework, will require careful judgements on the part of the assessor. These, in turn, will require some knowledge of qualitative research and some expertise in using qualitative methods. Judgement will also be needed in deciding the weight to attach to particular indicators in order to assess its 'fitness for purpose' – that is, how well it addresses the objectives for which it was undertaken. For example, in a study carried out to evaluate the implementation of a new scheme, it may well be more important to have a detailed account of how practice has affected outcomes, or an accessibly written report, than to have a thorough literature review.

2.4.5 Use for assessing proposals

As was noted in Section 2.3, the framework has been designed to assess *outputs* from qualitative inquiry. However, selected questions and indicators from the framework could be used to assess *proposals* for designing and conducting a qualitative evaluation (see particularly Qs 2, 5, 6, 7, 9, 10 and 16–18). This, of course, will require changing from past to present the tense within which questions are phrased. It is important to stress, however, that other questions not included in this framework, will also be relevant to proposals (for example, feasibility, timescale, resources, experience of research team). This framework is therefore *not* intended as a comprehensive aid for proposal assessment.

FRAMEWORK FOR ASSESSING QUALITATIVE EVALUATIONS

Study being appraised: ....

	a) Appraisal questions	b) Quality indicators (possible features for consideration)	c) Notes on study being appraised
1 **F I N D I N G S**	How credible are the findings?	Findings/conclusions are supported by data/study evidence (*i.e. the reader can see how the researcher arrived at his/her conclusions; the 'building blocks' of analysis and interpretation are evident*) Findings/conclusions 'make sense'/have a coherent logic Findings/conclusions are resonant with other knowledge and experience (*this might include peer or member review*) Use of corroborating evidence to support or refine findings (i.e. other data sources have been used to examine phenomena; other research evidence has been evaluated: see also Q14)	
2 **F I N D I N G S**	How has knowledge/ understanding been extended by the research?	Literature review (where appropriate) summarising knowledge to date/key issues raised by previous research Aims and design of study set in the context of existing knowledge/understanding; identifies new areas for investigation (*for example, in relation to policy/practice/substantive theory*) Credible/clear discussion of how findings have contributed to knowledge and understanding (*e.g. of the policy, programme or theory being reviewed*); might be applied to new policy developments, practice or theory Findings presented or conceptualised in way that offers new insights/alternative ways of thinking Discussion of limitations of evidence and what remains unknown/unclear or what further information/research is needed	

	a) Appraisal questions	b) Quality indicators (possible features for consideration)	c) Notes on study being appraised
3 **F I N D I N G S**	How well does the evaluation address its original aims and purpose?	Clear statement of study aims and objectives; reasons for any changes in objectives Findings clearly linked to the purposes of the study – and to the initiative or policy being studied Summary or conclusions directed towards aims of study Discussion of limitations of study in meeting aims (*e.g. are there limitations because of restricted access to study settings or participants, gaps in the sample coverage, missed or unresolved areas of questioning; incomplete analysis; time constraints?*)	
4 **F I N D I N G S**	Scope for drawing wider inference – how well is this explained?	Discussion of what can be generalised to wider population from which sample is drawn/case selection has been made Detailed description of the contexts in which the study was conducted to allow applicability to other settings/contextual generalities to be assessed Discussion of how hypotheses/propositions/findings may relate to wider theory; consideration of rival explanations Evidence supplied to support claims for wider inference (*either from study or from corroborating sources*) Discussion of limitations on drawing wider inference (*e.g. re-examination of sample and any missing constituencies: analysis of restrictions of study settings for drawing wider inference*)	
5 **F I N D I N G S**	How clear is the basis of evaluative appraisal?	Discussion of how assessments of effectiveness/evaluative judgements have been reached (*i.e. whose judgements are they and on what basis have they been reached?*) Description of any formalised appraisal criteria used, when generated and how and by whom they have been applied Discussion of the nature and source of any divergence in evaluative appraisals Discussion of any unintended consequences of intervention, their impact and why they arose	

	a) Appraisal questions	b) Quality indicators (possible features for consideration)	c) Notes on study being appraised
6 **D E S I G N**	How defensible is the research design?	Discussion of how overall research strategy was designed to meets aims of study Discussion of rationale for study design Convincing argument for different features of research design (*e.g. reasons given for different components or stages of research; purpose of particular methods or data sources, multiple methods, time frames etc.*) Use of different features of design/data sources evident in findings presented Discussion of limitations of research design and their implications for the study evidence	
7 **S A M P L E**	How well defended is the sample design/target selection of cases/documents?	Description of study locations/areas and how and why chosen Description of population of interest and how sample selection relates to it (*e.g. typical, extreme case, diverse constituencies etc.*) Rationale for basis of selection of target sample/settings/documents (*e.g. characteristics/features of target sample/settings/documents, basis for inclusions and exclusions, discussion of sample size/number of cases/setting selected etc.*) Discussion of how sample/selections allowed required comparisons to be made	
8 **S A M P L E**	Sample composition/case inclusion – how well is the eventual coverage described?	Detailed profile of achieved sample/case coverage Maximising inclusion (*e.g. language matching or translation; specialised recruitment; organised transport for group attendance*) Discussion of any missing coverage in achieved samples/cases and implications for study evidence (*e.g. through comparison of target and achieved samples, comparison with population etc.*) Documentation of reasons for non-participation among sample approached/non-inclusion of selected cases/documents Discussion of access and methods of approach and how these might have affected participation/coverage	

	a) Appraisal questions	b) Quality indicators (possible features for consideration)	c) Notes on study being appraised
9 D A T A C O L L E C T I O N	How well was the data collection carried out?	Discussion of: • who conducted data collection • procedures/documents used for collection/recording • checks on origin/status/authorship of documents Audio or video recording of interviews/discussions/conversations (*if not recorded, were justifiable reasons given?*) Description of conventions for taking fieldnotes (*e.g. to identify what form of observations were required/to distinguish description from researcher commentary/analysis*) Discussion of how fieldwork methods or settings may have influenced data collected Demonstration, through portrayal and use of data, that depth, detail and richness were achieved in collection	
10 A N A L Y S I S	How well has the approach to and formulation of the analysis been conveyed?	Description of form of original data (*e.g. use of verbatim transcripts, observation or interview notes, documents, etc.*) Clear rationale for choice of data management method/tool/package Evidence of how <u>descriptive</u> analytic categories, classes, labels etc. have been generated and used (*i.e. either through explicit discussion or portrayal in the commentary*) Discussion, with examples, of how any <u>constructed</u> analytic concepts/typologies etc. have been devised and applied	
11 A N A L Y S I S	Contexts of data sources – how well are they retained and portrayed?	Description of background or historical developments and social/organisational characteristics of study sites or settings Participants' perspectives/observations placed in personal context (*e.g. use of case studies/vignettes/individual profiles, textual extracts annotated with details of contributors*) Explanation of origins/history of written documents Use of data management methods that preserve context (*i.e. facilitate within case description and analysis*)	

	a) Appraisal questions	b) Quality indicators (possible features for consideration)	c) Notes on study being appraised
12 **A N A L Y S I S**	How well has diversity of perspective and content been explored?	Discussion of contribution of sample design/case selection in generating diversity Description and illumination of diversity/multiple perspectives/alternative positions in the evidence displayed Evidence of attention to negative cases, outliers or exceptions Typologies/models of variation derived and discussed Examination of origins/influences on opposing or differing positions Identification of patterns of association/linkages with divergent positions/groups	
13 **A N A L Y S I S**	How well has detail, depth and complexity (i.e. richness) of the data been conveyed?	Use and exploration of contributors' terms, concepts and meanings Unpacking and portrayal of nuance/subtlety/intricacy within data Discussion of explicit and implicit explanations Detection of underlying factors/influences Identification and discussion of patterns of association/conceptual linkages within data Presentation of illuminating textual extracts/observations	
14 **R E P O R T I N G**	How clear are the links between data, interpretation and conclusions – i.e. how well can the route to any conclusions be seen?	Clear conceptual links between analytic commentary and presentations of original data (*i.e. commentary and cited data relate; there is an analytic context to cited data, not simply repeated description*) Discussion of how/why particular interpretation/significance is assigned to specific aspects of data – with illustrative extracts of original data Discussion of how explanations/theories/conclusions were derived – and how they relate to interpretations and content of original data (*i.e. how warranted*); whether alternative explanations explored Display of negative cases and how they lie outside main proposition/theory/hypothesis etc.; or how proposition etc. revised to include them	

	a) Appraisal questions	b) Quality indicators (possible features for consideration)	c) Notes on study being appraised
15 R E P O R T I N G	How clear and coherent is the reporting?	Demonstrates link to aims of study/research questions Provides a narrative/story or clearly constructed thematic account Has structure and signposting that usefully guide reader through the commentary Provides accessible information for intended target audience(s) Key messages highlighted or summarised	
16 R E F L E X I V I T Y & N E U T R A L I T Y	How clear are the assumptions/ theoretical perspectives/ values that have shaped the form and output of the evaluation?	Discussion/evidence of the main assumptions/hypotheses/theoretical ideas on which the evaluation was based and how these affected the form, coverage or output of the evaluation (*the assumption here is that <u>no</u> research is undertaken without some underlying assumptions or theoretical ideas*) Discussion/evidence of the ideological perspectives/values/philosophies of research team and their impact on the methodological or substantive content of the evaluation (*again, may not be explicitly stated*) Evidence of openness to new/alternative ways of viewing subject/theories/assumptions (*e.g. discussion of learning/concepts/constructions that have emerged from the data; refinement restatement of hypotheses/theories in light of emergent findings; evidence that alternative claims have been examined*) Discussion of how error or bias may have arisen in design/data collection/analysis and how addressed, if at all Reflections on the impact of the researcher on the research process	

	a) Appraisal questions	b) Quality indicators (possible features for consideration)	c) Notes on study being appraised
17 **E T H I C S**	What evidence is there of attention to ethical issues?	Evidence of thoughtfulness/sensitivity about research contexts and participants Documentation of how research was presented in study settings/to participants (*including, where relevant, any possible consequences of taking part*) Documentation of consent procedures and information provided to participants Discussion of confidentiality of data and procedures for protecting Discussion of how anonymity of participants/sources was protected Discussion of any measures to offer information/advice/services etc. at end of study (*i.e. where participation exposed the need for these*) Discussion of potential harm or difficulty through participation, and how avoided	
18 **A U D I T A B I L I T Y**	How adequately has the research process been documented?	Discussion of strengths and weaknesses of data sources and methods Documentation of changes made to design and reasons; implications for study coverage Documentation and reasons for changes in sample coverage/data collection/analytic approach; implications Reproduction of main study documents (*e.g. letters of approach, topic guides, observation templates, data management frameworks etc.*)	

PART 2: THE FOUNDATIONS AND DEVELOPMENT OF THE FRAMEWORK

3 THE CRITERIA DEBATE

When undertaking the task of devising a framework for assessing the quality of qualitative evaluations, we as a team had to confront the whole issue of whether or not it was feasible, or even desirable, to set standards or criteria. And if such a framework could be drawn up, what types of criteria should be adopted? Would 'qualitative criteria' be different from or similar to those accepted as appropriate for judging the quality of other kinds of evaluations, particularly those using experimental or other quantitative methods?

In this, and the following chapters, we discuss the debates and views on which we drew in developing our framework and show, in text boxes, the decisions we made about the scope, approach and content of our framework.

3.1 Evaluation and research

Before reviewing some of the positions discussed in the literature, however, we must first clarify what we mean by qualitative policy evaluation and the relationship between evaluation and research. Some writers and evaluators maintain that evaluation is a distinct activity from research (see, among others, House, 1980; Lincoln and Guba, 1986; Scriven, 1986). In some respects this distinction is seen as being based on the differences between 'pure' or 'basic' and 'applied' research. According to Patton (2002), for example, basic research aims to generate new theories or test existing ones, whereas applied research uses knowledge (acquired through research) to contribute to understanding or to the resolution of a contemporary issue or problem. But it is also the appraisal or normative element in evaluation which is seen to distinguish it from research (Fournier, 1995). An overview of the main ways in which evaluation is thought to differ from research is given by Shaw (1999):

- Evaluation addresses practical problems, whereas research addresses theoretical issues.
- The end products are different: evaluation culminates in action while research culminates in description.
- Evaluation has a normative component and makes judgements of worth or merit; research describes.
- Evaluation addresses short-term issues; research looks at longer-term ones.
- Not all evaluation methods are research-based.
- Evaluation is not bound by particular disciplinary boundaries, whereas research is usually disciplinary.

One writer who expresses particularly strong opinions on the difference between evaluation and research is Martyn Hammersley. Hammersley (2002; 2003) argues for

a distinction between what he calls *practical research* and *scientific research*. The key difference between these kinds of inquiry relates to the audiences to which research findings are directed and the criteria against which they are validated. Practical research aims to produce knowledge of practical use to practitioners or policy-makers who assess the findings in terms of relevance, timeliness and validity – validity being judged according to the plausibility of the findings in relation to *practical knowledge* and *experience*. Scientific research, on the other hand, aims to contribute to a cumulative body of knowledge and its primary audience is fellow researchers who assess the validity of the findings in terms of what is already *accepted as knowledge in the academic community* and in relation to *the credibility of the evidence presented*.[4]

While acknowledging that there are *some* differences between evaluation and research, Shaw (1999) himself does not believe that – in practice – the two are very separate, and claims there is considerable overlap. He argues that evaluation does not always involve normative appraisal and that the short-term as against long-term distinction does not hold up because policy-makers and politicians also look at evidence over a period of time, not just at single studies. The practical as against theoretical distinction is also, he argues, overdrawn. Shaw prefers the term 'qualitative inquiry' to encompass both qualitative evaluation and qualitative research. Patton (2002), despite making a distinction between basic and applied research, also adopts this approach and, in his encyclopaedic text book, *Qualitative Research and Evaluation Methods*, often uses 'qualitative inquiry' to refer to both evaluation and research.

This debate about the relationship between evaluation and research arises partly because the term 'evaluation' is used in different ways. Shaw (1999) argues that evaluation is sometimes seen as synonymous with large-scale programme evaluation, and that evaluation theorists often neglect other kinds of evaluation. He himself, together with other writers, notably Patton (2002), proposes a much broader view of evaluation, including policy research and studies of practice, as well as programme evaluation.

> For the purposes of developing a framework to assess the quality of qualitative evaluations we have decided to adopt a broad conception of evaluation, to refer both to background research in relation to a policy issue, and more specific appraisals of particular practices or interventions. Of course, not all evaluation involves research and so we confine our discussion, and the development of our framework, to *policy evaluations which are based on empirical research and which use qualitative research methods.* For shorthand purposes, we sometimes use the term qualitative evaluation and qualitative research interchangeably, but we always mean studies which use qualitative research methods to explore

---

[4] Hammersley (2003, personal communication) considers *practical* and *scientific* research as forms of *specialised research* whose primary aim is to produce knowledge. He then distinguishes *specialised research* from *research subordinated to another activity* where the main aim is to address a problem rather than necessarily to produce knowledge. In this schema, evaluation may be a form of *practical research* or of *research subordinated to another activity*.

policy issues, practice, or particular initiatives or schemes.

3.2 The nature and contribution of qualitative research in policy evaluation

In order to understand the debate about criteria for assessing the quality of qualitative evaluations, it is important to examine the nature of qualitative research and the way it is used in policy evaluation. This is because different positions adopted by writers and researchers in relation to the idea of criteria are influenced by perceptions of the essential nature of qualitative research, its distinctiveness in relation to quantitative research, and its contribution to evaluation. We begin by looking at definitions of qualitative research.

3.2.1   What is qualitative research?

There are many different definitions of qualitative research in the literature and writers approach the question of definition in a number of different ways. One approach is simply to define qualitative research in terms of a collection of research methods. These methods include those which attempt to capture naturally occurring data where behaviour is enacted in its natural setting, and generated data which involves a 'reconstruction' (Bryman 2001), a recounting or re-telling of beliefs, experiences or behaviour. Methods that use naturally occurring data include:

- participant observation
- observation
- analysis of existing documents or visual media (these may be contemporary or historical sources)
- discourse analysis
- conversation analysis.

Methods that involve generated data include:

- biographical methods such as life histories
- unstructured diary-keeping and journals (where these have been written specifically for a research project)
- individual interviews
- focus groups.

Some authors define qualitative research in terms of key characteristics which are associated with these kinds of methods (see, for example, Bryman, 1988; Denzin and Lincoln, 2000; Hammersley, 1992; Hammersley and Atkinson, 1995; Holloway and Wheeler, 1996; Mason, 2002; Miles and Huberman, 1994; Patton, 2002; Sandelowski, 1986). Some key characteristics of qualitative research discussed in the literature are summarised in Box 3.1 on the following page.

Of course, there is a great deal of debate in the literature about the extent to which some of these features are actually method-specific, rather than characteristic of all qualitative methods, whether or not all these features are actually desirable, and the extent to which they clearly distinguish qualitative from quantitative research (see, for example, Agar, 1980; Bryman, 1988; Fetterman, 1989; Hammersley, 1992; Murphy *et al*, 1998; Phillips; 1990; Silverman, 1993; 2000; Schwandt, 1997). There is also the consideration that some qualitative research, especially contract research, cannot be as 'prolonged', flexible and open-ended as the above list might suggest (Murphy *et al*, 1998).

Box 3.1 Key features of qualitative methods

- a concern with *meanings*, especially the subjective meanings of participants
- a commitment to viewing (and sometimes explaining) phenomena *from the perspective of those being studied*
- an awareness and consideration of *the researcher's role and perspective*
- *naturalistic inquiry* in the *'real-world'* rather than in experimental or manipulated settings
- prolonged *immersion* in, or *contact* with, the research setting
- the absence of methodological orthodoxy and the use of a *flexible (emergent) research strategy*
- the use of *non-standardised, semi-structured or unstructured methods* which are sensitive to the *social context* of the study
- the capture of data which are *detailed, rich and complex* (for example, the use of 'thick description')
- the setting of data in *context*
- the collection and analysis of data that are mainly in the form of *words and images* rather than numbers
- a commitment to retaining *diversity and complexity* in the analysis
- a respect for the *uniqueness* of each case as well as themes and patterns across cases
- a mainly *inductive* rather than deductive analytical process
- attention paid to *emergent categories and theories* rather than sole reliance on *a priori* concepts and ideas
- *development* rather than testing of hypotheses
- a concern with *micro*-social processes
- *explanations offered at the level of meaning*, or in terms of local 'causality' (why certain interactions do or do not take place) rather than 'surface workings' or context-free laws

Another way of defining qualitative research is to refer to a number of traditions or schools associated with qualitative research and which make use of predominantly qualitative methods.[5] Taking the disciplines of philosophy, anthropology, sociology, psychology, and linguistics, the following traditions can be identified: phenomenology; ethnomethodology; conversation analysis; discourse analysis; protocol analysis; symbolic interactionism; grounded theory; ethogenics;

---

[5] There is some overlap between 'traditions' and 'methods'; for example, conversation analysis and discourse analysis are listed both as *qualitative methods* and as *qualitative traditions*.

28

constructivism; and critical theory (including Marxist and neo-Marxist research, and some forms of feminist research). A very brief summary of the key features of these traditions is given in Box 3.2.

Box 3.2 Traditions within qualitative research

Research tradition	Disciplinary origins	Aims
Ethnography	Anthropology/ sociology	Understanding the social world of people through immersion in their community to produce detailed description of culture and beliefs
Phenomenology/ethnome thodology	Philosophy/sociology	Understanding the 'constructs' people use in every day life to make sense of their world
Leading to **Conversation analysis**	Sociology/linguistics	Analysing the way in which talk is structurally organised, focusing on sequencing and turn-taking which demonstrate the way people give meaning to situations
Discourse analysis	Sociology/linguistics	Examining the way knowledge is produced within a particular discourse and the performances, linguistic styles and rhetorical devices used in particular accounts
Protocol analysis	Psychology	Examining and drawing inference about the cognitive processes that underlie the performance of tasks
Symbolic interactionism	Sociology/social psychology	Exploring behaviour and social roles to understand how people interpret and react to their environment
Leading to **Grounded theory**	Sociology	Developing 'emergent' theories of social action through the identification of analytical categories and the relationships between them
Ethogenics	Social psychology	Exploring the underlying structure of behavioural acts by investigating the meaning people attach to them
Constructivism	Sociology	Displaying 'multiple constructed realities' through the shared investigation (by researchers and participants) of meanings and explanations
Critical theory	Sociology	Identifying ways in which material conditions (economic, political, gender, ethnic) influence beliefs, behaviour and experiences (and in some cases using new understanding to facilitate change)

Source: Adapted from Snape and Spencer (2003).

Finally, some authors define qualitative research in terms of a set of distinctive philosophical assumptions about the nature of reality (ontology) and how we can find out about it (epistemology). Among practising qualitative researchers, however, there are many different ontological and epistemological positions and these are discussed in more detail later in this report (see Section 4.1).

We also explored views about the key contributions made by qualitative research in our interviews with policy-makers, research managers and commissioners, funders, academics and practitioners. These views were fairly consistent across the different groups we interviewed, and three roles in particular were stressed. First, exploring issues in depth, highlighting and understanding complexity. Here interview participants talked about qualitative research '*getting beneath the surface*', '*getting inside the black box*', '*getting beneath the data*' and exploring '*what lies beneath*' behaviours, attitudes and decisions. Second, an exploratory and generative role. As well as talking about using qualitative research to generate hypotheses, people talked about the ability of qualitative research to explore issues which had not previously been the subject of much research, perhaps phenomena that were newly being seen as important – the example was given of how the development of AIDS highlighted a need to understand male homosexuality.

> *One [role] is this opening up of an unknown area, which isn't peculiar to evaluative research but it's finding out how a world or an arena or a context operates at an early stage of realising that there is a problem for government … as one major policy-maker or ultimate user of the work … So really very basic exploration of the nature or phenomenon that is not well understood by decision-makers.*
> Academic/research practitioner

The third contribution of qualitative research was its ability to provide insight into the social worlds of research participants, to understand their own subjective meanings and how they arise or are constructed. Here, interview participants talked about the value of qualitative research in allowing researchers and research users to '*see through the eyes of the people you are studying*', giving insight into their perspectives, motivations, assumptions, perceptions, frames of reference, language and views of the world. Government research managers sometimes stressed the impact this understanding makes on policy-makers, and policy-makers themselves valued it highly.

> *I often commission qualitative research when it's about either users or stakeholders and it's when I want to understand ... how a user is likely to respond to either this initiative or this policy. I want to know how they see the world, what assumptions they make about it, what they're doing, what the Government is trying to achieve, how they're likely to respond ... Getting that understanding of why, how do people see the world and how are they likely to react to anything in the future. So it's a wonderful vehicle ... if you want to*

> *understand the motives of people and their views of the world.*
> Research manager

> *It's ... really important as a way of understanding things in that, if you like, kind of hearing what people are saying without it being filtered too much and in a way that you can then hopefully interpret carefully, is a real privilege and insight into people's lives, you know. So at its best I think it's just hugely illuminative of different perspectives and different views.* Funder

Other key features of qualitative research highlighted were its ability to preserve and explore context (at the individual level and in the sense of understanding broader social and organisational contexts), highlighting and exploring diversity, identifying and looking in detail at sub-groups, and exploring sensitive issues.

3.2.2   The contribution of qualitative research to evaluation

Turning now to the role of qualitative research in evaluation, we can see that these distinctive features of qualitative research are very relevant. The focus in qualitative research on exploring meanings and behaviour in depth, identifying diverse perspectives, capturing processes and contexts, and using flexible methods, means that qualitative research is seen as being capable of making a distinctive contribution to policy evaluation. According to Finch (1986), it is these features that make qualitative research attractive to policy-makers. Finch (1986) herself sees a broad role for qualitative policy evaluation – '*to describe and understand the real effects of policies, to compare the assumptions upon which policies are based with social experience, and to assist in a considered assessment of their viability and appropriateness*' (p. 158). She also argues for qualitative research to have a role in developing and changing policies.

In practice, there are many different models of the potential role of qualitative research in qualitative evaluation. Rist (2000), for example, sees qualitative research as relevant throughout the different phases of the policy-making process, contributing in numerous ways to policy formulation, implementation and accountability. Patton (2002) also suggests a wide range of applications for qualitative enquiry, including, for example:

* personalising and humanising evaluation so that it is open to stakeholders' ideas
* harmonising evaluation values across different stakeholders
* in developmental, participatory, and democratic evaluations
* when there is a need for non-obtrusive measures
* where an issue is poorly understood
* for rapid reconnaissance
* for capturing and communicating stories
* for auditing and monitoring the extent to which something is actually happening on the ground

- for capturing unintended consequences
- for generating new insights.

Williams (1986) argues that qualitative evaluation is appropriate when:

- evaluation issues are not clear in advance
- official definitions of the evaluand are not sufficient and insider perspectives are needed
- thick description is required
- formative evaluation is needed
- outcomes include complex actions in natural settings
- there is a need to know how the evaluand is operating in its natural state
- there is time to study the natural cycle
- diverse data sources are available
- there are resources to search for negative instances.

Davies (2000) proposes a not too dissimilar set of situations where qualitative research can play an important role, including:

- helping define policy questions
- looking in detail at how a programme or trial was actually implemented
- helping determine appropriate outcome measures by looking at 'subjective' outcomes
- helping clarify what counts as effective or successful
- identifying and exploring unintended consequences
- contributing to service delivery and policy development by describing processes and contexts
- informing and illuminating quantitative studies, for example, by contributing to the design of structured instruments, assessing the fairness of comparisons in experimental studies, or unpacking variation within aggregated data.

In the field of health care assessment, for example, Popay and Williams (1998) outline two current models of the role of qualitative research. The first is the '*enhancement model*' (see Black, 1994) in which qualitative research can help with understanding of why interventions work, improve the accuracy and relevance of quantitative studies, identify appropriate variables to be measured in quantitative research, explain unexpected results from quantitative work, and generate hypotheses to be tested using quantitative methods**.** The second model they call the '*difference*' model which sees a role for qualitative research in exploring 'taken for granted' practices, understanding lay behaviour, exploring people's perceptions of the quality and appropriateness of interventions, understanding organisational culture, and evaluating complex interventions.

Running through these different models are recurring motifs, for example, that a qualitative approach is valuable where issues or problems are poorly understood, where a researcher wants to explore a policy context, where evaluation criteria are not clear or alternative criteria are sought, where it is important to understand practices in detail or it is important to discover how a scheme was actually implemented, or where an evaluator wants to know about people's subjective experiences of outcomes.

In some of these models, qualitative inquiry is seen as a junior partner or '*handmaiden*' to quantitative inquiry (Dean *et al*, 1969) or appropriate only for certain kinds of evaluation (Knapp, 1999). In others models there is a 'horses for courses' approach in which methods are chosen according to the aims of the study (Munhall, 1993; Patton, 2002; Seale, 1999; Silverman, 1997; Walker, 1985). Finally, some models advocate qualitative inquiry as the preferred methodology because of its ability to explore meanings, processes, and local causality. Shaw (1999), for example, believes that qualitative methods should be the methods of choice where timescales and resources allow, and that qualitative evaluation can address policies, programmes, projects, and elements of service or practice. He also maintains that qualitative inquiry can be used for *summative* as well as *formative* evaluation (Scriven, 1991) at any stage, from drawing board, feasibility, and implementation to outcomes. In his view, evaluation is not just limited to measurement of results, but includes gaining understanding of public issues, and of present and past attempts to address them. For a fuller discussion of perceptions of the relative strengths of qualitative and quantitative inquiry, see Murphy *et al*, 1998.

Despite the wide range of possible applications of qualitative inquiry, there is a recurrent theme in the literature that policy evaluation has been dominated by experimental studies and other quantitative approaches and that qualitative evaluation is a relatively neglected area, apart from the work of a few pioneers (see, for example, Guba and Lincoln, 1981; Stake, 1995; Stake and Turnbull, 1982). Some writers attribute this neglect to the supremacy of a particular paradigm (variously described as positivist, or [naïve] realist), and to a widespread perception that there is a hierarchy of methods, in which quantitative are considered the 'best' (Chamberlain, 2000; Drewett, 1997). The dominance of quantitative methods is often dated from the middle of the last century when they were widely used to evaluate the 'effects' and 'effectiveness' of large-scale social programmes, particularly in America. This continued up until the latter half of the century when this model of evaluation began to be challenged by the growth of action research and by alternative views of evaluation purposes, which stressed reformist, democratic, collaborative, educative or emancipatory goals.

More recently, quantitative methods are again perceived as taking precedence with the development of the evidence-based practice movement in a range of policy fields (Popay and Williams, 1998). Within this movement, accountability, validity and reliability are the guiding ideals, and the emphasis on randomised control trials (RCTs) has again implied a hierarchy of methods, with RCTs at the apex and qualitative methods almost at 'the bottom' somewhere above anecdotal evidence (Hammersley, 2001). Shaw (1999) maintains that while RCTs are perceived as the '*Rolls Royce*' of evaluation methods (Chelimsky, 1997), qualitative methods are the '*Morris Minor*'. Even with a 'horses for courses' model, Shaw believes that qualitative evaluation is given only a modest role:

> *Proponents of the horses-for-courses approach to evaluation believe –*
> *to extend the metaphor – that, when the ground is hard or firm,*
> *outcome-oriented designs can be jockeyed into position, whereas*

when the ground is soft we should opt for mounts which can safely negotiate an understanding of sticky institutional processes. (Shaw, 1999: 2)

While some researchers welcome the promotion of evidence-based practice and believe that scientific knowledge can be used to facilitate progressive change (for example, Oakley, 2000; Pawson and Tilley, 1997a; 1997b), others are concerned that it is based on a 'social engineering' model. They believe that evidence-based practice takes a naïve view of the relationship between evidence and practice, failing to appreciate fully that empirical evidence requires translation and judgement. (For a detailed discussion of different models of the role of research in informing social policy, for example, social engineering, enlightenment and cognitive resources models, see Bulmer, 1982; Hammersley, 2002; Silverman, 1993.)

Whether or not research does indeed influence policy is a matter of debate within the literature. Weiss (1988a; 1988b), for example, argues that research is not always used in policy formulation and that evaluators need to understand how policy is actually made – as a result of multiple influences and actions. From her study of policy-makers, Weiss concluded that there is a 'trickle down' effect in the way that evaluation findings enter the knowledge base. Although findings can be used as guidance for improving and developing policy, policy-makers also use them in other ways: as a warning that things are going wrong; as a new way of looking at issues; and as a way of mobilising support for a policy.

Rist (2000) also maintains that policy research, no matter how rigorously conducted, will not have more influence unless it is more directly linked to the way the policy process actually takes place. Patton (1997) argues vigorously in favour of utilisation-focused evaluation which involves end-users in the conception and presentation of evaluations in order to enhance their utility and actual use. Writing in 1986, Finch suggested that qualitative research traditionally had not had much impact on policy in Britain for a number of reasons. She highlighted a strong emphasis on statistical research; the dominance of a Fabian social engineering model; and an anti-quantitative position in academic sociology, which shunned involvement in social reform until the burgeoning of action research.

Some aspects of how qualitative research contributes to evaluation were described in the interviews we carried out. Qualitative research was strongly linked with process evaluation, because of its ability to explore how a service or policy is delivered and experienced in practice, the mechanisms by which it can produce the desired effects, the barriers and facilitators to this, and the circumstances under which successful operation might be transferable. One policy-maker described this as '*tak[ing] the schemes to bits*'. But it was also seen as important in understanding the various impacts or outcomes of policies: the meanings they have to those experiencing them, unintended consequences and un-anticipated issues. Qualitative research was also seen as valuable in identifying solutions or strategies, and in helping to develop appropriate interventions for different settings and populations.

The two issues are, how does it feel on the ground to the practitioners that you want to deliver this? … And secondly how does it feel to the recipient? … To my mind it's inconceivable that we don't back up the testing of interventions, which may well be primarily survey driven … with qualitative research which actually gets under the skin of the people who we want to deliver programmes, the practitioners, and actually differentiates between the receptiveness of different individuals and populations. So we've got to do both [quantitative and qualitative research] if we are to come up with sensible policy ideas, it seems to me. Policy-maker

3.3 Is it possible or desirable to establish quality criteria for qualitative inquiry?

Having established what we mean by qualitative policy evaluation, we then had to consider whether it was feasible and desirable to establish criteria for assessing the quality of qualitative inquiry. This section presents the range of views and positions held among researchers, academics, commissioners and policy-makers.

From our review of the literature it is clear that the subject of 'qualitative criteria' is a hotly debated issue. Not only are there different positions on this question, but a number of authors have analysed these positions and put forward different classificatory schema (see, for example, Angen, 2000; Bryman, 2001; Cutcliffe and McKenna, 1999; Denzin and Lincoln, 1994; Emden and Sandelowski, 1998; Goodwin and Goodwin, 1984a; and 1984b; Hammersley, 1992; Murphy *et al*, 1998; Schwandt, 1996; Shaw, 1999; Sparkes, 2001; Whittemore *et al*, 2001). Our reason for referring the reader to these sources, rather than discussing them in this review, is that different schema highlight slightly different positions and, more confusingly, there is little consistency (apart from at the extremes) in terms of the positions to which particular authors are assigned. So, for example, while some consider Lincoln and Guba to have completely reformulated quality criteria to fit an interpretivist approach, others maintain that they have retained the same basic criteria as used in quantitative research and remained within a positivist paradigm.

Broadly speaking, for the purposes of this review, we think it is more useful to think of 'positions' on criteria as ranging at one extreme from a rejection of the notion of criteria altogether to, at the other, a retention of concepts common to quantitative and qualitative research. In between these two extremes are a wide range of different positions. For example, there are some who *reject the notion of criteria* but accept a looser notion of guiding principles and ideals; some who *propose alternative criteria* or principles for qualitative research which are quite different from those used in quantitative research; and others who *subscribe to parallel* but slightly different *criteria*.

The rejection of criteria altogether is most commonly associated with Smith (1984; 1990). Smith's objection to the idea of criteria is based on the idea that qualitative research is philosophically distinct from quantitative. Qualitative research is *idealist*

(social reality does not exist independently of individual human constructions) and *anti-foundational* (because there are many different constructions of reality, there can be no certain, 'correct' or privileged understanding). He argues that the idea of criteria, in the sense of standards for judging good or bad research, is essentially foundational, based on the assumption that there are 'correct' procedures which can be followed, whereas Smith maintains that '*the anti-foundationalist nature of interpretive inquiry means that no epistemological privilege can be attached to any procedure for doing or criterion for judging this approach to inquiry*' (Smith, 1984: 389). Attempting to define criteria for qualitative research is therefore logically impossible. However, more recently Smith has acknowledged that quality can be judged in qualitative research, albeit in terms of open-ended and evolving 'characteristics' rather than clearly specified criteria:

> *As we approach judgement in any given case, we have in mind or bring to the task a list, for lack of a better term, of characteristics that we use to judge the quality of that production. The use of the term list should not be taken to mean that we are referring to something like an enclosed and precisely specified or specifiable shopping or laundry list. … To the contrary, for us a list of characteristics must be seen as always open-ended, in part unarticulated, and, even when a characteristic is more or less articulated, it is always and ever subject to constant reinterpretation. Moreover, the items on the list can never be the distillation of some abstracted epistemology, they must inevitably be rooted in one's standpoint.* (Smith and Deemer, 2000: 888)

At the other extreme, the position of accepting common criteria for both qualitative and quantitative research, mainly in the form of validity and reliability, is generally attributed to Le Compte and Goetz (1982) and Kirk and Miller (1986). However, although these authors reject the idea that qualitative research is philosophically distinct from quantitative research and requires different criteria, they do recognise that it is conducted quite differently and discuss ways in which the concepts of validity and reliability can be applied to qualitative research.

The development of parallel criteria (and one of the most widely discussed contributions to the criteria debate) is most commonly associated with Lincoln and Guba. Like Smith, Lincoln and Guba also believe qualitative or 'interpretivist' research is based on different ontological and epistemological assumptions from quantitative or 'scientific' research but, unlike Smith, they do not reject the notion of criteria. Instead they propose that qualitative research should be judged on its own terms, and they formulated an alternative set of criteria to assess the *trustworthiness* of a piece of qualitative research (see Guba and Lincoln, 1981; and Lincoln and Guba, 1985).

As Box 3.3 shows, however, these criteria address parallel concerns to the concepts of validity, reliability and objectivity which dominate debates about criteria in quantitative research. So, the traditional concept of *internal validity*, which Lincoln and Guba translate as truth value, is replaced by the concept of *credibility* – whether or not the participants studied find the account true. *External validity*, or the extent to

which findings are more generally applicable, is replaced by fittingness or *transferability*, which is based on the idea that accounts may be transferable to other specified settings through the provision of thick description about both the sending and the receiving contexts (rather than on a notion of context-free generalisations). *Reliability*, or the consistency of findings, is replaced by the notion of *dependability*, which is achieved through an auditing process called an 'audit trail', in which the researcher documents methods and decisions, and assesses the effects of research strategies (rather than being concerned about replication). Finally, *objectivity*, or a concern with neutrality, is replaced by *confirmability* – the extent to which findings are qualitatively confirmable through the analysis being grounded in the data and through examination of the 'audit trail'.

Box 3.3: Lincoln and Guba's 'naturalistic' criteria

Aspect	Scientific term	Naturalistic term
Truth value	Internal validity	Credibility
Applicability	External validity or generalisability	Transferability
Consistency	Reliability	Dependability
Neutrality	Objectivity	Confirmability

Source: Adapted from Guba and Lincoln (1981) and Lincoln and Guba (1985).

Lincoln and Guba's alternative criteria are constantly cited and have sparked a tremendous debate within the literature. For many writers, their contribution is still considered the 'gold standard' even though it has provoked as much criticism as acclaim (see, for example, Whittemore *et al*, 2001).

There are also completely different approaches to the issue of criteria which largely avoid notions of validity and reliability, and are not even concerned with underlying issues about the extent to which a study can be believed or how rigorously it was conducted. Agar (1986), for example, argues that qualitative research should reject concerns about validity in favour of intensive personal involvement, abandoning of scientific control, and adoption of improvisation and flexibility. Others propose criteria which largely relate to relationships with participants or the impact of the study, advocating criteria linked to, for example, ethics, democracy, empathy, education or emancipation. An example from a feminist, materialist perspective is that of Roman and Apple (1990) who propose four criteria – the extent to which the study:

- resonates with the 'lived experience' of those being studied
- contributes to the understanding of participants and enables them to transform their experiences of subordination
- reduces the divide between the researcher's and 'ordinary' ways of describing and understanding phenomena

- leads to the researcher's theoretical and political commitments being informed by participants' experiences.

Interestingly, Guba and Lincoln (1989), responding to criticisms that their criteria remained within a positivist paradigm, later added a fifth criterion of *authenticity*, which they saw as particularly appropriate for evaluations, and on which they placed increasing emphasis. Authenticity included the extent to which:

- different views were fairly represented
- researchers had a more sophisticated understanding of the phenomenon under study than at the outset, and that this understanding also involved a greater appreciation of the understandings of others
- the evaluation process had stimulated action and empowered participants to act.

Running through these different positions on criteria for judging the quality of qualitative inquiry are questions about whether or not it is possible to have 'qualitative' criteria at all, and, if it is possible, what kind of criteria should be adopted and how these should relate to the criteria used to assess quantitative research. In addressing these questions, writers appear to be struggling with a number of key issues, for example:

- *What is meant by 'criteria'?*
  How people feel about the idea of criteria is partly influenced by how they understand the term. Smith (1984), for example, distinguishes different meanings of the word 'criterion': a fixed standard against which to judge good from bad, right from wrong, or a 'characterising trait', which does not have such judgemental connotations. So, while Smith rejects the idea of fixed standards, he is not opposed to 'characterising traits' or an evolving set of guiding ideals.

  Other writers also prefer the notion of guiding principles rather than some kind of fixed standards, although they would not share Smith's other views about judging the quality of qualitative research (Mason, 2002; Seale, 1999). Debate around the notion of criteria also stems from the fact that there is some confusion in the literature between *criteria themselves* and *ways of assessing them* (Whittemore *et al*, 2001). For example, Hammersley (1992) argues that credibility and plausibility are not themselves criteria, but ways of addressing the criterion of validity.

- *Whether or not qualitative research rests on particular philosophical assumptions which make criteria problematic.*
  Smith based his objection to criteria on the notion that qualitative research is essentially relativist and anti-foundational and that this makes the idea of criteria problematic. Other writers argue that there is no single philosophical basis for qualitative research and that a range of different ontological and epistemological assumptions are adopted by qualitative researchers (Altheide and Johnson, 1994; Hammersley, 1992; Phillips, 1990). Nevertheless, there is widespread acknowledgement that underlying philosophical assumptions are often invoked as a basis for concerns about criteria and as a justification for alternative quality criteria in qualitative research (Blaikie, 1991; Engel and Kuzel, 1992).

- *Whether qualitative methods themselves pose problems in terms of setting or applying criteria.*
  It is argued that key features of qualitative methods, namely that they are non-standardised, unstructured, flexible and open to amendment, make them difficult to appraise. Davies and Dodd (2002), for example, maintain that qualitative methods cannot be judged in the same way as quantitative: '... *if rigor is understood only in terms of a structured, measurable, systematised, ordered, uniform and neutral approach, then other research methods that allow flexibility, contradictions, incompleteness, or values will appear "sloppy", epitomising everything that is "nonrigor" and therefore lacking in credibility'* (p. 280).

  Unlike in quantitative research where validity can apply to particular instruments, validity is not a property of a particular method in qualitative research (Maxwell, 1992; Sandelowski, 1993). There is no methodological orthodoxy in qualitative research – there are no set procedures or methodological rules – yet *criteria are essentially procedural* (Schwandt, 1996). Because qualitative methods are flexible there is a danger that focusing too much on methods – what Chamberlain (2000) refers to as 'methodolatry' or the '*privileging of methodological concerns over other considerations*' (p. 285) can act as a straitjacket. Nevertheless, Chamberlain and others acknowledge that the very flexibility of qualitative methods means that some assurances are needed that the researcher conducted the study thoroughly and professionally. Whittemore *et al* (2001) see this issue as a conflict between *rigour* and *creativity*.

- *Whether other kinds of criteria that are not strictly methodological should supplement or even replace methodological ones.*
  Some writers propose non-methodological criteria for assessing qualitative research. Sometimes this is because of the difficulties of judging qualitative methods discussed above. Other times it is because qualitative research has attracted researchers with particular moral, political or philosophical values, who wish to use research for educative, reformist or emancipatory aims. They believe that qualitative research should be evaluated according to criteria which reflect these aims. Greene (1996), for example, argues that research should generally be evaluated in terms of its social and political consequences.

- *How authors deal with the concept of validity.*
  Much of the debate around criteria is actually directed at the concepts of validity and (to a lesser extent) reliability, and how these terms are interpreted. We discuss this in more detail in Section 5.2, and show how some conceptions of validity refer to the conduct or research, some to the 'truth' or credibility of findings, and some to broader notions of value or worth.

Having reviewed the literature, we now turn briefly to views about criteria expressed in our interviews with research commissioners, research managers, policy-makers, academics and other practising researchers. Among those interviewed there was much agreement that it is appropriate to talk about standards or criteria in qualitative research, that it is important to be concerned about standards and that there is value

in making quality criteria explicit. This was seen to be as important in qualitative research as in quantitative, but it was felt that there remained more debate about quality in qualitative research and that formulating criteria could be more complex in qualitative research.

There were some, however, who were unhappy with the concept of criteria or standards if this meant there was to be a clear-cut judgement of whether a piece of research is either 'good' or 'bad'. This concern also stemmed from a view that the world of qualitative research is too diverse – in terms of approaches, methods and applications – for standards to be relevant except at a very high level of abstraction. And there was also a belief, among academics, that the constructionist nature of qualitative research means that standards, particularly if they are prescriptive, will not be helpful. If knowledge is seen as constructed through the research interaction, there cannot be rigid criteria for how it is produced.

Despite the diversity of views on the feasibility or desirability of criteria, we feel there is a fair degree of support for the idea of providing guidelines for assessing the quality of qualitative evaluations, as long as these are not rigidly procedural and take account of key features of qualitative research. As was clear in the presentation of our framework we have attempted to address concerns about retaining flexibility and creativity, while still encouraging transparency and rigour, by adopting a non-prescriptive approach, based around a series of guiding principles and open-ended appraisal questions.

4 THE SCOPE OF A FRAMEWORK

This chapter addresses key decisions made about the scope of our framework. Having decided that it was worthwhile devising a framework for assessing the quality of qualitative evaluations, we as a team then had to consider how inclusive such a framework could be. Could it address different research paradigms within qualitative research, different kinds of evaluation, and different qualitative methods? And could it cover all stages of research (from proposals to reports) or should it focus on one stage only?

4.1 Addressing different research paradigms and philosophical positions

In this section we briefly describe the philosophical positions that underpin different approaches to research, and explain how we addressed these philosophical issues when defining the scope of our framework. The issues here are complex, but they provide an important context to our framework.

From our review of the literature, it became clear that views about the desirability of criteria are heavily influenced by the research paradigm or philosophical assumptions of the writer. It is also evident that paradigms and assumptions influence which particular criteria or principles are favoured (Patton, 1999; 2002). Addressing this issue, however, was far from easy.

In the first place, research paradigms are often presented as a set of opposing camps which tends to stereotype and to exaggerate differences. Secondly, paradigms are labelled and described in confusingly different ways and the philosophical assumptions on which they are supposedly based vary between different accounts. Finally, the identification and labelling of paradigms seems to suggest that *practising researchers* actually operate consistently within them rather than perhaps varying their approach for different projects (Patton, 1999; 2002; Seale 1999).

Box 4.1 lists just some of the paradigm labels and shows how they appear to be aligned on different sides of a quantitative/qualitative divide.

Box 4.1: 'Quantitative' and 'qualitative' research paradigms

Quantitative	Qualitative
scientific	naturalistic
positivist	interpretivist/hermeneutic
realist	idealist/relativist/constructivist
objectivist/materialist	subjectivist
foundational	fallibilistic/anti-foundational
experimental	ethnographic

Of course, these labels are very crude. Some refer to broad research approaches (positivism), some to types of research method (experimental) or research tradition (ethnographic), others to broad philosophical assumptions (realist, fallibilistic). These labels also fail to reflect the diversity of positions within a paradigm – for example, there are different forms of positivism, realism and constructivism (as discussed below), and many different approaches to ethnography.

In order to unpack this complexity, we need to give a *very brief* account of some of the basic philosophical assumptions which are supposedly linked to different paradigms. Firstly, there are different *ontological* positions or views about *the nature of the world.* Some believe that there is a reality that exists independently of human beliefs and understandings (*realism*). Within this overall *realist* perspective, there are those who go further and argue that this independent reality consists only of *matter* (*materialism*). Set against *realism* and *materialism* is the contrasting view that there is no external reality beyond human constructions and that what exists fundamentally is *ideas* (*idealism*). Within *idealism*, however, there are further distinctions. In *subjective idealism* or *relativism* it is claimed that there is no single shared reality, only a series of alternative constructions and understandings. In *objective idealism*, on the other hand, it is accepted that there is a world of collectively shared understandings.[6]

---

[6] Salner (1989) for example, argues that social reality is not the property of individuals and that there are inter-subjective meanings; Kvale (1988) describes extremes of post-modern relativism as 'indifference' to making sense.

In addition to ontological assumptions, we must also consider *epistemological* beliefs, or views about *the nature of knowledge* and *how it is possible to know about the world*. Key ideas here concern:

- the relationship between the researcher and the researched
- the relationship between facts and values
- the extent to which knowledge can be certain
- the kinds of methods which are appropriate for studying the world.

While some researchers believe that the world is independent of and unaffected by the researcher, others maintain that the researcher and the world impact on each other and that studying something changes it. Some feel it is possible to be 'objective' because findings are separate from a researcher's perceptions or assumptions; others, however, maintain that findings are inevitably influenced by the researcher's values because of the selective nature of research (in focus, data collection and analysis), and that the 'subjective' nature of research should be acknowledged, or even celebrated.

Views of knowledge may be based on different models. In a *foundational* model there is an assumption that it is possible to produce accurate accounts which correspond with reality because '*there is some foundation (sense data, direct observations) … from which knowledge can be derived by logical means.*'[7] A *fallibilistic* model, on the other hand, rests on the idea that it is possible to produce knowledge claims which attempt to be as accurate as possible but they are provisional and may be amended in the light of new findings. In a *sceptical* model, however, there are no 'accurate' or 'privileged' accounts, only alternative understandings.

Finally, on the issue of which methods are appropriate in the social sciences, there are those who believe in the superiority of 'natural science methods' (with their emphasis on hypothesis testing, the use of variables to simplify phenomena and operationalise concepts, and the search for law-like generalisations). Others claim that different methods are required for studying the social world which reflect the complexity of social phenomena and allow for emergent categories and explanations.

However, even though ontology and epistemology deal with separate philosophical issues, in practice, in the social science literature, philosophical positions often contain *both ontological and epistemological assumptions* – that is, they are based on views about the nature of the world *and* how we can know about it. For example, Madill, Jordan and Shirley (2000), in their discussion of different kinds of *realism*, refer to the nature of reality and our access to it. According to these writers:

- *Naïve realism* maintains that reality exists independently of human constructions and can be known directly and 'correctly'.
- *Scientific realism* is based on the view that it is possible for knowledge to approximate closely an external 'reality'.

---

[7] Hammersley (2003, personal communication)

- *Critical realism* accepts that knowledge of reality is mediated by our perceptions and beliefs.

Bhaskar (1978) also links ontological and epistemological assumptions when he argues in favour of:

- *transcendental realism* which assumes that there is a mind-independent world (*ontological realism*) but recognises that acquiring knowledge is a human activity which is culturally determined (*epistemological relativism*).

Hammersley (1992) proposes:

- *subtle realism* in which it is accepted that there is no direct access to reality, and that people's views, beliefs and behaviour are constructions, but that these constructions can still provide us with information about the phenomena to which they refer.

Similarly, under the umbrella of *idealism*, writers discuss the concept of constructivism[8] and distinguish different forms:

- *contextual constructivism*, in which it is claimed that our knowledge of the world is provisional and context-dependent but that we can know and make assumptions about particular settings

- *radical constructivism*, which aims to debunk accounts of the social world by showing that it is impossible to make any assumptions or offer any kind of account of reality (see Shaw, 1999; Madill, Jordan and Shirley, 2000 for a discussion of these positions).

Having unpacked these assumptions we can begin to appreciate the unsatisfactory nature of crude depictions of qualitative as against quantitative research. In the opposing camps model, particular philosophical assumptions, as well as different methodological approaches, are firmly associated with qualitative and quantitative research. Box 4.2 gives an adaptation of Guba and Lincoln's model (for a more complex and recent depiction of the '*paradigm wars*', see Oakley, 2000). No allowance is made for the fact that qualitative research, in practice, is not necessarily tied to a particular set of philosophical assumptions, or that there are subtleties in philosophical positions (for example, different kinds of realists, or overlaps between subtle realists and objective idealists) which make a nonsense of placing realists and idealists in opposing quantitative and qualitative camps. In addition, the model fails to reflect the diversity of qualitative methods (for example, not all of them are strictly naturalistic).[9]

---

[8] Crotty (1988) makes a distinction between *constructivism*, which refers to the meaning-making activity of the individual mind, and *constructionism*, which refers to the collective generation and transmission of meaning.
[9] Of course, not all discussions contrast qualitative and quantitative research, but, instead, contrast different approaches *within* qualitative research. See, for example, Denzin and Lincoln (1994) who compare '*positivism*',

Box 4.2: 'Scientific' vs 'naturalistic' paradigms

	Paradigm	
	Scientific	**Naturalistic**
Assumptions about reality	Singular, convergent, fragmentable (variables)	Multiple, divergent, inter-related
Assumptions about the inquirer–subject relationship	Independent	Inter-related
Assumptions about the nature of truth statements	Generalisations, nomothetic (law-like) statements, focus on similarities	Working hypotheses, idiographic (understanding particular events), focus on differences
Preferred techniques	Quantitative	Qualitative
Quality criterion	Rigor	Relevance
Source of theory	A priori	Grounded
Causality	Can x cause y?	Does x actually 'cause' y in a natural setting?
Knowledge types	Propositional	Propositional and tacit (statements through language plus intuitions, apprehensions)
Stance	Reductionist (limited conditions for control)	Expansionist (holistic/complex)
Purpose	Verification	Discovery
Instrument	Paper and pencil or physical device	Inquirer (often)
Timing of specification of data collection and analysis rules	Before inquiry	During and after inquiry
Design	Preordinate	Emergent
Style	Intervention	Selection (sift through naturally occurring events)
Setting	Laboratory	Nature (natural)
Treatment	Stable (standardised)	Variable
Analytic units	Variables	Patterns
Contextual elements	Control	Invited interference (welcome rather than attempt to control)

Source: Adapted from Guba and Lincoln (1981).

The people we interviewed also rejected dichotomous distinctions between qualitative and quantitative research and commented on the fact that, ontologically epistemologically and in practical terms, they share common ground. Qualitative researchers count, both in their analysis of data and in their reporting of it. Quantitative as well as qualitative researchers make judgements about and interpret data patterns. Surveys generate 'qualitative' data through open-ended questions.

'*post positivism*', '*critical theory*' and '*constructivism*', or Henwood and Pidgeon (1994) who propose three main strands of qualitative research based on *empiricism, contextualism* and *constructivism.*

They may be used to explore subjective meanings, or be built around or aim to explore the social constructions of particular phenomena. Quantitative research also often involves categorical as well as numerical data.

> *First of all I don't accept a distinction between qualitative and quantitative methods. I think it's a completely false dichotomy, I think it's a nonsense. Except in those very rare occasions where you have a fixed zero point and scale where you know what the interval is between the points and you can measure things in that rather rigorous fashion which applies very little in social sciences, where if you're talking about attitudes you're talking about categoric data or you're talking about more and less maybe, and that is constructed.*
> Academic/research practitioner

> *I think there are different kinds of research methods and there are different kinds of research questions, and what's important is that the methods fit the question. And one of the big problems about a lot of research is that the two don't meet, so that you have research questions and you have people cloning all the kinds of findings as a result of the research, but the design of the research really didn't enable them to answer the question. I think all so-called quantitative research involves judgement and everything is ultimately subjective, so, you know, the results of a trial, the hard outcomes, have a qualitative dimension. And every so-called qualitative research project that I've ever come across has engaged in quantification in one way or another. You know, 'a few', 'some of', 'most of'.*
> Academic/research practitioner

We also sought advice from the people we interviewed about what would be feasible in terms of the philosophical scope of our framework. There were different views about how broad-ranging a useful framework could be, but it was generally felt that it would be impossible to address the full range of epistemological and ontological traditions. There were also different views about what might be outside the scope. Where this was discussed, interviewees talked about spanning only the range of positions that would be found among those conducting government-funded research, and excluding:

- approaches that dismiss objectivity as a futile aspiration or see reality as completely unfixed
- research that is primarily designed to be catalytic or emancipatory
- research that rejects any notion of separation between researcher and researched
- research that is heavily rooted in ethnomethodology, symbolic interactionism or post-modernist approaches.

However, some of our interviewees were happy to define their philosophical position. Our sample included a 'realist', a 'subtle realist', an 'interaction constructionist', someone who believed in multiple realities and in capturing the '*smells and sounds of the market place*'. Others were sometimes reluctant to label

themselves, admitting that they were quite 'eclectic'. One practitioner described himself as:

A bit of a whore really! ... I probably began life in a fairly positivist way, because that's what I was taught, certainly from an economic point of view, although I did anthropology which gave me a completely different perspective. Then I got sucked into a fairly constructivist world view for quite a while and then began to be very irritated by it, partly because relativism gets on my nerves. I want to make some choices and I want to think there is a better and a worse way. And I've ... probably shifted in a fairly realist direction in the last five years. Academic/research practitioner

Another researcher described her difficulty in pinpointing her position:

I wouldn't [describe it] in terms of, you know, the big labels, really ... because I take bits from them. I mean, I think I have an approach which is theoretically grounded ... it's probably interpretivist … There's certain things I'm not … I am not a Ray Pawson sort of realist … although I have a great deal of respect for that position. Academic/research practitioner

---

If we accept that qualitative research is not uniform but contains many different paradigms and is based on diverse ontological and philosophical assumptions, then we must acknowledge that a single framework cannot address qualitative inquiry *per se*, but must specify its scope and boundaries. Since we cannot tie our framework to particular paradigms within qualitative research (because of conflicting accounts of their base) we have decided to adopt an *elemental* approach, specifying the range of philosophical and methodological assumptions with which we believe our framework is compatible and those which lie outside its scope. Box 4.3 outlines those assumptions. This means that someone thinking about using our framework will first need to consider whether they subscribe to the assumptions on which it is based. We hope that researchers, commissioners and policy-makers who subscribe to these assumptions will find the framework helpful, but we are aware that it may not be appropriate for certain approaches to qualitative evaluation.

Box 4.3 Assumptions 'in' and 'out' of scope of framework

Assumptions about:	In scope	Out of scope
The nature of reality	• Reality is mediated through human constructions – critical realism, subtle realism OR • Constructions of reality are shared: 'objective idealism', contextual constructivism	• An external reality exists independently of human constructions and is accessible directly and exactly: naïve realism OR • There is no (shared) reality, only alternative individual human constructions: radical constructivism/relativism
The relationship between researcher and researched	• The world is affected by the researcher • The researcher and the world are inter-dependent	• The world is unaffected by the researcher OR • The world is inseparable from the researcher
The relationship between facts and values	• Neutrality is not attainable but a guiding ideal • Subjective perspectives should be articulated and documented	• Research can be freed from all influence by values OR • Attempts at value-free research are undesirable or impossible • Subjective involvement should be maximised and celebrated
The nature of knowledge	Fallibilistic model: • It is desirable to produce knowledge claims which attempt to be as accurate as possible but they are provisional and may be amended in the light of new findings	Foundational model: • It is possible to produce accurate accounts which one knows with certainty correspond directly with reality OR Sceptical model: • There are no privileged accounts, only alternative understandings.
Appropriate methods	• There are no 'valid' or 'invalid' methods • Methods should be appropriate to the aims of an evaluation • Methods should be flexible • It is possible to use methods more or less rigorously	• There are 'valid' and 'invalid' methods • There is a hierarchy of methods • Methods are based on set procedures and rule-following OR • Human beings and behaviour can only be understood by a process of intuition which is tacit and impossible to explicate

4.2 Addressing different kinds of evaluation

A second issue we had to address when determining the scope of our framework was its applicability to different kinds of evaluation. We have already discussed our decision to adopt a broad definition of evaluation, to include contextual research around a policy issue, practice evaluation, and appraisals of particular schemes or interventions. However, we still had to decide whether our framework should try to be applicable to a wide range of approaches to evaluation or focus on a more restricted set.

When reviewing different types of evaluation we encountered similar kinds of problems as in our review of qualitative research paradigms, namely that there is a proliferation of labels, and that the basis of the labelling is not consistent. For example:

- Scriven (1986) distinguishes between formative and summative evaluation.
- Guba and Lincoln (1981) list five different kinds of evaluation: the countenance model; context–input–process–product model; a goal-free model; a connoisseurship model; and a responsive model.
- Fournier (1995) proposes a rather similar typology: causal; goal-free; consumer; pluralistic; connoisseurial.
- Patton (2002) outlines four evaluation models: goal-free evaluation; transaction models (responsive and illuminative evaluation); connoisseurship models; and utilisation-focused evaluation.
- Shaw (1999) differentiates between non-partisan, pluralist/stakeholder, reformist and openly ideological evaluation, but also contrasts 'enlightenment' and 'instrumental' evaluation models.
- Greene (1998) argues that evaluations should be classified according to their philosophical basis, proposing the following types: post-positivist; pragmatist; interpretivist; and critical/normative science. Pawson and Tilley (1997a) would add realistic evaluation.

Although there are some overlaps, for example goal-free and connoisseurship models appear in more than one schema, classifications are often based on different considerations. For example, evaluations may be classified according to *their purpose*, such as providing background knowledge, aiding development and implementation, or assessing outcomes. Some classifications are based on *who the evaluation is for*, such as managers, policy-makers, or users. Alternatively *the role of the evaluator* – expert, outsider, or insider – may determine how an evaluation is classified. In some schema it is *the role of the stakeholders* – representative or participatory – which is of interest. Some writers distinguish different approaches to evaluation in terms of *the methodological orientation* – quantitative, qualitative, mixed methods, *the ideological basis* – managerial, consumerist, reformist, emanicipatory, or *the philosophical position* on which the evaluation is based, such as realist, constructivist and so on.

As with research paradigms, we have decided to adopt an elemental approach to determining the scope of our framework in relation to different types of evaluation. In terms of the elements identified above, we feel that the relevance and applicability of our framework will not be affected or limited by the purpose of an evaluation, nor by the question of who the evaluation is for. The role of stakeholders and the ideological basis of an evaluation *may* be a limiting factor in cases where evaluations are committed to participatory or emancipatory aims, *at the expense of methodological rigour.* Since our framework is directed at evaluations that utilise *qualitative research,* methodological orientation is an important factor, as is the role of the evaluator, because our framework is based on the assumption that the evaluator conducts empirical research and so some forms of the connoisseurship model may lie outside its scope. The philosophical assumptions which are compatible with our framework have already been identified at the end of Section 4.1

4.3 The applicability of the framework to different qualitative methods

The third issue around the scope of our framework is its applicability to different qualitative methods.

4.3.1 Method-specific or generic conceptions of quality

Where writers accept the notion of assessing quality, some of their discussions refer explicitly to particular qualitative methods rather than more generically to qualitative research as a whole. Of course, some discussions about ensuring quality in particular methods are of a 'how to' nature and are concerned with the development of skills. We have not reviewed these kinds of text books but have focused on sources which discuss *conceptions of quality* or propose ways of *judging the quality* of studies based on a particular method.

Below we give a few examples to illustrate the diversity of perceptions of quality in relation to different qualitative methods. In some cases quality is associated with the validity or credibility of findings and the extent to which the research process is documented in sufficient detail for this to be judged; in others quality is discussed in relation to the rationality of the research process.

Ethnography:
In ethnographic research, Agar (1986) maintains that quality issues can be addressed through making analysis more explicit, distinguishing between: basic data units; evolving schema for determining meaning; and the process of interpretation. Mackenzie (1994) suggests that ethnographic studies should ensure quality through the careful reporting of aims; rationale for the choice of method; sampling procedures; and framework of analysis.

Case studies:
In case studies, Stake (1995) proposes the following criteria for assessing quality:

- Is the report *easy to read*?
- Does it *fit together*?
- Is there a *conceptual structure*?
- Are the *issues developed* in a scholarly way?
- Is the case study adequately *defined*?
- Is there a *sense of story*?
- Is the reader provided with some *vicarious experience*?
- Have *quotations* been used effectively?
- Are *headings, figures and appendices* used effectively?
- How well is it *edited*?
- Has the writer made *sound assertions*?
- Has adequate attention been paid to various *contexts*?
- Were sufficient *raw data* presented?
- Were *data sources* well chosen?
- Do observations and interpretations appear to have been *triangulated*?
- Is the *role and point of view of the researcher* apparent?
- Is *empathy* shown for all sides?
- Are *personal intentions* examined?
- Does it appear that individuals were *put at risk*?

Narrative analysis:
For narrative analysis, Bailey (1996) argues that quality is a question of *validation* rather than validity and rests on the reader having access to some of the original material. Polkinghorne (1988) on the other hand thinks that quality in narrative analysis should be understood as *verisimilitude* in which the results have the appearance of truth.

Conversation analysis:
For conversation analysis, quality requires detailed verbatim transcription and validity is judged through the observability of conversational structures, whether or not the analysis deals with deviant cases (those not following normal conversational rules), and whether or not analysis incorporates context into categorisation (Peraklya, 1997).

Discourse analysis:
For discourse analysis, Potter (1996) also argues that full transcription with interactional detail is vital, and that some original material should be reproduced so that the reader can assess the interpretations which are made.

Interview-based studies:
In interview-based studies quality can be judged in terms of the extent to which an iterative design was adopted for sampling and the coverage of interviews so that emerging ideas can be validated (Rubin and Rubin, 1995). Kvale (1996) suggests that there are many different ways of judging the validity of interviews; for example, interviews can be read *experientially, veridically* or *symptomatically*. Only in a veridical reading is the researcher concerned with the interviewee as a witness or informant.

Documentary analysis:
In document analysis, Platt (1981) talks about establishing authenticity of the documents. She describes the document as a 'surrogate researcher' and argues

that it is important to establish if a document was written by someone who directly witnessed the 'event', whether the reporter spoke the language, what role they played in the society. She also discusses the issue of multiple documents dealing with the same event and an exploration of diversity or conformity in reports.

Citizens' juries:
With citizens' juries, Coote and Lenaghan (1997) argue that these should be judged according to *fairness* in the sense that the process was open and not 'rigged', *incorruptability* whereby it is evident that the process was free from pressure or lobbying, *authenticity* – by which they mean clarity about aims, and *effectiveness*, or the ability to reach meaningful conclusions which make an impact on the decision-making process.

Action research:
In action research, Waterman (1998) argues that quality should be assessed using different forms of validity. He proposes *dialectic validity*, which recognises tensions within the field and the research process, *critical validity*, which looks at the responsibilities of the action researcher, and *reflexive validity*, which calls for the researcher to look at a range of possible sources of bias.

From this brief overview, it is clear that there are many different views as to what constitutes quality in qualitative inquiry. Some aspects of quality could refer to most kinds of qualitative research, for example, a clear account of decisions made about the research process, an explicit account of how analysis was conducted and some reproduction of original material, and some evidence that attention was paid to negative cases. Other aspects, however, are more method-specific, such as the production of detailed verbatim transcripts, or procedures for establishing the authenticity of documents.

The question of whether or not a framework could apply across a wide range of qualitative methods was also discussed in our interviews with policy-makers, commissioners, research managers and practitioners. Two rather different views were expressed on this issue. On the one hand, it was felt that it was only at a very abstract or general level that quality standards could be said to apply to all methods; on the other hand, there was some concern about the idea of frameworks developed specifically for individual methods.

4.3.2   *Qualitative research methods used in government evaluations*

In order to help us determine the scope of a framework that could be used by government researchers and policy-makers, we decided to review a sample of recent qualitative studies commissioned by central government to establish which are the most commonly used methods. As described in Chapter 1 of this report, a search of evaluative research from five government departments was carried out (these included DWP, DfES, DoH, ODPM and HO). It should be noted again, however, that this search was based on the information available on the relevant government websites, which were generally abstracts or research summaries, but, in a small number of cases, full reports.

A total of 298 evaluative[10] studies found across the five websites reported using qualitative research methods. Box 4.4 presents a summary of the qualitative methods reported.[11] The search confirmed that the field of qualitative government evaluative research is dominated by four main methods – depth interviews, focus groups, documentary analysis and observation. The table reflects the spread of these methods used on their own, in combination with each other, and in combination with quantitative methods. Other qualitative methods such as conversation or discourse analysis were not specifically reported, and in only one case were deliberative fora used – in the form of a 'citizen workshop'. Based on those studies where the information was available, a breakdown of the 'type' of evaluation being carried out is presented. Irrespective of the type of study undertaken, the range of qualitative methods remained consistent.

While a breakdown by government departments is not provided here, there was some variation between them in the distribution of methods used. The most common pattern was for qualitative methods to be used in conjunction with quantitative – this accounted for 66–85% of the reports for all but one department (DWP). Furthermore, just one government department (DoH) accounted for almost all of the reports where either documentary analysis or observation was carried out. Across the departments, interviews and focus groups were overwhelmingly the main qualitative methods noted.

> In the light of this review and the views expressed by our interviewees we have decided to adopt a two-tier approach. At a more abstract level, the guiding principles of the framework are applicable across all qualitative methods, but as the framework is worked out in more detail, we focus on four main methods: interviews, focus groups, observations and the use of documents.

---

[10] Evaluative studies were those for which a key word search with the word 'evaluation' produced a hit – 'evaluation' tended to appear either in the report title or research summary.

[11] The summary table of methods used in government evaluative reports presented here is limited. Inclusion in the table was dependent on sufficient information being available on each report. Where adequate descriptions of methods used were not available, this has resulted in gaps in the information provided. For example, some studies gave insufficient information about the combination of methods used and so have been classified as 'other'. The numbers provided are based on the information available in research summaries/abstracts, rather than from thorough examination of each report.

Box 4.4 Summary of templates of government evaluative research reporting to have used qualitative methods

TOTAL: 298 evaluations

	Interviews only	Focus groups only	Interviews and focus groups only	Qualitative case study including interviews/focus groups/site visits	Interviews/focus groups + document analysis	Interviews/focus groups + quantitative methods	Qualitative case study (including interviews/ focus groups/ site visits) + quantitative methods	Other
	41	**5**	**25**	**18**	**1**	**134**	**59**	**15**
Study included observation	7		4	1		38	4	
Study included document analysis	6		2	1		41	11	
Type of study								
Appraisal of specific scheme, project, policy, practice	37	3	20	18	1	123	55	1
General/ background	4	2	2			8	4	
Other			1					

4.4 The applicability of the framework to different stages of research conduct

The final issue we had to resolve about the scope of the framework was whether it could or should apply across all stages of a research study (from proposal or funding application, through conduct, to outputs) or whether it should focus on one stage only.

Our literature review did not give us a specific lead here, except to say that much of the discussion of quality relates to the *conduct* and *reporting* of qualitative research. The existing frameworks (as Chapter 6 discusses) were developed variously for assessing written outputs, conduct or research proposals. Among our interviewees, there were different views about how a framework might be used, although by no means all interviewees felt they themselves would make use of one. The different applications discussed were:

- commissioning – using a framework as a set of issues to be addressed by those tendering for work, as a guide for those adjudicating between tenders, or to be built into specifications and contracts

- refereeing grant applications

- as a training tool – for policy-makers, research managers or researchers

- refereeing papers sent to journals

- an aide for researchers during the conduct of research or for those supervising their research; particularly for inexperienced researchers, but it was noted that it could be a useful 'reminder' to those with more experience

- for peer review of reports at the end of commissioned studies, where this was the practice of a government department.

It was felt that the different applications envisaged had some implications for the type of format or features that would be appropriate. But, more generally, there was some doubt about whether a single framework could effectively address all stages of a research study, spanning commissioning, conduct and evaluation of outputs.

> Given the brief for this research, we decided that we should focus our framework on the written outputs of studies – reports, papers and journal articles. However, as Chapter 2 noted, the framework can also be applied to assessing at least some aspects of research proposals.

5 THE CONTENT OF THE FRAMEWORK: ADDRESSING QUALITY

In this chapter we discuss our decision-making around the content of the framework. Deciding which aspects of 'quality' should be included in our framework was undoubtedly the most difficult task we faced. As we have seen, conceptions of what counts as quality in qualitative inquiry, and how it should be assessed, vary tremendously, depending on the position people adopt in relation to the idea of criteria; underlying philosophical and ideological assumptions; and the particular qualitative method under consideration. In addition, there is the question of whether or not there are other aspects of quality that are relevant when applied to the field of policy evaluation.

We begin this chapter by briefly displaying the diversity of possible criteria relating to quality. We then look in detail at the concepts of validity, reliability and objectivity, describing different definitions and positions, and the complications these raise for assessments of quality. We also discuss criteria relating specifically to evaluations. Finally, we draw together key concerns about quality and the approaches or steps which are frequently proposed to address them. These issues help to explain how we arrived at the content of our framework.

5.1 The diversity of criteria

In order to demonstrate comprehensively the extent of the diversity in criteria, we have only to refer to Patton (2002) and his comparison of '*alternative sets of criteria for judging the quality and credibility of qualitative inquiry*' (p. 544). Patton identifies five different sets of criteria: traditional scientific; social constructionist/constructivist; artistic and evocative; critical change; and evaluation standards. These alternative quality criteria are summarised in Box 5.1.

Our interviews also confirmed that there are differing perceptions of quality among research commissioners, managers and practitioners. However, some key concepts came up recurrently, with much emphasis, and were often the first responses to our initial questions about what matters most:

- validity, reliability and generalisability (some used these terms happily in relation to qualitative research, although often stressing that the meanings are subtly different from those applying to quantitative research; others rejected the terms but felt that at least some of the concepts behind them were applicable)
- methodological rigour and justification
- 'scientific' quality (the rigour of research questions, design, conduct and theorising)
- logic of inquiry

- credibility, or linkage between claim and evidence
- the quality of the theoretical perspective or contribution.

The emphasis given to these concepts was very influential on the structure and content of our framework.

5.2 Simplifying complexity: addressing the 'holy trinity' of validity, reliability and objectivity

In order to avoid being swamped by these competing conceptions of quality and to help us decide which principles or criteria to adopt for our framework, we found that we had to tackle the 'holy trinity' of validity, reliability, and objectivity, and the myriad ways in which these concepts have been conceptualised and redefined. In the course of unpacking these terms we have discovered that:

- Whether people accept or reject the terms validity, reliability, or objectivity is *not necessarily* a good guide as to how they feel about criteria or how they feel quality should be addressed.

- There are a number of underlying dimensions of validity and reliability which, when extrapolated, help to simplify the complexity and show that there is perhaps more consensus about underlying notions of quality than the debate would suggest.

- Varying conceptions of validity, reliability, and objectivity relate to how the inquiry was conducted; the status of claims made as a result of the inquiry; relations with participants; and the impact or contribution of the study.

5.2.1 Acceptance or rejection of terms

The reason that rejection or acceptance of the terms validity, reliability and objectivity is not a good guide to how people feel about criteria is that there is actually a wide range of different positions among 'rejectors' and 'accepters'. So, for example, among those who *reject these concepts* there are those for whom this refutation is linked to a rejection of criteria altogether (Smith, 1984). Then there are those who *reject some concepts but retain others*. In this case, it is usually *some form* of validity which is retained, and reliability (Hammersley, 1992; Harding, 1987; Reichardt and Cook, 1979) or objectivity (Cutcliffe and McKenna, 1999; Sandelowski, 1986) which is rejected. Others *reject all the terms.* However, this does not necessarily mean that underlying concerns about transparency of conduct, thoroughness, rigour, or the defensibility of findings are not considered important (Bailey, 1996). The terms may be rejected because they are difficult to apply in ways that resemble their application in quantitative research, or because they have certain connotations and are associated with a 'positivist' paradigm. Perhaps the most

Box 5.1 Alternative quality criteria

Traditional scientific criteria	Social constructivist criteria	Artistic and evocative criteria	Critical change criteria [neo-Marxist, some feminist]	Evaluation standards and principles
• Objectivity (attempts to minimise bias) • Validity of the data • Systematic rigour of fieldwork practices • Triangulation (for consistency of findings) • Reliability of coding and pattern analysis (multiple coders) • Correspondence of findings to reality • Strength of evidence supporting causal hypotheses • Generalisability • Contributions to theory	• Subjectivity acknowledged and embraced • Trustworthiness and authenticity – fairness and coverage of others' perspectives • Triangulation (for capturing multiple perspectives) • Reflexivity and praxis – understanding one's own background and how to act in the world • Particularity – doing justice to unique cases • Contributions to dialogue – encouraging multiple perspectives	• Opens the world to us in some way • Creativity • Aesthetic quality • Interpretive vitality • Flows from self – embedded in lived experience • Stimulating • Provocative • Connects and moves the audience • Voice is distinct and expressive • Feels 'true', 'authentic', 'real' • Case studies become literary works, blurring of boundaries	• Critical perspective – increases consciousness about injustice • Identifies nature and sources of inequalities and injustice • Represents the perspective of the less powerful • Makes visible the ways in which those with more power exercise and benefit from this power • Engages those with less power respectfully and collaboratively • Builds capacity of those involved to take action • Identifies potential change-making strategies • Clear historical and values context • Consequential or catalytic validity	• Utility – if not going to be useful to some audience, then no point doing it • Feasibility – if not practically or politically do-able then no point • Propriety – fair and ethical • Accuracy • Systematic inquiry • Integrity/honesty; respect for people • Responsibility to general public welfare

Source: Adapted from Patton (2002).

frequently cited example of someone who *apparently* rejects validity while retaining its underlying concerns is Wolcott (1990, 1994). On the one hand Wolcott claims that he cannot see any place for validity in his work but then sets out ways in which he tries to '*satisfy the implicit challenge of validity*' – such things as:

- listening more than talking
- recording accurately
- beginning writing early and sharing ideas with others in the setting
- letting readers '*see for themselves*'
- reporting fully
- being candid
- seeking feedback
- trying to achieve a balance through rigorous subjectivity
- writing accurately.

Similarly, the *acceptance* of the terms validity and reliability may mean that the writer is using them in a fairly traditional way, such as Kirk and Miller (1986) or in a way which retains concerns with rigour or accuracy, such as Hammersley (1992), Silverman (1997) or Merrick (1999). In some cases, however, the meaning of the terms has been *completely redefined*. Taking validity as an example, we discover that this now has a huge array of meanings. Adcock and Collier (2001) found '*37 different adjectives[12] that have been attached to the noun "validity" by scholars wrestling with issues of conceptualisation and measurement*' (p. 530).

5.2.2   *Validity and its many meanings*

To add to the confusion, some writers use the term validity to refer both to the criterion itself and to ways of assessing it, using validity and validation interchangeably. Others reject the term validity but retain the concept of validation, referring to ways of demonstrating 'goodness'. Some alternative conceptions of validity and validation include:

- *descriptive validity* (Maxwell, 1992; Miller and Fredericks, 1995), *descriptive adequacy* (Hammersley, 1991), or *validity at the individual level* (Sykes, 1990), which refer to researchers actually capturing what they intended to study and accurately reporting what they have seen or heard

- *validity of data generation* (Mason, 2002) or *procedural trustworthiness* (Stiles, 1993), which refer to the rigour and defensibility of the conduct of an inquiry

- *validity of interpretation* (Mason, 2002), *theoretical validity* (Maxwell, 1992; Miller and Fredericks, 1995), or *explanatory adequacy* (Hammersley, 1991), which refer to the defensibility of claims said to arise from the research

- *interpretive validity* (Maxwell, 1992; Miller and Fredericks, 1995), which – in this rather different conception – means that the researcher employs an emic

---

[12] The 37 adjectives attached to validity, found by Adcock and Collier (2001), were: 'a priori, apparent, assumption, common sense, conceptual, concurrent, congruent, consensual, consequential, construct, content, convergent, criterion-related, curricular, definitional, differential, discriminant, empirical, face, factorial, incremental, instrumental, intrinsic, linguistic, logical, nomological, postdictive, practical, pragmatic, predictive, rational, response, sampling, status, substantive, theoretical, and trait' (p. 530)

perspective using the participants' own categories for understanding their own behaviour

- *validity as an incitement to discourse* (Lather, 1995), in which the researcher leaves gaps and questions to be answered rather than giving a definite account

- *substantive validation* (Angen, 2000) or *reflexive validity* (Stiles, 1993), in which researchers reflect on and document their relationship to the subject matter; confusingly, this kind of reflexivity is also offered as an alternative formulation for reliability and as a way of dealing with subjectivity in qualitative research

- *ethical validitation*, which requires the researcher to provide '*practical, generative, possibly transformative and hopefully non-dogmatic answers*' (Angen, 2000: 389)

- *dialectic validity*, which recognises tensions within the field and the research process (Waterman, 1998)

- *critical validity*, which looks at the responsibilities of the (action) researcher (Waterman, 1998)

- *pragmatic* validity, where the criterion is whether or not findings of an inquiry are adopted or lead to change (Kvale 1996)

- *catalytic* or *emancipatory validity* (Stiles, 1993), where a study is judged according to the extent to which it promotes greater understanding of their situation among participants, and enables them to bring about change for themselves.

We can see that some of these conceptions of validity relate to *methods* or the *research process* – how rigorously the research was carried out and whether it captured the social world of participants fully and faithfully. Some refer to the *status of the findings* – whether interpretations of the data should be believed or taken seriously. Some refer to the *quality of relations with participants*. And finally some relate to the *impact or contribution* of the inquiry.

Where validity is seen as concerned with **methodological and interpretive issues**, these conceptions actually stem from concerns that underpin traditional 'scientific' notions of validity. As Hammersley (1991) indicates, when Campbell (1957; Cook and Campbell, 1979) adopted the term (internal) validity, he used it in two quite different ways: to refer to whether or not an *instrument is valid* and measures what it purports to measure; and whether or not a *particular measurement is valid*, or a *particular proposition can be made*. This had led to validity being seen as a *procedural* or *methodological issue* concerning the effectiveness of methods, and an *interpretive issue* concerning the '*truth*' or '*credibility*' of empirical findings.

As outlined earlier, there are no standardised instruments or methods in qualitative research and so validity cannot be a property of a particular method. Concerns about methods have consequently been translated into concerns about

rigour, and the idea that this has to be *demonstrated* through careful documentation so that others can see for themselves how the research was carried out. Chamberlain (2000), for example, argues that qualitative research needs to '*explicate how we collected our data – the aim being to demonstrate the care, thoroughness and professionalism of our practices*' (p. 291). Regardless of whether or not researchers subscribe to this as a form of 'validity', there is fairly widespread support for the idea of rigorous and transparent conduct of qualitative research, even if this is based on a range of different ontological and epistemological assumptions.

Turning now to validity in terms of **the truth or credibility of a claim,**[13] we find that there is much less agreement about this, and indeed a great deal of controversy in the qualitative research world. Part of the problem surrounds the question of:

- whether or not researchers believe that an interpretation or knowledge claim can actually be '*true*' in the sense of a direct, or more or less accurate *correspondence* with reality
- whether a claim is considered true if it makes logical sense (a *coherence* criterion of truth)
- if there is agreement that it is true (a *consensus* view of truth)
- if it works in practice and expected outcomes follow (a *pragmatic* view of truth).

Hammersley (1992), for example, believes that it is possible for knowledge claims to be more or less accurate, and consequently sees validity as a question of whether or not a claim represents accurately those features of the phenomena it is intended to describe, explain or theorise about. For him, credibility (whether the evidence for the claim is convincing) and plausibility (how the claim fits with what is already known) are **ways of establishing** rather than **definitions** of validity. Others, however, take a different position from Hammersley, adopting credibility and plausibility as **alternative ways of thinking about validity**, either because they believe it is not possible to establish the 'truth' of a claim in the sense of some kind of correspondence with 'reality', or because they maintain that there can be different understandings of reality (Lincoln and Guba, 1985; Sandelowski, 1986; Whittemore *et al*, 2001).

There is a sense that validity is about communication – Kvale (1989) actually uses the term '*communicative validity*' – and that the reader wants to know how the researcher 'got there'. Bailey (1996) describes validation as a process of persuasion in which the researcher is judged as to whether of not he or she has done a rigorous job, rather than a process of claiming a single truth.

---

[13] For a truly excellent discussion of truth and validity, see Enerstvedt (1989). See also Flew (1977) who argues that it is essential to distinguish between truth and validity. '*What is true, or false, is propositions. What is valid or invalid is arguments. These notions, and these distinctions, are absolutely basic. To say that an argument is true or that a proposition is valid is as uncomprehending, or as inept, as to say that someone played scrum half in a soccer match or scored a lot of tries at cricket*' (p. 9). Hammersley (2003, personal communication) suggests that Flew is talking about validity in deductive logic where validity can be determined formally and there is a clear distinction between truth and validity which is not the case in empirical research.

Interestingly, concerns with credibility bring us back again to **methodological** notions of validity – in the sense of how a study was conducted, which is perhaps why there is so much consensus about the importance of rigour and transparency. Finally, some writers and researchers, such as radical constructivists, do not even believe in the idea of credibility since there are no privileged, only alternative understandings.[14]

Other conceptions of validity, those which focus on **relations with participants**, such as participatory, ethical or dialectic validity, and those which relate to the impact or contribution of research, such as catalytic validity, appear to be based on rather different understandings of the term. Democratic, reformist, or emancipatory versions of validity seem to rest on the notion of validity as *goodness*, or *worthwhileness.* By contrast, validity when defined as the rigour of methods or the status of claims seems to rest on notions of validity as *soundness* or *well-foundedness.*

Different notions of validity were also found among the policy-makers, academics, research commissioners, managers, funders and practitioners interviewed as part of this study. The terms used, and the basic concepts to which they referred, also varied. Some people, while recognising the particular ontological and epistemological foundations of qualitative research, argued that concepts of validity and truthfully representing reality apply to qualitative research as to any form of scientific inquiry.

> *I mean there clearly is a difference between people who are engaged in an activity which says we're trying to construct explanations of the world which are contingent explanations and … what we're trying to do all the time, is to discover whether the ideas we've got are true or not, we're trying to actually offer some kind of true account of what happens in the world. We understand there are all sorts of complexities about meaning, and human action is uniquely difficult. Nevertheless, that's what we're trying to do. And then other people who say what they want is a kind of account of the world which is coherent in some sort of sense, is interesting in some sort of sense but nevertheless is not contingent. Now one for me is social science and the other isn't, and if you're doing something called social science and I think if we're talking about how can we use qualitative methods in the business of government, then I think we're in the business of social science, not at the other end of the extreme.* Research manager

Others, however, preferred notions of 'credibility' and 'trustworthiness', incorporating notions of persuasiveness, believability and convincing accounts,

---

[14] For a lighthearted critique of constructivist evaluation, see Pawson (1996: 214) who describes the way contemporary evaluation 'dances to a variety of tunes, none more recognisable than the constructivist three-step' which involves *seeking out the subject* (capturing the knowledgeable subject in the act of producing change); *sizing up the stakeholders* (searching for points of collective understanding), and *parleying with the participants* (deciphering the standpoints of stakeholders, and generating shared understandings through joint constructions albeit within 'hermeneutic, dialectic circles').

and shored up by openness about methods and how the data has been used to construct conclusions. Plausibility was also discussed, either as part of the assessment of credibility or seen as a different and preferable concept. It involved a judgement of how persuasive or believable the claims made are, or how far they resonate with common sense and with what is already known – a judgement which might be made by the research community, by policy-makers or practitioners, or by others with experience or knowledge of the field.

There was some discussion of the validity of data –whether it conveys the subjective meanings of respondents, whether it captures people's experiences and understandings, whether responses given are '*truthful*' and a situation created in which the respondent is able to be forthcoming. But most of the discussion was about the validity of interpretations and conclusions (discussed further in Section 5.4).

5.2.3   Notions of reliability

Reliability is generally seen as a very difficult concept in qualitative research, especially when it is defined as '*the extent to which studies can be replicated*' (Le Compte and Goetz, 1982: 35). Reliability is problematic partly for methodological reasons. Qualitative research rarely uses any strict 'experimental' controls. The design of a study is often responsive to the setting and therefore emergent and iterative. There are no standardised methods and, indeed, the researcher is considered the primary instrument, so it is difficult to disentangle 'researcher effects' from 'methods effects' (Robson, 1989). Reliability is also problematic for ontological and epistemological reasons. Some researchers believe that the social world and constructions of it are not stable. Some believe that by studying a setting you change it. Consequently the concept of *reliability as replication* is often rejected. However, there are other notions of reliability, whether or not this term is actually used, which need to be considered.

One conception of reliability which is quite widely discussed is the notion of *consistency.* Le Compte and Goetz (1982) give the most extensive account of this approach to the problem of reliability and distinguish between *external* and *internal* reliability. External reliability relates to the issue of '*whether independent researchers would discover the same phenomena or generate the same constructs in the same or similar settings*' (p. 32). Reliability in this sense can be addressed through clear documentation of:

- the status position adopted by the researcher
- the nature of the data sources
- the social situations in which the study was carried out
- the methodology adopted
- the theories and ideas that informed the study.

Internal reliability '*refers to the degree to which other researchers, given a set of previously generated constructs, would match them with the data in the same way as did the original researcher*' (p. 32). Reliability in this sense can be addressed through:

- recording data mechanically

- using low inference indicators
- using a team of researchers
- using participant researchers
- peer examination.

Silverman (2000) discusses reliability in terms of the degree of consistency with which instances are assigned to the same category by the same researcher on different occasions, or different researchers on the same occasion. Some writers think of reliability predominantly in terms of inter-coder reliability (for example, Ambert *et al*, 1995; Armstrong *et al*, 1997).

An alternative conception of reliability is *auditability*, *dependability* or *reflexivity* (for example Lincoln and Guba, 1985; Sandelowski, 1986). Here there is no assumption that another researcher would or could necessarily apply concepts in the same way, or necessarily come up with the same interpretation. But if the researcher gives a clear account of how the research was conducted, readers can see *how the findings were derived*. Sandelowski (1986) lists twelve ways of achieving auditability through description, explanation or justification of:

- how researchers became interested in the subject matter
- how they view the phenomenon being studied
- the purposes of the study
- how participants or pieces of evidence came to be included and how participants were approached
- the impact the participants and the researcher had on each other
- how the data were collected
- how long the data collection period lasted
- the settings in which data were collected
- how data were reduced or transformed for analysis, interpretation and presentation
- how various elements of the data were weighted
- the inclusiveness and exclusiveness of the categories developed to contain the data
- the specific techniques used to determine the truth value and applicability of the data.

The same range of notions of reliability came up in our interviews, although the issue tended to be given less attention than validity (the distinction between the two concepts was not always consistent). Reliability was discussed in varying contexts: the reassurance that another researcher investigating the same issue would derive the same observations, or the same findings, and reassurance that another researcher working with the same data set would derive the same findings.[15]

Where reliability was seen as relevant to qualitative research, the importance of leaving a clear 'audit trail' was stressed, so that another researcher could replicate the research process. However, it was said that it is very rare in practice in social

---

[15] The ESRC Qualidata Archive, for example, makes some qualitative studies available for secondary analysis.

research to attempt to replicate a study, and that the concept is therefore less relevant than in the natural sciences. There was much scepticism about whether it would be realistic to expect replication – especially in data collection as opposed to analysis – given the specificity of the interview in terms of the researcher (their epistemological and ontological perspectives, and their role as data collection instrument), and the relationship between researcher and respondent. At most it was thought that similar rather than identical data and conclusions could be expected. But there was a preference instead for a reflexive and reflective approach to the conduct of research and, overall, reliability seemed not to be seen as a central aspect of quality in qualitative research for the people we interviewed.

5.2.4 The problem of objectivity and bias

While some writers argue that objectivity is a problem for quantitative as well as qualitative research (Scriven, 1972), there is considerable agreement that objectivity is particularly problematic in qualitative research. This can be for methodological reasons – based on the involvement of the researcher in the research setting and inevitable selectivity of data collection, analysis and interpretation. Or it can be for epistemological reasons – concerning the extent to which the researcher and the phenomenon being studied are separable. Chen (1994) argues there are additional issues in evaluation because the evaluator may be torn between advocacy and a more neutral role.

One response to the issue is to consider any notion of neutrality as misguided and to propose that the researcher becomes actively and empathically engaged with participants in the setting, celebrating rather than trying to reduce subjectivity. This position tends to be associated with some feminist methodologies (see Koch and Harrington, 1998; or Marcus, 1994). A more common position is a call for reflexivity on the part of researchers in which they not only document the research process but also declare the values that have guided the research, and discuss the impact of their role and presence and also how the research process has affected them (Lincoln and Guba, 1985; Sandelowski, 1986; Cutcliffe and Mckenna, 1999; Meyer, 2001). Patton (2002) firmly believes that that a clear exposition of the research process is a way of overcoming debates about objectivity since:

> *It opens up the possibility of getting beyond the meaningless abstractions of objectivity and subjectivity and moving ahead to carefully selecting descriptive methodological language that best described your own inquiry processes and procedures. That is, don't label those processes as 'objective', 'subjective', 'trustworthy', 'neutral', 'authentic', or 'artistic'. Describe them, and what you bring to them, and then let the reader be persuaded, or not, by the intellectual and methodological rigour, meaningfulness, value and utility of your result.* (Patton, 2002: 576)

Seale (1999) argues that objectivity is not achievable in the naïve sense of facts being neutrally produced but that it is an attitude of mind on the part of researchers who try and stand back as far as possible in terms of their own values. Hammersley and Gomm (1997) take this somewhat further and focus on the idea

of error and attempts to avoid bias. Because researchers do not have perfect access to the phenomena they are studying, there are a number of different ways in which error can arise and threaten the credibility of a study. They distinguish between systematic and haphazard, culpable and non-culpable error. Bias is then defined as:

> *systematic and culpable error that the researcher should be able to recognise and minimise, as judged by the researcher him or herself (in retrospect) or by others. This then allows us to distinguish between motivated and unmotivated bias, according to whether it stems from other goals than the pursuit of knowledge.* (Hammersley and Gomm, 1997: 9)

Silverman (2000) is particularly concerned about partisan or ideologically motivated research because this can lead to researchers deliberately seeking out evidence to back up their position. Dingwall (1980) argues that the desire to champion the underdog is '*inimical to the serious practice of ethnography, whose claims to be distinguished from polemic or investigative journalism must rely on its ability to comprehend the perspectives of top dogs, bottom dogs, and indeed lap dogs!*' (p. 874).

When we discussed objectivity and bias in our interviews, there were different shades of emphasis with a general view that complete objectivity is not achievable but that steps can be taken to ensure a '*reasonable*' level is attained. Objectivity was said to require openness throughout the research process, not allowing pre-conceptions or '*prior relevances*' to drive the design, data collection and analysis. Particularly in evaluative research, which seeks to produce a collective account by drawing together different actors, there should be openness about which perspective, or set of relevances, is being privileged.

Reflexivity was stressed here too. For the people we interviewed this involved:

- demonstrating an awareness of how biases may emerge
- thoughtfulness about and attempts to minimise the impact of the researcher on the data collected
- attempts to address bias through systematic and comprehensive analysis, and reflectiveness about the research methods, the decisions made, and the consequent limitations of the study.

There was some interest, particularly among academic respondents, in researchers being explicit about their own background, their interest in the research subject, and the assumptions, pre-conceptions or biases they have brought to their research. But there was also scepticism about whether a '*confessional*' statement of the researcher's own characteristics and pre-conceptions was useful, and there was some doubt that research can be '*authenticated by telling a tale about yourself*'.

5.2.5   *Generalisation or 'external validity'*

Whether or not it is possible to generalise from qualitative studies, and whether this should be adopted as a way of judging quality, is widely debated in the

literature. Schwandt (1997), for example, argues that generalisation is impossible because evaluation is context-specific and that there are no context-free meanings. Others, however, argue that generalisation in the sense of wider applicability is not only possible but also desirable (Mason, 2002; Silverman, 2000). While there are a number of different ways in which generalisations can be made, and confusingly these are given rather different labels, there is a common view that the basis of generalisation is different in qualitative as compared with quantitative research and based on *assertional* rather than *probabilistic* logic (Stake, 2000, Kvale, 1996). In other words, generalisation does not rest on statistical representativeness and sampling theory but on careful comparison and unassailable analysis (Mitchell, 1983).

One type of 'qualitative' generalisation described in the literature is *representational* (Lewis and Ritchie, 2003), or generalisation within a case. Even where an evaluator is not concerned to make wider inferences to other settings, there is still an issue about the extent to which the findings – which are inevitably selective in terms of participants interviewed or events observed – can be taken as representative of the phenomenon or setting studied.

Then there are various kinds of *empirical* or *inferential* generalisation in which the findings from one setting are generalised to another. This is sometimes seen as *naturalistic*, resting on personal experience and tacit knowledge. For example, Stake (1978) argues '*what becomes useful understanding is a full and thorough knowledge of the particular, recognising it also in new and foreign contexts. That knowledge is a form of generalisation, arrived at by recognising the similarities of objects and issues in and out of context and by sensing the natural co-variations of happenings.*'

Sometimes it is discussed in terms of careful *case-by-case transfer* which is achieved through 'thick description'. Contexts (and 'treatments' in the case of some evaluation studies) are described in detail and the researcher, or reader, compares the similarities and differences between them and judges whether the findings from one context are applicable to another (Kennedy, 1979; Mitchell and Bernauer, 1998; Lincoln and Guba, 1985; Mason, 2002). Another view of empirical generalisation is given by Hammersley (1992) who describes it as a process by which findings from one case are judged as relevant to a wider group or an aggregate because the case is deemed typical, atypical, extreme or vanguard.

A third broad type of generalisation is *analytical* or *theoretical.* In some cases this is a matter of generating analytical concepts which can then be applied more widely, in different contexts, for example, as in the idea of formal grounded theory (Glaser and Strauss, 1967; Strauss and Corbin, 1998). Other writers are more hesitant about the idea of theoretical generalisation in the sense of generating theory from a single study, and prefer to think of generalising from a case to a theory through the application of analytical concepts or theoretical ideas (Seale, 1999; Mitchell, 1983).

Finally, Schofield (1990) proposes an interesting and rather different perspective, distinguishing between generalisations based on:

- '*what is*', which attempts to establish the typical, common or ordinary, seeking the best fit between what happens in the chosen case and in the wider society

- '*what may be*', where one is generalising from the leading edge in terms of what may happen in the future
- '*what could be*', locating situations which are seen as ideal, for example, cases of good practice, and envisaging possibilities rather than just mapping how things are at the moment.

There were also different perspectives on generalisation in the interviews we carried out. First, although a non-generalisable study may be of value (as an exemplar, because of the intrinsically important nature of the case, or if findings are at least thought-provoking) it was said that most research needs to be justified by the concept of its wider relevance. That the findings have some wider resonance – to the sampled population, or to a different context – was widely seen as an essential element of quality.

> *It seems to me that it's very rare for any research, qualitative research, to have as its end point a description and an understanding of a particular phenomenon or setting. We do it to get insights which we believe are transportable to settings which are similar and measurable degrees different from the objects of study. Therefore it seems to me that everything is about generalisability.* Academic/research practitioner

Theoretical generalisation was also discussed and seen by some as particularly relevant to qualitative research, but there were also some concerns about it. Some people argued that it cannot be a substitute for considerations of the representativeness of the data itself, and that theory, in this context, should be understood as 'theoretical ideas' rather than as 'validated knowledge'. There were also different emphases in terms of whether respondents were envisaging the wider inference being drawn by the researcher, or by the reader, or both. But some discussion of the scope, and limitations, for wider inference was seen as an important aspect of quality.

5.3 Identifying 'evaluation' criteria

Because our brief is to devise a framework for assessing qualitative evaluations we felt it was important to look at some of the criteria and standards which have been drawn up by the 'evaluation establishment'. We review the recommendations of a number of different bodies and committees, all of them American.

In 1982, the Evaluation Research Society Standards Committee proposed 55 standards which were grouped into six main categories:

- Twelve standards referred to *the formulation* and *setting-up* of an evaluation and concerned clarifying the design and inquiry process with the client, safeguarding the interests of participants and resolving any conflicts of interest.

- Six standards related to the *structure* and *design* of the evaluation, including sampling procedures and choice of methods or instruments, to ensure that it could produce defensible inferences.

- Twelve standards concerned *data collection* and *preparation* and referred to training of data collectors, monitoring and documenting data collection, 'controlling bias' and assessing validity and reliability.

- Nine standards dealt with *data analysis* and *interpretation* which basically called on evaluators to choose appropriate analysis procedures and to justify their approach to analysis.

- A further ten standards related to *communication* and *disclosure* and covered issues like determining authority for the release of findings, and presenting findings clearly.

- Finally, six standards referred to the *use of results* and considered the need for evaluators to think about the information needs of users, deal with misunderstandings and be cautious in their recommendations.

The Joint Committee on Standards for Educational Evaluation devised a set of principles in 1980, and a second edition was agreed in 1994. Both versions are based on 30 standards which are grouped under four key categories:

- *Utility* – an evaluation should ensure that it serves the information needs of users.
- *Feasibility* – an evaluation should be realistic, prudent, diplomatic and frugal.
- *Propriety* – an evaluation should be conducted legally, ethically and with respect for the welfare of those affected by the results.
- *Accuracy* – an evaluation should be conducted in a rigorous and well-documented way so that conclusions are defensible, valid and reliable.

Yet another set of guidelines was drawn up by in 1995 by the American Evaluation Society which issued a set of five '*guiding principles*', backed up by a series of underlying '*normative statements*'. The guiding principles included:

- *Systematic inquiry* – evaluations should be based on high technical standards, collaboration with clients to ensure appropriate procedures, and candid communication.
- *Competence* – evaluators should develop their expertise, have appropriate skills, and practice within them.
- *Integrity* and *honesty* – evaluators should be ethical and candid.
- *Respect for people* – evaluators should respect the dignity of participants and other stakeholders.
- *Responsibilities for general and public welfare* – evaluators should think about the impact of the inquiry and its long- as well as short-term effects, and involve audiences who have a right to know the findings.

(For a detailed discussion and comparison of these different professional standards, see Stufflebeam, 2000).

In the interviews, people felt that the quality issues they saw as relevant to qualitative research would apply as much to qualitative evaluations, but some felt that evaluative research also brings forth some additional issues, or places a greater emphasis on those that would apply elsewhere:

- the feasibility of design
- timeliness
- clarity about the basis on which judgement of effectiveness are made
- the utility of findings, the quality and feasibility of recommendations
- the depth of understanding of outcomes
- clarity about the context of the evaluation, and about responsibility for decisions about its aims and goals
- research designs which take into account the perspectives of all actors.

It was also said that, more generally, all aspects of rigour are of greater importance in evaluative studies because of their explicit purpose of informing policy-making, and thus contributing to change which will have real impacts on people's lives.

5.4 Addressing quality criteria

From this review of the literature, with additional material from our interviews, we have identified a number of recurring and underlying themes that underpin notions of quality in qualitative inquiry. Some of them are linked to definitions and redefinitions of validity, some arise out of debates about reliability, objectivity and generalisation, some reflect a concern that evaluation research should be professionally conducted, and some are based on concerns about evaluation as an ethical or moral enterprise. These themes (see Box 5.2), which deal with aspects of the conduct of a study and its outcomes, are discussed below, together with a *brief* account of *some* examples of more frequently proposed ways of addressing them. It should be noted that ways of ensuring quality may be relevant to two or more concerns and therefore certain ones recur in the five sections of the right-hand column of the table.

Box 5.2 Key quality issues and concerns

Underlying concerns in relation to 'quality'	Ways of ensuring quality, drawn from literature review and interviews with research commissioners and managers, funders, policy-makers, academics and practitioners
The defensibility of approach	A clear logic of inquiry Clarity of research questions Theoretical framework/rationale to research questions Design responsive to 'real life' context Defensible rationale for choice of methods – 'fitness for purpose' Defensible sampling strategy Logic and clarity to sample selection criteria Comprehensive and balanced sample coverage Detailed sample profile Consideration of implications of sample coverage
The rigour of	Generation or collection of in-depth data

conduct	Careful recording of data 'Narrative' of data collection process Contextual documentation Systematic and thorough analysis including atypical cases and emergent issues Explication of practical and intellectual analysis process, the origin and application of concepts Approaches to analysis which aid exploration of linkages and depth In-depth interrogation of data, explanatory as well as descriptive Data considered in individual/social/organisational contexts Multiple coding Low-inference indicators Auditing – documenting and discussing the research process and its implications; discussing the values and assumptions that underpin the study The skills of the researcher and appropriate staff qualifications, experience, training and supervision

Underlying concerns in relation to 'quality'	Ways of ensuring quality, drawn from literature review and interviews with research commissioners and managers, funders, policy-makers, academics and practitioners
The relationship of the researcher to the researched	Ethics – showing steps have been taken to secure consent and to safeguard privacy, confidentiality and well-being of participants Participatory and interactive approaches – involving participants in the design, conduct and interpretation of an evaluation Democracy – ensuring that all relevant stakeholders are included and able to take part in the evaluation Reflexivity – showing awareness of the impact of the researcher on the researched, and vice versa; recognising how values, assumptions and presence of researcher may impact on data Fieldwork skills: empathy, sensitivity, rapport, openness to emergent issues Recognition of multiple subjectivities, privileging subjective meanings
The credibility of claims	Triangulation Member checking, respondent validation, validation or corroboration through other evidence, plausibility Peer review Display and explanation of diversity in data Negative cases Alternative explanations Display and interpretation of data which 'feels' balanced, comprehensive, persuasive, and conveys complexity Access to or presentation of original data/giving reader vicarious experience/retaining participants' categories to give reader access Unbiased and balanced selection of original data Indication of how selected cases and illustration link to the data as a whole Demonstrable link between data and conclusions, and between different assertions or conclusions Discussion of how concepts and categories emerged from data Discussion of scope and limitations, displaying doubt and tentativeness
The broader impact and contribution of the study	Relevance and utility of findings to policy Facilitation of utilisation through involvement of users in planning the evaluation, assessing feasibility of recommendations, and dissemination Timeliness Use of research to enhance participants' understanding and facilitate change Enhancement of wider inference through accessible reporting (well structured, coherent, accessible, showing 'flair') and detailed description of: the aims of study and their policy or theoretical context; the research context; the scheme or intervention; the inquiry process; the relationship of sample to the wider phenomenon Discussing conditions of relevance/generalisability Linkage of findings to existing research or theory Making a contribution to theory Generating new understanding or insight, creative interpretation, discussion of patterns and connections between concepts

5.4.1 Defensibility of approach

A recurrent theme concerns the importance of a clear *logic of inquiry* so that a study is designed in such a way as to ensure that it can address its aims (Adler and Adler, 1994; Fournier and Smith, 1993; Mackenzie, 1994; Mason, 2002; Patton, 2002). Some argue that having a clear theoretical orientation to guide the study is also important (Mitchell and Bernauer, 1998; Silverman, 2000; Mason, 2002).

A case should also be made for the *choice of particular methods* which demonstrates their appropriateness (Patton, 2002). There is some concern expressed in the literature about an over-dependence among qualitative researchers on interview data in cases where other methods, such as observation, might be more fitting (Dingwall, 1997; Silverman, 1997), leading to what Atkinson and Coffey (2002) have called 'the interview society'. (See Hammersley (2003) for an account of different ways in which interviews can be used: as witness accounts; as a source of self-analysis; as an indirect source of evidence about people's attitudes or perspectives; and as a '*source of evidence about the construction work of the informant*' (p. 3).)

Apart from the choice of methods, *sampling* is another key component of research design, and many writers call for a clear account of how a 'purposive' or 'theoretical' sample was drawn up, how the design relates to the research questions, how the sample is presumed to relate to a wider population (typical, atypical and so on) and how it will enable key comparisons to be made (Kennedy, 1979; Mason, 2002; Mitchell and Bernauer, 1998; Patton, 2002; Strauss and Corbin, 1998).

These three aspects of approach – logic of inquiry, choice of methods and sampling – were also stressed in our interviews. The logic of inquiry was conceptualised as the quality and clarity of the conceptualisation of the research questions; recognition of the type of data or knowledge that would be required to address them; the use of methods which are able to generate that type of knowledge, and outputs that address the questions. A theoretical framework to the research design was seen as important by research practitioners and academics (noted much less by government research managers and policy-makers). This involves discussion of how theoretical propositions have informed the design of the study and explicit statements of the theoretical basis of the study (which might relate to the conceptualisation of the research issue or to the theoretical basis of the methods). The feasibility of the overall design and its responsiveness to real-life contexts were also seen as important.

Across the groups of people we interviewed, the importance of appropriate choice of research methods was stressed: a '*fitness for purpose*' approach in which methods are driven by the research questions, not by the ideological or practical preferences of researchers. As in the literature, there was some criticism of over-reliance on interviews as a research method, particularly in policy-related research.

The importance of an overall sampling strategy which will provide depth and insight was stressed. What this meant varied. For some it meant building in maximum variation, or scope for comparison between sub-groups or between sites. Others talked about the importance of a strategy that provides maximum opportunity to generalise the research findings. The importance of an underlying logic or theoretical justification for the selection of respondents was stressed: clarity about the criteria or characteristics driving selection, and justification for focusing on them in relation to the research questions. Securing appropriate and defensible sample coverage was seen as critical: a sample which is balanced, diverse and maps the range of the population it seeks to represent. People also talked about the importance of showing the profile of the sample, showing how it relates to the broader population, discussing its representativeness, being honest about groups or dimensions that are missing, and showing awareness of how the achieved sample coverage might limit the claims that can be made.

5.4.2 The rigour of conduct

Full and faithful recording is seen as a key aspect of quality for data which are captured by the researcher, such as interviews, focus groups, and observations. Rich, detailed and complex data are seen as a hallmark of qualitative research (Rubin and Rubin, 1995; Mason, 2002). Ways of ensuring this depth include ,for example, tape and video recordings, full transcriptions, 'thick description', and fieldnote conventions that distinguish description from interpretive commentary (Adler and Adler, 1994; Geertz, 1993; Hammersley and Atkinson, 1995; Le Compte and Goetz, 1982; Peraklya, 1997; Potter, 1996; Silverman, 2000). Documenting the context of data sources and describing case-study settings is also proposed as an important aspect of rigour (Jensen, 1989; Howe and Eisenhardt, 1990; Lincoln and Guba, 1985; Rodgers and Cowles, 1993).

A clear account of the analysis procedure, and in particular how concepts were derived and applied, is a recurrent theme in the literature (Ambert *et al*, 1995; Johnson, 1999; Mackenzie, 1994; Miles and Huberman, 1994; Seale, 1999; Silverman, 2000; Strauss and Corbin, 1998). Some writers advocate that qualitative research should go beyond categorisation, and search for patterns and explanations, using the sampling rationale to identify sub-groups, or by giving an account of local causal processes (Barbour, 2001; Miles and Huberman, 1994; Mitchell and Bernauer, 1998; Shaw, 1999). In the course of the analytic process, some argue in favour of using more than one researcher to categorise data (multiple coders) and recommend that initial categories remain descriptive and loose rather than abstract (Barbour, 2001; Le Compte and Goetz, 1982; Seale 1999).

By far the most frequently discussed way of ensuring rigour and dealing with the issue of subjectivity in qualitative research, however, is through careful documentation of the research process, an account of the researcher's values and theoretical orientation, and an assessment of the researcher's role and impact (Ambert *et al*, 1995; Lincoln and Guba, 1985; Meyer, 2001; Merrick, 1999; Miles and Huberman, 1984; Sandelowski, 1986; Stiles, 1993). This procedure is sometimes referred to as auditability or the audit trail, and sometimes as reflexivity. Some writers, however, maintain that auditing or reflexivity cannot fully capture the researcher's decision making, for example, the role of intuitive thinking or 'hunches' (Cutcliffe and McKenna, 1999).

Finally, for some writers, the skills and craftsmanship of the researcher are a key component of rigour, especially since the researcher is the primary instrument in qualitative research. Training is considered particularly important, and qualitative research is seen to involve considerable conceptual as well as social skill (Kvale, 1996; Mason, 2002; Patton, 2002; Salner, 1989).

Discussions of rigour of conduct in the interviews similarly revolved around data collection, analysis and transparency.

Data collection was seen as a particularly difficult area to make transparent and to judge. The actual collection of data was discussed primarily in relation to conducting interviews, where people stressed the importance of in-depth examination of relevant issues, full probing, openness to unexpected issues and their full examination, and not steering the interview excessively. Transcripts, video and audio recording were seen as important. Having more than one researcher involved in the fieldwork stage – but not *too* many – was sometimes seen as helpful, particularly among government commissioners and managers of research, for reflective learning and to ensure fieldwork was not unduly influenced by the interests of one researcher. But the best measure of the quality of fieldwork was generally said to be the reporting of findings itself – whether they appeared to demonstrate that issues had been explored comprehensively and in depth, whether it appeared to have moved beyond the first level of response given, giving a '*gut feel*' that the research team had got to the heart of the issues.

The judgement of quality was also said to be influenced by how well the researcher has described the conduct of fieldwork: how respondents were approached; where the fieldwork took place; what was prioritised in data collection, how, and how far, respondents were involved in that process of prioritisation; how the data were generated or what one academic described as '*a narrative of how the interviews were conducted*' – how topics were introduced, what stimuli were used by the researchers, what emerged spontaneously and what was prompted or emerged only with probing.

In terms of analysis, the emphasis in the interviews was on systematic and comprehensive coverage of the data: the inclusion of negative or atypical cases in the analysis; checking the fit of hypotheses or theories across the data set and refining them in the light of other cases; considering alternative interpretations; and ensuring that variance in the data set can be explained. It also involved in-depth exploration of and real intellectual engagement with the data, moving between different levels of abstraction, seeing linkages, similarities and differences.

As with data collection, it was said that ultimately the quality of analysis is demonstrated in outputs in two ways: the presentation of findings, and the documentation of the analysis process. In terms of the presentation of findings, the judgement was whether these appear thorough, balanced, persuasive, comprehensive and in-depth. In terms of documentation, a description of the process involved in data analysis was required, but more emphasis was placed on describing the principles involved. This meant explaining how analytical constructions such as categories and codes were formed and used.

The importance of transparency in and thorough documentation of the research process was stressed throughout our interviews. It was seen as important both as itself a feature of good research and to enable others to judge the quality of a study and its findings. Transparency was seen to involve:

- giving '*an honest account*' of the conduct of the research
- a full description of what was actually done and the processes involved in relation to sampling, data collection and analysis
- an explanation for and justification of decisions made through these stages
- discussion of the implications of those reasons, the strengths and weaknesses of the design and conduct
- openness about restrictions on the conduct and compensating features.

This was seen as critical to the 'trustworthiness' of the research and the claims made on the basis of it (see further below)

> *I mean I see it as the researcher–writer role as trying to convince the reader that this is a sensible piece of work, you know, within the constraints of whatever context you're working in and that you've done the best you can and you've thought of these sorts of issues and are coming clean about what you could and couldn't do and you're not making unreasonable claims.*
> Academic/research practitioner

5.4.3   The relationship of the researcher to the researched

Key issues here include the ethical and respectful treatment of participants. There is general agreement that ethical research has to be sensitive to particular settings and cannot be prescribed in the abstract. Some researchers are concerned with issues such as informed consent, privacy, confidentiality, protecting participants from harm, and the integrity of the researcher (Bryman, 1988). Many of the evaluation standards drawn up by professional bodies deal with the rights of stakeholders and the duty of the evaluator to behave in an ethical manner (Stufflebeam, 2000).

Davies and Dodd (2002) call for '*attentiveness, empathy, carefulness, sensitivity, respect, reflection, conscientiousness, engagement, awareness and openness*' (p. 279). The term reflexivity is sometimes used in this context to refer to the active engagement of researchers with participants (Koch and Harrington, 1998; Marcus, 1994).

Other researchers stress the importance of participatory or democratic relationships with participants, for example, Reason and Rowan (1981), and Kushner (2000). Evaluation guidelines emphasise the need to ensure that participants from different backgrounds are consulted and that an evaluation is sensitive to cultural differences.

Schwandt (1996) argues that qualitative inquiry should be assessed according to the extent to which it:

- generates knowledge that complements or supplements lay knowledge

- enhances or cultivates *critical* intelligence in the parties to the research encounter
- enables the training or calibration of human judgement by enhancing the capacity for practical wisdom.

There was little discussion of ethics in our interviews – rather more of reflexivity, as discussed above – although it was sometimes raised as an important aspect of research conduct and one that is intrinsic to quality: *'a good piece of research is an ethical piece of research.'* One policy-maker stressed the value of research that has a positive developmental impact on those participating in it, but did not feel this should be a requirement.

> *Well, a very minimal quality indicator is that it won't give bad experience for the people participating, you know, obviously but it would be very good if it could be a good experience for them … So I certainly would look for that in qualitative research, that the people who are the subjects of it get something out of it and even better if they can be engaged in data collection, you know, in those discussions, because I think that can be really developmental. But I wouldn't put, I think it would be unfair to put it as a requirement but it's certainly a bonus and the requirement is that it's not awful.* Policy-maker

The importance of good fieldwork skills which help to built effective relationships with respondents was stressed in the interviews. This meant listening, sensitivity, awareness, empathy, being able to make attempts to remain neutral, not inhibiting the interviewee, but also being able to shape an interview with diplomacy and tact. Again, it was felt these are ultimately demonstrated in the quality of the written output. Our interviewees, particularly academics, also stressed the importance of recognising multiple subjectivities and privileging the subjective meanings of research participants.

5.4.4   The credibility of claims

According to Hammersley (1990; 1992), defensibility or credibility is judged differently according to the type of claim being made. Descriptions, for example, are judged according to how well concepts or categories fit the phenomena being described. Explanations, on the other hand, account for why things occurred or co-occurred and have theoretical assumptions built into them, so the reader needs to have information about the 'relevances' which guided the selection of explanatory factors, and to see evidence that the co-occurrence cannot be explained by other factors.

Fournier and Smith (1993) and Fournier (1995) argue that claims are not only backed by evidence but also by different kinds of *warrant* or *logic of inference* and that these vary in different kinds of research. The types of warrant used in qualitative research include:

- *structural corroboration* in which pieces of data fit coherently together and support other pieces

- *multiplicative replication* in which others agree and 'see the same thing'
- and *referential adequacy* in which others, when presented with the data, 'can see what the researcher is talking about'.

These different types of logic underpin a number of the ways in which qualitative researchers attempt to establish the 'validity' or credibility of a claim.

One frequently discussed method for strengthening a claim is *triangulation*. It is based on the idea of taking a number of different 'readings', for example, through different methods, data sources, researchers, or theoretical assumptions. There is much debate in the literature, however, about the value of triangulation and the extent to which it is possible to arrive at a single definitive account. Some writers argue that its use lies more in providing a broader, richer account, or even in generating alternative accounts (Fielding and Fielding, 1986; Greene, 1994; Mason, 2002; Miles and Huberman, 1994; Patton, 2002; Seale, 1999; Silverman, 2000).

Another approach is *member checking*, or *respondent validation,* which involves checking findings with the people who took part in the research. Again, there is much debate and disagreement about member checking, since there could be many reasons for participants agreeing or disagreeing with an account that are actually unrelated to its credibility. Reservations about member checking include, for example, the idea that there is no static truth so you cannot 'go back' again, the possibility that participants may not understand technical terms, that their response may be 'politically' motivated, or that there is no intrinsic reason for believing that respondents' interpretations are somehow superior to those of the researcher. Consequently, member checking may simply be a courtesy or a way of gaining a more rounded picture (Abma, 1997; Angen, 2000; Bloor, 1997; Eisner, 1983; Hammersley, 1992; Seale, 1999; Silverman, 2000; Stiles, 1993).

Peer review is also suggested as a validation technique. Although peer review lies at the heart of Hammersley's (1992) notion of plausibility, there are arguments against the effectiveness of this approach. For example, Potter (1996) suggests a number of reasons why peers who have not actually been involved in the research might not agree with or understand its findings, and Sandelowski (1998) argues an external reviewer may judge a study from a different philosophical position or set of values.

There is much greater consensus in the literature, however, about defending claims through the constant comparison of cases and explanations. Particular strategies include: searching for and analysing *negative cases*, *checking the meaning of outliers* and cases which are at the margins of a particular classification, and *demonstrating that alternative explanations have been considered* and found less convincing than the one proposed (Adler and Adler, 1994; Gliner, 1994; Mason, 2002; Miles and Huberman, 1994; Potter, 1996; Seale, 1999; Silverman, 2000).

Giving the reader *access to some of the original data* is a recurrent theme in the literature. Abma (1997) argues that the researcher should give sufficient description so that readers feel they have actually been in the research setting. Glassner and Loughlin (1987) suggest that retaining participants' categories gives the reader access to the participants' social world. However, the most common argument in favour of including some original data in reports is to provide evidence in support of a claim. It is also seen as important to indicate how the

examples or cases were chosen, and how they relate to the data as a whole, to avoid the charge of anecdotalism (Miles and Huberman, 1994; Seale, 1999; Silverman, 2000).

These issues were all discussed in great detail in our interviews. The importance given to the defensibility of interpretation meant that people stressed the importance of showing a '*traceable relationship*' or a '*trail*' between the data and conclusions. It was acknowledged that there will always be a step between them, but that the researcher should demonstrate how they made that step, engaging the reader in a '*journey of learning*' or a '*dialogue of analysis*'.

There was much discussion of the role of verbatim extracts of data. Quotations were seen as having a role beyond assuring the credibility of findings – they were seen as important to illustrate, to provide examples, to convey breadth and range, and to show the nature of the data and the way in which people expressed themselves. But they were also seen as essential 'evidence' for the conclusions drawn, and a distinction was drawn between illustrative quotations and those that are essential to the argument being made. There was discussion about the need to explain the basis for selection of quotations, avoiding over-reliance on just some respondents, tagging quotations with descriptors so that this can be judged, and selecting quotations based on counts.

Establishing the defensibility of interpretation was also helped by thorough analysis of the findings: in-depth discussion of the data, evidence that negative or untypical cases and alternative explanations or perspectives have been included. It was said that with experience the reader develops a '*gut instinct*' that the findings are drawn from the whole data set, rather than reflecting only what the researcher wanted the reader to see. Cogency, a logic in the telling of the account and linkages between assertions were also felt to make research more credible. Expressions of doubt or tentativeness about findings and conclusions added credibility. One academic stressed the linkage between the credibility of conclusions, the quality of analysis and the logic of design:

> *I want to trust the argument, you know … I want to be convinced that I should believe it. So … I want to see some skilful analysis, some analysis that looks meaningful that does seem to be derived from research questions and that is making sense of the sampling decisions and so on… in a sense what I think works best is where it's driven by the argument rather than 'This was the methodology that led to the argument.' So the argument is what you're confronted with really. But then in the process of that you're showing, well as I said, how it applies, whether it applies to everybody, whether it applies to just a group of interviewees or whatever. You know a sense of the researcher getting inside the differences … I'd want to see that it was all joined up really, that the claim came out of a sound argument that was derived from data generated through a good research design, you know.* Academic/research practitioner

Validation and triangulation were also discussed as ways of providing reassurance about the credibility of findings. The more important the decision that

might be made on the basis of the study, the greater the need for validation (particularly by reference to other research), and in this context some government-based respondents doubted that qualitative evidence *alone* would ever be sufficiently strong to have much influence on policy-making.

Most attention was focused on the use of other forms of evidence. This was described variously as the search for correspondence, rather than convergence, between data sets; as the consideration of how the conclusions sit alongside other existing evidence in assessing their robustness; and as the development of a more persuasive picture if a range of perspectives have been taken into account. However, there was some scepticism about whether supporting evidence using one method can be useful in assuring the validity of data from another method, and particularly about what reliance can be placed on data from interviews.

> *The only way in which triangulation can make sense is if we believe there is a decontextualised unitary phenomenon out there that we're going to access. Now … to my mind, that's, as I understand it, part of the logic of quantitative research. It shouldn't be the logic of qualitative research. As I say, because we're so aware of how phenomena get constructed on a term-by-term basis. The idea you could somehow get a better or fuller picture by adding to say what you've got on video, people's comments at the time, is to me very misleading.*
> Academic/research practitioner

Respondent validation was seen as useful to establish the validity of the researcher's constructions and interpretations, as well as being an opportunity to fill gaps in the data collected. But there was scepticism about its value (since people challenge reports of their behaviour for all sorts of reasons) and feasibility. Other forms of validation were also discussed briefly: validation within the data set (repeated patterns, interpretations supported by other parts of the data); theoretical validation; validation within the team; and validation by other experts in the field (the assessment by other researchers or by practitioners or policy-makers of the credibility or plausibility of the findings).

5.4.5   The broader impact or contribution of the study

Hammersley (1992) proposes that qualitative research should be judged in terms of its relevance or contribution to knowledge as well as on the basis of validity. The contribution of some evaluation research depends on the direct utilisation of findings. Patton (1997) argues in favour of *utilisation-focused evaluation* which enhances the application of research and achieves this through the clear identification of end-users and their involvement in designing the study, analysing the data, and generating recommendations.

Some writers judge the contribution of a study in terms of its educative or emancipatory impact on participants. Key considerations here include the accessibility of findings and the degree to which they extend participants' own understandings and enable them to change their situation, (Guba and Lincoln, 1989; Roman and Apple, 1990; Schwandt, 1996).

Not all qualitative evaluation can be, or is intended to be, applied in a such a direct way but requires 'translation' (Hammersley, 2002). An initial consideration here is the way findings are presented and, in particular, the quality of written outputs such as reports, articles or monographs. Finch (1986), for example, argues that reports should be understandable and presented in an eye-catching form. Ambert *et al* (1995) suggest that qualitative research needs to be presented without jargon so that reviewers and readers who may not be familiar with particular methods still find it accessible.

In order to be able to assess the wider relevance of qualitative findings, the aims of a study need to be clearly stated in the report and set in a broader context. This context can be practice knowledge, policy debate, or a set of theoretical assumptions (Bloor, 1997; Mason, 2002; Mitchell and Bernauer, 1998; Silverman, 2000). So that comparisons can be made and transferability or wider applicability can be assessed, reports should also give a detailed description of the setting or case; the relationship of the chosen case or sample to the wider population or phenomenon of interest; and a clear account of how the study was conducted (Kennedy, 1979; Lincoln and Guba 1985; Mason, 2002; Mitchell and Bernauer, 1998; Sandelowski, 1986).

We also explored issues of impact, contribution and relevance with the people we interviewed. The relevance and usefulness of the research questions being addressed was a particular concern to government policy-makers and research managers, who saw meeting the information needs of policy-makers as critical in assessing the quality of reports. The utility of findings – drawing out meaning in policy terms and the feasibility and value of recommendations – was also seen as important here. Timeliness was also stressed by policy-makers and research managers as an aspect of utility.

> *Timeliness is especially critical for a policy research programme. Because we're not just interested in the output, we actually need the output to feed into policy development or policy evaluation. And normally there are very crucial deadlines. And if you miss the deadline … you know, there was no point in commissioning the research in the first place, sometimes, because you've really missed that window of opportunity. And I can't stress that too much to researchers … If you are going to try and effect change then, you know, it's absolutely critical to us that you deliver on time.* Research manager

But it was acknowledged that researchers are not always ideally placed to comment on policy implications, and that the process by which research informs policy is a subtle one involving '*incremental nudges*' rather than immediate translation. There was also some concern that government has unrealistically high expectations of how research contributes to policy. A distinction was drawn between practical evaluation – with a focus on determining what works and whether one approach is better than another – and conditional evaluation. The latter was seen as a more realistic expectation of research:

> *Conditional evaluation says, well, if this is your goal and if these are the criteria that are going to shape your judgement then the*

> *factual information that we have would point in this direction.*
> *That seems to me to be the limit of what a researcher can provide*
> *in terms of the expertise of the researcher, conditional evaluation.*
> *I don't think researchers can provide practical evaluations.*
> Academic/research practitioner

Academics and research practitioners generally put more emphasis on the extent to which studies contribute to theory, and saw this as an important aspect of quality. It meant interpretation which is conceptually and theoretically driven rather than descriptive research or '*mere journalism*'; theories derived from rather than imposed on data; and linking study findings with wider theory and research.

People also stressed the importance of reports providing a richer understanding of a topic or phenomenon and of its relevant discussions and concepts, and providing new understandings and insights.

The quality and accessibility of writing was commented on across all the groups we interviewed: well structured with clear signposting, coherent storytelling; clarity of language; an artistic, creative quality in the use of words, and succinctness.

Finally, in terms of the wider relevance of research findings, two issues were commented on in particular. First was the importance of detail in the account of context and findings, so that wider relevance could be judged by others.

> *It's basically saying there need to be sufficient hooks that the*
> *reader can hang on to from their own relevances, and enough*
> *information there for them to be able to evaluate the strength of*
> *those hooks or the degree of connection of those hooks. And*
> *therefore I think the sorts of information that we therefore need to*
> *convey is enough information that would enable the reader to say*
> *'That is relevant to the setting that I'm now interested in or the*
> *setting that I'm drawing my knowledge from', or 'it isn't'.*
> Academic/research practitioner

Second was to use theory, and a higher level of abstraction in the conceptualisation of findings so that research would have wider applicability and make a more fundamental contribution.

> *I think that qualitative research is required to [use theory] more*
> *explicitly because it is the way we move from the specific to the*
> *general, and although I think that quantitative research should*
> *also be using theory to move from the specific to the general, they*
> *have a stronger argument for doing it probabilistically. I don't*
> *think that's adequate. I think that my judgement would be that to*
> *move on to the kind of high quality research … then you need to*
> *have the theoretical imagination being brought to bear, whatever*
> *your research paradigm. And that for me separates out probably*
> *very good adequate research from excellent research where you*
> *can have a much greater degree of confidence about the potential*

applicability of that beyond the specific. Academic/research
practitioner

Our review of the literature, supported by material from the interviews,
led us to shape our framework on four guiding principles – contribution,
defensibility of design, rigour of conduct, and credibility of claim, from
which a number of appraisal questions and quality indicators are derived.
The literature review and interviews were very influential in shaping the
detailed content of our framework.

Not all the aspects of quality discussed above are included. For example,
our framework does not address democratic, reformist or emancipatory
notions of quality because judgement about these is highly personal,
depending on the researcher's or readers' own values. Only a brief
reference is made to ethics, partly because these are so context specific and
partly because there are already a number of professional codes or
guidelines (for example, those drawn up by the British Sociological
Association; British Psychological Association; Social Research
Association; Market Research Society).

Some aspects of conduct cannot be addressed directly because of our
decision to gear our framework to outputs rather than the research process
itself, so issues like researchers' skills can only be assessed in relation to
the product of a study. Similarly, the feasibility of a study is not discussed
in our framework because this issue would have to have been resolved at
the commissioning or design, rather than output, stage.

Finally, although the timeliness of an evaluation is very important to
policy-makers, it is not something that requires any detailed guidelines
and it is likely to be judged in relation to considerations that lie outside the
quality of the study itself.

Having described in detail the material from the literature review and interviews,
we now turn to our review of existing frameworks, which was also very
influential in shaping the approach to, and content of, our framework.

6 FORMALISED CRITERIA FOR QUALITY ASSESSMENT

At the time of conducting this research, a number of 'frameworks' for assessing the quality of qualitative research were in existence. Twenty-nine of these have been reviewed as part of this study. These are listed in Box 6.1 and a full list of references is given in Appendix 14.

Most of the reviewed frameworks were generated in the last ten years although the first dates back to 1987 (Cobb and Hagemaster, 1987). However, the thinking that underpinned the frameworks had often been developed much earlier, usually by authors who had been formative in developing qualitative research.

The frameworks, as they are collectively termed here, are variously described as 'criteria for evaluation', 'guidelines' for assessment, a 'critical appraisal checklist' or simply 'questions to ask' when appraising or 'making sense of' qualitative studies. But all had a primary concern with identifying features that would define 'good practice' in the conduct of, or outputs from, qualitative research. This chapter presents an overview of the origins and content of these pre-existing frameworks and their surrounding commentary.

6.1 Background to the frameworks

6.1.1 The origins of the frameworks

A high proportion (25) of the frameworks were developed in the fields of medical or health services research (particularly nursing research) or related disciplines (e.g. medical sociology, clinical psychology). The reasons why such attention should be given to the 'quality' of qualitative research in this arena were discussed by many of the authors. They included the rapid growth in the use of qualitative methods in the health field without agreed quality criteria (Devers, 1999; Mays and Pope, 2000; Popay *et al*, 1998); the difficulty of placing articles on qualitative research in medical and health related journals and the need for more informed reviews (Blaxter, 1996; Britten *et al*, 1995; Elliott *et al*, 1999; Hoddinott and Pill, 1997); and, more recently, the need for critical review of qualitative research studies to inform evidence-based practice (Long and Godfrey, 2003; Campbell *et al*, 2003).

But often these reasons were underpinned by the more fundamental need to provide greater understanding of the nature and practice of qualitative research in an arena which, previously, had been dominated by quantitative research (Boulton and Fitzpatrick, 1997; Inui and Frankel, 1991). In particular, there was concern that, in the absence of such understanding, qualitative research studies were being judged only by quality standards applicable to quantitative or 'scientific' research (Forchuk and Roberts, 1993; Muecke, 1994). Not only were many of these wholly inappropriate for qualitative research, but they were also inevitably causing qualitative studies to 'fail' on quality grounds (Beck, 1993).

Box 6.1 Existing quality frameworks (2002)

Developers	Originating field	Assessment purposes for which developed	Orientation	Significant features
Cobb and Hagemaster 1987	Nursing	Proposals	Empirically based	*One of the earliest attempts to formalise criteria in the health field*
Howe and Eisenhart 1990	Education	Qualitative (and quantitative) studies	Philosophically based: generic	*Proposes a set of 'logics in use' criteria that would apply across the full range of qualitative and quantitative methods*
Inui and Frankel 1991	Medical	Journal articles/papers/ reports	Empirically based	
Beck 1993	Nursing	Validity and reliability of qualitative studies	Philosophically based: generic	*Focuses on credibility, fittingness and auditability*
Forchuk and Roberts 1993	Nursing	Journal articles/papers/ reports	Philosophically based: generic	
Kuzel, Engel, Addison, Bogdewic 1994	General practice / family medicine	Qualitative studies	Empirically based	*Identifies a series of questions for researchers to address in carrying out research as well as features of good qualitative inquiry*
Leininger 1994	Nursing	Qualitative studies	Philosophically based: generic	*Identifies criteria that have been developed over 'three decades' and widely used by students*
Muecke 1994	Health	Ethnographic studies	Philosophically based: specific	*Compares criteria for anthropologic and health sciences ethnographies*
Britten, Jones, Murphy, Stacy 1995	Primary care	Qualitative studies	Empirically based	
Miller and Fredericks 1995	General scholarship	Qualitative evidence	Empirically based	*Identifies 'rules' for assessing qualitative evidence and questions to aid researchers in more self-reflection*
Secker, Wimbush, Watson and	Health promotion	Qualitative studies	Philosophically based: specific	*Compares criteria for sociological and marketing approaches to qualitative research*

Milburn 1995				
Blaxter 1996	Medical sociology	Journal articles/papers/ reports	Empirically based	*Developed in liaison with the Medical Sociology Group of the British Sociological Association*
Boulton, Fitzpatrick and Swinburn 1996	Health	Systematic review of journal articles	Empirically based	*Framework used to evaluate papers from medical journals which had used qualitative methods*

Developers	Originating field	Assessment purposes for which developed	Orientation	Significant features
Boulton and Fitzpatrick 1997	Health	Journal articles/papers/ reports	Empirically based	
Drisko 1997	Social work	Qualitative studies	Philosophically based: generic	*Designed to aid assessment in relation to qualitative paradigm*
Greenhalgh and Taylor 1997	Medical	Journal articles/papers/ reports	Empirically based	
Hoddinott and Pill 1997	General practice	Review of qualitative articles	Empirically based	*Designed to assess qualitative interviewing methodology*
Murphy, Dingwall, Greatbach, Parker and Watson 1998	Health technology assessment	Qualitative studies	Philosophically based: specific	*Set in the context of an extensive review of the literature on qualitative research methods.*
Popay, Rogers and Williams 1998	Health	Standards for systematic review	Philosophically based: generic	*Designed to assess specific ontological contribution of qualitative research*
Elliott, Fischer and Rennie 1999	Clinical psychology	Journal articles/papers/ reports	Philosophically based: specific	*Identifies 'publishability guidelines' shared by qualitative and quantitative and those specifically pertinent to qualitative research*
Healy and Perry 2000	Market research	Validity and reliability of qualitative research	Philosophically based: specific	*Proposes criteria for qualitative research in 'realism' paradigm and compares with criteria for other paradigms*
Mays and Pope 2000	Health	Qualitative studies	Philosophically based: generic	*Focuses on quality criteria concerned with validity and relevance*
Treloar, Champness, Simpson and Higginbottom 2000	Clinical epidemiology and public health	Qualitative studies	Empirically based	*Devised to aid assessment of qualitative studies by clinical epidemiologists*
CASP	Health	Qualitative	Empirically	*Designed as a questionnaire*

		studies	based	*to be used for assessment*
Waterman, Tillen, Dickson and de Koning 2001	Health	Action research,[16] studies – reports and proposals	Philosophically based: generic	*Presented in the context of an extensive literature review on action research*

---

[16] This framework was included because although action research is not method-specific, a high proportion is qualitative in form.

Developers	Originating field	Assessment purposes for which developed	Orientation	Significant features
Whittemore, Chase and Mandle 2001	Nursing and health	Validity of qualitative research	Philosophically based: specific	*Analyses the development of validity criteria by formative authors*
Sandelowski and Barroso 2002	Health	Articles/papers/ reports	Philosophically based: generic	*Argues for a shift in the debate on quality away from epistemic criteria to aesthetic and rhetorical concerns*
Campbell, Pound, Pope, Britten, Pill, Morgan and Donovan 2003 – in press	Health	Systematic review of articles/papers/ reports	Empirically based	*Applied criteria to a series of studies on diabetes as part of a systematic review*
Long and Godfrey 2003 – in press	Health and social care	Systematic review of articles/papers/ reports	Philosophically based: generic	*Applied criteria to a series of studies as part of a systematic review of rehabilitative service interventions and their outcomes*

The four frameworks that were prepared outside of the health-related arena had a variety of origins. Drisko's (1997) criteria were developed for evaluating qualitative social work research, for much the same underlying reasons as given by health researchers. Howe and Eisenhart (1990), whose framework is generated within the arena of education research, wanted to develop standards that would rise above the methodological and philosophical debates that dominated quality assessment in that field. Healy and Perry (2000), whose framework focused on marketing research, were specifically concerned with quality criteria for assessing the validity and reliability of qualitative research. Finally, Miller and Fredericks (1995) produced criteria for assessing 'the context of justification' for general scholarship.

6.1.2 Definitions of qualitative research

A number of the authors give definitions of qualitative research in their framework articles. These range from fairly brief descriptions of the methods and approaches used in qualitative research (Boulton and Fitzpatrick, 1997; Greenhalgh and Taylor, 1997; Inui and Frankel, 1991; Kuzel *et al*, 1994; Long and Godfrey, 2003) to more detailed expositions of its foundations and philosophical

perspectives (Devers, 1999; Elliott *et al*, 1999; Healy and Perry, 2000; Popay *et al*, 1998). In other cases, authors defined a particular form or method of research involving a qualitative approach (for example, ethnography (Muecke, 1994); action research (Waterman *et al*, 2001)). Whatever the level of the definition, there is general agreement that qualitative research encompasses a wide array of methods, traditions and approaches.

On the question of definition, some of the authors note that there is considerable misunderstanding about what constitutes qualitative research in the health field. As Cobb and Hagemaster (1987) state, definitions can include '*anything without numbers*'. This misunderstanding is further exemplified by Boulton *et al* (1996) who were among the few framework developers to have systematically applied their quality criteria (see below). Boulton *et al* (1996) looked at 210 papers that had appeared in medical journals which claimed to report a qualitative study or to use qualitative methods. They *excluded* 160 of these on the grounds that the reported studies '*would not normally be considered qualitative*' (p. 172). There was a range of reasons for these exclusions: studies which had used semi-structured or open questions but only to rate respondents on pre-determined scales for subsequent quantitative analysis; surveys which had included some open-ended questions; and research which had used qualitative methods only to develop instruments for later large scale studies but not reported the qualitative component.

6.1.3   The purpose of the frameworks

The frameworks differ in terms of the purposes for which they were derived and this inevitably affects the focus of the items included and the way the quality criteria are specified. Some were developed for assessing the quality of the study itself (for example, Beck, 1993; Mays and Pope, 2000; Treloar *et al*, 2000) or its validity or reliability (Healy and Perry, 2000; Whittemore *et al*, 2001); a number for assessing written reports, papers and articles (for example, Elliott *et al*, 1999; Forchuk and Roberts, 1993; Greenhalgh and Taylor, 1997; Miller and Fredericks, 1995; Sandelowski and Barroso, 2002); and some for reviewing research proposals (Cobb and Hagemaster, 1987; Waterman *et al*, 2001, who covered both reports and proposals).

Among those concerned with the outputs of the research, five (Boulton *et al*, 1996: Campbell *et al*, 2003; Hoddinott and Taylor, 1997; Long and Godfrey, 2003; Popay *et al*, 1998) had a further specific interest in developing standards for the systematic review of articles and written papers within the context of health research. There is a growing literature on the development of evaluation or appraisal 'tools' for such purposes although it is noted by some to be still 'in its infancy' (NHS Centre for Reviews and Dissemination, 2001)

Although a number of the authors saw quality frameworks as having considerable value for general teaching and educational purposes, their primary use is likely to be for the assessment of written output of qualitative studies. Indeed, Sandelowski and Barroso (2002) argue that the report itself is the only means through which research studies, whether qualitative or quantitative, can be evaluated:

> *The 'production of knowledge' cannot be separated from the 'communication of knowledge' by which 'communities' of responsive readers are created and then come to accept a study as valid (Shapin, 1984: 481) … although we tend to distinguish between epistemic and aesthetic criteria, they are in practice indistinguishable as the sense of rightness and feeling of comfort readers experience reading the report of a study constitute the very judgements they make about the validity or trustworthiness of the study itself (Eisner, 1985).* (Sandelowski and Borroso, 2002: page unavailable, web text)

In the context of discussing the assessment of research outputs, the length of journal articles was frequently raised. Some authors were of the view that it was difficult to achieve the full range of 'quality' features when reporting qualitative research in articles or shorter papers. Others felt that with thought, effective presentation and attention to the significant features of qualitative research, essential evidence for appraisal could be provided:

> *In many instances, researchers find it difficult to report sufficient qualitative raw data to allow readers to form their own interpretations. In a print-format, raw data such as texts and memos can require considerable 'page space'; nonprint reports such as audio/videotapes of participant observation can be a technical challenge to produce and to integrate with a manuscript. All too often these editorial challenges lead authors to offer a single illustration in the participant's own words. With limited raw data, readers have neither opportunity to ascertain the accuracy and completeness of the researcher's interpretations, nor a basis on which to develop their own interpretations.* (Drisko, 1997: 3–4)

> *All of these desiderata are, without doubt, difficult to achieve. It is no accident that the basic 'unit of scholarship' in qualitative research is considered by many to be the book, not the article. A book is the meal; articles no more than tasty hors d'oeuvres. Whatever the degree of difficulty, we nevertheless believe that meaningful material from qualitative research can be presented in journal article length.* (Inui and Frankel, 1991: 486)

6.1.4 The status of the frameworks

Few of the authors claimed that the criteria they had developed were either definitive or comprehensive. Indeed, many of them were at pains to explain that the quality standards they were describing were 'evolving' or 'in developmental form' or, as Sandelowski and Barroso (2002) describe it, 'work in progress':

> *Our own view and that of a number of individuals who have attempted, or are currently working on, the task (of developing critical appraisal criteria) is that such a checklist may not be*

> *exhaustive or as universally applicable as various guides for*
> *appraising quantitative (medical) research, but that it is certainly*
> *possible to set some ground rules … you should note, however*
> *that there is a great deal of disagreement and debate about the*
> *appropriate criteria for critical appraisal of qualitative research,*
> *and the ones given here are likely to be modified in the future.*
> (Greenhalgh and Taylor, 1997: page unavailable, web text)

> *This contemporary synthesis of validity criteria in qualitative*
> *research facilitates the decision-making process for investigators*
> *and the evaluative process for consumers of research. Further*
> *development of validity criteria requires ongoing dialogue.*
> (Whittemore et al, 2001: 535)

Such views are clearly a reflection of the relative recency of formalised assessment standards for qualitative research and the ongoing debate that surrounds it.

Perhaps the most assured claim to a definitive list of criteria came from Leininger (1994) who identified and defined '*six central and important criteria to evaluate qualitative paradigm studies*' (p. 104). She noted that these had been developed over time and '*used by a number of nursing and anthropology students for several years*' (p. 104). Leininger also sees it as important that these are '*applicable to all research methods used within the qualitative paradigm*' (p 104), a feature that is discussed more fully below.

Some of the more recently developed frameworks also make some tentative claims towards comprehensiveness. Long and Godfrey (2003), for example, state that their appraisal tool '*aims to be comprehensive*' (page unavailable, in process of publication) in including both characteristics of the study and how it was done through various stages of the qualitative research process. However, it is recognised by Long and Godfrey, as indeed by the current research team, that the learning and wisdom provided by early framework developers has been formative in the process of more definitive collections.

6.1.5   Application of the frameworks

Relatively few of the framework developers provided explicit evidence of the systematic application of their criteria to qualitative studies. Indeed, some of the developers suggest that testing of their criteria are needed to further extend or refine the criteria suggested (Waterman *et al*, 2001).

Among those who reported on the application of their criteria, three had used a framework as an aid for systematic reviews of qualitative evidence (Campbell *et al*, 2003; Long and Godfrey, 2003; Sandelowski and Barosso, 2002). Boulton *et al* (1996) used their list of 'criteria of good practice' to review the impact, nature and quality of qualitative research articles reported in medical journals. Hoddinott and Pill (1997) used a more limited set of criteria concerning qualitative interviewing to assess quality of practice.

There are a number of specific results of interest but there were recurring conclusions about the quality of reporting. In particular, it was noted that reporting of methods or 'the way the study was done' was limited or incomplete (Campbell *et al*, 2003; Long and Godfrey, 2003; Hoddinott and Pill, 1997); and that there was often a failure to adequately report the processes involved in analysis and interpretation (Boulton *et al*, 1996). Similar findings were reported by O'Connor *et al* (2001) who assessed qualitative articles in four pastoral counselling journals using the criteria developed by Mays and Pope (2000). Levels of agreement in the assessments made were variable –this was in part attributed to the 'poor' quality of methodological reporting (Campbell *et al*, 2003).

6.2 Philosophical influences on the development and form of frameworks

Chapter 4 of this report describes the wide-ranging philosophical debates that have surrounded qualitative research. These debates have particular salience in the context of quality standards, both in terms of whether quality criteria can be set or agreed at all and in the form they should take. This section provides a brief overview of where the framework developers positioned themselves in relation to such questions.

6.2.1 Ontological and epistemological orientations

The authors varied considerably in the extent to which they discussed the ontological and epistemological bases of qualitative research and the extent to which these had influenced the quality criteria developed. At one end of the spectrum were those who detailed the philosophical roots of qualitative research and saw these as formative to the quality criteria they defined. Indeed, some were critical of guidelines that ignored the ontological origins of qualitative research and the nature of the knowledge it sought to provide:

> *The debates that have taken place concerning standards in qualitative research have tended to focus almost exclusively on techniques, with little attention being paid to differences in the nature of knowledge that sociologically informed qualitative research entails and the philosophical underpinnings of the methods being deployed.* (Popay *et al*, 1998: 342)

At the other extreme were those who noted that qualitative research is based on different principles and methods than quantitative research – and therefore needs different criteria to assess it – but without explicit discussion of its philosophical base:

> *This single phrase ['qualitative research'] apparently embraces all those instances of inquiry in which investigators adopt no metric and therefore do not summarise their results by measuring or counting. Some have*

> *described this domain of inquiry as meaning rather than theory driven.*
> (Inui and Frankel, 1991: 485)

6.2.2   Whether quality criteria for qualitative research can be defined

As might be expected, a number of the authors referred to the long-standing debate about whether or not it is possible to determine 'agreed' quality criteria for qualitative research (see Chapter 3). Some of the authors summarised the ontological origins of this debate; others referred to the uncertainty about whether any criteria could be agreed; yet others noted concerns that prescribed criteria might act as a straitjacket and inhibit creative and imaginative uses of qualitative research. Although, by definition all of the framework developers had eventually concluded that quality criteria could be identified, many of them expressed some reservation. In particular, it was emphasised that guidelines should not be rigidly or prescriptively applied (for example, Elliott *et al*, 1999; Miller and Fredericks, 1995: Waterman *et al*, 2001) and must '*resist the temptation to provide a rigid checklist of rules that qualitative research must follow if it is deemed to be valid*' (Murphy *et al*, 1998: 178). Similarly, others referred to the lack of agreement about what the core quality criteria should be and warned against overly rigid subscription to any one set of criteria:

> *The hotly contested debate about whether quality criteria should be applied to qualitative research, together with the differences of view between 'experts' about what criteria are appropriate and how they should be assessed, should warn against unthinking reliance on any one set of guidelines.* (Mays and Pope, 2000: 98).

A number of other authors share concerns about the potential effects – or dangers – of formalised guidelines and checklists. Barbour (2001), for example, summarises her main concerns as follows:

- *Checklists can be useful in improving qualitative research methods, but overzealous and uncritical use can be counterproductive.*
- *Reducing qualitative research to a list of technical procedures (such as purposive sampling, grounded theory, multiple coding, triangulation and respondent validation) is overly prescriptive and results in 'the tail wagging the dog'.*
- *None of these 'technical fixes' in itself confers rigour; they can strengthen the rigour of qualitative research only if embedded in a broader understanding of qualitative research design and data analysis.*
- *Otherwise we risk compromising the unique contribution that systematic qualitative research can make to health services research.* (Barbour, 2001: 1115)

Similarly, Williams (2001) argues that the constant search for rigour merely results in longer and longer checklists and he questions whether this is the way qualitative research wants to go. He notes that historically qualitative researchers have addressed the issue of rigour through the '*important process of documenting reflection*' (p. 514). Power (2001) suggests that qualitative researchers should not be '*shielding behind a protective wall of checklists and quasi-paradigmatic research techniques*' (p. 514).

In an article exploring the potential effects of quality frameworks, Chapple and Rogers (1998) share these concerns. But they also worry about the danger that people might be discouraged from attempting qualitative health studies '*because of a proliferation of criteria and standards that emphasise either the need for sociological and anthropological theory, or the need for computer skills or other techniques such as conversation analysis*' *(*p.560). They also suggest that formalised criteria might make it more, rather than less, difficult to get journal articles published because of the level of detail needed for their assessment.

In contrast to such positions, there were those who were fairly strident in their call for quality criteria for qualitative research. Elliott *et al* (1999), for example, suggest three specific reasons why guidelines are needed for qualitative research surrounding clinical psychology. These are, first, 'to help to legitimise qualitative research'; second, to 'foster more valid scientific reviews of qualitative research'; and, third, to improve the quality of the research being conducted. Others, like Leininger (1994), argue that the generation of quality criteria relevant to qualitative research is long overdue and that without them '*we will continue to have noncredible, inaccurate, and questionable findings for qualitative studies*' (p. 95). And several authors note the importance of having standards relevant to qualitative research so that invalid criteria from other research approaches are not applied inappropriately '*in the absence of any attempt to develop standards, there is a danger that qualitative research evidence will be misunderstood and judged inferior by those whose field of vision is firmly fixed on a hierarchy of evidence that makes randomised control trials (RCTs) the gold standard*' (Popay *et al*, 1998: 342).

6.2.3   The need for paradigm specific criteria

Framework developers were divided about whether or not quality criteria for qualitative research should be paradigm specific. Although there was virtually universal agreement that qualitative research could not – and should not – be judged by quality concepts developed for quantitative research, there was variation in view about the extent to which 'conventional' quality concepts can be translated for qualitative research. In such discussions, reference is often made to the differing schools of thought that appear in the wider theoretical literature (see Chapter 3).

Not unexpectedly, the views of individual framework developers are strongly related to how they perceive the influence of ontology and epistemology on the formulation of quality criteria. Leininger (1994), for example, who identifies six qualitative specific criteria for appraising qualitative research, argues the case as follows:

> *It is imperative to use qualitative evaluation criteria to assess qualitative methods within the qualitative paradigm … qualitative researchers should not rely on the use of quantitative criteria such as validity and reliability to explain or justify their findings. Such dependence reflects a lack of knowledge of the different purposes, goals, and philosophical assumptions of the two paradigms. Because the paradigms are so radically different, a misuse of criteria of each paradigm poses critical problems and*

greatly curtails the development of credible and valid outcomes.
(Leininger, 1994: 96).

Other developers who saw a need for paradigm specific criteria also avoided the use of quantitative criteria in their frameworks – and most particularly the use of concepts such as validity, reliability, generalisation and objectivity (Elliott *et al*, 1999; Forchuk and Roberts, 1993; Popay *et al*, 1998; Secker *et al*, 1995). For example, Beck (1993) explains:

> *In this article, the quantitative terms of internal validity, external validity, and reliability are referred to as credibility, fittingness and auditability, respectively. Guba and Lincoln (1981) suggested the renaming of the scientific terms to these naturalistic terms, which are more appropriate to qualitative research.* (Beck, 1993: 264)

In contrast, others saw it as essential to the development of standards that such concepts remain central to notions of quality in qualitative investigation. Within this latter group, most of the authors re-conceptualised or interpreted quantitative concepts in qualitative terms (Blaxter, 1996; Mays and Pope, 2000; Cobb and Hagemaster, 1987).

Among those who were not in favour of paradigm-specific criteria for qualitative research, some authors emphasised the importance of commonality of quality concepts for enhancing understanding of qualitative research among the wider research community. Indeed, some felt that it was the avoidance of quantitative terms that led others to believe that there was no interest in the underlying concepts they portray – and certainly did not help communication between researchers from different backgrounds:

> *While we agree that there is a need for a set of criteria which do not attempt to force qualitative studies into a quantitative framework, we also agree with Agar (1980) and Miles and Huberman (1984) that qualitative researchers must make their work understandable to the 'dominant culture' in the scientific community. Our major evaluative categories, then, do not depart greatly from what might be found in any scientific proposal; however, the questions forming the content of each category address the concerns unique to qualitative research.* (Cobb and Hagemaster, 1987: 139)

6.2.4   Paradigm generic versus tradition specific criteria

A further division arises over whether criteria can be developed to cover the full range of qualitative approaches or whether they need to be specifically related to their underlying philosophy. Some of the framework developers were of the view that it was possible to develop criteria for the qualitative research paradigm as a whole. In some cases this was based on a belief (either explicitly or implicitly stated) that there were sufficient shared assumptions across different qualitative

traditions for common quality criteria to apply. While these authors recognised that different qualitative movements define qualitative objectives in different ways, it was felt that there was enough unity for common standards to be derived (Beck, 1993; Mays and Pope, 2000). In other cases, authors took the more insistent line that criteria *needed* to be defined at a level at which they could be generically applied (for example, Cobb and Hagemaster, 1987). Whittemore *et al* (2001) achieve this through defining primary and secondary criteria, the former having general application to qualitative research, the latter varying with the approach or method chosen.

Other authors who generated generic criteria for qualitative research saw the responsiveness of the study to its philosophical origins as one of the key assessment criteria (Drisko, 1997; Forchuk and Roberts, 1993; Long and Godfrey, 2003; Popay *et al*, 1998; Sandelowski and Barroso, 2002; Waterman *et al*, 2001). It was therefore a factor built into their frameworks for appraisal:

> *Once a paradigm has been selected, we should evaluate the merits of the research assumptions and goals within the philosophy/epistemology that underpins the paradigm. Once a method has been selected, we should evaluate the merits of the data collection and analysis within the parameters of the chosen paradigm. Overall evaluations have their place, but these should not be confused with evaluations within specified parameters and purposes.* (Drisko, 1997: 185)

> *A problem with identifying determinants for guidance on the assessment of action research proposals and reports is that it can easily develop into a technical exercise, and the philosophical persuasions that lie behind action research, such as social improvement, can quickly become neglected. It is important for action researchers to stipulate their philosophical approach and to indicate how it has influenced the development of their action research project.* (Waterman *et al*, 2001: 44)

In contrast, others argued that qualitative research does not represent a single paradigm and that different approaches and traditions within it require different quality criteria. However, one of the difficulties that arises in identifying quality criteria for different qualitative traditions surrounds how the varying paradigms or approaches are conceived and defined. Nevertheless, among those who see the need for distinctive criteria, some attempted to define criteria for different traditions. For example, Devers (1999), in an article commenting on frameworks, compares the quality criteria that have been suggested for 'positivist' and two 'post-positivist' traditions; Secker *et al* (1995) consider differences in quality standards for sociological (based on interpretivism) and marketing (based on empiricism) approaches. Others selected a specific tradition within which they defined quality criteria (for example, Healy and Perry (2000) and Murphy *et al* (1998) define criteria broadly within the 'realism' paradigm); or they discussed how the tradition or approach being used would – or might – effect quality criteria (for example, Elliott *et al*, 1999; Whittemore *et al*, 2001). Yet others were less specific about how quality criteria might vary but defined what they saw as the

key traditions (for example, Healy and Perry (2000) identify four 'scientific research paradigms' relevant to qualitative research: positivism; critical theory; constructivism; and realism).

In the view of some authors, this plethora of approaches and movements and their varying basis for definition has led to uncertainty about the parameters of qualitative research and the core principles on which it is based. In the eyes of some, this has also impeded any attempts to develop quality criteria:

> *Authors often vacillate between the philosophical and methodological meanings [of qualitative research] without being clear about the one to which they are referring or without specifying the relationship between the two. As discussed further on, the link between paradigms, theoretical perspectives and particular qualitative methods is real but imperfect.* (Devers, 1999: 1155)

From a more empirical viewpoint, Treloar *et al* (2000) argue:

> *The formulation of standard conventions for assessing qualitative research has been hampered by the variety of theoretical approaches (and their preferred methods) which inform empirical studies.* (Treloar et al, 2000: 347)

A similar view is expressed by Howe and Eisenhart (1990) who are concerned with standards for educational research. They argue that the educational field requires 'a division of labor' that requires various forms of expertise '*Such a division can take two forms; give-and-take collaboration or insular fragmentation*' (p. 8). In order to promote the former, they have developed a set of 'logics in use' criteria that would apply not only to all qualitative traditions but also to quantitative research.

6.2.5   *Methodological specificity*

As is indicated in the discussion above, there is some link between qualitative traditions and the different methods they use, although this is not always explicit. However, in some cases, authors developed frameworks specifically for particular methodological approaches. Muecke (1994), for example, developed criteria for the evaluation of ethnographic research, distinguishing as she did so between what she terms anthropologic and health services (or focused) ethnographies. Yet others alluded to the need for standards which were related to different methods although without specifying how these might be constructed. For example, Elliot *et al* (1999) suggest that methods such as empirical phenomenology, grounded theory, discourse analysis, or conversation analysis might need the development of specialised guidelines.

6.3 The content of the frameworks

This section provides an overview of the content of the frameworks reviewed. In particular, it compares the different approaches taken by framework developers and highlights some of the distinctive features of individual sets of quality standards.

6.3.1 The philosophical base of the frameworks

The extent to which explicit attention is given to the philosophical base of the framework has a key impact on the nature and formulation of quality criteria identified. The stance taken by the authors was identified in Box 6.1, defined as either philosophically or empirically based. **Philosophically based** frameworks are distinguished by explicit attention to the ontological or epistemological base of qualitative research in their development. **Empirically based** are those where there is little reference to the philosophical base of qualitative research in their development, and there is more emphasis on features of sound investigation. Within the former group we have further identified those who see criteria as specific to different approaches within qualitative research – termed **specific**; and those which have generated quality criteria for the qualitative paradigm as a whole – termed **generic**.

6.3.2 Specification of methods

As previously noted, there is relatively little explicit reference to the methods for which the framework applies. It is, however, sometimes implicit, either through the philosophical orientation of the framework or through the nature of the criteria identified. For example, a number include specific criteria concerning generated data (for example, interviews, focus groups) but, make little explicit reference to other forms of data collection (for example, documentary analysis, observation).

In the context of such discussion it should be noted that some authors made specific reference to quality features that would apply to *any* social research study, irrespective of the paradigm within which it is conducted. These surround issues like clarity of purpose, effective matching between aims and design, clarity and coherence of written presentation and so on. For example, some of the authors (Elliott *et al*, 1999; Popay *et al*, 1998) explicitly delineate between criteria that apply to both qualitative and quantitative research and those that only have relevance for qualitative research:

> *... many of the prima facie criteria used to assess the quality of research are similar no matter which method is adopted ... However, beyond such minimalist criteria, a judgement about whether what is presented as good or bad requires the invocation of criteria that is more tailored to the particular features of the work in question.* (Popay *et al*, 1998: 344)

Others acknowledge that there are overlapping issues but do not distinguish between the generic and qualitative specific criteria (for example, Cobb and Hagemaster, 1987).

6.3.3 Guiding principles and central concepts.

A number of the framework developers identify central criteria that are formative in the development of their specific quality standards. These are core quality standards from which more specific criteria flow. Popay *et al* (1998) describe this as their 'primary marker' which they specify as follows:

> *'Does this research, as reported, illuminate the subjective meaning, actions, and context of those being researched?' … the extent to which the research adopts a verstehen approach to knowledge, illuminating the meanings people attach to their behaviours and experiences.* (Popay et al, 1998: 345)

Other authors have centred on core criteria first identified by qualitative 'methodologists' who have written extensively about the theory and practice of qualitative research and helped to develop parameters of good practice. Murphy *et al* (1998) and Mays and Pope (2000), for example, use ideas first generated by Hammersley (1990) to focus on criteria which assess research according to:

> *…its <u>validity</u> [our underlining] defined as the extent to which the account accurately represented the social phenomena to which it referred, and its <u>relevance</u> [our underlining], defined in terms of the capacity of the research to help some group of practitioners solve the problems they faced.* (Mays and Pope, 2000, p.93)

Similarly as noted above, Beck (1993) identifies the central concepts of 'credibility', 'fittingness' and 'auditability', as first described by Lincoln and Guba (1985), to derive more detailed quality criteria. Long and Godfrey (2003) used key questions identified by other authors (most notably Popay *et al*, 1998; Hammersley, 1990; and Lofland and Lofland, 1971) to focus on factors related to the 'context' of the study and 'the way the data were collected and analysed' as central structures for framing their evaluative tool.

Certain guiding principles were also stated by some authors as influencing the coverage of their criteria. Leininger (1994), for example, states two major principles which she notes as often '*violated by researchers due to ignorance or denial of the* (qualitative) *paradigm's premises and purposes*' (p. 101). These concerned the need to understand and respect the philosophical premises and epistemic roots of qualitative and quantitative research for credible and sound outcomes to be achieved; and a warning that methods cannot be sensibly mixed across qualitative and quantitative paradigms.

It is perhaps evident that the authors who identified core criteria, concepts or guiding principles of these kind were largely the developers of philosophically

based frameworks. And without exception, although related, they believed in quality standards that were paradigm specific.

6.3.4 Specification of criteria

One of the difficulties in comparing the detailed content of frameworks arises because of the very different ways in which quality criteria are specified. They differ in three key respects:

- **Level of conceptualisation**

  The criteria vary considerably in terms of the breadth of conceptualisation with which they are specified. At one extreme there are those which set out to delineate standards at a sufficiently high level of abstraction that the criteria would have application to a range of methods and/or philosophical orientations. For example, Howe and Eisenhart (1990), identify five 'general standards' that can apply to 'educational research of all kinds' (for example, '*the effective application of specific data collection and analysis techniques*', '*overall warrant*'). Drisko (1997) suggest 'six broad criteria' for evaluating qualitative reports which includes features like '*specifying the goals and audience*' and '*identifying biases*'.

  At the other extreme, are those which have selected highly specific criteria as important indicators of quality for particular parts of the research process. For example, Hoddinott and Pill (1997) identify eight 'methodological criteria' for assessing papers using 'qualitative interviewing' (for example '*how were the respondents recruited*', '*were the authors explicit about how the research topic was explained to the respondents*').

- **Form of specification**

  A related feature concerns the form in which criteria or standards are specified. Some are presented as a series of questions to be answered, often with a series of sub questions or indicators (for example, Boulton and Fitzpatrick, 1997; CASP, 2001; Greenhalgh and Taylor, 1997; Treloar *et al*, 2000), others as a more discursive account of the issues to be addressed (for example, Elliott *et al*, 1999; Popay *et al*, 1998; Healy and Perry, 2000; Whittemore *et al*, 2001) and some as both (Mays and Pope, 2000; Long and Godfrey, 2003; Waterman *et al*, 2001). In some cases, specific examples are included, either to illustrate a point or to exemplify good and bad practice.

  One of the more unusual, although effective, ways of delineating questions and their background, is contained in a framework recently developed by Sandelowski and Barroso (2002). The appraisal guide is divided into 14 sections, each concerned with features of qualitative research reports that readers 'are likely to want to see'. For each, it directs readers/reviewers to the kind of information they need to look for before specifying some appraisal parameters.

- ***Guidance given to users***

  There is also variation in the guidance given to assessors, both generally in how to use the framework and in terms of what should be expected – or might be evident – by way of good practice. Although almost all of the frameworks contain some general background on its intended purpose and use, some go a little further. Others provide explanatory detail (either in the form of indicators or examples) of the kinds of features that will help in the assessment of the extent to which the criteria have been met. This is usually in the form of selected features that would provide reassurance, or further questions to address, rather than as a prescription of what is essential.

Examples of different styles of framework presentation are shown in Appendix 11.

6.3.5 Structure and order of criteria

There is also variation in the structure and order of presenting appraisal criteria. Essentially, the frameworks divide into three broad types. First, and most common, is to present the criteria in some form of processional order. This is usually related to the qualitative research process itself (for example, Cobb and Hagemaster, 1987; Britten *et al*, 1995; Secker *et al*, 1995; Treloar *et al*, 2000; Waterman *et al*, 2001), although sometimes to the order in which a report is assimilated and assessed (for example, Sandelowski and Barrosso, 2002).

A second way of grouping and ordering criteria is to use core concepts or principles for quality assessment as a structure. So for example, Whittemore *et al* (2001), who delineate between *criteria* for assessing the validity of qualitative research (both primary and secondary) and *techniques* to diminish validity threats, present their framework within this structure. Beck (1993) who, as noted before, is primarily concerned with the credibility, fittingness and auditability of qualitative research presents her criteria under these three headings.

Finally, some developers choose a structure which helps to emphasise features of assessment that they see as important. For example, Healy and Perry (2000) use *ontology*, *contingent validity*, *epistemology*, *methodology*, *analytic generalisation* and *construct validity*. Kuzel *et al* (1994) distinguish between *values which guide an inquiry*, *characteristics of the inquiry and report*, *techniques of inquiry*, and *outcomes* to list their criteria .

6.3.6 Length

Because of the different levels and ways in which the main criteria for assessment are specified, it is inappropriate to compare the length of frameworks simply on the basis of the number of key questions addressed. Nevertheless, even allowing for this there is still considerable variation in the density of the frameworks in terms of the number of specific features of a qualitative study that are included for appraisal. To some extent this is related to the particular focus of the framework. For example, Hoddinott and Pill's (1997) was designed to appraise only the interviewing stage of a qualitative study and is therefore short. But even among those which aim to cover the full qualitative process, the frameworks range

considerably in terms of the number of criteria identified for assessment and their related indicators.

6.3.7 Coverage

Because of the varying features described above it is very difficult to provide a full digest of the coverage of the 29 frameworks reviewed. Although, we have attempted to provide a summary profile, this does little justice to the very detailed – and often elegant – way in which the quality criteria, markers and indicators were expressed. However, our main interest here is to highlight some of the main similarities and differences in content, particularly between different types of frameworks.

A summary profile of content in Appendix 13 is structured within three broad, although inter-related frames. These cover the status of the research, the conduct of the research, and the nature and status of the evidence. Within each we have identified

- broad criteria or questions which shape more detailed quality criteria
- specific criteria which are linked to the broad criteria
- some illustration of how the specific criteria may be manifest or demonstrated, drawn from individual frameworks.

As will be evident, most of the broad criteria could be applied to any form of research, whatever its philosophical or methodological base. The particular requirements of qualitative research therefore only begin to emerge once the more specific criteria or means of demonstration are given.

6.3.8 Common features

A dominant area for assessment across many of the frameworks concerns the **_nature and status of qualitative evidence_**. This is in part because a significant number of the frameworks were developed to assess the quality of articles and written papers but also because many of the developers had a prime concern with the status of the knowledge or findings generated by qualitative studies.

There is also a relatively high level of consensus about some of the broad quality criteria that should be met in qualitative studies. This is most notable for items like:

- explicitly stated aims/objectives of the research
- appropriate use of qualitative methods
- appropriate sample design for qualitative research and explicit selection criteria
- clarity about the analytic process

and, perhaps most widely included:

- clarity about how the evidence and conclusions are derived.

6.3.9   Contrasting features

Although there is some degree of consensus about broad quality criteria, different emphasis is placed on specific features of content, depending on the purpose of the framework, any core concepts or principles within which it operates and the way it is formulated. But, on a general level, the greatest contrasts can be found between philosophically focused and empirically focused frameworks. This can take a number of forms.

First, the philosophically based frameworks are more likely than those that are empirically based to include paradigm specific criteria. This occurs around features like:

- understanding of the researcher's connection to research subject, study and subject matter
- capturing and conveying subjective meaning
- evidence of saturation
- displaying context
- provision of thick description
- negative case analysis

all of which are rarely mentioned by empirically based frameworks but are the subject of some discussion by those which are philosophically based.

Second, there are differences in the conceptualisation of certain concepts like validity, reliability, generalisation etc. Here, for example, is what two rather different frameworks have to say on the subject of generalisation:

> <u>Analytic generalisation</u> (that is, theory building) rather than statistical generalisation (that is theory testing) … given the complexity of realism's world, realism research must be primarily theory-building, rather than the testing of the applicability of a theory to a population, which is the primary concern of positivism. Realism researchers do not say this theory-testing should not be done, they merely say that the theory has to be built, and confirmed or disconfirmed, before its generalisability to a population is tested. (Healy and Perry, 2000: 123)

> <u>How generalisable are the results reported</u>? Generalisability is a particular problem in qualitative work. Typically, the demands of qualitative work mean that only one setting can actually be observed or the interviews are conducted with a small number of people. This means that the researcher cannot draw upon the kinds of representative sampling which strengthen the claims survey researchers make about the generalisability of their findings … However, the qualitative researcher cannot simply give up on generalisability. In observational studies, readers should be given enough information to allow them to consider how relevant the setting which has been studied is to other

settings in which they are interested. So, for example, … [several examples given of kinds of information required] … In interview based studies, qualitative researchers do not normally aim for representative samples. Often a sample is deliberately chosen to represent a range of characteristics and views, with consequent over-representation of 'minority' groups. Again, readers need to be told the characteristics of an interviewed sample in order to judge its relevance to their own context. (Britten *et al*, 1995: 110)

Third and related, the same or similar concepts are variously defined and labelled within different frameworks. The greatest haze surrounds issues concerning validity, reliability, trustworthiness, credibility, transferability congruence and coherence. This is in part a consequence of attempts to use definition and language which is unique to qualitative research rather than import terminology from the quantitative paradigm. But even allowing for this there is still considerable diversity in the way terms are used, even among philosophically based authors. For example, in Beck's (1993) discussion of credibility, she includes concepts others have defined as auditability, reflexivity, validity, validation and trustworthiness (see examples, Appendix 12).

Finally, there is variation in what needs to be in place to demonstrate that specific criteria have been met. For example in relation to r*eflexivity,* empirically based frameworks are more concerned with reflections on the research process and philosophically based frameworks place more emphasis on reflections on the role of the researcher.

Some further examples of these differences are shown in Appendix 12, where the treatment of similar subjects by differently orientated frameworks is displayed.

Before leaving this analysis of differences in criteria, it should be noted that other comparisons have been made of some of the frameworks covered in this review. Devers (1999) compared the criteria identified by Mays and Pope (2000), Inui and Frankel (1991) and Kuzel *et al* (1994) terming the former 'positivist' and the latter two 'post-positivist' as she did so. Oakley (2000) compared eight frameworks, four which she described as using 'quantitative' criteria (Cobb and Hagemaster, 1987; Mays and Pope, 1995; Boulton *et al*, 1996; Blaxter, 1996 (Oakley refers to this as the Medical Sociology Group framework)) and four which she describes as identifying 'qualitative' criteria (Guba and Lincoln, 1989; Leininger, 1994; Muecke, 1994; Popay *et al*, 1998) for judging the trustworthiness of qualitative research. Although there is much of interest in the detail of the comparisons made, some general conclusions are very resonant with the discussion above. In particular, both authors found 'some overlap and some differences' in coverage, both between individual frameworks but more notably between those with different ontological or epistemological orientations or from different traditions.

6.3.10 Scoring the criteria

There is a notable absence in the frameworks of any attempts to score or mark the assessments of qualitative studies, either for individual criteria or for the piece of work as a whole. A few of the framework developers give explicit reasons why

scoring systems have not been included (for example Waterman *et al*, 2001) but for most it is clear from other commentary why this is so. In summary, most of the frameworks are intended to guide systematic assessment, not provide definitive prescriptions; and the relative weight of individual features will vary with the philosophical orientation, purpose and methods of the research. In addition, comment was also made that the difficulties surrounding similar attempts to 'mark' or 'score' quality criteria for *quantitative* research had discouraged, rather than persuaded, framework developers to do the same for qualitative research.

6.3.11 The need for professional judgement

In many of the frameworks, a heavy emphasis is placed on the need for judgement in using the criteria suggested. Although many of the developers do not state who the assessors should be, it is implicit in some that expertise in qualitative methods was not expected. For example, phrases like 'those unfamiliar with qualitative research' or 'to help' editors or reviewers, appear in many articles, particularly those developed in the earlier days of formalising quality criteria. But more recently, framework developers have begun to be explicit about the need for assessors and reviewers to have some grounding in qualitative methods:

> *To effectively use the tool, the appraiser requires knowledge, understanding and experience of the qualitative paradigm and, in particular, the problems and difficulties faced in undertaking qualitative research.* (Long and Godfrey, 2003: page unavailable, web text)

> *The appraisal of qualitative studies requires discerning readers who know and take account of what their reading preferences are and who are able to distinguish between non-significant representational errors and procedural or interpretive mistakes fatal enough to discount findings. The appraisal of qualitative studies also requires discerning readers able to distinguish between a report that says all of the right things, but which contains no evidence that these things actually took place.* (Sandelowski and Barroso, 2002: page unavailable, web text)

6.4 Interview commentary on quality frameworks

The desirability and potential uses of formalised criteria were discussed in the interviews conducted for this study. The thoughts expressed covered many of the issues described above and opinion was to some extent divided about their value and use. In the main, however, there was a general view that formalised criteria would be of value to make shared assumptions about quality more explicit, to help to define what is and what is not legitimate and to make more systematic what can currently be somewhat ad hoc or crude assessments of quality. It was thought that reviewing quality standards in qualitative research might help to increase the legitimacy of, or confidence in using, qualitative research within government, and that documenting quality markers that apply across qualitative research methods might help to expand the range of methods used. More specifically, one funder felt that the issue of quality is not currently being addressed directly enough by government, universities or research councils, and that frameworks could be helpful in promoting issues of quality and the need to improve research practice.

Although there was only limited knowledge of existing frameworks (apart from among academics who had developed their own), there were some concerns about overly formalised quality checklists. In particular, there was anxiety that something may be lost in their specification. Interviewees constantly stressed that judgement must remain at the heart of considerations of quality standards or criteria. The importance of not stifling creativity, flair and innovations was raised. It was seen as important that standards or criteria should be applied flexibly, recognising that if a piece of work does not meet all of even the most abstract quality criteria it may nevertheless be of merit and of high quality.

The difficulty of formalising or codifying standards was also noted. Most recurrently, it was thought critical that a framework is designed to be flexibly and not mechanistically applied. It should be seen as a way of structuring and guiding reflections on quality rather than making it a '*bureaucratisation of the research process*'; what was wanted was '*a tool and not a regulation … an aid rather than a determinant*'. Capturing and codifying the essence of quality in qualitative research in an operationalisable way was said to be an extremely difficult task and there were concerns that formalising quality assessments could produce perverse results.

It was also noted that government departments sometimes knowingly commission research which cannot be of the highest quality. This is usually for practical reasons such as the need for early insights, initial exploration or scoping of issues prior to a major study. There would still be concerns about quality, and it would be important to make the study as robust as possible given the constraints of time or money, and to be clear about these constraints. But it was therefore stressed that standards should be applied flexibly with a sense of what is appropriate to the circumstances:

It seems to me there is no standard that you could say 'this is the way it has always got to be.' I think it's got to be what's appropriate in particular circumstances and those judgements are always down to people like me I think … Not a checklist, more sort of guidelines that you've got to decide are appropriate or not appropriate to your study. Research manager

There was also a recurrent view that research that is not methodologically of the highest order can sometimes produce valuable insights. But there was also a caveat to such views, with some interviewees noting that at least minimum technical or methodological standards had to be attained for research to be useable in any context.

Among those more familiar with existing frameworks, it was noted that some lack specificity, with aspirations or imperatives that are too vague to be really helpful. There was also a degree of frustration about the proliferation of frameworks and some doubts about the extent to which they are actually used.

Among research managers and policy-makers there was sometimes a preference for a checklist or a single page of standards or criteria, or more broadly for something that was very concise, very clear and practically rather than theoretically oriented. These groups sometimes saw a broader educative role for a framework, and there were suggestions that it should address issues such as the types of research questions for which qualitative research is appropriate or how to get the most out of the research.

Overall, particularly among academic interviewees and research practitioners, there was concern that judgements about quality require in-depth knowledge of qualitative research, and that no framework can replace the need for real expertise on the part of those making judgements.

7 THE DEVELOPMENT OF THE FRAMEWORK

The previous chapters have described the resources we drew on in developing our framework, Chapters 3 to 5 outlining key issues arising from the literature review and interviews, and Chapter 6 (and Appendix 13) outlining key issues from the review of existing frameworks.

This chapter describes the iterative process we went through in developing our own framework, and outlines three main stages of development. Stage 1 represents the process of developing a first draft of the framework, based on our review of the literature, our interviews, and our assessment of other frameworks. Much of this stage has already been discussed in detail in Chapters 3 to 6, and so here we give just a brief summary of the key decisions we took and the way in which our initial framework drew on these sources. Stage 2 refers to the development of a second draft in the light of discussions and comments at a workshop held for those who had taken part in the interviews earlier. Finally, Stage 3 refers to further revisions made as a result of applying the framework to a range of qualitative reports. The version of the framework given at the beginning of this report is the product of these three stages.

7.1 Stage 1

We began work on the development of the framework at a stage when we had completed most of the literature reviewing, reviewed a number of the existing frameworks, and carried out and analysed around half of the interviews. We had deliberately constructed the timetable so that some reviewing of literature, interviewing and framework reviewing would run concurrently with the development of our framework. We felt it was important that these activities were informed by each other, rather than being separate stages. Members of the research team produced documents summarising the key questions and the potential quality criteria that these activities had so far produced and the team met for a long session to begin to formulate our framework.

We first had to address some fundamental questions about the scope, status and orientation of our framework. These have been described in previous chapters but we briefly reiterate them here and outline how our decisions drew on the activities so far undertaken. Our thinking and our approach were refined as the work continued, but some critical early decisions were made. The issues we addressed were:

- **The feasibility and desirability of producing a framework**
  Although the literature review in particular had highlighted the substantial differences in approaches to the issue of criteria, we felt there was strong evidence of a shared concern about rigour. Views from the people we interviewed were also diverse, but again we felt there was broad support for the idea of guidelines or aids to thinking about quality. We noted, however, the very important caveat – highlighted in the interviews and also by authors of other frameworks – that guidelines or frameworks should not be rigidly

procedural, must resist over-prescription, and must retain scope for judgement, discretion and context-specific assessments.

- **The types of items to include**
  We decided early on that we needed to address the concepts of validity, reliability and objectivity. But we wanted to ensure that our formulation of these concepts – and the way we addressed them in the framework – was driven by the particular nature of qualitative research. We therefore decided to focus on the issues and concerns which lay behind the concepts, but not to use the terms themselves.

- **The philosophical orientation of the framework**
  Our review of the literature supported our early view that we would need to delineate the traditions, paradigms, schools or approaches within qualitative research to which the framework would be applicable. We were aware that this issue is dealt with differently by existing frameworks, some of which specify the traditions to which they apply, or the broader assumptions on which they are based, while others do not refer to philosophical orientation in any form. Our desire for specificity was largely supported by the views of interviewees. But we felt it was clear from the literature review and interviews that it would be difficult to link our framework with specific labelled paradigms or approaches. Our 'elemental' approach (see Chapter 4) emerged early on as a way forward.

- **Orientation to evaluations**
  It was clear that our framework needed to be oriented to evaluations, since this was a key part of our brief. We faced the same difficulty in scoping its applicability to different evaluation approaches as we had to different philosophical traditions, but felt that our elemental approach would again be the way forward. We were encouraged by our literature review, and by the interviews, to take a wide definition of evaluation. We drew particularly on our review of criteria and standards drawn up by people from the 'evaluation establishment', and on the recommendations of the people we interviewed, in identifying criteria that would be particularly important in evaluations.

- **Scoping the methods to which the framework would apply**
  We had doubts early on whether a single framework could address the full range of methods used in qualitative research without becoming vague or vacuous, and again were largely supported in this by the views of the people we interviewed. We decided to take a two-tier approach, with higher-level principles which were applicable across all qualitative methods but more detailed criteria focusing on in-depth interviews, focus groups, documentary analysis and non-participant observation. This choice of methods was later supported by our review of the methods used in government evaluations (see Section 4.3).

- **Identifying the stage of research to which the framework would apply**
  Our review of existing frameworks highlighted that they differ in terms of the purposes for which they were developed, and the stage of research to which they apply (proposals, conduct, outputs etc.). We felt, and were supported here by the people we interviewed, that we needed to be specific about the

orientation of our framework, and felt that our brief suggested it should focus on outputs.

Having broadly agreed these principles, we began to generate items for inclusion in the framework drawing on the literature review, review of frameworks and interviews so far. (These were reviewed in earlier chapters. Box 5.2 in Chapter 5 summarises issues emerging from the literature review and interviews, and Appendix 13 documents issues arising from the review of existing frameworks.)

From this long and comprehensive list we found ourselves initially instinctively, and then more explicitly as the study proceeded, focusing on potential items which were heavily recurrent in these resources or strongly advocated by individual authors or interviewees. We also later decided to prioritise items which related to more than one of our four 'guiding principles' (see Chapter 2 and below) or which we saw as generally emblemic of good quality research. (Examples of this are the inclusion of negative cases, different ways of corroborating research findings, and the description and documentation of approaches.)

We also had in mind whether potential criteria were strongly relevant to applied, government-funded research. For example, in our final framework we amended references to full literature reviews because we felt these would not always be relevant (in a new subject area or where a new policy is being evaluated) and that a summary of key issues may be sufficient. We also changed references to 'main hypothesis' reflecting the fact that not all research is hypothesis-driven. We should also acknowledge that our own personal professional judgements came into play. Where there were strong differences in view in the literature and in the interviews – for example, on the role of quantification in qualitative research – we drew on our professional judgement. In this case, we decided not make quantification an indicator of quality.

The framework that we produced at the end of this first stage of work had a tiered structure with three levels:

- **quality concepts** – these were a set of 12 guiding principles informed by central and recurrently stated tenets of 'robust' research. They were intended to be formulated at a sufficiently high level of abstraction to apply to a diversity of qualitative approaches, and to cover, between them, all the stages of a qualitative study. They were

    –   sets aims and purpose in context
    –   gives logic of enquiry design
    –   shows openness to emergent issues
    –   offers transparency about conduct
    –   provides understanding of subjective meanings
    –   provides understanding of context
    –   provides faithful re-presentation of data
    –   conveys depth, diversity, subtlety and complexity
    –   shows sound interrogation of evidence
    –   presents well-founded argument
    –   offers reflection on research process

–   has utility or relevance.

- **quality markers** – these were intended to point to the specific aspects of a research or evaluation study which might be examined to assess how the quality concept had been addressed. We decided that we had to be selective in the markers we included, and a key criterion for selection was that they related to more than one of the quality concepts. In the version of the framework we produced at this stage, we indicated, against each marker, the quality concepts to which they linked. We also identified the stage of the research study (e.g. design, sampling, analysis etc.) to which each marker most strongly related.

  We had identified 50 quality markers, but the intention was that we would subsequently reduce this to approximately 20 and we identified a selection to which we were most strongly wedded.

- **reference indicators** – these were intended to describe in some detail what might need to be demonstrated to assure an assessor that a quality marker had been addressed adequately. While we wanted to avoid being over-prescriptive, we also felt there needed to be some specificity about how a marker might be appraised. In the version of the framework we produced at this stage, we had taken the quality markers to which we were most strongly attached and shown the reference indicators that we proposed for each.

  To give an example, one of our quality markers was 'clarity about the basis of sample composition'. This related to several of the quality concepts we had identified: sets aims and purpose in context; gives logical of inquiry design; offers transparency of conduct; conveys depth, diversity, subtlety and complexity; and shows sound interrogation of evidence. The reference indicators we associated with this marker were

  – description of population of interest and how sample relates to it (e.g. typical, extreme case, diverse constituencies etc.)
  – description of target groups and reasons for selected characteristics
  – discussion of the basis of inclusions and exclusions
  – discussion of how sample will enable required comparisons
  – discussion of how sample will affect wider inference/generalisation
  – documentation of how achieved and target samples relate and implications for study coverage
  – documentation of reasons for non participation among sample approached
  – discussion of methods of approach and how these might have affected participation.

This version of the framework was then presented to and discussed at the workshop.

7.2 Stage 2

The workshop, as noted in Chapter 1, involved 12 participants, all of them people who had been interviewed earlier or their close colleagues. An explanation of our approach and the framework so far developed was sent to participants in a briefing paper in advance of the workshop.

The workshop began with discussion of the overall approach and rationale of the framework. Participants then worked in small groups and plenary sessions on reviewing and commenting on the quality concepts and then the quality markers, and then discussing the structure of the framework and overall reflections on it. (The workshop programme is shown at Appendix 10.)

The key issues that emerged from the workshop and that influenced the development of the second iteration of the framework were:

- **the need to give more emphasis to some fundamental aspects of quality**. Three of these had been highlighted in the briefing paper as framing our approach: rigorous and transparent conduct of research; claims which are well-founded, plausible and credible; and that the study contributes to wider debates and knowledge. The importance of defensibility of design was also discussed here. This led us to identify the four central principles that underpin the final version of framework – that research is contributory, defensible in design, rigorous in conduct and credible in claim (see Chapter 2).

- **the suggestion that we re-formulate the quality concepts as questions rather than statements of requirements**, but as open-ended questions to avoid a dichotomous yes/no or pass/fail approach. The resulting phraseology of 'how much …', 'how well …', 'how adequately …' begs further questions, but the team decided this was unavoidable and a more useful way forward than a prescriptive approach, and the concepts were thus re-formulated.

- **the need to amend items to ensure that where possible they apply to the range of qualitative methods the framework sought to address**, and to indicate where an item related to only one method. This led us, for example, to give less coverage to 'providing understanding of subjective meanings', which had been one of our 12 quality concepts but which was relevant to methods using generated but not existing data.

- **the need for more focus on ethics**. We had not included ethical research conduct in our framework as we were unsure how easily it could be assessed in a research output and felt it was better addressed by reference to existing sets of ethical guidelines. But participants saw its omission as problematic and we therefore included it specifically in the next iteration of the framework.

- **the structure of the framework.** The structure of the first version of the framework was generally thought to be a little unwieldy, over-detailed in some places. Our attempts to identify linkages between quality markers and concepts was not thought to be particularly helpful. There was also some debate about how distinctive or consistent the different levels of the framework – concepts, markers and indicators – were, and of which items

belonged at each level.

There were also lots of helpful suggestions for where items could be amalgamated or reformulated.

The second iteration of the framework was heavily influenced by these concerns and suggestions. It was structured as follows:

- **guiding principles**. These were the four key principles – that research is contributory, defensible in design, rigorous in conduct and credible in claim.

- **appraisal questions.** These were 22 open-ended questions, largely based on what had been listed as quality concepts and quality markers but responding to comments about priority and formulation.

- **quality indicators.** Around 4–6 quality indicators were identified for each appraisal question, based on markers and reference indicators from the previous version but responding to comments about importance, applicability and the need to avoid over-prescription.

7.3 Stage 3

This second version of the framework was then taken forward by applying it to eight example studies. The purpose here was a developmental one. We wanted to identify areas where the content of the framework or the phraseology used needed to be modified, for example, because we had found a criterion hard to assess, or felt that the framework had missed what we saw as a crucial feature of the quality of the study. We were not assessing the consistency of judgements nor systematically testing the usefulness of the framework, which we see as requiring a separate process of piloting and testing (see Chapter 8).

Eight documents were selected to which the framework was applied. Selection was based on the following criteria:

- **different types of evaluation** – including studies evaluating specific policies and those looking more broadly at the context for policy operation or development
- **policy areas** – five key policy areas were covered (benefits and work; education; health; criminology; and sexuality)
- **the qualitative methods used** – encompassing in-depth interviews; group discussions; non-participant observation; documentary analysis; single and multi-method qualitative studies; studies that also had a quantitative component; and studies specifically described as case studies
- **authors** – covering academia, commercial research and independent research institutes (including one report by NatCen)
- **type of output** – full reports, shorter reports and journal articles were included.

Each document was assessed by two researchers working independently. The framework was drawn up in a table format with columns for documenting the actual assessment and comments on it, the features of the study taken into account in making the assessment, and any difficulties encountered in using the assessment question or indicator. In noting our assessment we tried to detail the evidence of each indicator or question being addressed, any conclusion drawn as to whether it was addressed adequately, the basis of that conclusion, and the degree of certainty about it.

This process led us to merge some appraisal questions and create one new one (giving more focus to the defensibility of the research design), moving from 22 to 18. We also amended our formulation of some of the quality indicators, giving more specific attention to issues such as the clarity of aims, how the design linked with the aims and what use was made of specific features of the design. We added to and refined the way we had included some qualitative specific issues such as the placing of data in context, the portrayal of depth and complexity and the exploration of diversity.

We also changed the order of the framework slightly, placing questions which related to findings (credibility, contribution, how well the study addressed its aims, scope for drawing wider inference and clarity of evaluative appraisal) first. This reflected the general approach we had taken to using the framework when we applied it. We found it was necessary to review these summative questions about the findings before looking back at what had been said and produced in relation to the different stages and conduct of the study.

Some further small changes to phraseology were made between the draft and final versions of the report. The draft version was circulated to those who had participated in the study and to the heads of research in government departments, and these last changes reflected their suggestions and the final thoughts of the research team.

Overall, some questions or indicators in our final version of the framework owed their presence more to one of the activities on which we drew (the literature review, review of frameworks, interviews, workshop or application of the framework) than to others. For example, the indicator relating to discussion of the impact of the research team's perspectives or values probably came more strongly from the literature review and existing frameworks; the point that findings should offer new insights or ways of thinking was underlined particularly in the interviews; the inclusion of the appraisal question about ethics came particularly from the workshop; the point that studies should make full use of features of their design came particularly from our testing of the framework. But most of the guiding principles, concepts and indicators were heavily recurrent and owe their presence to all the activities which contributed to the framework.

CONCLUDING COMMENTS

The framework this study has produced has, as we have shown, built heavily on existing frameworks or guidelines for assessing quality, as well as on the literature review and interviews with policy-makers, research managers and commissioners, academics and research practitioners. Like others who have produced similar documents, we see our work as developmental. We recognise that there will be debate and alternative views about the decisions we have made in shaping the structure, focus and content of the framework.

The importance of judgement and discretion in assessments of quality is strongly emphasised by authors of other frameworks, and was underlined by participants in our interviews and workshop. We think it is critical that the framework is applied flexibly, and not rigidly or prescriptively: judgement will remain at the heart of assessments of quality. Assessments of quality will always be, to some extent, context-specific. The context will affect the relative priority of different appraisal questions and of different quality indicators, as well as decisions about 'how much is enough' – the threshold at which a standard would be judged to have been secured or met. Some notion of the fitness of a study for the particular purpose for which it is being assessed needs to be retained. These points, and the complexity of considerations of quality generally, imply that some degree of expertise in and theoretical understanding of qualitative evaluation research is required on the part of assessors.

Two broader points about the role of qualitative research in government evaluations were made in the interviews we carried out. The policy-makers and government managers and commissioners of research we interviewed sometimes described a need for more support, education and guidance for non-research experts in their use of research, and of qualitative research in particular. Policy-makers themselves sometimes said they would like guidance about when and how to use qualitative research, and about the roles of different qualitative research methods. They hoped that a framework would serve an educative function as well as helping to shape assessments of quality. We share this hope, but their comments suggest, too, a continuing need for broader initiatives to support excellence in the use of research within government.

Finally, there is a need to consider a creative extension of the range of qualitative research methods used in government evaluations. The dominance of interviews in commissioned research is striking, and there is a strong view in the research literature and in our interviews that other methods are under-utilised. Making more use of other methods would have implications for the competencies required in research commissioners, managers and users, as well as for the expertise sought in those conducting research. But it would do much to strengthen the knowledge base for policy-making and policy evaluation.

REFERENCES CITED

Abma, T.A. (1997) 'Playing With/In Plurality: Revitalizing Realities and Relationships in Rotterdam', *Evaluation,* vol. 3, no. 1, pp. 25–8.

Adcock, R., and Collier, D. (2001) 'Measurement Validity: A Shared Standard for Qualitative and Quantitative Research', A*merican Political Science Review*, vol. 95, no. 3 pp. 529–46.

Adler, P.A. and Adler, P. (1994) 'Observational Techniques' in N. Denzin and Y. Lincoln (eds), *Handbook of Qualitative Research*, Thousand Oaks, CA: Sage.

Agar, M. (1980) *The Professional Stranger: An Informal Introduction to Ethnography*, New York: Academic Press Inc.

Agar, N.H. (1986) *Speaking of Ethnography*, Qualitative Research Methods Series, no. 2, London: Sage

Altheide, D.L. and Johnson, J.M. (1994) 'Criteria for Assessing Interpretive Validity in Qualitative Research' in N. Denzin and Y. Lincoln (eds), *Handbook of Qualitative Research*, Thousand Oaks, CA: Sage.

Ambert, A., Adler, P.A., Adler, P. and Detzner, D.F. (1995) 'Understanding and Evaluating Qualitative Research', *Journal of Marriage and the Family*, vol. 57, no. 4, pp. 879–93.

Angen, M.J. (2000) 'Evaluating Interpretive Inquiry: Reviewing the Validity Debate and Opening the Dialogue', *Qualitative Health Research*, vol. 10, no. 3, pp. 378–95.

Armstrong, D., Gosling, A., Weinman, J. and Marteau, T. (1997) 'The Place of Inter-Rater Reliability in Qualitative Research: An Empirical Study', *Sociology* (UK), vol. 31, no. 3, pp. 597–606.

Atkinson, P. and Coffey., A. (2002) 'Revisiting the Relationship Between Participant Observation and Interviewing' in Gubrium, J.F. and Holstein, J.A. (eds), *Handbook of Interview Research*, Thousand Oaks, CA: Sage.

Bailey, P.H. (1996) 'Assuring Quality in Narrative Analysis', *Western Journal of Nursing Research*, vol. 18, no. 2, pp. 186–94.

Barbour, R.S. (2001) 'Checklists for Improving Rigour in Qualitative Research: A Case of the Tail Wagging the Dog?', *British Medical Journal*, vol. 322, no. 7294, pp. 1115–17.

Beck, C.T. (1993) 'Qualitative Research: The Evaluation of its Credibility, Fittingness and Auditability', *Western Journal of Nursing Research*, vol. 15, no. 2, pp. 263–6.

Bhaskar R (1978) *A Realist Theory of Science*, Sussex: Harvester Press.

Black, N. (1994) 'Why We Need Qualitative Research?' *Journal of Epidemiology and Community Health*, vol. 48, pp. 425–6.

Blaikie, N.W.H. (1991) 'A Critique of the Use of Triangulation in Social Research', *Quality and Quantity*, vol. 25, no. 2, pp. 115–36.

Blaxter, M. (1996) 'Criteria for the Evaluation of Qualitative Research Papers', *Medical Sociology News*, vol. 22, no. 1, pp. 68–71.

Bloor, M. (1997) 'Techniques of Validation in Qualitative Research: A Critical Commentary' in Miller, G. and Dingwall, R. (eds), *Context and Method in Qualitative Research*, London: Sage.

Boulton, M. and Fitzpatrick, R. (1997) 'Evaluating Qualitative Research', *Evidence Based Health Policy and Management*, vol. 1, no. 4, pp. 83–5.

Boulton, M. Fitzpatrick, R. and Swinburn, C. (1996) 'Qualitative Research in Health Care II. A Structured Review and Evaluation of Studies', *Journal of Evaluation in Clinical Practice*, vol. 2, no. 3, pp. 171–9.

Boulton, M. Fitzpatrick, R. and Swinburn, C. (1997) *Evaluating Qualitative Research*, Harcourt Brace & Company Ltd.

Britten, N. Jones, R. Murphy, E. and Stacy, R. (1995) 'Qualitative Research Methods in General Practice and Primary Care', *Family Practice*, vol. 12, no. 1, pp. 104–14.

Bryman, A. (1988) *Quantity and Quality in Social Research*, London: Unwin Hyman.

Bryman, R. (2001) *Social Research Methods*, Oxford: Oxford University Press.

Bulmer, M. (1982) *The Uses of Social Research. Social Investigation in Public Policy Making*, London: George Allen and Unwin.

Campbell, D. (1957) 'Factors Relevant to the Validity of Experiments in Social Settings' *Psychological Bulletin*, vol. 54, no. 4, pp. 297–312.

Campbell, R., Pound, P., Pope, C., Britten, N., Pill, R., Morgan, M. and Donovan, J. (2003) 'Evaluating Meta-Ethnography: A Synthesis of Qualitative Research on Lay Experiences of Diabetes and Diabetes Care', *Social Science and Medicine*, vol. 56, no. 4, pp. 671–84.

Critical Appraisal Skills Programme (CASP) (2001) *10 Questions to Help You Make Sense of Qualitative Research*, Oxford: CASP.

Chamberlain, K. (2000) 'Methodolatry and Qualitative Health Research', *Journal of Health Psychology*, vol. 5, no. 3, pp. 285–96.

Chapple, A. and Rogers, A. (1998) 'Explicit Guidelines for Qualitative Research: A Step in the Right Direction, a Defence of the 'Soft' Option, or a Form of Sociological Imperialism?', *Family Practice*, vol. 15, no. 6, pp. 556–61.

Chelimsky, E. (1997) 'Thoughts for a New Evaluation Society', *Evaluation*, vol. 3, no. 1, pp. 97–118.

Chen, H. (1994) 'Current Trends and Future in Program Evaluation', *Evaluation-Practice*, vol. 15, no. 3, pp. 229–38.

Cobb, A.J. and Hagemaster, J.N. (1987) 'Ten Criteria for Evaluating Qualitative Research Proposals', *Journal of Nursing Education* , vol. 26, no. 4, pp. 138–43.

Cook, T. and Campbell, D. (1979) *Quasi-Experimentation*, Chicago: Rand McNally.

Coote, A. and Lenaghan, J. (1997) *Citizens' Juries: Theory into Practice*, London: IPPR.

Crotty, M. (1988) *The Foundations of Social Research: Meaning and Perspective in the Research Process*, London: Sage.

Cutcliffe, J.R. and McKenna, H.P. (1999) 'Establishing the Credibility of Qualitative Research Findings: The Plot Thickens', *Journal of Advanced Nursing*, vol. 30, no. 2, pp. 374–80.

Davies, P. (2000) 'Contributions from Qualitative Research' in H.T.O. Davies *et al*, *What Works? Evidence-Based Policy and Practice in Public Services*, Bristol: Policy Press.

Davies, D. and Dodd, J. (2002) 'Qualitative Research and the Question of Rigor', *Qualitative Health Research*, vol. 12, no. 2, pp. 279–89.

Dean, J., Eichhorn, R., and Dean, L. (1969) 'Limitations and Advantages of Unstructured Methods' in G. McCall and J. Simmons (eds), *Issues in Participant Observation: A Text and Reader*, Reading, MA: Addison-Wesley.

Denzin, N.K. and Lincoln, Y.S. (eds) (1994) *Handbook of Qualitative Research, First Edition*, Thousand Oaks, CA: Sage.

Denzin, N.K. and Lincoln, Y.S. (eds) (2000) *Handbook of Qualitative Research, Second Edition*, Thousand Oaks, CA: Sage.

Devers, K.J. (1999) 'How Will We Know "Good" Qualitative Research When We See It? Beginning the Dialogue in Health Services Research', *Health Services Research*, vol. 34, no. 5, pp. 1153–88.

Dingwall, R. (1980) 'Ethics and Ethnography', *Sociological Review*, vol. 28, no. 4, pp. 871-91.

Dingwall, R. (1980) 'Ethics and Ethnography', *Sociological Review*, vol. 28, no. 4, pp. 871-91.

Dingwall, R. (1997) 'Accounts, Interviews and Observations' in G. Miller and R. Dingwall (eds), *Context and Method in Qualitative Research*, London: Sage.

Drewett, A. (1997) 'Evaluation and Consultation: Learning the Lessons of User Involvement', *Evaluation*, vol. 3, no. 2, pp. 189–204.

Drisko, J.W. (1997) 'Strengthening Qualitative Studies and Reports: Standards to Enhance Academic Integrity', *Journal of Social Work Education*, vol. 33, no. 1, pp. 185–97.

Eisner, E. (1983) 'Anastasia Might Still be Alive, but the Monarchy is Dead', *Educational Researcher*, vol. 12, no. 5, pp. 13–24.

Elliott, R., Fischer, C. and Rennie, D.L. (1999) 'Evolving Guidelines for Publication of Qualitative Research Studies in Psychology and Related Fields*', British Journal of Clinical Psychology*, vol. 38, no. 3, pp. 215–29.

Emden, C. and Sandelowski, M. (1998) 'The Good, the Bad and the Relative, Part One: Conceptions of Goodness in Qualitative Research', *International Journal of Nursing Practice*, vol. 4, no. 4, pp. 206–12.

Enerstvedt, R.G. (1989) 'The Problem of Validity in Social Science', in S. Kvale (ed.) *Issues of Validity in Qualitative Research*, Lund, Sweden: Studentlitteratur.

Engel, J.D. and Kuzel, A.J. (1992) 'On the Idea of What Constitutes Good Qualitative Inquiry', *Qualitative-Health-Research*, vol. 2, no. 4, pp. 504–10.

Fetterman, D.M. (1989) *Ethnography: Step by Step*, Newbury Park, CA: Sage.

Flew, A. (1977) *Thinking Straight,* Buffalo, NY: Prometheus Books.

Fielding, N., and Fielding, J. (1986) *Linking Data*, London: Sage.

Finch, J. (1986) *Research and Policy: The Uses of Qualitative Methods in Social and Educational Research*, London: Falmer.

Forchuk, C. and Roberts, J. (1993) 'How to Critique Qualitative Research Articles', *Canadian Journal of Nursing Research*, vol. 25, no. 4, pp. 47–56.

Fournier, D.M. (1995) 'Establishing Evaluative Conclusions: A Distinction Between General and Working Logic', *New Directions for Evaluation*, vol. 68, pp. 15–32.

Fournier, D.M. and Smith, N.L. (1993) 'Clarifying the Merits of Argument in Evaluation Practice', *Evaluation and Programme Planning*, vol. 16, no. 4, pp. 315–23.

Geertz, C. (1993) *The Interpretation of Cultures*, London: Fontana.

Glaser, B.G. and Strauss, A. (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Chicago: Aldine.

Glassner, B. and Loughlin, J. (1987) *Drugs in Adolescent Worlds: Burnouts to Straights*, New York: St Martin's Press.

Gliner, JA. (1994) 'Reviewing Qualitative Research – Proposed Criteria For Fairness And Rigor', *Occupational Therapy Journal Of Research*, vol. 14, no. 2, pp. 78–90.

Goodwin, L., and Goodwin, H. (1984a) 'Qualitative vs Quantitative or Qualitative and Quantitative Research?', *Nursing Research*, vol. 33, pp. 378–80.

Goodwin, L., and Goodwin, H. (1984b) 'Are Reliability and Validity Relevant in Qualitative Evaluation Research?', *Evaluation and the Health Professional*, vol. 7, pp. 413–26.

Greene, J. (1994) 'Qualitative program evaluation: practice and promise' in N. Denzin, and Y. Lincoln (eds) *Handbook of Qualitative Research,* Thousand Oaks, CA: Sage.

Greene, J.C. (1996) 'Qualitative Evaluation and Scientific Citizenship: Reflections and Refractions', *Evaluation*, vol. 2, pp. 277–89.

Greene, J. (1998) 'Qualitative Programme Evaluation: Practice and Promise' in N.K. Denzin, and Y.S. Lincoln, *Collecting and Interpreting Qualitative Materials*, Thousand Oaks, CA: Sage.

Greenhalgh, T. and Taylor, R. (1997) 'How to Read a Paper: Papers that Go Beyond Numbers (Qualitative Research)', *British Medical Journal*, vol. 315, pp. 740–43.

Guba E. and Lincoln Y. (1981) *Effective Evaluation*, San Francisco: Jossey-Bass.

Guba, E. and Lincoln, Y. (1989) *Fourth Generation Evaluation*, Newbury Park, CA: Sage.

Hammersley, M. (1990) *Reading Ethnographic Research*, New York: Longman.

Hammersley, M (1991) 'A Note on Campbell's Distinction Between External and Internal Validity', *Quality and Quantity*, vol. 25, pp. 381–7.

Hammersley, M. (1992) *What's Wrong with Ethnography?*, London: Routledge.

Hammersley, M. (2001) 'On "Systematic" Reviews of Research Literatures: A "Narrative" Response to Evans and Benefield', *British Educational Research Journal*, vol. 27, no. 5, pp. 543–54.

Hammersley, M. (2002) *Educational Research, Policy Making and Practice.* London: Paul Chapman Publishing, Sage.

Hammersley, M. (2003) 'Can and Should Educational Research be Educative?', *Oxford Review of Education*, vol. 29, no. 1, pp. 3–25.

Hammersley, M. and Atkinson, P. (1995) *Ethnography: Principles in Practice.* London: Routledge.

Hammersley, M. and Gomm, R. (1997) 'Bias in Social Research', *Sociological Research Online*, vol. 2, no. 1, http//www.soc.surrey.ac.uk/socresonline

Harding, S. (1987) *Feminism and Methodology* Bloomington: Indiana University Press.

Healy, M. and Perry, C. (2000) 'Comprehensive Criteria to Judge Validity and Reliability of Qualitative Research within the Realism Paradigm', *Qualitative Market Research: An International Journal*, vol. 3, no. 3, pp. 118–26.

Henwood, K. and Pidgeon, N. (1994) 'Beyond The Qualitative Paradigm – A Framework For Introducing Diversity Within Qualitative Psychology', *Journal Of Community & Applied Social Psychology*, vol. 4, no. 4, pp. 225–38.

Hoddinott, P. and Pill, R. (1997) 'A Review of Recently Published Qualitative Research in General Practice. More Methodological Questions than Answers?', *Family Practice*, vol. 14, no. 4, pp. 313–9.

Holloway, I. and Wheeler, S. (1996) *Qualitative Research for Nurses,* Oxford: Blackwell Science.

House, E. (1980) *Evaluating with Validity*, Beverly Hills: Sage.

Howe, K.K. and Eisenhart, M. (1990) 'Standards for Qualitative (and Quantitative) Research: A Prolegomenon', *Educational Researcher*, vol. 19, no. 4, pp. 2–9.

Inui, T.S. and Frankel, R.M. (1991) 'Evaluating the Quality of Qualitative Research: A Proposal Pro Tem', *Journal of General Internal Medicine*, vol. 6, no. 5, pp. 485–6.

Jensen, G. (1989) 'Qualitative Methods in Physical Therapy Research: A Form of Disciplined Inquiry', *Physical Therapy*, vol. 69, pp. 492–500.

Johnson, M. (1999) 'Observations on Positivism and Pseudoscience in Qualitative Nursing Research', *Journal of Advanced Nursing*, vol. 30, no. 1, pp. 67–73.

Kennedy, M. (1979) 'Generalizing from Single Case Studies', *Evaluation Quarterly*, vol. 3, no. 4, pp. 661–78.

Kirk, J. and Miller, M. (1986) *Reliability and Validity in Qualitative Research*. Newbury Park, CA: Sage.

Knapp, M.S. (1999) 'Ethnographic Contributions to Evaluation Research: The Experimental Schools Programm Evaluation and Some Alternatives' in R. Bryman, and R.C. Burgess, *Qualitative Research*, Volume IV, Sage.

Koch, T. and Harrington A. (1998) 'Reconceptualizing Rigour: The Case for Reflexivity*', Journal of Advanced Nursing*, vol. 28, no. 4, pp. 882–90.

Kushner, S. (2000) *Personalizing Evaluation* London: Sage.

Kuzel, A.J., Engel, J.D., Addison, R.B., and Bogdewic, S.P. (1994) 'Desirable Features of Qualitative Research', *Family Practice Research Journal*, vol. 14, no. 4, pp. 369–78.

Kvale, S. (1988) 'Postmodern psychology - A contradictio in adjectio?', Paper presented at the *Seventh International Human Science Conference* in Seattle, June 8-12, 1988.

Kvale, S. (1989) 'To Validate is to Question' in S. Kvale (ed.) *Issues of Validity in Qualitative Research*, Lund, Sweden: Studentlitteratur.

Kvale, S. (1996*) InterViews: An Introduction to Qualitative Research Interviewing*, Thousand Oaks, CA: Sage.

Lather, P.(1995) 'The Validity of Angels: Interpretive and Textual Strategies in Researching the Lives of Women with HIV/AIDS', *Qualitative Inquiry*, vol. 1, pp. 41–68.

Le Compte, M. and Goetz, J.(1982) 'Problems of Reliability and Validity in Ethnographic Research', *Review of Educational Research*, vol. 52, no. 1, pp. 31–60.

Leininger, M. (1994) 'Evaluation Criteria and Critique of Qualitative Research Studies' in J.M. Morse (ed.), *Critical Issues in Qualitative Research Methods*, Thousand Oaks, CA: Sage.

Lewis, J. and Ritchie, J. (2003) 'Generalising from Qualitative Research' in Ritchie J and Lewis J (eds) *Qualitative Research Practice*, London: Sage.

Lincoln, Y. and Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills: Sage

Lincoln, Y. and Guba, E. (1986) 'Research, Evaluation and Policy Analysis: Heuristics and Disciplined Inquiry', *Policy Studies Review*, vol. 5, no. 3, pp. 546–65.

Lofland, J. and Lofland, L.H. (1971) *Analyzing Social Settings*, Belmont CA: Sage.

Long, A.F. and Godfrey, M. (2003) 'An Evaluation Tool To Assess the Quality of Qualitative Research Studies', *International Journal of Social Research Methodology Theory and Practice* (in press).

Mackenzie, A.E. (1994) 'Evaluating Ethnography – Considerations for Analysis', *Journal Of Advanced Nursing*, vol. 19, no. 4, pp. 774–81.

Madill, A., Jordan, A. and Shirley, C. (2000) 'Objectivity and Reliability in Qualitative Analysis: Realist, Contextualist and Radical Constructionist', *British Journal of Psychology*, vol. 91, no. 1, pp. 1–20.

Marcus, G.E. (1994) 'What Comes (Just) After "Post"? The Case of Ethnography' in N. Denzin and Y. Lincoln, (eds), *Handbook of Qualitative Research*, Thousand Oaks, CA: Sage.

Mason, J. (2002) *Qualitative Researching*, London: Sage.

Maxwell, J.A. (1992) 'Understanding and Validity in Qualitative Research', *Harvard Educational Review*, vol. 62, no. 3, pp. 279–300.

Mays, N. and Pope, C. (1995) 'Rigour and Qualitative Research. King's Fund Institute, London. Tutorial', *British Medical Journal*, vol. 311, pp. 109–12.

Mays, N. and Pope, C. (2000) 'Quality in Qualitative Health Research' in C. Pope, and N. Mays, *Qualitative Research in Health Care*, London: BMJ Books.

Merrick, E. (1999) 'An Exploration of Quality in Qualitative Research: Are "Reliability" and "Validity" Relevant?' in M. Kopala and L.A. Suzuki (eds), *Using Qualitative Methods in Psychology*, Thousand Oaks, CA: Sage.

Meyer, C.B. (2001) 'A Case in Case Study Methodology', *Field Methods*, vol. 13, no. 4, pp. 329–52.

Miles, M.B. and Huberman, A.M. (1984) 'Drawing Valid Meaning from Qualitative Data: Toward a Shared Craft', *Educational Researcher*, vol. 13, pp. 20–30.

Miles, M.B. and Huberman, A.M. (1994) *Qualitative Data Analysis: An Expanded Sourcebook*, London: Sage.

Miller, S. I. and Fredericks, M. (1995) 'Can There Be "Rules" for Qualitative Inquiry?', *Journal of Thought*, Spring, pp. 61–71.

Mitchell, J.C. (1983) 'Case and Situational Analysis', *Sociological Review*, vol. 31, no. 2, pp. 187–211.

Mitchell, R. and Bernauer, T. (1998) 'Empirical Research on International Environmental Policy: Designing Qualitative Case Studies', *Journal of Environment and Development*, vol. 7, no. 1, pp. 4–31.

Morse, J.M. (1986) 'Quantitative and Qualitative Research: Issues in Sampling' in P. Chinn (ed.), *Nursing Research Methodology*, Rockville: Aspen.

Muecke, M.A. (1994) 'On the Evaluation of Ethnographies' in J.M. Morse (ed.), *Critical Issues in Qualitative Research Methods*, Thousand Oaks, CA: Sage.

Munhall, P. (1993) 'Epistemology in Nursing' in P. Munhall and B.C. Oiler (eds), *Nursing Research: a Qualitative Perspective* New York: National League for Nursing.

Murphy, E., Dingwall, R., Greatbatch, D., Parker, S. and Watson, P. (1998) 'Qualitative Research Methods in Health Technology Assessment: A Review of the Literature', *Health Technology Assessment*, vol. 2, no. 16, pp. 1–274.

NHS Centre for Reviews and Dissemination, University of York (2001) *Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for those Carrying Out or Commissioning Reviews* CRD Report Number 4 (2nd edition).

Oakley, A. (2000) *Experiment in Knowing. Gender and Method in the Social Sciences*, Cambridge: Polity Press.

O'Connor, T.S., Koning, F., Meakes, E., McLarnon-Sinclair, K., Davis, K. and Loy, V. (2001) 'Quantity and Rigor of Qualitative Research in Four Pastoral Counselling Journals', *Journal of Pastoral Care*, vol. 55, no. 3, pp. 271–80.

Patton, M.Q. (1997) *Utilisation – Focused Evaluation* (3rd edition), Newbury Park, CA: Sage.

Patton, M.Q. (1999) 'Enhancing the Quality and Credibility of Qualitative Analysis', *Health Services Research*, vol. 34, no. 5, pp. 1189–1208.

Patton, M.Q. (2002) *Qualitative Research and Evaluation Methods*, Thousand Oaks, CA: Sage Publications Ltd.

Pawson, R. (1996) 'Three Steps to Evaluation Heaven', *Evaluation*, no.2, pp. 213 – 220.

Pawson, R. and Tilley, N. (1997a) *Realistic Evaluation*, Newbury Park: Sage.

Pawson, R. and Tilley, N. (1997b) 'An Introduction to Scientific Realist Evaluation' in E. Chelimsky and W. Shadish (eds), *Evaluation for the 21st Century*, Thousand Oaks: Sage.

Perakyla, A. (1997) 'Reliability and Validity in Research Based on Tapes and Transcripts' in D. Silverman (ed.), *Qualitative Research: Theory, Method and Practice*, London: Sage, pp 201–20.

Phillips, D. (1990) 'Post-Positivistic Science: Myths and Realities' in E. Guba (ed.), *The Paradigm Dialogue*, Newbury Park, CA: Sage.

Platt, J. (1981) 'Evidence and Proof in Documentary Research 1 : Some Specific Problems of Documentary Research', *Sociological Review*, vol. 29, no. 1, pp. 31–52.

Polkinghorne, D.E., (1988) *Narrative Knowing and the Human Sciences*, Albany: State University of New York Press.

Popay, J. and Williams, G. (1998) 'Qualitative Research and Evidence-Based Healthcare', *Journal of the Royal Society of Medicine*, Supplement, vol. 191, no. 35, pp. 32–7.

Popay, J., Rogers, A., and Williams, G. (1998) 'Rationale and Standards for the Systematic Review of Qualitative Literature', *Health Services Research, Qualitative Health Research*, vol. 8, no. 3, pp. 341–51.

Potter, J. (1996). 'Discourse Analysis and Constructionist Approaches: Theoretical Background' in Richardson, J.T.E. (ed.), *Handbook of Qualitative Research Methods for Psychology and the Social Sciences*, Leicester: BPS Books.

Power, R. (2001) 'Checklist for Improving Rigour in Qualitative Research. Never Mind the Tail (Checklist), Check Out the Dog (Research)', *British Medical Journal*, vol. 323, pp. 514–15.

Reason, P. and Rowan, J. (1981) *Human Inquiry: A Sourcebook of New Paradigm Research*, Chichester: Wiley.

Reichardt, C. and Cook, T. (1979). 'Beyond Qualitative Versus Quantitative Methods' in T. Cook and C. Reichardt (eds), *Qualitative and Quantitative Methods in Evaluation Research*, Beverly Hills, CA: Sage Publications.

Rist, R.C. (2000) 'Influencing the Policy Process with Qualitative Research' in Denzin, Norman K. and Lincoln, Y. S. (eds), *Handbook of Qualitative Research*, Thousand Oaks, CA: Sage.

Robson, S. (1989) 'Validity and Reliability in Qualitative Research: The Concept of Fit', *Proceedings of a one day seminar on 'Reliability and Validity in Qualitative Research', MRDF.*

Rodgers, B.L. and Cowles, K.V. (1993) 'The Qualitative Research Audit Trail: A Complex Collection of Documentation', *Research in Nursing and Health*, vol. 16, no. 3, pp. 219–26.

Roman, L. and Apple, M. (1990) 'Is Naturalism a Move Away from Positivism? Materialist and Feminist Approaches to Subjectivity in Ethnographic Research' in E. Eisner and A. Peshkin (eds), *Qualitative Inquiry in Education: The Continuing Debate*, New York: Teachers College Press.

Rubin, H.J. and Rubin, I.S. (1995) *Qualitative Interviewing: The Art of Hearing Data*, Thousand Oaks, CA: Sage.

Salner, M. (1989) 'Validity in Human Science Research' in S. Kvale (ed.), *Issues of Validity in Qualitative Research*, Lund, Sweden: Studentlitteratur.

Sandelowski, M. (1986) 'The Problem of Rigor in Qualitative Research', *Advances in Nursing Science*, vol. 8, pp. 27–37.

Sandelowski, M. (1993) 'Rigor Or Rigor-Mortis – The Problem of Rigor in Qualitative Research Revisited', *Advances in Nursing Science*, vol. 16, no. 2, pp. 1–8.

Sandelowski, M. (1998) 'Writing a Good Read: Strategies for Re-presenting Qualitative Data', *Research in Nursing and Health*, vol. 21, no. 4, pp. 375–82.

Sandelowski, M. and Barroso, J. (2002) 'Reading Qualitative Studies', *International Journal of Qualitative Methods*, vol. 1, no. 1, (electronic journal).

Schofield, J.W. (1990) 'Increasing the Generalizability of Qualitative Research' in E.W. Eisner and A. Pershkin (eds), *Qualitative Inquiry in Education*, New York: Teachers College Press.

Schwandt, T.A. (1996) 'Farewell to Criteriology', *Qualitative Inquiry*, vol. 2, no. 1, pp. 58–72.

Schwandt, T.A. (1997) *Qualitative Inquiry: A Dictionary of Terms*, Thousand Oaks, CA; Sage.

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.

Scriven, M. (1986) 'New Frontiers of Evaluation', *Evaluation Practice*, vol. 7, no. 1, pp. 7–44.

Scriven, M. (1991) *Evaluation Thesaurus* (4th edition), Newbury Park CA: Sage.

Seale, C. (1999) *The Quality of Qualitative Research*, London: Sage.

Secker, J., Wimbush, E., Watson, J. and Milburn, K. (1995) 'Qualitative Methods in Health Promotion Research: Some Criteria for Quality', *Health Education Journal*, vol. 54, pp. 75–87.

Shaw, I. F. (1999) *Qualitative Evaluation*, London: Sage.

Silverman, D. (1993) *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction*, London: Sage.

Silverman, D. (ed.) (1997) *Qualitative Research: Theory, Method and Practice*, London: Sage.

Silverman, D. (2000) *Doing Qualitative Research: A Practical Handbook*, London: Sage.

Smith, J.K. (1984) 'The Problem of Criteria for Judging Interpretive Inquiry', *Educational Evaluation and Policy Analysis*, vol. 6, no. 4, pp. 379–91.

Smith, J.K. (1990) 'Goodness Criteria: Alternative Research Paradigms and the Problem of Criteria' in E.G. Guba, *The Paradigm Dialogue*, London: Sage.

Smith, J.K. and Deemer, D. (2000) 'The Problem of Criteria in the Age of Relativism' in N. Denzin and Y. Lincoln (eds), *Handbook of Qualitative Research*, California: Sage.

Smith, M.C. (1984) 'Research Methodology: Epistemologic Consideration', *Image*, vol. 16, pp. 42–6.

Snape D. and Spencer L. (2003) 'The Foundations of Qualitative Research' in J. Ritchie and J. Lewis (eds), *Qualitative Research Practice*, London: Sage.

Sparkes, A.C. (2001) 'Myth 94: Qualitative Health Researchers Will Agree About Validity', *Qualitative Health Research*, vol. 11, no. 4, pp. 538–52.

Stake, R.E. (2000) 'Case Studies' in N. Denzin and Y. Lincoln (eds), *Handbook of Qualitative Research*, California: Sage.

Stake, R.E. (1978) 'The Case Study Method in Social Inquiry', *Education Researcher*, vol. 7, pp. 5–8.

Stake, R. (1995) *The Art of Case Study Research*, Thousand Oaks: Sage.

Stake, R., and Turnbull, D. (1982). 'Naturalistic Generalizations', *Review Journal of Philosophy and Social Science*, vol. 7, pp. 1–12.

Stiles, W.B. (1993) 'Quality Control in Qualitative Research', *Clinical Psychology Review*, vol. 13, no. 6, pp. 593–618.

Strauss, A.L. and Corbin, J. (1998) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, Thousand Oaks, CA: Sage.

Stufflebeam, D. L. (2000) 'Professional Standards and Principles for Evaluations' in D.L. Stufflebeam *et al* (eds), *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, Boston: Kluwer Academic Publishers.

Sykes, W. (1990) 'Validity and Reliability in Qualitative Market Research: A Review of the Literature', *Journal of the Market Research Society*, vol. 32, no. 3, pp. 289–328.

Treloar, C., Champness, S., Simpson, P.L., and Higginbotham, N. (2000) 'Critical Appraisal Checklist for Qualitative Research Studies', *Indian Journal of Pediatrics*, vol. 67, no. 5, pp. 347–51.

Walker, R. (1985) 'An Introduction to Applied Qualitative Research' in Walker, R. (ed.), *Applied Qualitative Research*, Aldershot: Gower.

Waterman, H. (1998) 'Embracing Ambiguities and Valuing Ourselves: Issues of Validity in Action Research', *Journal Of Advanced Nursing*, vol. 28, no. 1, pp. 101–5.

Waterman, H., Tillen, D., Dickson, R. and de Koning, K. (2001) 'Action Research: A Systematic Review and Guidance for Assessment', *Health Technology Assessment*, vol. 5, no. 23, pp. 1-157.

Weiss, C. (1988a) 'Evaluation for Decisions: Is Anybody There? Does Anybody Care?', *Evaluation Practice*, vol. 9, no. 1, pp. 5–19.

Weiss, C. (1988b) 'If Programme Decisions Hinged Only on Information', *Evaluation Practice*, vol. 9, no. 3, pp. 15–28.

Whittemore, R., Chase, S.K. and Mandle, C.L. (2001) 'Validity in Qualitative Research', *Qualitative Health Research*, vol. 11, no. 4, pp. 522–37.

Williams, D.D. (1986) 'When is Naturalistic Evaluation Appropriate?' in Williams, D. D. (ed.), *Naturalistic Evaluation: (Programme Evaluation Series 30)*, San Francisco: Jossey-Bass Wiley.

Williams, B. (2001) 'Checklist for Improving Rigour in Qualitative Research. Including Personal Reflections Might Help', *British Medical Journal*, vol. 323, pp. 515.

Wolcott, H.F. (1990) 'On Seeking - and Rejecting - Validity in Qualitative Research' , in E.W. Eisner, and A. Peskin, (eds) *Qualitative Inquiry in Education*: *The Continuing Debate*, New York: Teachers College Press, Columbia University.

Wolcott, H.F (1994) *Transforming Qualitative Data: Description, Analysis, Interpretation,* Thousand Oaks, CA: Sage.

TEXTS CONSULTED BUT NOT CITED

Annells, M. (1997) 'Grounded Theory Method, Part 1: Within The Five Moments Of Qualitative Research', *Nursing Inquiry*, vol. 4, pp. 120–29.

Bloor M. 'Techniques of Validation in Qualitative Research: A Critical Commentary', in G. Miller and R. Dingwall (eds), *Context and Method in Qualitative Research*. London: Sage, 1997.

Brink, H.I. 'Validity and Reliability in Qualitative Research', *Curationis*, Jun, vol. 16, no. 2, pp. 35–8.

Brink, P.J. (1991) 'Issues of Reliability and Validity' in Morse, Janice, M. (ed.), *Qualitative Nursing Research: A Contemporary Dialogue* (rev. ed.) Thousand Oaks, CA: Sage Publications.

Britten, N. and Fischer, B. (1993) 'Qualitative Research and General Practice', *British Journal of General Practice*, July, pp. 270–71.

Brockopp, D.Y. and Hastings-Tolsma, M.T. (1989) *Fundamentals of Nursing Research*, Boston: Scott, Foresman.

Buchanan, D.R. (1992) 'An Uneasy Alliance: Combining Qualitative and Quantitative Research Methods', *Health Education Quarterly*, Spring, pp. 117–35.

Bullock, C., Ory, J. (2000) 'Evaluating Instructional Technology Implementation in Higher Educational Environment', *The American Journal of Evaluation*, vol. 21, no. 3, Fall, pp. 317.

Burgess, R.G. (2000) 'Some Issues and Problems in Cross-Cultural Case Study Research', *Studies in Qualitative Methodology*, vol. 6, pp. 43–52.

Campbell, J. (2002) 'A Critical Appraisal of Participatory Methods in Development Research', *International Journal Of Social Research Methodology: Theory & Practice*, vol. 5, no. 1, pp. 19–29.

Chen, H.T. (1996) 'A Comprehensive Typology for Program Evaluation', *Evaluation Practice* vol. 17, no. 2, pp. 121–30.

Clarke, L. (1992) 'Qualitative Research: Meaning and Language', *Journal of Advanced Nursing*, vol. 17, no. 2, pp. 243–52.

Coyne, I.T. (1997) 'Sampling in Qualitative Research. Purposeful and Theoretical Sampling: Merging or Clear Boundaries?', *Journal of Advanced Nursing*, vol. 26, 723–30.

Cresswell, J.W. (1998) *Qualitative Inquiry and Research Design: Choosing Among Five Traditions*, Thousand Oaks, CA: Sage.

Cunningham, J.B. (1997) 'Case Study Principles for Different Types of Cases*', Quality and Quantity*, vol. 31, no. 4, pp. 401–23.

Davies, P.T., Petrosino, A. and Chalmers, I.G. (1999) 'The Effects of Social and Educational Interventions', *Report of a meeting held at the School of Public Policy, University College London, 15–16 July (available from P.T. Davies, Department for Continuing Education, University of Oxford).*

Dingwall, R. (1992) 'Don't Mind Him – He's from Barcelona: Qualitative Methods in Health Studies' in Daly, J., McDonald, I., Willis, E., (eds), *Researching Health Care*, London: Tavistock/Routledge.

Dingwall, R., Murphy, E., Watson, P., Greatbatch, D. and Parker, S. (1998) 'Catching Goldfish: Quality in Qualitative Research', *Journal of Health Services Research and Policy*, vol. 3, pp. 167–72.

Feyerabend, P. (1975) *Against Method*, London: Verso.

Field, P. and Morse, J. (1985*) Nursing Research:: The Application of Qualitative Approaches*, Rockville MD: Aspen.

Fielding, N.G. (1999) 'The Norm and the Text: Denzin and Lincoln's Handbooks of Qualitative Method', *British Journal of Sociology*, vol. 50, no. 3, pp. 525–34.

Flick, U. (1992) 'Triangulation Revisited: Strategy of Validation or Alternative?', *Journal for the Theory of Social Behaviour*, vol. 22, no. 2, pp. 175–97.

Gatrell, J., Popay, J., Thomas, C., Williams, G. (Forthcoming) 'Theorising Inequalities In Health: The Place Of Lay Knowledge' in M. Bartley et al (eds), *The Sociology of Inequalities in Health, Sociology of Health and Illness Monograph 5.*

Giorgi, A. (1988) 'Validity and Reliability From a Phenomenological Perspective' in W. Baer (eds), *Recent Trends in Theoretical Psychology*, New York: Springer-Verlag.

Hamberg, K., Johansson, E., Lindgren, G. and Westman, G. (1994) 'Scientific Rigour in Qualitative Research-Examples From a Study of Women's Health in Family', *Family Practice*, vol. 11, no. 2, pp. 176–81.

Hammersley, M. (1995) 'Theory and Evidence in Qualitative Research', *Quality and Quantity*, vol. 29, no. 1, pp. 55–66.

Hughes, J. and Sharrock, W. (1997) *The Philosophy of Social Research*, London: Longman.

Knaack, P. (1984) 'Phenomenological Research', *Western Journal of Nursing Research*, vol. 6, no. 1, pp. 107–14.

Knafl, K.A. and Howard, M.J. (1984) 'Interpreting and Reporting Qualitative Research', *Research in Nursing & Health*, vol. 7, pp. 17–24.

Koch, T. (1994) 'Establishing Rigour in Qualitative Research: The Decision', *Trail Journal of Advanced Nursing*, vol. 19, no. 5, pp. 976–86.

Leininger, M.M. (1990) 'Technomethods: The Philosophic and Epistemic Bases to Explicate Transcultural Nursing Knowledge', *Journal of Transcultural Nursing*, vol. 1, pp. 40–51.

Lilford, R.J., Richardson, A., Stevens, A., Fitzpatrick, R., Edwards, S., Rock, F. and Hutton, J.L. (2001) 'Issues in Methodological Research: Perspectives From Researchers and Commissioners', *Health Technology Assessment*, vol. 5, no. 8, pp. 1–57.

Locke, L.F., Silverman, S.J. and Spirduso, W.W. (1998*) Reading and Understanding Research*, Thousand Oaks, CA: Sage Publications, Inc.

Marshall, C. (1990) 'Goodness Criteria: Are They Objective Judgement Calls?' in E.G. Guba, *The Paradigm Dialogue*, London: Sage.

Marshall, C. and Rossman, G.B. (1989) *Designing Qualitative Research*, London: Sage.

Maxwell, J. (1996) *Qualitative Research Design: An Interpretative Approach*, Thousand Oaks, CA: Sage.

Medical Sociology Group (1996) 'Criteria for the Evaluation of Qualitative Research Papers', *Medical Sociology News*, vol. 22, no. 1, pp. 68–71.

Melia, K.M. (1996) 'Rediscovering Glaser', *Qualitative Health Research*, vol. 6, no. 3, pp. 368–78.

Miles, M.B. and Huberman, A.M. (1984*) Qualitative Data Analysis: A Sourcebook Of New Methods*, Beverly Hills, CA: Sage.

Miller, S.I. (1990) 'Confirmation and Qualitative Evidence-Instances: Justifying the Use of Qualitative Research Methods', *Quality and Quantity*, vol. 24, no. 1, pp. 57–63.

Mitchell, G.J. and Cody, W.K. (1993) 'The Role of Theory in Qualitative Research', *Nursing Science Quarterly*, vol. 6, no. 4, pp. 170-78.

Morse, J.M. (1991) 'Approaches To Qualitative-Quantitative Methodological Triangulation', *Research in Nursing and Health*, vol. 14, pp. 187–96.

Mullen, E. J. (1995) 'Pursuing Knowledge Through Qualitative Research', *Social Work Research*, vol. 19, no. 1, pp. 29–32.

Noblit, G.W. and Hare, R.D. (1988) *Meta-Ethnography: Synthesizing Qualitative Studies*, Newbury Park: Sage.

Oiler, C. (1982) 'The Phenomenological Approach in Nursing Research*', Nursing Research*, vol. 31, no. 3, pp. 178–81.

Papworth, MA. and Milne, DL. (2001) 'Qualitative Systematic Review: An Example From Primary Prevention In Adult Mental', H*ealth Journal of Community & Applied Social Psychology*, vol. 11, no. 3, pp. 193–210.

Parse, R.R., Coyne, A.B., and Smith, M.J. (1985) *Nursing Research: Qualitative Methods*, Bowie: Brady Communications.

Patton, M. (1988) 'The Evaluator's Responsibility for Utilization', *Evaluation Practice*, vol. 9, no. 2, pp. 5–24.

Patton, M.Q. (1988) 'Paradigms and Pragmatism' in D.M. Fetterman (ed.), *Qualitative Approaches to Evaluation in Education: The Silent Scientific Revolution*, New York: Praeger.

Patton, M.Q. (1990) *Qualitative Evaluation and Research Methods* (2nd edition), Beverly Hills, CA: Sage.

Plummer, K. (1983) *Documents of Life: An Introduction to the Problems and Literature of a Humanistic Method*, London: Allen and Unwin.

Popay, J. and Williams, G. (1995) 'Qualitative Research and the Gingerbread Man: Guest Editorial', *Health Education Journal*, vol. 54, no. 4, pp. 389–92.

Pope, C. and Mays, N. (1995) 'Reaching the Parts Other Methods Cannot Reach: An Introduction to Qualitative Methods in Health and Health Services Research', *British Medical Journal*, vol. 6996, no. 311, pp. 42–5.

Pope, C. (1991) 'Trouble in Store: Some Thoughts on the Management of Waiting Lists', *Sociology of Health and Illness*, vol. 13, pp. 193–212.

Ramos, M.C. (1989) 'Some Ethical Implications Of Qualitative Research', *Research in Nursing and Health*, vol. 12, pp. 57–63.

Ritchie, J. and Lewis, J. (eds) (2003) *Qualitative Research Practice,* London: Sage

Robertson, M. and Boyle, J. (1984) 'Ethnography: Contributions to nursing research', *Journal of Advanced Nursing*, vol. 9, pp. 43–9.

Robinson, C.A. and Thorne, S.E. (1988) 'Dilemmas of Ethics and Validity in Qualitative Nursing Research', *The Canadian Journal of Nursing Research*, vol. 20, pp. 65–76.

Rogers, A. and Pilgrim, D. (1995a) 'Experiencing Psychiatry: An Example of Emancipatory Research' in G. Wilson (ed.), *Researching Users' Views Of Community Care Services*, London: Chapman Hall.

Sackett, D.L. and Wennberg, J.E. (1997) 'Choosing the Best Research Design for Each Question' Editorial, *British Medical Journal*, vol. 315, pp. 1636.

Sandelowski, M. (1995) 'Sample Size in Qualitative Research', *Research in Nursing Health*, vol. 18, no. 2, pp. 179–83.

Sandelowski, M. (1999) 'Time and Qualitative Research', *Research in Nursing Health*, vol. 22, no. 1, pp. 79–87.

Schatzman, L. and Strauss, A. (1973) *Field Research: Strategies For A Natural Sociology*, Englewood Cliffs, NJ: Prentice-Hall Methods of Social Science Series.

Scheff, T. (1995) 'Single Case Analysis In The Health Sciences', *European Journal of Public Health*, vol. 5, no. 2, pp. 72–4.

Schwandt, T.A. (1997) 'Evaluation as Practical Hermeneutics', *Evaluation*, vol. 3, no. 1, January, pp. 69–84.

Seale, C. and Silverman, D. (1997) 'Ensuring Rigour in Qualitative Research', *European Journal of Public Health*, vol. 7, pp. 379–84.

Seers, K. (1999) 'Qualitative Research' in D. Dawes *et al*, *Evidence-based Practice, A Primer for Health Care Professionals*, Edinburgh: Churchill Livingstone.

Silverman, D. (1989) 'Telling Convincing Stories: A Plea For More Cautious Positivism In Case Studies' in Glassner, B., Moreno, J.D., (eds), *The Qualitative-Quantitative Distinction In The Social Sciences*, Dordrecht: Kluwer Academic.

Sim, J. (1998) 'Collecting And Analysing Qualitative Data: Issues Raised by the Focus Group', *Journal Of Advanced Nursing*, vol. 28, no. 2, pp. 345–52.
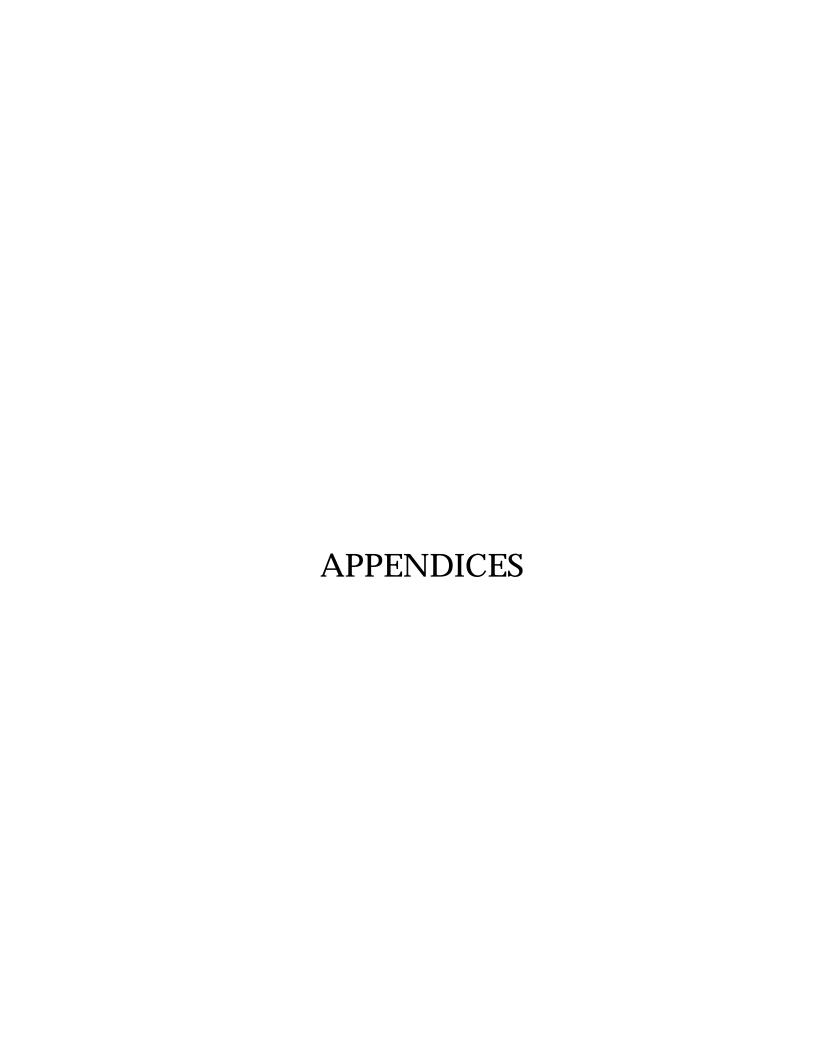
Stafford, M.R. and Stafford, T.F. (1993) 'Participant Observation and the Pursuit of Truth', *Journal of The Market Research Society*, vol. 35, no. 1, pp. 63–76.

Stanley, L. (1990) 'Doing Ethnography, Writing Ethnography: A Comment on Hammersley', *Sociology*, vol. 24, no. 4, pp. 617–27.

Weiss, C.H. (1982) 'Policy Research in the Context of Diffuse Decision Making' in Rist, R.C. (ed.), *Policy Studies Annual Review*, Beverly Hills, CA: Sage.

Yin, R. K. (1997) 'Case Study Evaluations: A Decade Of Progress?', *New Directions For Evaluation*, vol. 76, pp. 69–78.

Yin, R.K. (1992) 'The Case Study Method as a Tool for Doing Evaluation', *Current-Sociology*, vol. 40, no. 1, pp. 121–37.

APPENDICES

Appendix 1
Inclusion/Exclusion Criteria for Documents in the Literature Review

First set of criteria used for inclusion in initial central bibliography

1.   Refers to quality criteria
2.   Refers to debate about validity and reliability
3.   Contains debate about qualitative methods
4.   Refers to quality in evaluation studies
5.   Refers to use of qualitative methods in particular substantive fields
6.   Refers to use of qualitative methods in government research/studies
7.   Contains examples of use of qualitative methods in evaluation

Second set of criteria for including/excluding references for review

1. gives an actual *quality framework*
2. contains a *methodological debate* about one or more aspects of *quality* (e.g. appropriate aims and scope, appropriate design, sampling rationale, skilled data collection, rigorous and transparent analysis, the nature of qualitative evidence, audit trail, validity, reliability, generalisability etc. – i.e. items which might be included in a framework
3. discusses philosophical issues relating to quality in qualitative methods/evaluation
4. is an empirical *policy evaluation study*, using qualitative methods, which
   a. generates a discussion of quality criteria
   b. or discusses methodological rigour
   c. or is methodologically very aware
   d. or uses qualitative methods in an innovative way
5. is a *key qualitative methods text book* which discusses quality issues not just a cook book
6. discusses qualitative evaluation
   a. giving definitions of *types of evaluation studies*
   b. reviewing the *role of different types of qualitative methods* in policy evaluation, mainstream and minor methods etc.
   c. outlining any *contextual or political constraints* on quality (participatory/dialogic; ethical issues etc.)
7. discusses *implications for quality criteria of particular qualitative methods* (e.g. validity of interview as against observational data)

Further criteria for rejecting references

1. *already included*/cited (duplication)
2. *empirical study which is not a (policy) evaluation* and/or which does not (appear) to contain reflection on rigour or methods or generate a discussion on criteria for quality etc.

3. examples of qualitative research contributions in substantive fields as **justification of usefulness** or track record etc./summary of findings from other studies to debate within substantive field (rather than contribution of particular methods to policy evaluation)
4. **academic 'filler' references** to show pedigree – i.e. that author knows the key texts in this field
5. author(s) **citing their own earlier work** when key ideas already in reviewed text
6. very broad **methods text book** or cook book, or how to do evaluation
7. discussion of issues/**debates within a particular substantive field** (e.g. health, education etc.)
8. **quantitative issues** (but not criteria)
9. the differences/overlaps between quant and qual/the **quant/qual debate** or merits of combining
10. **argument already made** in (lots of) other references included (either more articulately or by more eminent authors)/very old references now subsumed in others
11. too **broadly theoretical**/discipline-based (anthropology, sociology, psychology, philosophy), feminism etc.
12. **specific fieldwork issues** (e.g. access), computer assisted analysis, power relations; or narrow methodological point

Appendix 2
Template for Reviewing Literature

1. Full reference details [Harvard style – author(s), date, title, publisher]

2. Type of document [book, chapter in book, article, published report, unpublished report etc.]

3. Reviewer

4. Brief summary of main ideas (and implications for framework)

5. Implications/possible issues for Topic Guides

6. Explicit discussion of philosophical underpinnings [in relation to quality, or qualitative evaluation, realist vs relativist etc.]

7. Quality issues/criteria discussed

STATUS OF RESEARCH/EVALUATION
a) Stance/approach of author (of article if known)/importance of making this transparent (if criteria)

b) Aims/purpose and scope (whether appropriate to qualitative methods; whether important or essential to have theoretical framework/basis)

c) Design strategy and appropriateness to aims

CONDUCT OF THE RESEARCH (spell out if/how criteria are affected by different qualitative methods)

d) Sampling (type, transparency, rationale)

e) Data collection (rigour in etc.)

f) Analysis

g) Auditability/audit trail

h) Reflexivity

i) Ethical considerations and procedures

j) Skills of the researcher

OUTPUTS OF THE RESEARCH/EVALUATION (spell out if/how criteria are affected by different qualitative methods)

k) Visibility of 'analytical hierarchy' – data, evidence, conceptualisation, interpretation, explanation and conclusions

l) Validity and validation

m) Reliablity

n) Generalisability/transferability

o) Credibility/coherence

p) Whose perspective – emic or etic

q) Contextualisation

r) Clarity

s) Relevance

t) Other

8. Type(s) of evaluation discussed/defined/used – formative or summative etc., dialogic or participatory (spell out if these place any practical/political/contextual limitations on 'quality')

9. Types of qualitative methods used in evaluation and their contribution, or methods used if a particular study is referred to

10. References to be cited (second hand) in review

11. References to be chased up and reviewed

Appendix 3
Template for Overview of Methods Used in Government Evaluative Research

Reference (report title, author, evaluating institution):

Document type consulted (synopsis, full report etc.):

Policy/service under evaluation:

Funder:

Research summary (if available):

Qualitative methods used [focus groups, in-depth interviews, observation, participant observation, documentary analysis, diaries, conversation/discourse analysis (documents/oral data?)]:

Quantitative methods used as well:

Location of information (website, search criteria etc.)

Search Path Followed for Carrying Out Overview of Qualitative Methods in Government Evaluative Research

In carrying out the overview of qualitative methods used in government evaluative research, four government websites (DWP, DfES, DoH and HO) and hard copies of research summaries provided to the team in the Strategy Unit in the Cabinet Office for one department (ODPM) were examined. Searches were carried out in Sept/Oct 2002.

Website searches followed the following routes:

DfES: www.dfes.gov.uk – Publications – research publications – evaluation (published). Produced 125 returns.

DWP: www.dwp.gov.uk – statistics and research – research – research summaries. Looked at Summaries of Reports 71–173.

Home Office: www.homeoffice.gov.uk – Research and Statistics – Publications – Home Office Research Studies. Looked at studies 165–248.

DoH: 2 search routes:
1. www.doh.gov.uk – Research and development – ReFeR – the DH Research Findings electronic Register – search – searched with keyword 'evaluation'. 386 returns.
2. www.doh.gov.uk – Publications – search POINT – search with keyword 'evaluation'.

ODPM: examined a range of studies from hard copy provided by the team from the Strategy Unit in the Cabinet Office.

Appendix 4
Description of Depth Interviews

Sample

It was agreed at an early stage in the project that the sample of people interviewed should include policy-makers in central government, government-based research managers and commissioners, representatives from funding councils and foundations, people who had developed frameworks, academics and research practitioners.

We decided to select policy-makers and research managers/commissioners from each of the five departments which make most use of qualitative research in evaluations: DWP, DfES, the Home Office, DoH, and Department for Transport, Local Government and the Regions (as it then was, now ODPM). For managers or commissioners, we initially approached the heads of research departments who either suggested a colleague or were interviewed themselves. Policy-makers were nominated by the Strategy Unit in the Cabinet Office who liaised with researchers in each of the departments. Five interviews were undertaken with commissioners and five with policy-makers.

We also approached senior people in four key foundations and councils which fund qualitative work. One had had little involvement with qualitative research in that capacity but was interviewed as an academic and practitioner, while the other three were interviewed in their capacity of funders of research.

The interviews with these groups generally took place in the first round of fieldwork, in part to ensure a focus on government evaluations from the start.

The remaining interviews were with research practitioners, academics and people who had developed frameworks. These are obviously overlapping groups. Our eventual sample of 16 included 14 people based in academia (most of whom were, to varying degrees, practitioners as well as writers about qualitative research methods, and four of whom had developed frameworks), and two research practitioners not based in academia. In selecting people to approach in these groups we aimed to build in as much diversity as possible in terms of: the substantive subject area of people's work; the research methods they were associated with; whether or not they conduct government-funded evaluations and, for framework developers, the focus of their framework.

The policy-makers and commissioners we approached sometimes nominated a colleague instead of being interviewed themselves and one academic we approached was on long-term leave, but otherwise everyone approached was able to take part. One interview with a funder involved two interviewees, and one with government commissioners involved three, where the people we approached thought it would be helpful to involve colleagues in the interview.

Interviews mostly took place at people's offices, although some academics chose to be interviewed at home. Interviews lasted between one and over two hours, the

longer interviews generally being those with academics. The topic guides used in the interviews are presented below. Our general approach was to use open questions where possible, particularly to generate spontaneous mentions of issues seen as critical to quality in qualitative research. We followed these up with more specific questions about why those issues were seen as central, what is required and what needs to be demonstrated to assess whether the standard has been met. We asked specifically about other criteria – particularly where we wanted to explore thoughts about issues arising from our work in the literature review so far, or to explore issues raised by interviewees in earlier work or in their own frameworks. We also asked whether the focus on evaluation influences what is seen as critical to quality, and sought views and advice on specific issues relating to the scope and structure of our framework. All the interviews were tape-recorded and transcribed verbatim.

We worked from verbatim transcripts in our analysis. Our approach was to summarise what had been said by each respondent under a series of headings. For the early round of interviews (mostly commissioners and policy-makers but also including some practitioners), the headings were:

- Role, background and types of research involved with
- Epistemological stance
- Scope, use, value of qualitative research
- Quality criteria:
  – general
  – design and approach
  – conduct and practice (sampling, fieldwork, analysis)
  – research team
  – outputs (documentation, validity/validation, reliability, credibility, generalisability, reporting)
- Knowledge and use of frameworks
- Other issues.

For the later stage of interviews we used the same headings as for the literature review (see Appendix 2). The headings at both stages were used flexibly so that any material seen as relevant to the study could be included in our summaries.

Appendix 5
<u>Letter of Introduction for Academics/Framework Developers</u>

<div align="center">Reference P6055/R2/A</div>

Dear

<div align="center">**Quality Standards in Qualitative Evaluation Methods**</div>

Thank you very much for agreeing to be interviewed for this study which the National Centre for Social Research is conducting on behalf of the Strategy Unit in the Cabinet Office.

I am writing to confirm that I/xxx and I will come to your office at [insert time] on [insert date]. I/We would be grateful for up to an hour and a half of your time. A short information sheet about the study is enclosed.

The objective of the study is to produce a framework within which the quality of qualitative evaluation studies can be assessed. Although the focus is on quality standards in qualitative *evaluations* we think it will be important to place the discussion in the context of qualitative research more generally. In the interview we will want to explore the criteria which you see as critical in making judgements about the standard of qualitative studies and why those criteria matter. *[(Academics) It would also be useful to hear about any experience you may have of frameworks or good practice guidelines for assessing qualitative research, and your views on their value./(Framework developers) It would also be useful to hear more about the work you did on quality criteria and any further thoughts you have on good practice guidelines for assessing qualitative research.]* Your thoughts on these and other related issues will be most helpful in informing the development of a quality standards framework.

We are interviewing a range of research commissioners, policy-makers, research practitioners and academics as part of this study. We are particularly keen to include people who have written extensively about the theoretical underpinnings of qualitative research to explore the essential requirements of quality assessment.

I/we look forward to meeting/seeing you on xxxx. Meanwhile, many thanks for making time to be involved in the study.

Yours sincerely

xxxxxxxxx

Appendix 6
Letter of Introduction for Research Commissioners/Funders/Users

Reference P6055/R2/A

Dear

Quality Standards in Qualitative Evaluation Methods

Thank you very much for agreeing to be interviewed for this study which the National Centre for Social Research is conducting on behalf of the Strategy Unit in the Cabinet Office.

I am writing to confirm that I/xxx and I will come to your office at [insert time] on [insert date]. I/We would be grateful for up to an hour and a half of your time. A short information sheet about the study is enclosed.

The objective of the study is to produce a framework against which the quality of qualitative evaluation studies can be assessed. In the interview we will want to explore the criteria which you see as critical in making judgements about the standard of qualitative studies. It would be useful to begin by hearing about the type of qualitative research studies with which you may have been involved, in terms of the methods and objectives. We would also like to ask you about the type of knowledge or understanding you see qualitative research as bringing, and the type of evidence you look to it to generate.

It would be useful to hear about any experience you may have of frameworks or good practice guidelines for assessing qualitative research, and your views on what can make them useful. The focus of this study is on quality standards in qualitative *evaluations*, but we think it will be important to talk about qualitative research more generally too. Your views on these issues will be very helpful in informing the development of a quality standards framework.

We will be interviewing a range of research commissioners, policy-makers, research practitioners and academics as part of this study. *[We are particularly keen to ensure that we include people who may have experience of using qualitative research as well as other types of evidence to inform policy development and evaluation, and their views on the features that make qualitative research sufficiently secure for this./We are particularly keen to ensure that we include people with a strategic involvement in the funding of qualitative research studies, and their views of the criteria by which qualitative research applications and studies should be judged./We are particularly keen to ensure that we include people who have a strategic involvement in commissioning and managing qualitative research studies, and their views on the criteria by which qualitative research studies should be judged.]*

I/we look forward to meeting/seeing you on xxxx. Meanwhile, many thanks for making time to be involved in the study.

Yours sincerely
xxxxxxxxx

Appendix 7
Introductory Note for Respondents

National Centre *for* Social Research   **Quality Standards in Qualitative Evaluation Methods**

The National Centre for Social Research is carrying out this study on behalf of the Centre for Management and Policy Studies. The purpose of the study is to identify a set of standards or a framework against which the quality of policy evaluation studies using qualitative research methods can be assessed, and to produce guidance on how such a framework could be applied in practice. The study will feed into a number of initiatives being co-ordinated by the Cabinet Office, and will link more broadly with other initiatives on the development of evidence-based policy-making.

The study involves five elements:

- A systematic review of literature on qualitative research methods, including a review of the ways in which different methods and approaches are used in evaluations. This will also involve reviewing existing quality standards frameworks.

- Thirty in-depth interviews with a range of qualitative research practitioners, academics and researchers who have written about qualitative research from a theoretical perspective, commissioners and funders of qualitative research, and policy-makers who have used qualitative research evidence in the development and evaluation of policies.

- The development of a proposed set of standards or a quality framework, which incorporates as fully as possible the various perspectives generated by the first two stages.

- Two workshops, held with the groups involved in the in-depth interviews, to review, critique and refine the framework.

- The application of the framework to a small number of research reports to assess its utility, followed by further refinement of the framework.

A full report will be published by the Cabinet Office, and it is also intended that the framework, and accounts of other elements of the study, will be disseminated through peer-reviewed journals and the Internet. The Cabinet Office expects to use the project findings to inform best practice guidance and professional training courses for government evaluators and policy-makers.

For further information or to discuss any aspect of the project, please contact:

Lucy Dillon, Senior Researcher, National Centre for Social Research
Email l.dillon@natcen.ac.uk Direct line 020 7549 9546

Jane Lewis, Director, Qualitative Research Unit, National Centre for Social Research
Email j.lewis@natcen.ac.uk Direct line 020 7549 9541

Rebecca Stanley, Cabinet Office
Email Rebecca.Stanley@cabinet-office.x.gsi.gov.uk Direct line 020 7276 2210

Phil Davies, Cabinet Office
Email Phil.Davies@cabinet-office.x.gsi.gov.uk Direct line 020 7276 1862

Topic guide for qualitative research commissioners, funders and users

National Centre *for* Social Research

P6055

Quality Standards in Qualitative Evaluation

TOPIC GUIDE

Introduction:
- Interviewer to introduce themselves and the National Centre for Social Research
- Introduce study
  Include:
    - *Objectives*
    - *Explain study is working towards a framework*
    - *Stages*
    - *Range of people to be interviewed*
    - *Reinforce that we are interviewing them as part of a range of perspectives which will feed in – which is why their perspective is so important*
    - *Outputs*
    - *How disseminated*

- Discuss confidentiality
    - *Discuss whether they would be prepared to be named in a list of people who participate in the study*
    - *Explain that data will not be attributed*
- Ask for permission to use the tape recorder

1. Background information – general
- Job title/responsibilities
- Department/section/unit/policy area
- Length of time in post
- Brief overview of disciplinary background, training and career path
- Other relevant posts/responsibilities currently/recently *(in particular to try to get a sense of any other 'hats' they wear)*

2. **Background – interactions with research** (as investigator, commissioner, user etc.)
- How the role described above brings them into contact with research
  (All research, not just qualitative)
- Types of <u>qualitative</u> research with which involved current or past
- Probe: methods, objectives, uses, researchers
- Involvement with <u>evaluation</u>
- Key studies involved with recently
- Extent of contact with proposals/conduct (or management of)/reports and other outputs.

Explain that our interest is in qualitative <u>evaluations,</u> but that we are going to start by talking about qualitative research generally

3. **Key contributions that qualitative research can make**
   (i.e. – in their view, what is it here to do?)

   (There are a number of ways into this question – select from the following:)

   - Link with the studies they have mentioned above

   - What kind of policy questions has qualitative research informed?
   - What is qualitative research good at showing/explaining?
   - What kind of evidence/knowledge do they want to get from qualitative studies?
   - What would they hope to learn from a qualitative study?

   Probe:
   - How this is influenced by beliefs about/conception of qualitative research/key defining characteristics
   - How this influences quality criteria used to judge qualitative research

   - Evaluation:
     - What they are looking for qualitative studies to bring to an <u>evaluation</u>
     - Whether contribution is different for evaluations

       Probe:
       (To explore their definition of evaluation)
       - What they are thinking about when they are talking about evaluation

   Use as a way in to quality criteria:

Aspects that make them feel safe/confident with a piece of qualitative research
   Factors that create:
   - Confidence/nervousness

4. Quality criteria
Broad opening question:
What do they see as the key criteria for assessing the quality of qualitative research?
Explore.

Probe as needed:
Explore their **criteria** *(one by one)*
(Probes to find out more about each criterion they mention)
- Precisely what they mean by it
- Why that criterion is important
- What needs to be done to create/secure it
- How they judge whether these requirements have been met/will be met (proposal)
- How easy or difficult is it to make that judgement

Explore their criteria with prompts as needed:
(Prompts to stimulate them to think of more criteria:)

Explore three broad types of criteria:	
Criteria that relate to the <u>design/approach</u> <u>of qualitative research</u> *(Link to what they said in Section 3 about the key contribution qualitative research can make)* • Whether qualitative evidence is the type that is needed • The match between qualitative research and the aims of the study *(Explore both re match of qualitative research generally and match of method used)*	Criteria that relate to the <u>practice</u> of qualitative research *(i.e. <u>how</u> it was done)* • the way the research is conducted *Prompt key research stages if appropriate:* sampling, data collection, analysis
Criteria that relate to the <u>research team</u> • Expertise • Experience/track record • Stance/role	Criteria that relate to the <u>outputs, reporting</u> <u>and documentation</u> • What are the key criteria by which reports are judged • What makes the findings/conclusions persuasive and credible *(Include interpretation and auditability)* • What they want to see documented about approach/practice
Other prompts:	
Standards & expectations in qualitative research: *Prompt where appropriate, if haven't already mentioned e.g.:* • Reliability/consistency/replicability • Generalisability/transferability to other settings/ability to draw wider inference • Objectivity/lack of bias/neutrality • Validity/credibility/verification/integrity • Reflexivity/reflection on role or impact of researcher/reflection on methods used and limitations	Examples *(Use as a vehicle for displaying criteria)* • **Example of a <u>good</u> piece of research** • *(Note to interviewer: use just for qualitative studies, or if appropriate, repeat for both qualitative and quantitative)* • Probe: • What it was about it that made it good • **Example of a <u>bad</u> piece of research**

• Timeliness	*(Note to interviewer: use just for qualitative studies, or if appropriate, repeat for both qualitative and quantitative)* *Probe*: What it was about it that made it bad

Is there anything about <u>evaluation</u> that makes the criteria different?

5. Universality of criteria

Extent to which quality criteria can be applied across different:

- Methods
- Traditions
- Uses/purposes *(e.g. evaluations vs non-evaluations, or commissioned vs theoretical research)*
- How do criteria need to be different for proposals and conduct/reporting
- Transferability of criteria between qualitative and quantitative studies

6. Frameworks

(Interviewer to say a bit about frameworks and that there are a number of ways they could be done)

- Familiarity with existing frameworks/guidelines/checklists of good practice
  - Description of those familiar with
  - Any experience of trying to apply
  - Whether they have a departmental or institutional framework/guidance – for use with proposals

- Developing a framework
  - Views on whether possible to produce a framework
  - Possible uses
  - Ideal form
  - Whether it would have value for <u>them</u> in own work
    - Would they use it?
    - How?
    - What would make it easy or difficult to use

If appropriate: It may be worth flagging up next stages, and that we may want to ask some people to contribute again in the workshops

![National Centre for Social Research logo]

P6055

Quality Standards in Qualitative Evaluation
ROUND 2 TOPIC GUIDE: ACADEMICS/NON-FW

Note: conduct of each interview will need to be particularly responsive to what
we know of each respondent from the literature review.

Introduction

- self and NatCen

- study
  - objectives, working towards framework
  - stages: literature review, interviews, workshops, application of
    framework
  - range of people interviewed: academics, theoreticians, practitioners, those
    involved in commissioning, funding, managing and using qualitative
    research
  - outputs: full report plus (probably) separate papers e.g. on framework

- focus on qualitative research reflects the fact that there remains more debate
  about criteria, and fewer consistently accepted standards, than quantitative
  research – rather than an assumption that the quality criteria are entirely
  different for qual research, or any desire to proselytise

- confidentiality
  - comments not attributed
  - would like to name those contributing in list in report

1. Background: general

- job title/role, responsibilities
- length of time in post
- other relevant current/recent responsibilities
- brief overview of disciplinary background, training and career path:
  - esp other posts which have brought into contact with research in different
  ways

2. Background: interactions with research

- how role/s described bring/s them into contact with research
- types of research with which involved
    - methods, objectives, uses, researchers
    - types of evaluation research and qualitative research in particular
- key studies with which currently involved
- extent of contact with
    - evaluation of proposals/applications
    - conduct of research
    - management or supervision of research
    - research synthesis/systematic review

3. Criteria for judging the quality of qualitative research/evaluations
This section will need to be very strongly focused around the key criteria/quality issues which have emerged from respondent's own writing.

Acknowledge their contributions to the literature

What do they see as the most critical criteria for assessing the quality of qualitative research?

Explore their <u>criteria</u> *(one by one)*
- Why does that matter?
- What needs to be done to create/secure it?
- What specifically would need to be demonstrated for criterion to be fulfilled?
- How can criterion be stated/described to facilitate making judgement?
- What difficulties would arise in making this judgement?

PROMPT
Any criteria not mentioned which we know they have discussed in their writing?

PROMPT
Any key areas missing from the 'criteria box'?

Criteria that relate to the **design/approach of qualitative research** • Whether qualitative evidence is the type that is needed • The match between qualitative research and the aims of the study	Criteria that relate to the **practice** of qualitative research *(i.e. <u>how</u> it was done)* • The way the research is conducted *Prompt key research stages if appropriate:* Sampling, data collection, analysis
Criteria that relate to the **research team** • Expertise • Experience/track record • Stance/role	Criteria that relate to the **outputs, reporting and documentation** • What are the key criteria by which reports are judged? • What makes the findings/conclusions

	persuasive and credible? *(Include interpretation and auditability)* • What they want to see documented about approach/practice
What matters in terms of … for assessing quality/standards • **Reliability**/consistency/replicability • **Validity**/credibility/verification/integrity • **Generalisability**/transferability to other settings/ability to draw wider inference **vs Context specificity** and how important is context to analysis • **Objectivity**/lack of bias/neutrality • **Reflexivity**/reflection on role or impact of researcher/reflection on methods used and limitations • **Timeliness** • **Relevance for policy and practice**	*7.3.1.1* Examples (use as a vehicle for displaying criteria if appropriate) • **Example of a <u>good</u> piece of research** • What it was about it that made it good • **Example of a <u>bad</u> piece of research** • What it was about it that made it bad

- Is there anything about <u>evaluation</u> that makes the criteria different/brings additional requirements?

- Linking to key criteria identified:
  - *priorities for inclusion*
  - *are there higher order criteria that need to be satisfied before others matter (i.e. primary markers)*
  - *any trade-offs, compromises, issues of lesser importance*

any other difficulties envisaged in stating/describing criteria in context of framework

4. Views about quality frameworks

- familiarity with existing frameworks/guidelines/good practice checklists
  - which, if any, aware of
  - any departmental or institutional guidance aware of
  - know of any attempts to develop or apply frameworks/criteria in unpublished/grey literature? what?

- views about frameworks seen
  - key distinctions if aware of several
  - views about content: any issues with which disagree, agree in particular
  - any frameworks which see as more or less useful
    e.g. in content, approach, format
  - any experience of applying/using frameworks
  - what if any impact have they had

- views about concept of a quality framework
  - how feel about the idea of frameworks for qualitative research anyway/are they appropriate? why? why not?
  - possible uses and applications
  - any application or use can envisage relating to own work
  - perceived shortcomings: what a framework can't do
  - challenges or difficulties envisaged in producing a useful framework

5. Limitations on scope of frameworks

- is it possible to devise a framework which people would agree with/use?
  - why/why not?

- epistemological/ontological position
- how critical to state this in drawing up a framework? *[keep to this aspect]*
- how far can a single framework span different traditions/approaches/positions?
- what are the key traditions/positions that need to be distinguished? *[must clarify what they see as the critical distinctions]*
- what issues produce key tensions or strains?
- what implications would this have for content? *[which traditions/approaches would require different criteria/frameworks]*

- different types of research/method
- how far can a single framework span different methods?
- how far can a single framework span different types of research? (e.g. evaluation cf. non-evaluation; applied cf. theoretical; qualitative cf. quantitative)
- where are the areas of tension or strain in trying to span different types of research/methods?
- how far can a single framework span different stages? e.g. proposal/application; conduct; outputs
- implications of these issues for content of framework *[which methods/types of research/stages of research would require different criteria/frameworks]*
-

6. Implications for our development of a framework

- further reflections on feasibility
- further reflections on value, use, role, applicability
- key features or general approach seen as desirable/undesirable in terms of format of framework

[Make a final check on any unpublished work on developing or applying frameworks that we need to be aware of?]

![National Centre for Social Research logo]

__P6055__

Quality Standards in Qualitative Evaluation
ROUND 2 TOPIC GUIDE: FRAMEWORK DEVELOPERS

Note: conduct of each interview will need to be particularly responsive to what we know of each respondent from the literature review.

Introduction

- self and NatCen

- study
  - objectives, working towards framework
  - stages: literature review, interviews, workshops, application of framework
  - range of people interviewed: academics, theoreticians, practitioners, those involved in commissioning, funding, managing and using qualitative research
  - outputs: full report plus (probably) separate papers e.g. on framework

- focus on qualitative research reflects the fact that there remains more debate about criteria, and fewer consistently accepted standards, than quantitative research – rather than an assumption that the quality criteria are entirely different for qual research, or any desire to proselytise

- focus of research is on qualitative evaluation but placing discussion in the context of qualitative research more generally

- confidentiality
  - comments not attributed
  - would like to name those contributing in list in report

1. Background: general

- job title/role, responsibilities
- length of time in post
- other relevant current/recent roles/activities within research

- brief overview of disciplinary background, training and career path:
  - esp other posts which have brought into contact with research in different ways

2. Background: interactions with research

- how role/s described bring/s them into contact with research
- types of research with which involved
  - methods, objectives, uses, researchers
  - types of evaluation research and qualitative research in particular
- key studies with which currently involved
- extent of contact with
  - evaluation of proposals/applications
  - conduct of research
  - management or supervision of research
  - research synthesis/systematic review
  - writing about qualitative research methods/social research

3. Frameworkers: experiences of developing own framework

This section will need to be very tailored to the individual respondent and their framework

- what prompted them to produce one
  - how aware of previous frameworks/criteria /checklist/guidelines etc.
  - any deficiencies perceived in other frameworks
  - how wanted to move on in producing own

- reflections on **content** of own framework compared with others
  - what see as distinguishing features of own
  need to be prepared to suggest what we see as distinguishing features, anything given unusual emphasis, anything not included that occurs in others: explore how deliberate this was, reasons, rationale
  - **reasons for including/excluding specific criteria**
  - any areas of difficulty relating to:
    – inclusion of items
    – labelling/terminology
    – description of item
    – specifying what is required to fulfil criterion
  - anything on reflection would add, reframe, change etc.

- reflections on **process** of producing framework
  - what was the process?
  - if collaborators: how much discussion, debate, difference of view among authors, what about?
  - any other issues raised for them by process of producing framework

- how explicit was own epistemological/ontological perspective in relation to production/content of framework?

- how did it impact on inclusion criteria, content, approach?

- application
  - whether see their own as a framework/set of criteria/set of guidelines and what is the difference between these
  - how do/did they intend their version be applied?
  - any experiences of applying own framework
  - examples of where used
  - how easy to use: what inhibits or facilitates this
  - level of specificity required in description of criteria to make it workable: challenges this raises
  - have others used theirs (any evidence of this/any feedback)?

4.   Views about other quality frameworks

- familiarity with other existing frameworks/guidelines/good practice checklists
  - which aware of
  - know of any other attempts to develop or apply frameworks/criteria in unpublished/grey literature? what?

- views about frameworks seen
  - key distinctions if aware of several
  - views about content: any issues with which disagree, agree in particular
  - any frameworks which see as more or less useful
    e.g. in content, approach, format
  - any experience of applying/using other frameworks
  - what impact have other frameworks had?

- general views about concept of a quality framework
  - in general, how do they feel about the idea of frameworks for qualitative research?
  - any other uses and applications not already discussed
  - perceived shortcomings: what a framework can't do
  - challenges or difficulties envisaged in producing a useful framework
  - desirable/undesirable features of format (added in light of suggestion to remove last section)

5.   Universality of frameworks
- do they see it as possible to produce a framework which receives reasonably general consensus?
  - how important is it to do that?
  - what are likely to be the major obstacles/differences of view?

- epistemological/ontological position
  - how critical to state this in drawing up a framework? *[keep to this aspect]*
  - how far can a single framework span different traditions/approaches/positions?
  - what are the key traditions/positions that need to be distinguished? *[must clarify what they see as the critical distinctions]*

- what issues produce key tensions or strains?
- what implications would this have for content? *[which traditions/approaches would require different criteria/frameworks]*

- different types of research/method
  - how far can a single framework span different methods?
  - how far can a single framework span different types of research? (e.g. evaluation cf. non-evaluation; applied cf. theoretical; qualitative cf. quantitative)
  - where are the areas of tension or strain in trying to span different types of research/methods?
  - how far can a single framework span different stages? e.g. proposal/application; conduct; outputs
  - implications of these issues for content of framework? *[which methods/types of research/stages of research would require different criteria/frameworks]*

6. Additional criteria for judging the quality of qualitative research/evaluations
This section will focus on criteria that have not been mentioned in the previous discussion about frameworks.

Are there any other criteria, not previously discussed, that they see as critical for assessing the quality of qualitative research?

- Explore their <u>criteria</u> *(one by one)*
  - Why does that matter?
  - What needs to be done to create/secure it?
  - What specifically would need to be demonstrated for criterion to be fulfilled?
  - How can criterion be stated/described to facilitate making judgement?
  - What difficulties would arise in making this judgement?

PROMPT
Any other criteria not mentioned which we know they have discussed in their writing
PROMPT
Any key areas missing from the 'criteria box'

Criteria that relate to the **design/approach of qualitative research** • Whether qualitative evidence is the type that is needed • The match between qualitative research and the aims of the study	Criteria that relate to the **practice** of qualitative research *(i.e. how it was done)* • The way the research is conducted *Prompt key research stages if appropriate:* Sampling, data collection, analysis
Criteria that relate to the **research team** • Expertise • Experience/track record • Stance/role	Criteria that relate to the **outputs**, **reporting and documentation** • What are the key criteria by which reports are judged • What makes the findings/conclusions persuasive and credible *(Include interpretation and auditability)* • What they want to see documented about approach/practice
What matters in terms of … for assessing quality/standards • **Reliability**/consistency/replicability • **Validity**/credibility/verification/integrity • **Generalisability**/transferability to other settings/ability to draw wider inference **vs Context Specificity** and how important is context to analysis • **Objectivity**/lack of bias/neutrality • **Reflexivity**/reflection on role or impact of researcher/reflection on methods used and limitations • **Timeliness** • **Relevance for policy and practice**	Examples (use as a vehicle for displaying criteria if appropriate) • **Example of a good piece of research** • What it was about it that made it good • **Example of a bad piece of research** • What it was about it that made it bad

- Is there anything about <u>evaluation</u> that makes the criteria different/brings additional requirements?
- linking to key criteria identified:
  - *priorities for inclusion*
  - *are there higher order criteria that need to be satisfied before others matter*
  - *any trade-offs, compromises, issues of lesser importance*

any other difficulties envisaged in stating/describing criteria in context of framework

7.   Implications for our development of a framework
- further reflections on feasibility
- further reflections on value, use, role, applicability
- key features or general approach seen as desirable/undesirable in terms of format of framework

[Make a final check on any unpublished work on developing or applying frameworks that we need to be aware of]

Appendix 9
Letter of Invitation to Workshop

email l.dillon@natcen.ac.uk
Direct line 020 7549 9551

18th October 2002

Re: Quality Standards in Qualitative Evaluation

Dear

During the summer, you kindly took part in an interview for the above study which we are conducting for the Cabinet Office. We all greatly appreciate your participation and the important contribution you have made to this study to date.

We are now writing to ask if you would be willing to help us again by taking part in the next stage of the study – a workshop to discuss the quality framework we are developing.

We expect there to be around 12 participants in the workshop, in which we aim to debate the approach we are adopting to developing quality criteria. Prior to the workshop, we will send participants a briefing paper explaining the background to our general strategy with examples of the quality concepts and indicators we have generated. The workshop will act as a forum for you and other people who took part in the interviews, to express your views on our strategy, the draft framework and quality concepts and indicators, and to make suggestions for its further development. We feel that this would be an excellent forum in which to debate these issues.

The workshop will take place on 22nd November between 10 am and 4 pm at the London Voluntary Resource Centre, 356 Holloway Road. As a small appreciation of the time involved in taking part in the workshop, we will be paying an honorarium of £150. We will, of course, also reimburse any travel expenses you incur.

We realise there are many demands on your time but would be delighted if you are able to take part in this workshop. We certainly would greatly value your reflections on the framework we are developing. However, if it is not possible for you to attend, we will of course understand and would just wish to thank you again for the time you have already given to the study.

We will contact you in the coming week to confirm whether you will be able to take part. In the meantime, if you have any queries, please do not hesitate to contact Lucy Dillon (Senior Researcher, NatCen) on 020 7549 9551.

With best wishes
Yours sincerely

Jane Lewis
Director, Qualitative Research Unit

Appendix 10
Programme for Workshop

Quality Standards in Qualitative Evaluation
Workshop Programme
22nd November 2002
at
London Voluntary Resource Centre, 356 Holloway Road, London N7 6PA

9.30–10 am	Coffee
10–10.15 am	Welcome and introductions Jane Lewis
10.15–10.45 am	Overview of approach and rationale of framework Jane Ritchie
10.45–11.45 am	Discussion of approach and rationale Liz Spencer
11.45–12 pm	Coffee
12–1.15 pm	Quality concepts. Small groups to discuss with feedback in plenary session
12–12.45 pm	Small group Group A: Chair- Liz Spencer Group B: Chair- Jane Lewis
12.45–1.15 pm	Plenary session Chair- Jane Ritchie
1.15–2 pm	Lunch
2–3.15 pm	Quality markers. Small groups to discuss with feedback in plenary session
2–2.45 pm	Small group Group A: Chair- Jane Ritchie Group B: Chair- Liz Spencer
2.45–3.15 pm	Plenary session Chair- Jane Lewis
3.15–4 pm	General reflections on the framework development Jane Lewis
4–4.30 pm	Tea

Appendix 11
Examples of Different Forms of Presentation

	Topic	Topic
Form of presentation	**On clarity about sample composition and selection**	**On visibility of link between data and conclusions**
Questions to be addressed	**Critical Appraisal Skills Programme (2001):** *(empirically based)* **4 Was the recruitment strategy appropriate to the aims of the research?** Has the researcher explained how the participants were selected? a) Have they explained why the participants they selected were the most appropriate to provide access to the type of knowledge sought by the study? b) Are there any discussions around recruitment? (e.g. why some people chose not to take part) (p. 2)	**Blaxter (1996):** *(empirically based)* 16. Is a clear distinction made between the data and its interpretation? • *Do the conclusions follow from the data? (It should be noted that the phases of research – data collection, analysis, discussion – are not usually separate and papers do not necessarily follow the quantitative pattern of methods, results, discussion* 17. Is sufficient of the original evidence presented to satisfy the reader of the relationship between the evidence and the conclusions? • *Though the presentation of discursive data is always going to require more space than numerical data, is the paper as concise as possible?*(p. 37*)*
Some amplification of specific criteria/ principles	**Popay, Rogers and Williams (1998):** *(Philosophically based: generic)* … the key question to be addressed in assessing standards of sampling is, 'Does the sample produce the type of knowledge necessary to understand the structures and processes within which the individuals or situations can be located?' Particular types of sampling will influence whether the criteria of adequacy in the level of subjective meaning is met. Rather than randomness and calculations based on statistical power, the process by which individuals or cases were theoretically or purposefully sampled needs adequate description. The use of key informants is frequently a preferred way of ensuring the selection of respondents with the appropriate knowledge. As Blumer (1979) has suggested: A half dozen individuals with such knowledge constitute a far better 'representative sample' than a thousand individuals who may be involved in the action that is being formed but who are not knowledgeable about that formation (p. 156) In mixed-methods studies, cases for in-depth study can also be sampled from a larger survey population, which may then serve as a context and a basis for empirical generalisations (pp. 346–47).	**Treloar (2000):** *(Empirically based)* Is there a clear progression from research question to conclusions drawn from the data? **Factors to consider** Clear analytic logic *Commentary* A report of qualitative research should contain a clear analytic story throughout the text *(reference cited).* Each decision and activity in the project should relate to this analytic logic and be apparent in the description of the process

Some detail of what needs to be manifest	**Forchuk and Roberts (1993):** *(Philosophically based: generic)*	**Drisko (1997):** *(Philosophically based: specific)*
	The study informants or participants, context and researcher must be described in relevant detail	CRITERION 6: CONSISTENCY OF CONCLUSIONS WITH STUDY PHILOSOPHY AND DATA
	The informants or participants, however few, must be described. Adequate sample size is generally determined through saturation and recurrent patterning, whereby the researcher finds that additional participants provide similar rather than dissimilar information. Therefore, although the required sample size can be estimated before the project begins, it cannot be truly known until data gathering is well underway.	The analysis, conclusions, and recommendations of a qualitative study should be consistent with the study philosophy and objectives, and limited to the presented data … qualitative reports will vary widely in the nature and extent of their conclusions and recommendations. Illustrating a subjugated viewpoint may require little discussion by the author; critical analysis of the same viewpoint, however, requires extensive discussion and analysis to yield meaningful conclusions. Qualitative evaluation requires careful analysis to demonstrate how the researcher derives recommendations from data and to establish limits to the transferability of the recommendations. Developing grounded theory requires careful and extensive analysis to assure the theory is credible, complete, and meaningful in content as well as to establish relevant contexts and settings for its application. Researchers must offer conclusions consistent with their chosen philosophical approach and method and frame interpretation of findings within the original study setting. Such points of coherence are important to the study's overall credibility and to the utility of conclusions and recommendations offered.
	The researcher must say how the study sites were determined and describe the selection of participants and their level of participation. Opportunistic samples are appropriate and sampling is purposeful, but these must be described in detail (Cobb and Hagemaster,1987). Particular informants (e.g. expert nurses) can be purposely included for their relevant knowledge or experience. Random selection of subjects for representativeness should not be used for qualitative research. Morse(1986) suggests that the criteria of appropriateness and adequacy be used to evaluate sampling: an appropriate sampling method has a good 'fit' with the study purpose (p. 185). The information provided by an adequate sample is of good quality and complete, and then provides sufficient information. Morse considers these factors to be more important than sample size (p. 185)	
	(Role of researcher also discussed – not reproduced here) (pp. 51–2)	*Two further paragraphs follow under this head dealing with issues concerning generalisation and the development of both general and local theories.* (page unavailable, web text)

Appendix 12

Examples of Differences in Criteria Between Empirically and Philosophically Based Frameworks

Empirically based	Philosophically based
Example 1: Boulton and Fitzpatrick (1997) ***On the impact of the research process*** *6. Were efforts made to minimise the impact of the research process on study findings?* What participants say to a researcher may be influenced by the way they perceive the study investigators. Was the study presented to the participants in an appropriate way? Were interviews held in an appropriate context? Did the interviewer neutral and independent relationship with the respondent? (p. 84)	**Example 1: Mays and Pope (2000) (*generic*)** ***On reflexivity*** Reflexivity means sensitivity to the ways in which the researcher and the research process have shaped the data collected, including the role of prior assumptions and experience, which can influence even the more avowedly inductive enquiries … (Researchers) can and should make their personal and intellectual biases plain at the outset of any research reports to enhance the credibility of their findings. The effects of personal characteristics such as age, gender, social class and professional status(for example, that of doctor, nurse, physiotherapist, sociologist etc.) on the data collected and the 'distance' between the researcher and those researched also need to be discussed.(p. 96)
Example 2: Greenhalgh and Taylor (1997) ***On Credibility*** ***Question 7: Are the results credible, and if so, are they clinically important?*** We obviously cannot assess the credibility of qualitative results through the precision and accuracy of measuring devices, nor their significance via confidence intervals and numbers needed to treat. It usually takes a little more than plain common sense to determine whether the results are sensible and believable, and whether they matter in practice (page unavailable, web text). One important aspect of the results section to check is whether the authors cite actual data. Claims such as 'general practitioners did not usually recognise the value of audit' would be definitely more credible if one or two verbatim quotes from the interviewees were reproduced to illustrate. The results should be independently and objectively verifiable – after all, a subject either made a particular statement or (s)he did not – and all quotes and examples should be indexed so that they can be traced back to an identifiable subject and setting (page unavailable, web text).	**Example 2: Beck(1993) (*generic*)** ***On Credibility*** In quantitative research, internal validity measures whether or not the manipulation of the independent variable really makes a significant difference in the dependent variable. Credibility in qualitative research measures how vivid and faithful the description of the phenomenon is. The informants, and also the readers who have had that experience, should recognise the researchers' described experiences as their own. To evaluate the credibility of a qualitative study, certain questions regarding the research design need to be answered. The questions identified in Table 1 can be used as a guide in critiquing this criterion. **TABLE 1: Evaluation of the Credibility of Qualitative Research** *covers 11 questions, including the following* 1. Did the researchers keep in-depth field notes regarding the researcher-informant relationships? 3. Did the researchers keep field notes of their actions, interactions, and subjective states during the investigation? 6. Were multiple methods of data collection (triangulation) used to determine the congruence of the results among them? 7. Were the readers provided with rich excerpts from the transcripts or field notes? 8. Did the researchers validate the findings with the informants? 10. Were data analysis procedures reviewed by a judge panel to prevent researcher bias and selective inattention (Morse, 1989)? 11. Do the readers view the findings as meaningful and applicable in terms of their own experiences? (p.284-285)

Appendix 13
Summary Profile of Existing Frameworks

FRAME 1: STATUS OF THE RESEARCH

Feature	Broad Criteria/question	Specific criteria	Examples of how may be manifest or demonstrated
Scope and purpose	Explicitly stated aims/objectives of the research	Clearly stated/formulated research questions Underlying purpose clear (e.g. description/theory building/ hypothesis testing etc.)	*Focused research questions/not too many* *Discussion of ways in which aims/functions changed during study and why*
	Relevance/importance of/need for investigation	Statement of why investigation is needed Clarity about how research will contribute to existing knowledge/ understanding	*Target audience for research identified* *Relevance to stakeholders clearly stated* *Discussion of ways in which research will bring new insights* *Discussion of ways in which will empower participant population/provide information that will be useful*
	Clarity about context of study	Clarity about connection to existing body of knowledge Clearly stated hypotheses/assumptions	*Systematic/comprehensive/critical review of relevant literature* *How do authors locate the need for the study within existing body of knowledge* *Clear outline of theoretical/conceptual framework within which working* *Explicitly stated or implied frame of reference*
	Clarity about phenomena under study	Details of nature of phenomena being studied	*Discussion of what is being studied and why* *How phenomena under study relate to theoretical/conceptual framework*
Design	Appropriate use of qualitative methods	Discussion of why qualitative methods are appropriate for research aims	*Discussion of how qualitative methods will bring type of knowledge/understanding required* *Demonstration of fit between purpose and style of investigation*
	Clarity about tradition/approach being used	Discussion of why particular approach chosen/methods used	*Discussion of ontological/epistemological basis of research and how methods fit within it*
	Appropriate design to meet study objectives	Discussion of why specific design features incorporated	*Discussion of specific features of design and their relevance/appropriateness/sensitivity for study objectives*

	Need for staged/phased/longitudinal research explained	Clarity about purpose of different stages and components of research	*Discussion of specific aims of different stages/phases/components* *Discussion of how different components of research will be linked/integrated*
	Clarity about basis/appraisal criteria of evaluative research	Description of how outcomes will be determined Rationale for evaluative criteria explained	*How and when will outcomes be defined; by whom?* *How and when will evaluative base be determined; by whom?*
Timescale	Manageable within timeframe/resources	Timetable/resources made explicit Rationale given for timeframe of study	*Discussion of implications of timetable/resources on conduct of research* *Discussion of why temporal structure of different phases was chosen*
Researcher	Understanding of researcher's connection to research study and subject matter	Perspectives on/relationship to subject matter explained Relationship with funding body known	*Relevant personal history stated* *What precipitated study in context of experience/underlying beliefs* *Potential conflicts of interest discussed*
	Clarity about philosophical orientation	Discussion of philosophical perspective within which conducting research	*Description of philosophical base and how it affects the way research is conducted*
	Does researcher have appropriate skills/experience to conduct the research	Knowledge and experience of qualitative paradigm and methods demonstrated	*Were professional roles and qualifications of researchers clearly stated*

Summary profile of existing frameworks (contd.)

FRAME 2 CONDUCT OF RESEARCH

Feature	Broad criteria/question	Specific criteria	Examples of how may be manifest or demonstrated
Sample	Appropriate for qualitative research	Discussion of form of sample design used and rationale for using	*Design theoretically/purposively justified to yield coverage required*
	Appropriate for aims of study	Clear rationale for sample design	*Design of sample justified in relation to objectives of study*
	Clarity about selection of settings	Description of areas/locations/settings of study	*Discussion of why areas/location/settings selected* *Description of areas/locations/settings* *Discussion of ways in which constraints of context/location of research will be met*
	Clarity about coverage of sample	Description of coverage of sampled individuals/cases/events	*Composition of sample explained* *Selection criteria made explicit* *Inclusions and exclusions explained*
	Comprehensive	Coverage of full range of identified constituencies/sample units/events Adequate size for design used/study objectives	*Coverage of confirming/contradictory cases/evidence* *Composition of achieved sample documented/discussion of how compares with target sample*
	Appropriate sample frame for aims of study	Methods of recruitment explained/appropriate for aims of study	*Discussion of how sample recruited/form of approach used* *Discussion of any prior relationship between researcher and sampled individuals explained* *Sites of recruitment fit needs of study*
Data collection	Appropriate to study's aims	Data collection methods explained/justified Sensitive/responsive to study sample/locations	*Were data collection methods justified/appropriate to nature of data required?*

	Rigorously/systematically conducted	Systematic documentation of tools/guides/researcher role Recording methods explicit Modifications justified/explained	*Evidence of sustained exposure to field work settings/engagement with participants* *Description of where interviews/observations took place* *Evidence that in-depth investigation took place/was facilitated/allowed saturation of data* *Evidence that approach was non-directive* *Discussion of how fieldwork methods changed/evolved and why*
	Role of researcher explained	Discussion of role of researcher in data collection	*Discussion of who did the interviewing/observations* *Discussion of relationship between researcher and participants during fieldwork*
Analysis	Appropriate for qualitative research study/aims	Analysis methods explained/justified in relation to type of data collected	*Evidence that analysis took place - not just data sorting* *Structure/process of analysis made clear*
	Clarity about analytic process	Documentation of analytic methods/tools used Discussion of how analytic process was conducted Evidence of rigorous/systematic analysis	*Clarity about how raw data organised and utilised/reduced* *Use of any software programmes described and justified* *Discussion/documentation of coding systems/conceptual frameworks and how evolved* *Discussion of how context of data was retained* *Evidence of more than one researcher involved and roles/interactions*
	Comprehensive	Analysis method takes account of all observations/contradictory evidence	*Evidence that all cases were analysed/reasons for any exclusions* *Evidence of saturation – discussion of any incomplete analytic areas* *Discussion of how conflicting evidence/inconsistency was dealt with*
Audit trail	Sufficient information for research process to be 'audited' by others (This overlaps with other sections but is distinguished by specific reference to facilitating 'auditing')	Full documentation of methods/systematic record of data collection/field notes of actions/interactions (as above) Sufficient discussion of research process such that others can follow 'decision trail'	*Description of social/physical and interpersonal contexts of data collection* *Provision of 'thick description'* *Discussion of decisions taken during conduct of research* *Discussion of researcher's experience in conducting study*

Reflexivity	Reflections on the researcher's impacts on data and its interpretation	Consideration of researcher's interaction with the research process, particularly - subject coverage - relationships with participants - analysis	*Evidence of self awareness of researcher* *Record of subjectivity of researcher* *Documentation of effects of research on researcher*
	Reflections on research conduct and process	Evidence of attention to problems encountered in design/data collection/analysis Discussion of impact of methods on data collected/efforts made to minimise bias	*Identification of any potential biases/gaps* *Discussion of limitations of research/modifications needed*
Ethical issues	Evidence of consideration of ethical issues	Documentation of how research was presented to participants Documentation of consent procedures Description of procedures surrounding confidentiality	*Evidence that protection of respondents was considered/described; how counsel/influence avoided* *Evidence of sensitivity towards participants and contexts of research* *Evidence/documentation of written 'contract' with participants* *Whether informed consent obtained* *Evidence of ethical committee approval*

Summary profile of existing frameworks (contd.)

FRAME 3 NATURE AND STATUS OF EVIDENCE

Feature	Broad criteria/question	Specific criteria	Examples of how may be manifest or demonstrated
Visibility of origins and logic of evidence	Clarity about how evidence and conclusions derived	Documentation of way evidence derived, particularly - presentation of original data - how categories/concepts/themes derived from original data - how(or why) particular interpretation/meaning/significance assigned - how conclusions, explanations and theories were developed	*Sufficient/representative/explicit presentation of original data (and how selected) for base of analytic constructions can be seen* *Evidence that 'subjective' phenomena treated as data/meanings explored in relation to participants' perspectives/participants 'given voice'*
	Conveys subjective meanings of participants		*Demystifies people studied* *Is data collected presented with sufficient depth that insights into meaning and perceptions of participants can be seen?*
	Display of diversity	Attention to/explanation of negative cases Multiple perspectives displayed	*Evidence that dis-confirming cases were considered and explained* *Avoidance of 'holistic bias' - i.e. making data look more patterned or conforming than it should*
	Display of context	Contexts of data collection settings described Individual data contextualised	*'Thick description' of settings and contexts of data collected* *Circumstances/characteristics/histories of participants conveyed in context*
	Tests of validity used/efforts made to assess validity	Multiple analyses to verify/check validity Analysis/evidence checked with participant groups/'makes sense' to participants Triangulation with other sources(qualitative/quantitative/both)	*Codes/concepts/themes checked by more than one researcher* *Expert/peer/external checking of analytic constructs and interpretation* *Use of supporting evidence* *Use of triangulation explained and interpreted* *Techniques of validation are tailored to study, not just 'text book' or rote descriptions*
Confirm-ability/reliability/consistency	Efforts made to assess consistency/reliability/ of results	Evidence that participant accounts have been faithfully represented Evidence of multiple checking of analytic output	*Analysis checked by more than one person* *Inter-observer reliability assessed* *Evidence of recurrence of themes/concepts/patterns* *Repeated documented evidence from primary sources*

		Evidence of replication to support results	*Providing 'quasi-statistics'* *Is saturation/completeness of evidence demonstrated?*
Generalis-ability/transfer-ability	Clarity about extent to which and evidence can be generalised beyond settings/participants of study	Sufficient evidence for typicality/specificity/generalisability to be assessed Wider uses explained and justified Limits of generalisability made clear	*Specification of theoretical relationships from which generalisations can be made* *Provision of sufficient information to see applicability to other settings* *Discussion of how hypotheses/propositions fit into other contexts*
Credibility/integrity/ plausibility	Provides credible evidence/meaningful illumination of the lives and contexts being researched	Provides evidence that is resonate with other knowledge and experience Results/conclusions clearly supported by evidence Evidence 'makes sense' Holds warrant	*Non-partisan discussion of evidence for and against researcher's arguments* *Provides evidence of believable and lived-through experience of those studied* *Interpretations of data are demonstrably plausible*
Contribution /value	Documentation of how evidence relates to existing knowledge Makes a contribution to field of study	Data/evidence located in relation to existing literature/information Provides new insights Increases understanding Discussion of relevance of results to practice/policy/knowledge Objectives/research questions fully addressed	*Meaning/understanding related to what already known* *Assessment of use to people working in the field* *Importance of findings for policy/practice* *Assessment of value/empowerment to participants* *Results offer new insights into the formulation of target phenomena* *Evidence of peer review*
Clarity	Coherently presented	Explicit findings Arguments/evidence clearly presented	*Tells 'a story'* *Coherent logic to presentation of findings*

Appendix 14
<u>Bibliography of Frameworks</u>

Beck, C.T. (1993) 'Qualitative Research: The Evaluation of its Credibility, Fittingness and Auditability', *Western Journal of Nursing Research*, vol. 15, no. 2, pp. 263–6.

Blaxter, M. (1996) 'Criteria for the Evaluation of Qualitative Research Papers', *Medical Sociology News*, vol. 22, no. 1, pp. 68–71.

Boulton, M. and Fitzpatrick, R. (1997) 'Evaluating Qualitative Research', *Evidence-Based Health Policy and Management*, vol. 1, no. 4, pp. 83–85.

Boulton, M., Fitzpatrick, R. and Swinburn, C. (1996) 'Qualitative Research in Health Care II. A Structured Review and Evaluation of Studies *Journal of Evaluation in Clinical Practice*, vol. 2, no. 3, pp. 171–9.

Britten, N., Jones, R., Murphy, E. and Stacy, R. (1995) 'Qualitative Research Methods in General Practice and Primary Care', *Family Practice*, vol. 12, no. 1, pp. 104–14.

Campbell, R., Pound, P., Pope, C., Britten, N., Pill, R., Morgan, M. and Donovan, J. (2003) 'Evaluating Meta-Ethnography: A Synthesis of Qualitative Research on Lay Experiences of Diabetes and Diabetes Care', *Social Science and Medicine*, vol. 56, no. 4, pp. 671–84.

Critical Appraisal Skills Programme (CASP) (2001) *10 Questions to Help You Make Sense of Qualitative Research*, Oxford: CASP.

Cobb, A.K. and Hagemaster, J.N.(1987) 'Ten Criteria for Evaluating Qualitative Research Proposals', *Journal of Nursing Education*, vol. 26, no. 4, pp. 138–43.

Drisko, J.W. (1997) 'Strengthening Qualitative Studies and Reports: Standards to Enhance Academic Integrity', *Journal of Social Work Education*, vol. 33, no. 1, pp. 185–97.

Elliott, R., Fischer, C.T. and Rennie, D.L. (1999) 'Evolving Guidelines for Publication of Qualitative Research Studies in Psychology and Related Fields', *British Journal of Clinical Psychology*, vol. 38, no. 3, pp. 215–29.

Forchuk, C. and Roberts, J. (1993) 'How to Critique Qualitative Research Articles', *Canadian Journal of Nurisng Research*, vol. 25, no. 4, pp. 47–56.

Greenhalgh, T. and Taylor, R. (1997) 'How to Read a Paper: Papers That Go Beyond Numbers (Qualitative Research)', *British Medical Journal*, vol. 315, pp. 740-43.

Healy, M. and Perry, C. (2000) 'Comprehensive Criteria to Judge Validity and Reliability of Qualitative Research Within the Realism Paradigm', *Qualitative Market Research*, vol. 3, no. 3, pp. 118–26.

Hoddinott, P. and Pill, R. (1997) 'A Review of Recently Published Qualitative Research in General Practice. More Methodological Questions than Answers', *Family Practice*, vol. 14, no. 4, pp. 313–19.

Howe, KK. and Eisenhart, M. (1990) 'Standards for Qualitative (and Quantitative) Research: A Prolegomenon', *Educational Researcher*, vol. 19, no. 4, pp. 2–9.

Inui, T.S. and Frankel, R.M. (1991) 'Evaluating the Quality of Qualitative Research: A Proposal Pro Tem', *Journal of General Internal Medicine*, vol. 6, no. 5, pp. 485–6.

Kuzel, A.J., Engel J.D., Addison, R.B. and Bogdewic, S.P.(1994) 'Desirable Features of Qualitative Research', *Family Practice Research Journal*, vol. 14, no. 4, pp. 369–78.

Leininger, M. (1994) 'Evaluation Criteria and Critique of Qualitative Research Studies' in J.M Morse (ed.), *Critical Issues in Qualitative Research Methods*, Thousand Oaks, Sage.

Long, A.F. and Godfrey, M. (2003) 'An Evaluation Tool To Assess the Quality of Qualitative Research Studies', *International Journal of Social Research Methodology Theory and Practice* (in press).

Mays, N. and Pope, C. (2000) 'Quality in Qualitative Research' in C. Pope and N. Mays (eds), *Qualitative Research in Healthcare*, London: BMJ.

Miller, S. I. and Fredericks, M. (1995) 'Can There be "Rules" for Qualitative Inquiry?', *Journal of Thought*, Spring, pp. 61–71.

Murphy, E., Dingwall, R., Greatbach,D., Parker, S. and Watson, P. (1998) 'Qualitative Research Methods in Health Technology Assessment: A Review of Literature', *Health Technology Assessment*, vol. 2, no. 16.

Muecke, M.A. (1994) On the Evaluation of Ethnographies in J.M. Morse (ed.), *Critical Issues in Qualitative Research Methods* ,Thousand Oaks, Sage.

Popay, J., Rogers, A. and Williams G. (1998) 'Rationale and Standards for the Systematic Review of Qualitative Literature in Health Services Research', *Qualitative Health Research*, vol. 8, no. 3, pp. 341–51.

Sandelowski, M. and Barroso, J. (2002) 'Reading Qualitative Studies', *International Journal of Qualitative Methods*, vol. 1, no. 1, (electronic journal).

Secker. J., Wimbush, E., Watson, J., and Milburn, K. (1995) 'Qualitative Methods in Health Promotion Research: Some Criteria for Quality', *Health Education Journal*, vol. 54, no. 1, pp. 74–87.

Treloar, C., Champness, S., Simpson P.L. and Higginbotham, N. (2000) 'Critical Appraisal Checklist for Qualitative Research Studies', *Indian Journal of Pediatrics*, vol. 67, no. 5, pp. 347–51.

Waterman, H., Tillen, D., Dickson, R and de Koning, K (2001) 'Action Research: A Systematic Review and Guidance for Assessment', *Health Technology Assessment*, vol. 5, no. 23.

Whittemore, R., Chase, S.K. and Mandle, C.L. (2001) 'Validity in Qualitative Research', *Qualitative Health Research*, vol. 11, no. 4, pp. 522–37.

The Government Chief Social Researcher's Office

The Government Chief Social Researcher's Office (GCSRO) is based in the Prime Minister's Strategy Unit and co-ordinates and promotes social research across government. It encourages departments to commission the right research at the right time in order to promote evidence-based policy making and the effective use of social research. It ensures that government research is of the highest quality and uses the most appropriate and up-to-date methods and techniques. GCSRO helps ensure that the government social research service has access to people with the right skills. The office maintains effective links with other professional groups within government as well as with the academic community and those engaging in applied social policy research and evaluation outside government. Sue Duncan, who is the government's Chief Social Researcher, heads GCSRO. Sue can be contacted by email at Sue.Duncan@Cabinet-Office.x.gsi.gov.uk

A web version of the research can be found on Policy Hub (http://www.policyhub.gov.uk). Policy Hub is a web resource launched in March 2002 that aims to improve the way public policy is shaped and delivered. It provides many examples of **initiatives, projects, tools** and **case studies** that support better policy making and delivery and provides extensive guidance on the **role of research and evidence** in the **evaluation of policy**.

Filename: Final draft report - accepted and edited proofs v3 JL.doc
Directory: C:\WINDOWS\Desktop
Template: C:\Program Files\Microsoft
    Office\Templates\Normal.dot
Title: IMPORTANT
Subject:
Author: SCPR
Keywords:
Comments:
Creation Date: 20/06/03 09:54
Change Number: 2
Last Saved On: 20/06/03 09:54
Last Saved By: nigel owens
Total Editing Time: 1 Minute
Last Printed On: 19/06/03 23:05
As of Last Complete Printing
    Number of Pages: 204
    Number of Words: 60,736 (approx.)
    Number of Characters: 346,198 (approx.)