

This article was downloaded by: [Hong Kong Polytechnic University]

On: 22 November 2008

Access details: Access Details: [subscription number 773506841]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



International Journal of Social Research Methodology

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713737293>

An evaluation tool to assess the quality of qualitative research studies

Andrew F. Long; Mary Godfrey

Online Publication Date: 01 April 2004

To cite this Article Long, Andrew F. and Godfrey, Mary(2004)'An evaluation tool to assess the quality of qualitative research studies',International Journal of Social Research Methodology,7:2,181 — 196

To link to this Article: DOI: 10.1080/1364557032000045302

URL: <http://dx.doi.org/10.1080/1364557032000045302>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

An evaluation tool to assess the quality of qualitative research studies

ANDREW F. LONG and MARY GODFREY

(Received 31 October 2001; accepted 10 September 2002)

Critical appraisal of research studies forms a central role within the application and uptake of evidence-based approaches within health and social care. While there are established checklists for evaluating quantitative research, this is not the case for qualitative research. This article outlines the process of developing an evaluative tool for qualitative studies, reflecting the uniqueness of the associated paradigm, and illustrates its use by reference to three qualitative research papers appraised within a systematic review of community-based rehabilitation services. At its centre are concerns with the context of the study and the way the data are collected and analysed. Use of such an evaluative template opens the way for the quality of qualitative research to be judged in a systematic manner and for qualitative research to take its rightful place within debates over what works, where, when and how within health and social care policy and practice.

Introduction

It is now almost taken for granted within the health and social welfare sector that practice should be grounded in best evidence. For any action or intervention designed to bring about change, at an individual or collective level, key questions are posed: does it work, for whom and why, and what are the elements that make it work, in what contexts and from whose perspective(s)? The appropriate research evidence base will in principle include studies employing a range of research design, both qualitative and quantitative, as well as systematic reviews.

While there has been considerable debate about the legitimacy of different types of evidence and the role of systematic reviews—for example, the appropriateness of different methodological approaches (Davies *et al.* 2000) and questions over the way knowledge accumulates, as additive, mosaic or building block (Hammersley 2001)—the ability to critically appraise research studies has become a core skill. Interest lies not just in whether the study is of high quality, but also whether its findings have relevance to the cultural and organizational context within which care is delivered and implemented with a view to enhance the quality and experience of treatment and care for users, patients and caregivers.

Andrew F. Long is Professor and Director of the Health Care Practice R&D Unit, University of Salford, Statham Building, Statham Street, Salford M5 4WT, UK; tel: 0161 295 2824; fax: 0161 295 2825; e-mail: a.f.long@salford.ac.uk

Mary Godfrey is Senior Research Fellow, Nuffield Institute for Health, University of Leeds, 71–75 Clarendon Road, Leeds LS2 9PL, UK.

A characteristic of the development of evidence-based practice within the health field has been the use of formal criteria for assessing the validity and usefulness of research findings. In the last 10–20 years this has extended from explicit questions over statistical aspects of design and analysis (Altman 1996) to a broader focus on the extent that the chosen research design adequately controls potential sources of bias. This has spawned the development of an array of checklists to aid the critical appraisal process (for example, Muir Gray 1996), in particular, for randomized controlled trials (RCTs) and sub-components of this design via the Consort statement and its revision (for example, Moher *et al.* 2001).

The critical appraisal and evaluation of qualitative studies has received far less attention. While this is in part attributable to the dominance of quantitative methods in health services research, it also reflects the state of qualitative enquiry. It is only in recent years that there has been an explicit focus on systematizing the process for managing and analysing qualitative data (see Miles and Huberman 1994). There is also increasing acknowledgement that unless the process of analysis is transparent, it is difficult to convince others that the findings are not simply anecdotal accounts (Wolcott 1994). Even more problematic is the fact that within the qualitative tradition there is no single definition of, or approach to, qualitative research. Indeed, the ‘qualitative’ descriptor is sometimes used to refer to data collection techniques that generate non-numerical data (for example, a semi-structured interview, in-depth interviews), in contrast to a more fundamental interpretation linked to a particular epistemological approach and research style (for example, symbolic interactionism, ethnography).

Qualitative research has a potentially powerful role to play in evidence-based practice and systematic reviews of effectiveness. It may ‘enhance’ understanding in an area or provide ‘different’ sorts of evidence and insight into a problem (Popay and Williams 1998, Davies 2000). For example, in-depth interviews with users can elicit lay experiences of the process of treatment and care, perspectives on what counts as a successful outcome and when, and contextual factors that may promote or dilute the achievement of desired outcomes. Observation and documentary analysis (for example, diaries) can also provide insight into the nature, content and ‘active ingredient’ (Medical Research Council 2000) of complex interventions. For example, what it is about the workings of a stroke unit that result in better survival outcomes? Or, what are the specific components of a rehabilitation model within an intermediate care service system that enhance the quality of life of older people and enables them to retain their independence? Such knowledge is critical to a successful implementation of a process of treatment or care package and understanding of cause and effect.

Moreover, knowing how users express their needs will assist the search and location of evidence for inclusion in systematic reviews. In our own review of outcomes and effectiveness of rehabilitation services for people with severe and enduring mental health problems (Long *et al.* 2002a), we drew on users’ conceptions of life quality from qualitative studies to search for studies that focused on domains and associated interventions designed to enable people to live ordinary lives in ordinary communities (Ritchie *et al.*

1988). Findings from qualitative research can thus both guide the use of quantitative measures of outcome and contextualize the knowledge arising from RCTs and other quantitative designs.

Against this background, this paper presents a tool developed to assist in the constructive evaluation of qualitative research studies in the context of systematic reviews of the effectiveness of health and social care interventions. It is firmly grounded in our belief that the criteria employed to evaluate qualitative studies should accord full significance to the epistemology and practice of qualitative research approaches. Alongside was an understanding that evaluation by definition involves a judgmental process.

Defining qualitative research

A qualitative study is defined here as one that uses qualitative methods in both the gathering and analysis of the data, that is, visual and verbal (conceptual or thematic) rather than numerical data manipulation. Its aim is to draw out the informants' understandings and perceptions as well as to explore the features of social settings and culture. It is informed primarily by an inductive model of reasoning. Qualitative research could thus encompass studies using methods such as interviewing, focus groups, ethnography, participant observation, documentary analysis and life histories. The collected data are intended to illuminate the meanings attached by individuals to events and situations (Bryman 1988) and to understand the dynamics of social life (Whyte 1997) and the linkages between processes and outcomes (Stake 1995). Emphasis is placed on understanding a phenomenon holistically, that is, from different perspectives and bounded by the context or setting within which it is located. The questions of 'how' and 'why' are paramount (Fitzpatrick and Boulton 1994).

At the same time, many papers identified as 'qualitative' in health and social care journals commonly report on small-scale open-ended interview studies, with the resultant data grouped into themes (Boulton *et al.* 1996). Few adopt a holistic approach, explore it in depth or adopt an iterative interplay between data collection and analysis and the use of the constant comparative method. Yet fewer adopt the highly systematic approach outlined by Miles and Huberman (1994) drawing on, for example, time ordered matrices and causal networks. Any evaluative tool has thus to be sufficiently flexible to cover this range of studies while yet be useful and informative.

Developing the content of the evaluative tool

Our starting point was the need to reflect the uniqueness of the qualitative research endeavour as a method for collecting and analysing data. At its most general this centres on a concern with meaning, context and depth (Temple 1998). Questions such as 'do the findings illuminate participants' experiences and/or understandings of the phenomenon under study?' (Popay *et al.* 1998) and 'are the informants' accounts, and the author's

interpretations plausible?’ (Hammersley 1990) formed an important beginning. Underlying these questions is the criticality of a transparency or openness in the author’s account of ‘how the study was done’ (Lofland and Lofland 1971). Thus, ‘to what extent has the author provided sufficient detail of their approach to convince the reader of its rigour and appropriateness?’

These questions led us first of all to a focus on factors related to the *context* of the study. This was broken down into a number of inter-related components.

The first component is the conceptual or theoretical framework in which the study is located and the way that this is reflected in the study. In our own systematic review on interventions to prevent ‘problematic’ grieving among older people following bereavement, the initial scoping of the literature revealed quite disparate and distinct strands of research. These tended to reflect a variety of disciplinary bases and underlying conceptual and theoretical frameworks. Moreover, within each area, a range of potentially competing models underpinned the research base (Long *et al.* 2002b). Central to making sense of the research findings was understanding and delineation of the theoretical framework that underpinned the study. A key evaluative question becomes ‘in what way does this study contribute to knowledge, theory and/or practice?’

A second aspect of context relates to the setting in which the study was undertaken. The significance of this issue is widely recognized in qualitative research. Central questions relate to ‘why this setting?’, ‘is it appropriate in order to examine the research question?’ and ‘is sufficient detail provided on the setting?’

A third area centres on the nature of the sampling strategy adopted. Here, unlike quantitative research studies where this is interpreted solely in terms of the population and an emphasis on representation, in qualitative research the reference is much wider. It embraces events, persons, times and settings purposively selected to illuminate the research questions under study. In some studies, typicality may be important. In others, it may be more illuminating to examine extreme cases as a way of throwing light on causal factors. Interest centres on the appropriateness of the strategy in relation to the study’s aims and objectives. Thus, ‘why were these informants or events chosen?’, ‘are key events and informants captured?’, and ‘were these informants and events appropriate to explore given the study’s aims?’ Attention lies in both depth (for example, the intensity of data collection) and width (for example, to capture key persons and events, and to explore their inter-relationships) and thus in explicating the context of the informants’ experiences and accounts.

A fourth and final component concerns the outcome criteria informing the study. Interest lies in the adequacy of the range and depth of perspectives on desired outcome. These might encompass just the user and carer, or the practitioner and service perspective. At other times, a contrast of two or more perspectives or deeper insight into one perspective is appropriate. This area is of critical importance in reviewing evidence on effectiveness. What counts as ‘success’ or beneficial effect, and over what time period, has major consequences on practice. An inappropriate measure of outcome or too

short a follow-up time may result in 'good' interventions being sidelined (Godfrey *et al.* 2000, Long *et al.* 2000). However, even outside of this domain, the issue of 'whose perspective is adopted in the study and why' is of wider relevance, given the significant opportunity that qualitative research provides in illuminating lay perspectives (cf. Davies 2000).

The way the *data were collected and analysed* formed a second set of issues to consider, again situated firmly within the qualitative paradigm. These were again broken down into a number of components, centring on issues of how, appropriateness and transparency.

In terms of data collection, insight is needed not just into what methods were used and how, but also in the subsequent presentation of the data. Questions centre on what data were elicited, the interview guide/questions and/or observational approach used and the role that the researcher adopted within the study settings and ways that this might affect what was recorded and observed. Interest also lies on the researcher's reflexivity, providing insight into the relationship between the researcher, setting, data production and analysis.

Similarly, in the data analysis, focus lies on the evidence presented. For example, 'was negative and confirmatory evidence sought?', and 'how do the emerging findings cohere with other studies and theory?' Exactly what questions to ask depends on how the analysis was done. In some instances, the use of data from different sources might be of interest; in others, the use of iterative analysis; and in yet others the way the themes or categories emerged and the adequacy of their elaboration through the use of data extracts.

The evaluative tool

Table 1 presents the resultant evaluative tool. It encompasses areas of context and data collection/analysis, both central to the qualitative paradigm and includes the additional areas of research ethics and policy and practice implications, both germane to all approaches to evaluate the quality of research studies. The policy and practice area is particularly relevant in the light of the importance of the uptake of appropriate research findings into practice and evidence-based practice in general.

The tool is inevitably lengthy (it aims to be comprehensive) and written at a high level of generality (it provides a guide on the aspects to explore). Importantly, it incorporates both descriptive ('what was done') and evaluative ('how well it was done') elements. Thus, it includes characteristics of the study (study type, sampling and setting) and how the study was done (rationale for the choice of setting, sample, data collection and analysis). The tool thus provides a framework for describing and thinking about qualitative research in a context where the purpose is evaluation.

To enhance the utility and communicability of the tool, a front end was designed to provide an overview of the study, in the form of an evaluative abstract (table 2). This indicates the purpose of the study, key findings, an evaluative summary of the strengths and weaknesses of the study and possible theory, policy and practice implications. The summary enables a

Table 1. An evaluative tool for qualitative research studies

Review area	Key questions
(1) PHENOMENON STUDIED AND CONTEXT	
Phenomena under study	● What is being studied? ● Is sufficient detail given of the nature of the phenomena under study?
Context I: Theoretical	● What theoretical framework guides or informs the study? ● In what ways is the framework reflected in the way the study was done?
Framework	● How do the authors locate the study within the existing knowledge base?
Context II: Setting	● Within what geographical and care setting is the study carried out? ● What is the rationale for choosing this setting? ● Is the setting appropriate and/or sufficiently specific for examination of the research question? ● Is sufficient detail given about the setting? ● Over what time period is the study conducted?
Context III: Sample (events, persons, times and settings)	● How is the sample (events, persons, times and settings) selected? (For example, theoretically informed, purposive, convenience, chosen to explore contrasts) ● Is the sample (informants, settings and events) appropriate to the aims of the study? ● Is the sample appropriate in terms of depth (intensity of data collection, individuals, settings and events) and width across time, settings and events? (For example, to capture key persons and events, and to explore the detail of inter-relationships) ● What are the key characteristics of the sample (events, persons, times and settings)?
Context IV: Outcomes	● What outcome criteria are used in the study? ● Whose perspectives are addressed (professional, service, user, carer)? ● Is there sufficient breadth (e.g. contrast of two or more perspective) and depth (e.g. insight into a single perspective)?
(2) ETHICS	
Ethics	● Was Ethical Committee approval obtained? ● Was informed consent obtained from participants of the study? ● Have ethical issues been adequately addressed?
(3) DATA COLLECTION, ANALYSIS AND POTENTIAL RESEARCHER BIAS	
Data collection	● What data collection methods are used to obtain and record the data? (For example, provide insight into: data collected, appropriateness and availability for independent analysis) ● Is the information collected with sufficient detail and depth to provide insight into the meaning and perceptions of informants? ● Is the process of fieldwork adequately described? (For example, account of how the data were elicited; type and range of questions; interview guide; length and timing of observation work; note taking) ● What role does the researcher adopt within the setting? ● Is there evidence of reflexivity, that is, providing insight into the relationship between the researcher, setting, data production and analysis?

Table 1. *(Continued)*

Review area	Key questions
Data analysis	● How are the data analysed? ● How adequate is the description of the data analysis? (For example, to allow reproduction; steps taken to guard against selectivity) ● Is adequate evidence provided to support the analysis? (For example, includes original/raw data extracts; evidence of iterative analysis; representative evidence presented; efforts to establish validity—searching for negative evidence, use of multiple sources, data triangulation); reliability/consistency (over researchers, time and settings; checking back with informants over interpretation) ● Are the findings interpreted within the context of other studies and theory?
Researcher's potential bias	● Are the researcher's own position, assumptions and possible biases outlined? (Indicate how these could affect the study, in particular, the analysis and interpretation of the data)

(4) POLICY AND PRACTICE IMPLICATIONS

Implications	● To what setting are the study findings generalizable? (For example, is the setting typical or representative of care settings and in what respects? If the setting is atypical, will this present a stronger or weaker test of the hypothesis?) ● To what population are the study's findings generalizable? ● Is the conclusion justified given the conduct of the study? (For example, sampling procedure; measures of outcome used and results achieved) ● What are the implications for policy? And for service practice?
--------------	--

reader to quickly grasp the essential details of a study and its potential value. This has considerable potential as a means of communicating key messages about the study in a reasonably succinct manner. It can provide one part of the information set, and knowledge base, for a policy maker. The reader retains the option of going inside the tool for a more detailed insight into a particular aspect of the study.

Table 2. Evaluative study overview

Review area	Key questions
STUDY OVERVIEW	
Bibliographic details	Author, title, source (publisher and place of publication), year
Purpose	● What are the aims of this paper? ● If the paper is part of a wider study, what are its aims?
Key findings	● What are the key findings?
Evaluative summary	● What are the strengths and weaknesses of the study and theory, policy and practice implications?

Each question is easily translated into a set of database fields. In our own work (Long *et al.* 2002a, b), the database record became a core source document, with a view to avoid as far as possibility the need to return to the original article in the evidence synthesis. Accordingly, comments for each question or database field needed to be self-standing. The areas and questions are thus purposely designed to avoid a dichotomous (yes/no) response, aiming to facilitate explanation of the judgements made. The evaluative comments are, by implication, informed judgements, based on the detail provided in the study report.

It is important that the tool is not seen as a list of methodological questions to ask of a qualitative study with a view to a comparison with some 'gold standard', whatever that might mean in the qualitative field. Instead, it provides a template to aid the formation of a judgement of the potential quality and contribution of a piece of qualitative research in illuminating a phenomenon. An understanding of the qualitative paradigm is essential in using the tool. Indeed, undertaking critical appraisal of any study requires understanding of the strengths and weaknesses of *at least* that particular research design *and* of the principles of rigorous research, along preferably with insight into the realities of doing research in practice.

Assessing adequacy and appropriateness

Most of the evaluative elements in the tool involve making a judgement about the 'adequacy' or 'appropriateness' of the approach undertaken and reported on in the research paper. It is arguable whether there are agreed standards amongst practitioners of the qualitative paradigm over what is adequate (Murphy *et al.* 1998). Indeed, there is considerable debate over what criteria should be employed in evaluating what is good qualitative research. Three distinct approaches are evident, each with considerable variation and nuances (Denzin 1994). These range from those who would wish to utilize the same criteria as used for quantitative studies to those, like ourselves, who propose that such criteria should reflect the uniqueness of the qualitative endeavour.

For example, an idealist or relativist school of thought (see Smith 1984) asserts that the very nature of qualitative research, specifically its reflexivity and relativity, means that it is irrelevant and inappropriate to develop criteria to judge its products. Here, all accounts are accorded equal weight. There is no measure against which one account may be regarded as a closer approximation of reality. An alternative stance, and one to which we would ascribe, has been variously described as 'subtle' or 'analytic' realism (for example, Hammersley 1992, Altheide and Johnson 1994). This acknowledges that we perceive the world from a particular perspective, but at the same time understands that the world also acts back upon us to constrain the points of view that are possible. The researcher must grasp at a world that is constantly changing. The purpose is to identify the processes and relations, which engender both continuity and change. Thus, appropriateness and adequacy reflect the extent to which we conduct research in a systematic and rigorous way to produce

plausible accounts that reflect what is being examined in as accurate a way as possible.

None of this negates the value of an evaluative framework. It may indeed be the case that qualitative researchers coming from different epistemological or theoretical positions (for example, phenomenology, discourse analysis or grounded theory) might evaluate the same study in different ways using the tool. However, proper use of the framework should make it plain why this was the case through the reviewer making explicit what was missing or present and its significance. This issue lies at the heart of constructive evaluation, in contrast to what is appropriately pejoratively described as 'hit-and-run' criticism (Campbell 1969, Bross 1970). The tool's main strength lies in the fact that it makes explicit the base on which judgements about individual studies are made and the theoretical framework in which the study is located.

Using the tool

To illustrate the use of the evaluative tool, three research reports, reviewed as part of a systematic review of what rehabilitative service interventions and associated mix of service and professional practices produce best outcomes (Long *et al.* 2002a), have been chosen. Each study illustrates a different type of qualitative study.

The first (Malone 1989) is an ethnographic study. This described the inner workings of 'Community Bound', an innovative non-professional community programme for the long-term mentally ill, being run in Austin, Texas. Its underlying philosophy was that persons with psychiatric impairment could be taught to live independently in the community if they had access to daily problem solving assistance from people they knew and trusted. The researcher adopted the role of observer-as-participant, generally sitting in the background and accompanying staff on errands, and appointments. The semi-structured interviews were taped for full transcription and notes made following each 30-minute observation period.

The second study (Torrey *et al.* 1995) provides an example of the use of qualitative data collection methods within a survey, itself nested within a natural experiment. The study explored the impact of an intensive supported employment approach on users, their families and clinical providers. It was undertaken in two community mental health centres in two small New England towns. The vocational rehabilitation programme aimed to achieve positive attitudes to work and enhance work adjustment within a wider context of increasing independent living and thus improve the individual's quality of life and health (Bond *et al.* 1997). Open-ended interviews were chosen as the data collection method.

The third (Perkins *et al.* 1997) illustrates the applicability of the tool for a descriptive case study of a service development. The study reported on a supported employment project by the Pathfinder Mental Health NHS Trust in London, UK. The approach was based on a Colorado programme (Shepherd and Porter 1991) and involved the recruitment of individuals who had experienced serious mental health problems to work within a

Table 3. Evaluative study overview for the selected papers

Study	Purpose	Key findings	Evaluative summary
Malone 1989	To describe a supported employment project designed to help people who have had serious mental health problems	The study identifies and clarifies six empirical categories which underlying the interactions between staff–client and client–clients relationships enabling people with long-term mental health problems to live in the community: survival skills; cooperation; hanging out; checking up; backing and supplementing. Differences in approach are also drawn out between lay staff and professionals operating within more structured programmes.	The study adopts a grounded theory approach. Clear links are made between the categories drawn out and the data illustrating those categories. These insights are located within the wider rehabilitation literature in a way which serves to draw out both the similarities and differences between the philosophy and working practices within Community Bound and professional, structured rehabilitation programmes. The data collection process is described in sufficient detail. However, there is no information offered about the process of analysis or how the author developed the categories from the large quantity of data ensuing from over 800 hours of observation and interviews. The description offered is rich in contextual detail and illuminating of the processes involved in delivery and receiving support and care. The findings were also tested out through consultation with experts in community mental health, peers and study participants.
Torrey <i>et al.</i> 1995	To explore the impact on users and their families of conversion from a day treatment to an intensive supported employment approach via the Individual	Positive comments were provided by all three stakeholder groups (users, their families and clinicians). Benefits included increased vocational support and staff availability, integration into community activities, independence and improvements in vocational skills. Losses were also identified,	This small-scale qualitatively oriented, interview-based, essentially retrospective study provides useful insight into the impact (positive and negative) of the conversion on the users and their families. Its focus on the perspective of three stakeholder groups (user, family and clinician) is particularly valuable. The sample comprises all eligible persons who were regular users of the day

Placement and Support (IPS)

in particular, of social opportunities. There is a suggestion that a small minority of users may have been worse off because of the loss of a structured daily activity.

treatment programme prior to conversion. Limited information is provided on the setting (but adequate for this type of data) and process of analysis of the open-ended data. While the discussion section makes reference to other data providing further insight into the process and impact of the conversion over the 2-year period, it is not clear how this was collected. The authors draw attention to limitations of the study (retrospective, no control group) and the possibility of an alternative explanation of the findings (users are fearful of any change and satisfied with whatever programme is available, but this does not cohere with other evidence arising from the evaluation of the IPS approach). This study suggests the value and need for an ethnographic study of the process of conversion.

Perkins *et al.*
1997

To describe the process of working within Community Bound, an innovative and non-professional community programme for the long-term mentally ill.

18 months after the first six user employees took up their half time posts, two have successfully applied for full time vacancies. The employees are very positive about the project. Benefits include the provision of work opportunities, feelings of self-worth and enhancement of quality of life. Problems faced include: personnel concerns about the recruitment of persons who had experienced mental health problems; and potential role conflict between the service user and service provider, with the latter being unsure how a user could be a service provider as well. The project is not expensive.

The paper provides a descriptive account of a service development rather than of a research study *per se*. This, however, makes it difficult to judge whether the presented findings are representative extracts from the data, and of what the data as such comprise. No account of a fieldwork or evaluation approach is provided. This makes the study findings and any assessment of the value of service development problematic. The reader would have increased confidence in the validity and credibility of the benefits identified if the study had been reported on in a way which provided insight into the way the data were generated and why. A fuller and more explicit account and evaluation is needed.

rehabilitation clinical team. Training and support were provided. The posts were ordinary positions within the team, not 'special posts' created for service users. Observational methods and written comments by service users were used.

Table 3 contains the summary overview of the three studies. The way these are written illustrates the potential power of the use of the tool where concern lies with the study's strengths and limitations. The focus on constructive evaluation is evident.

To illustrate the descriptive and analytical content of the tool, table 4 presents selected extracts from the wider database record for each of the three studies for two areas of 'context' and two areas of 'data collection and analysis'.

One of these extracts, that of the Perkins *et al.* (1997), serves to raise a common problem faced by those who try to pull together what is known in an area, in the form of a meta analysis of either method, data or theory (Paterson *et al.* 2001). This is the issue of the incompleteness of reporting on the way the study was done, thus limiting the ability to form a rounded judgement on the study's validity, relevance and usefulness. This problem is not unique to the reporting of qualitative research. It may arise from an oversight of the reporting author(s) and/or inadequacies of the refereeing process or because of a perceived or actual difficulty in making transparent the way that the study was done. In addition, especially the case for reports on qualitative research, there may be word-count limitations or the journal itself may emphasize the reporting of practice dimensions. If the study is of particular potential significance, for example, central to an area of study or suggesting new findings or directions for theory and research, the reviewer will need to contact the author to seek further detail on their methods. Alternatively, as in the Perkins *et al.* extract, the reviewer will record that there was insufficient detail provided to form a judgement in the particular field.

Concluding remarks

This article has outlined the process of the development of an evaluative tool for the appraisal of qualitative research studies and illustrated its use. Reflecting the uniqueness of the qualitative paradigm, the tool aims to provide a set of informative areas and questions both to summarize the core content of the study and to enable the assessment of the quality of the design in relation to its aims and outcomes. It is purposely elaborative and discouraging of yes/no responses, with a view to explain the rationale behind the judgements made. Since its development, the approach has been extended to develop an evaluative tool for use in studies using qualitative data collection as part of a quantitative research study, so called 'mixed method' studies (HCPRDU 2001).

Evaluating studies, be they qualitative or quantitative, involves a process of judgement. An intended consequence of the use of this tool is to reduce the variation in judgements and opinions about the quality and relevance of a study. The tool directs attention to a set of questions that have been

Table 4. Selected extracts from the database records

Study	Setting rationale	Appropriateness of sample	Adequacy of description of fieldwork	Adequate evidence to support analysis
Malone 1989	The very low hospital re-admission rate experienced by clients on the programme and the use of non-professional staff presented an opportunity to examine the process of care within an apparently successful scheme.	The study included all members of the programme. Fieldwork was both intense (4 hours per day) and deep (over 5 months). It would have been useful to know whether evening and weekend time periods had been included in time sampling.	The process of collecting the data was clear. The time spent (4 hours per day varying the time of day) doing fieldwork (direct observation by sitting in the office, tagging along on errands and appointments), observation and note taking (30 minutes of observation followed by note taking, fleshed out after each day's observations) is well described.	Sub-categories are identified and occasional illustrative quotes or fieldwork included. These could valuably have been more frequent. It is unclear how the data from the different sources are pulled together, whether they confirm one another or not, or whether the presented evidence is representative. At the same time, the data analysis is informative and convincing.
Torrey <i>et al.</i> 1995	The study took place in the context of conversion of day treatment programme to a supported employment approach, taking advantage of a 'naturally' occurring change.	The study was undertaken with a group seen as most vulnerable due to the loss of a structured day programme. It involved a one point in time open-ended interview with an appropriate group of informants. It would have been valuable to present observational data on the conversion process, its ongoing impact and user perceptions.	Only brief details are provided of the nature of the interview (15–30 minutes), its nature (semi-structured) and content (open-ended questions on the positive and negative aspects of the programme change; satisfaction questions taken from the Lehmann's Quality of Life Interview).	Quotations are presented and counts of reactions to change (measured on a 7-point scale). It is difficult to judge whether the evidence presented is adequate or not, given the lack of detail over how the data analysis was done.
Perkins <i>et al.</i> 1997	A tradition and managerial and service commitment exists in the Trust to enable service users to gain work experience in non-clinical settings. There is also a well established, funded, independent Patients Council and Advocacy Project.	The included persons were those who applied and were successful for the available posts.	The report is not written in standard research report style. It is unclear what the 'fieldwork' as such comprised, as opposed to a descriptive account of a programme in operation.	It is unclear whether the extracts provided are representative of the data collected. At the same time there is no reason to doubt their descriptive account of the programme.

systematically derived from the qualitative paradigm. Its format encourages the appraiser to clarify how and why they make their judgements.

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. It also requires knowledge of the qualitative paradigm and a constructive approach to critical appraisal, that is, not only pointing to weaknesses, but also strengths and ways that the quality of the study could be enhanced. We would recommend use of the tool in any context where the aim is either to develop qualitative (or other) research in a rigorous way and/or to promote and sustain reflexive practice in qualitative research.

Use of such an evaluative tool should serve to enhance the contribution of qualitative research within evaluations of the effects of interventions (see, for example, the work of the Campbell Collaboration Methods Group 2001). More broadly, the availability and use of this tool should have an additional positive consequence of encouraging high standards for qualitative research. In particular, qualitative researchers must give greater insight into their methods, indicating the rigour and systematic nature of the work. The pejorative description of qualitative research as 'anecdote' needs to be replaced. Properly conducted qualitative studies provide valuable information in the search for effective practices. Sorting the 'good' from the 'poor' through use of a set of systematic and comprehensive methodological questions, grounded in the qualitative paradigm, becomes even more essential.

Acknowledgements

This work originally arose within a study funded by the Department of Health, under the Outcomes for Social Care for Adults (OSCA) Initiative (1997–99), to whom grateful acknowledgement is made.

References

- Altheide, D.L. and Johnson, J. (1994) Criteria for assessing interpretive validity in qualitative research. In N.K. Denzin and Y.S. Lincoln (eds) *Handbook of Qualitative Research* (London: Sage).
- Altman, D.G. (1996) Better reporting of randomised controlled trials: the CONSORT statement. *British Medical Journal*, **313**, 570–571.
- Bond, G.R., Drake, R.E., Mueser, K.T. and Becker, D.R. (1997) An update on supported employment for people with severe mental illness. *Psychiatric Services*, **48**, 335–346.
- 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*, **2**, 171–179.
- Bross, I.D.J. (1970) Statistical criticism. In E.F. Tufte (ed.) *The Quantitative Analysis of Social Problems* (London: Addison–Wesley).
- Bryman, A. (1988) *Quantity and Quality of Social Research* (London: Unwin Hyman).
- Campbell, D.T. (1969) Reforms as experiments. *American Psychologist*, **24**, 409–429 [reprinted in Bynner, J. and Stribley, K.M. (eds) (1979) *Social Research: Principles and Procedures* (London: Longman)].
- Campbell Collaboration Methods Group (2001) <http://www.missouri.edu/~c2method/>
- Davies, H.T.O., Nutley, S. and Smith, P.C. (eds) (2000) *What Works? Evidence-based Policy and Practice in Public Services* (Bristol: Policy Press).

- Davies, P. (2000) Contributions from qualitative research. In H.T.O. Davies, S. Nutley and P.C. Smith (eds) *What Works? Evidence-based Policy and Practice in Public Services* (Bristol: Policy Press), pp. 291–316.
- Denzin, N.K. (1994) The art and politics of interpretation. In N.K. Denzin and Y.S. Lincoln (eds) *Handbook of Qualitative Research* (London: Sage), pp. 500–515.
- Fitzpatrick, R. and Boulton, M. (1994) Qualitative methods for assessing health care. *Quality in Health Care*, 3, 107–113.
- Godfrey, M., Randall, T., Long, A.F. and Grant, M.J. (2000) *Home Care: Review of Effectiveness and Outcomes* (Exeter: Centre for Evidence Based Social Services).
- Hammersley, M. (1990) *Reading Ethnographic Research: A Critical Guide* (London: Longman).
- Hammersley, M. (1992) *What's Wrong with Ethnography?* (London: Routledge).
- Hammersley, M. (2001) On 'systematic' reviews of research literatures: a 'narrative' response to Evans & Benefield. *British Educational Research Journal*, 27, 543–554.
- HCPRDU (2001) www.fhsc.salford.ac.uk/hcprdu/assessment.htm
- Lofland, J. and Lofland, L.H. (1971) *Analysing Social Settings* (Belmont, CA: Wadsworth).
- Long, A.F., Mercer, G. and Hughes, K. (2000) Developing a tool to measure holistic practice: a missing dimension in outcomes measurement within complementary therapy. *Complementary Therapies in Medicine*, 8, 26–31.
- Long, A.F., Godfrey, M., Randall, T., Brettell, A.J. and Wistow, G. (2002a) *Developing Evidence Based Social Care Policy and Practice. Final Report: Part 1: Effectiveness and Outcomes Of Rehabilitation for People with Severe and Enduring Mental Illness* (Salford and Leeds: Health Care Practice R&D Unit, University of Salford, and Nuffield Institute for Health, University of Leeds).
- Long, A.F., Godfrey, M., Randall, T., Grant, M.J. and Chapman, M. (2002b) *Developing Evidence Based Social Care Policy and Practice. Final Report: Part 2: Effectiveness and Outcomes of Preventive Services for older People: Risk Factors, Coping Strategies and Outcomes of Interventions in Bereavement* (Salford and Leeds: Health Care Practice R&D Unit, University of Salford, and Nuffield Institute for Health, University of Leeds).
- Malone, J. (1989) Concepts for the rehabilitation of the long term mentally ill in the community. *Issues in Mental Health Nursing*, 10, 121–135.
- Medical Research Council (2000) *A Framework for Development and Evaluation of RCTs for Complex Interventions to Improve Health* (London: Medical Research Council).
- Miles, M.B. and Huberman, A.M. (1994) *Qualitative Data Analysis* (London: Sage).
- Moher, M., Schulz, K.F. and Altman, D.G., for the CONSORT Group (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *Annals of Internal Medicine*, 134, 657–662.
- Muir Gray, J.A. (1996) *Evidence-Based Health Care* (Edinburgh: Churchill Livingstone).
- 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*, 2, 1–276.
- Paterson, B.L., Thorne, S.E., Canam, C. and Jillings, C. (2001) *Meta-Study of Qualitative Health Research* (London: Sage).
- Perkins, R., Buckfield, R. and Choy, D. (1997) Access to employment: a supported employment project to enable mental health service users to obtain jobs within mental health teams. *Journal of Mental Health*, 6, 307–318.
- Popay, J. and Williams, G. (1998) Qualitative research and evidence-based healthcare. *Journal of the Royal Society of Medicine*, 901, Supplement 35, 32–37.
- 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*, 8, 341–351.
- Ritchie, J., Morrissey, C. and Ward, K. (1988) *Keeping in Touch with the Talking: The Community Care Needs of People with a Mental Illness* (Birmingham: Birmingham City Council and Social and Community Planning Research).
- Shepherd, P.S. and Porter, R. (1991) Mental health consumers as case management aides. *Hospital and Community Psychiatry*, 42, 494–498.
- Smith, J. (1984) The problem of criteria for judging interpretive inquiry. *Educational Evaluation and Policy Analysis*, 6, 379–391.
- Stake, R. (1995) *The Art of Case Study Research* (Thousand Oaks, CA: Sage).
- Temple, B. (1998) A fair trial? Judging quality in qualitative research. *International Journal of Social Research Methodology*, 1, 205–215.

- Torrey, W.C., Becker, D.R. and Drake, R.E. (1995) Rehabilitative day treatment vs. supported employment: II. Consumer, family and staff reactions to a program change. *Psychosocial Rehabilitation Journal*, **18**, 67–75.
- Whyte, W.F. (1997) *Creative Problem Solving in the Field: Reflection as a Career* (Walnut Creek: Altamira Press).
- Wolcott, H.F. (1994) *Transforming Qualitative Data: Description, Analysis and Interpretation* (Thousand Oaks, CA: Sage).