

Special issue

Making qualitative judgements of quality

This issue of *Building Research Capacity* features responses to the Cabinet Office document *Quality in Qualitative Evaluation: A framework for assessing research evidence* by Liz Spencer, Jane Ritchie, Jane Lewis and Lucy Dillon of the National Centre for Social Research (2003). To bring *BRC* readers up to speed with issues raised by the three responses there is a brief summary of the framework document (p.1). Murphy and Dingwall (p.3) are mostly supportive of the document, praising it for its clarity in a field fraught with complex methodological and philosophical debate. They do, however, suggest more work can be done to make this framework a more effective tool for the non-expert, providing more detailed discussion of qualitative methods to assist evaluative judgement making. Kushner (p.5) claims that in the absence of a strong tradition of educational evaluation in the UK, this framework may place more significant limitations on the possibilities of evaluation than the promise offered in the document. While Kushner identifies uses and benefits of the document, he is concerned at the unreasonable demand on government "...to manage and guarantee the independence of an evaluation". Torrance (p.8) offers further criticism of the framework, and suggests that this framework is about evaluating the quality of qualitative evaluation, rather than a serious attempt to evaluate qualitative research on its own terms, i.e. qualitatively.

A Summary of 'Quality in Qualitative Evaluation: A framework for assessing research evidence'

Ruth Boyask, RCBN, Cardiff School of Social Sciences

An outcome of the Cabinet Office's responsibility for excellence and accountability in Government research and evaluation, this report was commissioned to provide a framework for assessing qualitative policy evaluation (Spencer, Ritchie, Lewis and Dillon, 2003). In reviewing literature surrounding the debates on the role of criteria in qualitative research, it examines the nature of qualitative evaluation, following writers like Shaw (1999) and Patton (2002) who take a broad view of evaluation, including policy research and more specific evaluations of practice and interventions as well as more traditional associations, like programme evaluation.

some ontological and epistemological understandings that suggest that, due to its subjective nature, there is an intractable conflict in assessing qualitative research against criteria, as well as some that would have qualitative research assessed on the same or similarly objective basis as quantitative research. The authors take a middle road where they identify the assumptions that are consistent with the scope of their framework. "This means that someone thinking about using our framework will first need to consider whether they subscribe to the assumptions on which

(Continued on page 2)

The application of criteria to qualitative evaluations

While the report makes references to the tensions implicit in applying a framework of criteria to qualitative methods, it comes down on the side of developing a set of guiding principles that retain "...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" (Spencer *et al*, 2003, p.43). The report identifies

Contents

Page 1	A Summary of Quality in Qualitative Evaluation: A framework for assessing research evidence (2003)
Page 3	A Response to Quality in Qualitative Evaluation: A framework for assessing research evidence
Page 5	Government Regulation of Qualitative Evaluation
Page 8	'Quality in Qualitative Evaluation' - a (very) critical response

(Continued from page 1)

it is based" (p.49). On the understandings of knowledge, facts and values that the framework is based upon, the report suggests that neutrality is an unattainable but ideal and guiding principle and that knowledge claims can be more or less accurate, albeit provisional. This would suggest that philosophical assumptions about knowledge, its value and nature that fall outside of the scope of the framework, cannot be assessed by the framework. The authors, however, suggest that "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" (p.52). If evaluations that are committed to participatory or emancipatory aims are considered as an example, it appears that within the framework, research underpinned by paradigms that fall outside of its scope will be seen as lacking. Another tension within debates surrounding qualitative research identified by the report relates to the applicability of a framework to qualitative methods in a generic sense, versus applicability to specific methods. After consulting with policy-makers, commissioners, research managers and practitioners and then reviewing 298 existing evaluative studies, the report chose to adopt a two-tiered approach to this issue. "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" (p.55).

Validity, reliability and objectivity

The report discusses the varied ways notions of quality are constructed in discourses surrounding qualitative research. The authors claim that in order to avoid becoming overwhelmed by the diversity

and discordance in qualitative methodologies, they have taken to addressing the "...holy trinity' of validity, reliability, and objectivity..." (p.59) and their role in determining quality in qualitative evaluation. The authors claim, that while again there is wide variance in how validity is constructed, there are some loose categories that can be applied. Through an investigation of literature they arrive at three concerns that underpin constructions of validity in qualitative research; where methodological or interpretive issues are at stake (e.g. the validity of methods to elicit a true interpretation), truth and credibility claims (e.g. the validity of knowledge claims and the problematic nature of truth) and validity in relation to participants (e.g. participatory and ethical concerns). Reliability is also examined in the light of literature, starting with the idea that reliability is problematic in qualitative research, particularly when it is seen in the light of the extent to which research studies can be replicated (Le Compte and Goetz, 1982, cited in: Spencer *et al*, 2003). With qualitative research design tending to be "...responsive to the setting and therefore emergent and iterative" (Spencer *et al*, 2003, p.64), it is difficult to conceive to what extent it could be replicated. Therefore the authors suggest that, while similar rather than identical findings may be reproducible, overall there is a preference in the literature for a reflexive and reflective approach where strict notions of reliability are not of central concern. Objectivity in qualitative research is also problematised, with naïve neutrality discounted, but the authors suggest that the general view is that while complete objectivity is not a realistic goal, "...steps can be taken to ensure a 'reasonable' level is attained" (p.67). Other issues that arise in the consideration of achieving valid, reliable and objective evaluations are for defensibility of approach and claim, rigour through rich and detailed data collection and sensitivity to building ethical and reflexive relationships to the researched.

Through their review of literature, and support from interviews, they claim that they came to shape the "...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" (p.82).

Other frameworks

In developing their framework, the authors reviewed twenty-nine other "frameworks" designed to assess the quality of qualitative research. They note that twenty-five of the existing frameworks were developed in the fields of medical or health related research, suggesting that this reflects the need for qualitative research to stand up to scrutiny within a predominantly quantitative field. They found that the frameworks were frequently explicit about their positionality, making claim to a particular method or tradition within qualitative research, as well as declaring the evolving or developmental nature of the framework and some referring more particularly to the philosophies that underpin it.

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 (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 strait-jacket 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 (p.90).

One of the debates that seemed to arise in the frameworks was whether criteria can be developed that is generalisable across qualitative approaches or whether the criteria needs to be specific to the operational research paradigm. Another of the issues that seems to have some currency amongst framework developers and users of the frameworks is that evaluation does not just require an adequate tool, it also requires professional judgement. Thus evaluators of qualitative research need some understanding of qualitative methods and methodologies. This concern was reiterated in interviews with

academics and researchers, whereas research managers and policy-makers appeared to prefer concise and standardised forms.

Using the framework

The framework is devised around 16 appraisal questions that are categorised by the following 7 identifiers; findings, design, sample, data collection, analysis, reporting, reflexivity and neutrality. For each appraisal question there is a list of quality indicators, sub-headed as "possible features for consideration". The authors claim that this framework should be applied flexibly rather than prescriptively and is open to further development and critique. They emphasise the role of professional judgement in making evaluations, so that the framework

should be used in conjunction with professional expertise and theoretical understanding of qualitative research. The policy-makers, government managers and commissioners of research interviewed in the study, however, were concerned that the framework could also play an educative role in providing guidance for the appropriate use of qualitative research.

Copies of the report can be obtained from the cabinet office. (www.policyhub.gov.uk).

A Response to 'Quality in Qualitative Evaluation: a framework for assessing research evidence'

Elizabeth Murphy, School of Sociology and Social Policy, University of Nottingham

Robert Dingwall, Institute for the Study of Genetics, Bio-risks and Society, University of Nottingham

Spencer and her colleagues have produced an admirably clear and concise discussion of quality assessment for qualitative evaluation and research. They have cleared a path through a dense thicket of philosophical and methodological debates. They offer a framework for evaluating qualitative research which both avoids the pitfalls of prescriptive check-lists and recognises the important place of informed judgement in decisions about the quality of such work. At times, however, clarity has been achieved at the expense of usability. Some elements of the framework are presented at such a high level of abstraction that they offer little help to the research commissioner or policy actor who needs to judge the quality of a research proposal or report.

This point can be illustrated by examining just one of the eighteen

key questions that the authors recommend for the appraisal of qualitative research. They argue, uncontroversially, that appraisers should consider the defensibility of a study's research design. They offer a number of 'quality indicators' that can be used in this process, including: 'Discussion of how overall research strategy was designed to meet the aims of the study'; 'Discussion of the rationale for study design'; and 'Convincing argument for different features of research design'. Again, these quality indicators are relatively uncontroversial. However, for the appraiser they raise as many questions as they answer. How is the appraiser to judge the rationale presented by the researcher? Are there any principles s/he might apply in evaluating the arguments used to defend features of the design? Further specification and operationalisation of the quality indica-

tors are necessary before they become usable by those charged with appraising research proposals and reports.

To assist the appraiser in forming a judgement on this matter, it would have been useful to have a more detailed discussion of the relative merits of the four research methods upon which the report focuses: in-depth interviews, focus groups, observation and documentary analysis. The report notes in passing (p.73) that some authors have criticised qualitative researchers' over-dependence on interview methods. The relevance of this issue is under-scored by the data presented on page 56 of the report which show that, while 95% of UK Government evaluative research using qualitative methods involve individual or group interviews, only 18% make use of observation.

(Continued on page 4)

(Continued from page 3)

There are very strong arguments in favour of treating observation as a 'gold standard' or 'method of choice' in qualitative research (Murphy and Dingwall, 2003). 'Observation' here should be taken to include both traditional participant observation and the direct audio or video recording of naturally occurring events. While neither observational nor interview data can be treated as straightforward reproductions of reality, the former minimise the 'chain of transformation'. Observational data are subject to just one transformation as researchers render the reality they observe into a form suitable for analysis. Interview data, by contrast, are transformed at least twice. The first transformation occurs through the interviewers' choice of questions and the second follows as informants select from the totality of their original experience and reformulate that experience into their responses. This second transformation is likely to be affected by a range of factors arising from the interactional context in which it is conducted. Interviews (and, to an even greater degree, focus groups) are, inevitably, occasions for impression management (Goffman, 1959). They are encounters in which individuals are called upon to present their experiences, feelings, practices and opinions in such a way as to preserve their standing as competent, sane and legitimate members of particular communities (Murphy et al., 1998).

None of this suggests that the use of interviews is never justified. In some circumstances they have significant advantages (Kleinman, Stenross et al., 1994). There are many important questions that simply do not lend themselves to observational research. This is particularly the case when the phenomenon of interest is not confined to a particular locale or occurs unpredictably in space or time. Likewise some activities take place in

private and semi-private settings, which may be inaccessible to observational researchers. In both these cases, the only realistic option may be to treat interviewees as proxy observers. Even where observational research would, in principle, be possible, it is often likely to be more expensive and more time consuming than an interview study. The practical imperative to deliver research findings in a timely fashion, and within a limited budget, may offset the relative benefits of observation over interviews.

Such pragmatic justifications for using interviews rather than observation may be entirely justifiable but we do need to recognise that their advantages are bought at a price. The resulting descriptions of events and interactions are second-hand reports, produced by untrained observers whose selection and organisation of such descriptions introduce uncontrolled and unspecifiable influences before the researcher gets to work. This does not necessarily mean that the data are irredeemably flawed but it does call for caution in the way in which they are analysed and the uses to which they are put.

In arriving at a judgement about the defensibility of a particular research design, research commissioners and policy actors may also need to consider the relative merits of documents as sources of data. The latter are, like observations, used relatively rarely in Government evaluative research. Spencer et al. (2003) report that only one-fifth of all such research includes some documentary analysis. Where documents are elicited specifically for the purposes of research (as when participants are invited to keep diaries), they raise many of the issues associated with interviews already discussed. However, most of the organisations naturally generate large numbers of documents, which can be a valuable source of research data. These documents - administrative records, publicity materials,

codes of practice and so on - are more akin to the 'given data' characteristic of observational research (Murphy and Dingwall, 2003). Nevertheless, like interviews, they must be seen as artfully constructed insofar as they incorporate an organisation's attempt to present itself to its own members and others in its environment.

There is, then, scope for the further development of the framework for assessing qualitative research evidence offered by Spencer and her colleagues. This should, however, not be allowed to detract from the authors' very significant achievement in developing a relevant and inclusive list of questions that need to be addressed in judging qualitative research. What is now required is further work to specify more precisely how such questions could, in practice, be answered.

References

- Goffman, E. (1959) *The Presentation of Self in Everyday Life*, Harmondsworth: Penguin
- Kleinman, S., Stenross, B., & McMahon, M. (1994) 'Privileging fieldwork over interviews: Consequences for identity and practice', *Symbolic Interaction* 17(1):37-50
- Murphy E., Dingwall R., Greatbatch E., Parker S. and Watson P. (1998) 'Qualitative research methods in health technology assessment: a review of the literature', *Health Technol Assessment* 2 (16).
- Murphy E. and Dingwall R. (2003) *Qualitative Methods in Health Policy Research*, New York: Aldine de Gruyter
- Spencer L., Ritchie, J., Lewis, J. and Dillon, L. (2003) *Quality in Qualitative Evaluation: a framework for assessing research evidence*, London: Government Chief Social Researcher's Office.

Government regulation of qualitative evaluation

Prof. Saville Kushner

Centre for Research in Education and Democracy

University of the West of England

In the USA there are university and independent centres dedicated to the practice of educational programme evaluation. These centres are active in mounting training programmes at Masters and Doctoral levels, publishing training materials and designing training web-sites. The methodological training of programme evaluators for the education sector is a vigorous industry, but beyond this, programme evaluation is an acknowledged independent discipline and professional practice, and as a community of practitioners it manages its own capacity-building. Many professionals in the USA have no difficulty in calling themselves 'programme evaluators'. Not so in Britain, however, which has no such centres and no postgraduate programme dedicated to programme and policy evaluation. There is, of course, the UK Evaluation Society which promotes learning and methodological deliberation specific to evaluation, but in this country the professional identity of programme evaluator has little currency – notwithstanding that many people from other disciplinary and methodological backgrounds are engaged in programme evaluation.

A recent publication by the Cabinet Office seeking to guide the government sponsorship of 'qualitative evaluation' is a significant event, therefore¹. Such a publication by a powerful government body would always have both promises and limitations for evaluators and the as yet underdeveloped and uninte-

grated professional community of evaluators is made vulnerable to the limitations and not best placed to capitalise on the promise. This article offers a brief review of the publication.

The Framework

The Framework sets out a matrix of 88 quality indicators against which all government departments are invited to judge the products of qualitative evaluation they have commissioned – as well as proposals and tenders. The indicators cover the chronology of the enquiry from conception and design to reporting and use, and, along the way, address ethics, conduct and even researcher reflexivity. So the Framework goes beyond assessing 'research evidence'. The indicators are supported by an evidence base made up of literature review and interviews with social researchers, evaluators and sponsors of programme evaluation. Clearly, some evaluators may feel it prudent to look to the Framework before they respond to invitations to tender, and pre-specify their evaluation design along the lines they are likely, eventually, to be scrutinised.

The publication takes its place among a range of measures being introduced by government departments to regulate/guide social and medical research – the Department of Health has published its detailed Research Governance Framework and the Department of Work and Pensions is producing its own guide to ethical practice – various Euro-

pean initiatives are generating ethical frameworks. Many of these measures are signs of the growing internalisation of social and medical research into the government administrative system and its ever-closer alignment with policy agendas. Policy-relatedness is an obvious benefit; the reciprocal danger of eroding evaluation independence is a dilemma – the political implications for a capacity-building agenda are complex where government plays a sponsoring role in enquiry. In the USA, for example, the Federal Department of Education recently published its determination to sponsor only research and evaluation grounded in randomised control trials, and many fear the consequent loss of a democratic and experiential knowledge base². In Britain, though we are becoming accustomed to government programme evaluation contracts stipulating methodological approaches – sometimes through the mechanism of ethics governance – these often invite mixed methodologies, now possibly enhanced by this Framework. Indeed, the contrast between this document and the methodological strictures being applied in the US is stark. The British Framework may be seen to be both stipulative and enabling.

What follow are some brief notes highlighting the benefits and limitations of the framework.

(Continued on page 6)

¹ 'A Framework for assessing research evidence' (Spencer et al, 2003 – www.strategy@cabinet-office.x.gsi.gov.uk)

² See <www.eval.org/doespage.htm> for the response of the American Evaluation Association to these plans.

(Continued from page 5)

Uses and benefits

(1) **Political legitimacy:** The existence of such a document will be helpful to government departments who may seek 'qualitative' information but who find it hard to argue the case in the context of their own internal accountability. This is an important consideration – apart from the welcome legitimacy the Framework offers to evaluators who seek authority to justify shifting from outcomes to process. We (evaluators and government) may be close to a point where we need to understand the frequent lack of outcome-productivity in reform and modernisation programmes in different ways, but with little of an information base on which to do so. Historically, this is the point at which governments express a need for methodologies which are designed to uncover the mechanisms of change rather than the productivity of programmes – one of the sources of the emergence of qualitative evaluation in the first place. It is the theorising of professionals and programme participants about change potential that is missing from outcomes-focused and 'strong' evaluation designs, but which is so essential to its accomplishment.

(2) **Resolving methodological confusions in accountability:** There is a paradox at the heart of government attempts to juggle the two competing agendas of *accountability* and *change*. At present, government sponsorship of evaluation often assumes that the dominant methodology of accountability in a low-trust environment (inspection

and monitoring against performance measures) serves to promote modernisation and reform agendas – that OfSTED or CHI, for example, promotes change and innovation in schools and hospitals. This tends not to be the case – though it may incidentally promote improvement and reform (i.e. '*do what you've always done, but do it better*'). Inspection for accountability requires of people that they look historically at what they have done, adopt the language of self-justification and certainty, and work within the confines of what is known and what is reliably do-able, and focuses on outcomes. Inspection generates risk and provokes risk-aversion strategies through its intolerance of shortcoming. On the other hand, what professional innovators require are methodologies which support futures thinking, risk-taking, a tolerance for failure as an opportunity to learn, and the language of aspiration and speculation – all those things discouraged by inspection. Change methodologies focus on learning rather than product. Whatever the distinctiveness and merits of qualitative evaluation – and we might better talk of case study³ - its interactive and proximal nature more easily provides a context for the creation of such safe environments. This framework might help to make these distinctions clearer.

(3) **Methodological enhancement:** Such a guide to both enquiry and its assessment will be useful in similar ways to Local Medical Ethics Committees whose members have struggled over the past fifteen or twenty years to come to terms with forms of research which lie outside of the frame of biomedical, epidemi-

ological and quasi-experimental enquiry – with too little success. The resistance of these key gatekeepers to emergent designs, studies of context, experiential data and respondent validation has held back our understanding of professional change issues in health and health-related practices. This framework will help them and those who submit to them.

(4) **Building social research capacity:** The demands of the indicators are such that we will quite possibly see reports to government sponsors considerably 'fattened' to include the extensive accounting of the enquiry process demanded by the Framework. Though these comprehensive accounts and justifications of the enquiry process will likely be confined to appendices, they may guarantee a growing evidence-base for research students and new researchers to learn the trade and its pitfalls. One side-effect of this qualitative accounting might be greater transparency of process and exchange in evaluative enquiry. The inclusion of a section on 'reflexivity and neutrality' – their legitimation as dimensions worthy of account – is bold and helpful to those who promote a more transparent and democratic accounting of the evaluation role.

(5) **Rescuing evaluators from contracts:** Relationships between evaluator and sponsor are too often and too closely regulated by contracts which prove to be clumsy instruments of arbitration in complex and sensitive matters such as ethical exchange, ownership of data and rights over publication. This Framework offers a more finely-

(Continued on page 7)

³There is no space to rehearse the difference here. Suffice it to say that *evaluation case study* – apart from its social research advocacy – is less a methodology and more the construction of a *site* within which methodological deliberations and decisions take place. Theorists of evaluation case study typically argue for both qualitative and quantitative methods as demanded by the context under study.

tuned instrument. The address on validity, in particular, may help to structure conversations between contractors and contractees allowing both to make appeals to the Framework for authority in a context where validation issues are too often reduced to contract clauses drawn up by people who have little direct experience of evaluation and its complex, negotiated reality.

Limitations

A key issue raised by this Framework is the trade-off between closer relationships between evaluation and policy deliberation, on the one hand, and evaluation independence, on the other. Independence is conventionally guaranteed by the principle that an evaluation contracted externally is 'sponsored' and not 'bought' by government. Both evaluator and government department can comfortably distance themselves from one another satisfying their accountability pressures and serving democratic process. The more government stipulates the conduct and the content of an evaluation the less able each party is to claim that distance. Responsibility for the evaluation falls to government whose claim to have sponsored independent scrutiny erodes. As I said earlier, there is both a stipulative and an enabling dimension to this Framework – the emphasis to be determined by its use. Here are a number of other limitations readers of this Framework may need to be aware of.

(1) **The question of warrant:** The evidence base is confined mostly to the epistemological and methodological literature – there is little or no address, for example, on the politics of enquiry, nor on the social and political warrant of the evaluator. This is a significant limitation given the almost universal view that evaluation both takes place in political milieu and is itself a political process. There is, for example, no discussion of the propensity for qualitative evaluation to uncover

multiple and often conflicting purposes, providing broader value frames within which official objectives and targets can be placed. Evaluators have to negotiate their warrant in each and every enquiry in ways that are not required of social researchers.

(2) **Displacing evaluation with applied social science:** The Framework is confined mostly to social research with little explication of the particular challenges of evaluation – a central claim is that programme and policy evaluation can be accomplished with the application of general social science (for a cautionary tale and argument to this view see Stake, 1986 – see, also, House, 1993 and Norris, 1990 for an account of the emergence of programme evaluation as a discrete discipline). The claim of evaluation to separate disciplinary status is dismissed. This is not to return to long-standing debates about the difference between evaluation and research, but to note that one of the foundations of evaluation as an independent discipline is that its community of practitioners has rehearsed such issues extensively and generated a discrete body of professional and theoretical knowledge as a result. Indeed, Scriven (2003), no theoretical slouch, argues that evaluation is a 'transdiscipline' – i.e. providing meta-methodological services to other disciplines who deal with judgement. As useful as this Framework may be, it is prudent to treat it for what it pretends to be – a guide to understanding the application of qualitative research methods to evaluation, and not a guide to qualitative evaluation practice. Otherwise, the Framework leaves both the social researcher and the sponsoring department exposed with an incomplete evaluation 'toolkit' in a field of action which is relatively unforgiving of error and naivety.

(3) **Portraying evaluation as an epistemology rather than as a practice:** A guide confined mostly

to questions of social knowledge production (most of the criteria used to distinguish qualitative enquiry deal with knowledge of context) sets aside the key dimension of programme and policy evaluation as a *professional practice* – not just as a process of theory production. As a professional practice, programme evaluation is founded upon an integration of knowledge and action – the contexts of discovery and explanation coincide, whereas social research, including ethnography – not exclusively but mostly – separates them out.

(4) **Mistaking 'policy evaluation' for 'programme evaluation':**

A final limitation of this Framework worthy of note is that the guide relates to what the authors see as the dominant case of 'policy evaluation'. This is at least as misleading as the claim that a guide to qualitative social research can serve as a guide to evaluation. There are, in fact, few public policy evaluations commissioned in Britain *in the public domain* (and few of those from academic researchers). We rely more heavily on the press than on evaluators here. For the most part, government departments commission *programme* evaluations – and even these are not always comprehensive programme evaluations, being confined, frequently, to impact studies. These are not esoteric or academic distinctions – they have considerable significance for practice. For example, policy evaluation would require the evaluator to enquire into policy sources and values, putting the administrative system under public scrutiny in the way that impact studies put practitioners under scrutiny. There are, notwithstanding the claims of the Framework authors, few such examples in Britain beyond the Audit Commission and the National Audit Office, and occasional forays such as the Scott ('arms to Iraq') and Hutton (David Kelly) enquiries which do stand as

(Continued on page 8)

policy evaluations.

References

House, E.R. (1993) *Professional Evaluation*, Thousand Oaks & London: Sage

Norris, N. (1990) *Understanding Educational Evaluation*, London: Kogan Page

Scriven, M. (2003) 'Evaluation in the new millenium: the transdisciplinary vision' in S.I. Donaldson & M. Scriven (Eds) *Evaluating Social Programmes and Problems: Visions for the New Millenium*, London: Lawrence Erlbaum Ass.

Spencer, L., Ritchie, J., Lewis, J. & Dillon, L. (2003) *Quality in Qualita-*

tive Evaluation: a Framework for Assessing Research Evidence, London: Cabinet Office

Stake, R.E. (1986) *Quieting Reform: Social Science and Social Action in an Urban Youth Programme*, Chicago: Univ. of Illinois Press

'Quality in Qualitative Evaluation' - a (very) critical response

Harry Torrance, Director, Education and Social Research Institute
Manchester Metropolitan University

Liz Spencer and colleagues, based at the National Centre for Social Research, were commissioned by the *Strategy Unit* of the Cabinet Office to produce this 'framework for assessing research evidence'. The NCSR is a "not-for-profit company", its work sponsored by "a range of public bodies" (www.natcen.ac.uk). The report *Quality in Qualitative Evaluation* actually comes in two parts - a 17 page summary including the *Quality Framework* and a 167 page full report including discussion of many of the issues raised by the framework. Presumably it's the short summary report which is expected to get the most attention - especially by policymakers and civil servants commissioning research on behalf of their departments.

This certainly seems to be the intention, since the framework has been "designed primarily to assess the *outputs* of qualitative enquiry...and...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 the training of social researchers" (summary report p.6, original emphasis). So, this is a guide for the commissioners of research when drawing up tender documents, managing the research, and reading reports. It

also has ambitions to influence the training of social researchers. Furthermore, if it becomes apparent that research commissioners are indeed using it to write tenders, then the framework will certainly influence the response of social researchers to tenders.

The full report claims that "there are no explicitly agreed standards regarding what constitutes quality in qualitative policy evaluation methods and no agreed criteria for judging the quality of qualitative evaluation research" (p.10). It continues that the report was commissioned to "address the need for agreed standards" (p.10). 'What need?' one might ask. There are many good reasons why seventy years of debate about anthropology, ethnography and the development of qualitative research methodology has not yet produced such agreement.

There isn't space here to review the full report in detail. Suffice to say that it reads like an introductory text on qualitative research methods. Paradigms are described, issues rehearsed, but all are resolved in a bloodless, technical and strangely old-fashioned counsel of perfection. Thus in addition to the obvious need for "Findings/conclusions [to be] supported by data/evidence" (Full re-

port p.22), qualitative reports should also include:

"Detailed description of the contexts in which the study was conducted..." (p.23);

"Discussions of how fieldwork methods or settings may have influenced data collected..." (p.25);

"Descriptions of background or historical developments and social/organisational characteristics of study sites..." (p.25);

"Description and illumination of diversity/multiple perspectives/alternative positions..."(p.26);

"Discussion /evidence of the ideological perspectives /values/ philosophies of the research team..." (p.27);

And so on and so forth across six pages and 17 quality "appraisal questions".

Of course no one would deny that these are important issues for social researchers to take into account in the design, conduct and reporting of research studies; but in evaluation reports to policymakers? And particularly in evaluation reports which, at other points in the document are also acknowledged to need to meet criteria of utility and feasibility (p.69)? With government departments being most unlikely to fund long-term ethnographic studies (or read 500 page reports) what we have here is a huge hostage to fortune, where

short, sharp, 'dirty' studies which nevertheless come up with findings which the government likes will be acceptable, but where, if the findings are not liked, we now have a 'framework' on which their sponsors can test them to destruction.

To be fair, the authors state that the "framework is designed to aid the informed judgement of quality...not to be prescriptive or to encourage the mechanistic following of rules" (full report p.7). And the 'Quality indicators' noted above are defined as "possible features for consideration" (p22 ff). For all we know the authors struggled long and hard behind closed doors to get as much complexity and ambiguity into the report as they could. If so, it is a pity that the full story is not included in the report. But this doesn't get around the problem of being a hostage to fortune - quite the reverse. It just means that these extended "possible features" are more likely to be employed if the evaluation report is not to a sponsoring department's liking. Furthermore, if the framework is intended to be "educational" for civil servants then by definition it is not going to be employed "to aid informed judgment": either it is needed, in which case it probably *will* be employed mechanistically (and even partially), or it is not needed, in which case informed judgment should suffice.

To try, further, to be fair to the authors, it might be argued that this document at least demonstrates that qualitative research is being taken seriously enough within government to warrant a guide book being produced for civil servants. This might then be said to confer legitimacy on civil servants who want to commission more exploratory qualitative work, on qualitative social researchers bidding for such work, and indeed on social researchers more generally who may have to deal with local ethics committees who are predisposed towards a more quantitative natural science model of investigation (especially in health-related studies from where, interestingly, most of

the references derive). But should we really welcome such 'legitimacy'? The dangers on the other side of the argument, as to whether social scientists need or should accede to criteria of quality endorsed by the state, are legion - that way lies the gulags. Indeed, one of the key quotes used to justify the use of qualitative methods comes from a departmental 'Research Manager':

I often commission qualitative research when it's about users or stakeholders and...I want to understand...how a user is likely to respond...I want to know how they see the world...it's a wonderful vehicle...if you want to understand the motives of people... (full report, p. 34).

So there we have it - social research as spying on the populace on behalf of the state.

My main argument with this publication however, is that, despite its title, it isn't about *evaluation*, and thus it does not address any of the really difficult practical and political questions which evaluation raises. The report is actually about evaluating the quality of qualitative research *per se*, not qualitative evaluation. This is not a pedantic distinction. The above quotation from a departmental research manager is reported without any recognition of the competing political pressures that such managers operate under, as if we can take for granted their *bona fides* without question and forever. Of course such agents *may* act in good faith and with good intentions, and we have all met many who do. But they will also be operating under policy constraints which, at best, move on, thereby rendering evaluation reports largely irrelevant, and at worst will brook nothing less than reports of unambiguous success. These are the pressures that, in turn, commissioned evaluators work under. Evaluation is not just social research conducted in a particular applied context, it is social research conducted in a high-stakes political context.

Nor are these pressures confined to debates between evaluators and sponsors about research designs and findings; they infuse every aspect of fieldwork relationships. Thus, for example, in summarising "key quality issues and concerns" the report states that researchers should show that "steps have been taken to secure consent and to safeguard privacy, confidentiality and well-being of participants" (full report p.71). *How* - in situations where real persons occupy real recognisable roles in the programme under scrutiny? We might not use their name, but it is quite possible that they will be recognised - especially if we also attend to the need to include "detailed description of the contexts in which the study was conducted" (p.23). Safeguarding confidentiality derives from traditional social research where respondents probably can be successfully anonymised and where the orientation of the report is in any case oriented towards building knowledge and theory within a discipline, not analysing the progress of a particular policy initiative. The report goes on to recommend "involving participants in the design, conduct and interpretation of an evaluation" (p.71). Well certainly, but on the face of it this completely contradicts the prior injunction to protect confidentiality - if participants are going to be around the table with evaluators and policymakers they are not going to be anonymous (nor indeed are the policymakers).

All of which is not to say that these issues are insoluble, far from it, they are at the core of enduring debates in the field of evaluation, and good evaluations should seek to take them on board (House and Howe 1999, Kushner 2000, MacDonald 1974, Norris 1990, Simons 1987, Stake 1983, Torrance 2003). This makes it all the more perplexing and frustrating that such literature has not been reviewed by the Cabinet Office report (despite the claim to have conducted a "comprehensive, systematic" review, full report p.11).

(Continued on page 10)

(Continued from page 11)

Some evaluation literature is included, but only at the level of semantics (whether or not there is any difference between evaluation and research) and it is dismissed as a source of experienced advice without the political issues being raised, far less addressed (pp.30-31).

I could go on. Spencer et. Al, 2003, hoist themselves on their own petard when, employing typically mealy-mouthed qualitative 'generalisations', they claim that their framework has "a fair degree of support" (p.43) and that "there was a general view that formalised criteria would be of value" (p.101). Such statements are based on the conduct of twenty nine interviews with social researchers and research commissioners (p.12) out of the thousands who were included in the last UK higher education. The thousands who were included in the last UK higher education research assessment exercise (1056 in the Sociology unit of assessment alone:
<http://www.hero.ac.uk/rae/overview>)

But the key issues with this report are the political ones - who has the resources to sponsor evaluations? Who defines the agenda of enquiry? Why are programmes and initiatives launched in the first place and what are they trying to achieve? Why are policymakers still so enamoured of such a centre-periphery model of government? How can the independence of evaluators from sponsors be negotiated? How can evaluation inform public debate rather than bureaucratic decision-making? How might vulnerable persons and communities be protected from the scrutiny of government, while we still try to learn something useful about how programmes operate that might

improve the experience and outcomes for participants? Raising and exploring these questions, not to produce a 'framework' to hold qualitative evaluation to account, but to provide an illustrative agenda for the negotiation of evaluation contracts, might have produced a more interesting and useful document.

Why were these questions ignored? Why were the key issues in evaluation theory and methodology not addressed? How did this peculiar report come to be written? I noted above the preponderance of references to qualitative methods in health-related studies and herein, perhaps, lies a clue. After winning the commission to write the framework, and no doubt under immense logistical and political pressure to get it completed in a very short space of time, did the authors rely on knowledge of the field and networks of colleagues with which they were most familiar - social research in general, and health-related studies in particular? Out of 190 references cited (p.111 ff) 97 are to discussions of qualitative methods *per se*, and a further 61 are to general discussions of social science theory and methodology (=84%). Only 28 (15%) are specifically to evaluation models and methods, and an electronic search of the full report in its on-line version revealed not a single reference to 'independent evaluation'. Furthermore, of the research topics and journal affiliations which are obvious from the titles, 54 of the references derive from health-related studies, with only 11 from criminology, 8 from education, and 4 from social work; this despite the continuous investment in research and evaluations of various kinds, including qualitative evaluations, by the Home Office and DfES going back over thirty

years.

So, we are left with a skewed report, even a biased one; certainly a very puzzling one. In a report which supposedly addresses problems of quality in qualitative evaluation, it is ironic indeed that the authors committed the cardinal sin of evaluation - responding only to the agenda of the sponsor and not identifying the concerns and issues of all relevant stakeholders.

Correspondence:
h.torrance@mmu.ac.uk

References

- House E. and Howe K. (1999) *Values in Evaluation and Social Research*, London: Sage
- Kushner S. (2000) *Personalising Evaluation*, London: Sage
- MacDonald B. (1974) 'Evaluation and the control of education' reprinted in Murphy R. and Torrance H. (Eds 1987) *Evaluating Education: issues and methods*, London: Harper and Row
- Norris N. (1990) *Understanding Educational Evaluation*, London: Kogan Page
- Simons H. (1987) *Getting to know school in a democracy: the politics and process of evaluation*, London: Falmer Press
- Stake R. (1983) 'Program evaluation: particularly responsive evaluation' in Madaus G. et. al. (Eds) *Evaluation Models*, Boston: Kluwer
- Torrance H. (2003) 'When is an 'Evaluation' not an Evaluation? When it's sponsored by the QCA? A Response to Lindsay and Lewis', *British Educational Research Journal*, 29, 2, 169-173

Building Research Capacity

ISSN 1475-4193, Issue 8

Editors

Chris Taylor, Project Manager
Ruth Boyask, Programme Co-ordinator
Rachel Prosser, Project Admin. Support

Cardiff University School of Social Sciences

Views expressed in this journal do not necessarily reflect those of the University, the Research Capacity Building Network, the Teaching and Learning Research Programme, nor the Economic and Social Research Council.

Qualitative research in teaching and learning: quality, innovation and future directions RCBN Seminar Series

Organised in collaboration with the Institute of Education, Manchester Metropolitan University.

The first seminar of this series was held in December 2003 and provided a forum for vigorous discussion about current issues regarding the quality, significance and reception of qualitative teaching and learning research. Ian Stronach, Saville Kushner and Ray Pawson were invited to present their reflections on the report *Quality in Qualitative Evaluation: A framework for assessing research evidence* (Spencer, Ritchie, Lewis & Dillon, 2003). These responses and the ensuing discussion prompted requests for written responses to the report that could be more widely disseminated, resulting in the three responses published in this issue of *BRC*. The seminar also generated discussion on the position of qualitative research in current evidence-based agendas for developing policy and practice, focusing on the proliferation of the use of systematic reviews and subsequent variations in systematic approaches. Throughout the discussion ideas were generated for the remaining three seminars in the series. The following is a list of seminars planned for 2004, generated through discussion at the seminar, and are of significance to all qualitative researchers with an interest in the teaching and learning field.

Developing Research Capacity in Qualitative Inquiry

12 May, Cambridge University

This one-day seminar explores some of the tensions and possibilities of developing research capacity in contexts of qualitative inquiry. The seminar has a particular focus on qualitative approaches to research and practice. Speakers have been invited to present in-progress papers on teaching and learning research that situates concepts of "partnership" at the centre of the relationship between research and practice.

Themes that will be addressed in this seminar include; how research capacity is constructed in educational research and the significance this has for qualitative educational researchers, initiatives in practitioner research in current teaching and learning research and the role university researchers play in supporting this.

Speakers include: **Ian Stronach** (Manchester Metropolitan University); **Jan Winter** (Bristol University); **Peter John** (University of Plymouth); **Colleen McLaughlin** (Cambridge University); **Alan Dyson** (University of Manchester).

Situating Qualitative Research in Evidence-based Research and Systematic Review Agendas

June, details to be confirmed

Creating a venue in which researchers can engage critically with a range of different approaches to evidence-based and systematic review methods and methodologies (e.g. systematic reviews, meta-analysis, Bayesian meta-analysis, realist synthesis, meta-ethnography) that have currency in educational research and policy development, and explore their congruencies and disparities with the interests, theoretical concerns and practice of qualitative researchers within education.

Speakers include: **Martyn Hammersley** (Open University); **Mary Dixon-Woods** (University of Leicester); **Liz Spencer** (University of Essex); **Harry Torrance** (Manchester Metropolitan University).

The Educational Future and Innovative Qualitative Research

November, Manchester Metropolitan University

Investigating innovative qualitative research methods and methodologies in the context of educational futures, with invited speakers and workshop facilitators (UK and abroad) engaged in cutting edge qualitative research in relevant social science disciplines.

For more information about this seminar series and other RCBN events please visit the RCBN website (www.cardiff.ac.uk/socsi/capacity) or email rcbn@cf.ac.uk

RCBN News

Change of staff within the RCBN

Rosie Lynne has recently moved from her post as our Project Administrator to take up a position in marketing in Tewkesbury. We are pleased to introduce Clare Davies as our new RCBN Administrator, who joins us from the Government Office South West, and Rachel Prosser (Administrative Support). All our contact details remain the same. For any information about the RCBN please feel free to email Clare or Rachel on rcbn@cardiff.ac.uk or call on 029 20875345. Please also remember that anonymous comments or suggestions can be made via the RCBN website: www.cardiff.ac.uk/socsi/capacity

Forthcoming RCBN events

Randomised Controlled Trials Series Randomised Trials in Educational Research (One-day Seminar)

20 May, London

The aims of this introductory session are to discuss the need for well-conducted, high quality research designs to identify the effectiveness (or otherwise) of educational interventions, compare quasi-experimental with fully randomised controlled trials, demonstrate the feasibility and values of trial design and discuss some of the barriers to conducting trials.

Design and Analysis of Randomised Controlled Trials (Two-day Workshop)

7-8th June, London

The aim of this two day workshop is to provide an understanding of key issues in the design, analysis and reporting of randomised controlled trials. The workshop is specifically intended for those considering that they may wish to undertake a randomised controlled trial in the near or distant future. The workshop will be problem based and participative, and participants will have the opportunity to work through the design of a trial of their choice.

Research Approaches to Revealing Tacit Knowledge

25 May, Institute of Physics, London

Led by Prof Michael Eraut, Stephen Steadman and Chris Fessey from the University of Sussex, this one-day event is an intensive workshop for researchers who are interested in developing understanding and skills in revealing tacit knowledge. The workshop will provide the opportunity to clarify thinking about uncovering that which is tacit, and in particular, to highlight the special difficulties that may arise. There will also be opportunities for researchers to reflect on the applicability of a range of available techniques and a practical component will allow participants to engage in activities to develop their skills and trial new research approaches.

Critical Reading for Self-Critical Writing— Introduction to a Structured Approach for Reviewing Literature

27 May, University of Bath

Key skills for social science researchers are the twin abilities to engage in a constructively critical way when reviewing literature and to mount a convincing argument in working up a written account of the review. This workshop has been designed particularly to support new and relatively inexperienced researchers in developing their capacity to engage critically with academic, policy and practical literature, especially in the course of contributing to the Teaching and Learning Research Programme. Event Leaders Mike Wallace and Louise Poulson have successfully pioneered a structured approach for learning to review literature critically with masters and doctoral students at the Department of Education in the University of Bath. They are co-authors of the recently published book *Learning to Read Critically in Teaching and Learning*, a teaching text based on their approach.

Other forthcoming events

**12 May Developing Research Capacity in
Qualitative Inquiry**
(1-day seminar)
New Hall, Cambridge University

**June Situating Qualitative Research in
Evidence-based Research and
Systematic Review Agendas**
(1-day seminar)

**For details of all RCBN events, and to request
places for those listed above, please visit our web-
site, or contact the RCBN Office (details below).**

ESRC Teaching and Learning Research Programme Research Capacity Building Network

Project Director

Gareth Rees (rees@cardiff.ac.uk)

Project Administrator

Clare Davies (rcbn@cardiff.ac.uk)

Correspondence

Cardiff University School of Social Sciences
Glamorgan Building
King Edward VII Avenue
Cardiff CF10 3WT

Tel. +44(0)29 2087 5345 Fax. +44(0)29 2087 4678
www.cardiff.ac.uk/socsi/capacity