

## **Commentary on: Quality in Qualitative Evaluation: a framework for assessing research evidence**

A research seminar organised by the Cabinet Office and the ESRC Research Methods programme

**Admiralty Arch, London 15 Sept 2003**

**Mary Dixon-Woods, Department of Health Sciences, University of Leicester**

### **Introduction**

Qualitative research is valuable because it can explore meanings and behaviours in depth, identify and characterise diverse perspectives, processes and contexts, and offer flexibility and access. In order for qualitative research to make this contribution, however, it has to be of satisfactory quality. It is no use to a policy maker if qualitative research tells her that the main reason teachers leave the profession within two years of training is because they have problems with stress, if the more important reason is that they are dissatisfied with pay levels.

I am not a policy-maker but I am interested in the quality of qualitative research for several reasons. First, as a practitioner. Second, as someone interested in the synthesis of evidence, and particularly in how qualitative evidence can be included in systematic reviews. Third, as someone who does an enormous amount of reviewing of grant applications, protocols, and manuscripts submitted to journals. Fourth, as someone who teaches about qualitative methods. Finally, I did spend a number of years in a policy-making environment as a civil servant, and recognise how important it is to have confidence in the evidence on which decisions are made. For all of these reasons, a soundly based means of appraising quality in qualitative evaluation is very welcome and I am delighted to have been asked to comment on this much-needed work by the National Centre for Social Research.

This work is much needed because we are now seeing a very unhelpful and wasteful proliferation of criteria in relation to qualitative research. My research group, for example, has identified over a hundred different attempts to identify quality standards for qualitative research. It is high time that we had a considered overview and synthesis of the type that we see here. I think this kind of work is valuable not only because it helps in the appraisal of existing research, but also because it helps to formulate and drive quality standards for new research.

The guiding principles of the framework – that research should be contributory in advancing wider knowledge or understanding, defensible in design, rigorous in conduct, and credible in claim – are entirely sound and completely appropriate. The documentation for the framework provides a clearly written, insightful, and very

useful overview of the use of qualitative research for evaluation, and is a valuable resource in its own right. For example the debate on whether it is possible or desirable to establish quality criteria for qualitative inquiry is handled in an even-handed and sensible way, while the discussion of validity and its many meanings is equally informative and helpful.

The process of developing the criteria seems to be particularly worthy of praise, with its rigorous and iterative approach. There is much to admire, too, in the presentation of the framework. I like the way that the function of each criterion has been labelled, for example as analysis or reporting. Too often there has been a failure to distinguish between what is a failure in reporting and what is a failure of study design or execution. For example, not discussing the ethical issues in a paper reporting conversation analysis study is a problem of reporting, while in the same conversation analysis study, failure to make verbatim transcripts using the conventional notation might be a fatal error of execution. I also like the way that the quality indicators are embedded in the framework with the questions, rather than being in a separate document as is sometimes the case. The emphasis on the need for professional judgement, and the caution about reducing appraisal to a series of pre-determined indicators, are also most welcome.

I do have some however's, however.

I think that there are probably too many criteria in this Framework. For example the question on ethics to me seems redundant, since all research and evaluation throws up ethical issues, and the extent to which these can be discussed is often very limited. I also think that many of the quality indicators are too detailed. I can think of no published journal article that would satisfy the criteria, and maybe only one or two full-length theses or books. I also think that, notwithstanding the cautions about the indicators, these kinds of measures do have a habit of transmogrifying themselves into a form of tyranny that their developers never intended.

Notwithstanding my comment about too many criteria, there are two very important criteria absent. The first criterion, and I believe it should be the first criterion in the framework, is the question about whether the evaluation is genuinely qualitative or not. Too often as a reviewer I am sent something purporting to be "qualitative" research which turns out to be simply some very badly done quantitative research. Two swallows never made a summer, but having just two participants doesn't make it a qualitative study either!

The second problem that I come across as a reviewer, and probably the most common one, doesn't seem to be adequately dealt with by the framework either but is perhaps my chief concern with this Framework. This is the problem of inadequate specification of study type and methodology. Too often, authors claim they have used a particular approach, but in fact they have not. It is not at all uncommon for authors to confidently assert at the outset of a paper that they are using discourse analysis, only to present findings that have nothing to do with discourse. But the worst offences occur in relation to grounded theory, where false claims abound. It is even more common to find a la carte approaches to particular methodological approach, where some elements of the approach are used but other key assumptions are violated. There are just too many examples of this kind of thing around, and

I don't think this is evidence of creativity and flexibility in qualitative research: I think this is evidence of a sloppy, ill-disciplined research which gives the appearance, at times, of methodological anarchy, where researchers feel able to do whatever they want yet still lay claim to using particular methodologies. This does qualitative research no favours at all. It is time that qualitative research stopped defending its appearance of chaos with high-handed defences of "interpretive research" and rejection of orthodoxies. So I think we need a criterion that asks if the authors have identified which methodological tradition they have used, and if there is evidence that in fact they executed their research within this methodology.

It is now I think we come to what is my major issue with the framework. Methods for appraising quantitative research have recognised that different study types (e.g. randomised controlled trials, case control studies, studies to evaluate screening programmes etc) may demand different criteria. This allows the very precise formulation of flaws that would be fatal or very damaging to the rigour of a particular study type. For example, failure to randomise properly in a randomised controlled trial is pretty much fatal. In qualitative research focus groups, interviews, participant observation, and so on, also constitute different study types and different methodological traditions – discourse analysis, grounded theory, interpretive phenomenological analysis, and so on, embody different sets of assumptions and involve different procedures. They may therefore may also need different criteria to allow their appraisal, and similarly have different types of fatal flaws. For example, there are often aspects of procedure that are crucial to the successful execution of a particular study design or a particular methodological tradition. A study claiming to use grounded theory that relies entirely on a priori concepts is, to my mind, fatally flawed. Equally I would query someone using a respondent validation approach to "validate" a discourse analysis, or someone who produced a content analysis with no evidence of explicit category specifications or counting.

This framework argues that it adopts a "paradigm neutral" approach, or what it calls an "elemental" approach. It was not felt possible to design a single set of quality markers that would be appropriate for all the different qualitative and evaluative traditions that exist. It is also pointed out that there is no methodological orthodoxy in qualitative research. The report points out that the assumptions on which they are based are too diverse – in some cases contradictory – and the conception of what matters in terms of 'quality' can vary with these assumptions. The elemental approach was intended to make the framework relevant to most types of qualitative evaluations, on the grounds that the framework will assess the assumptions within which an evaluation operates rather than the model of research used. I accept many of the arguments that the authors make about the difficulties of trying to devise method or design specific conceptions of quality. In the main, these refer to the underlying tensions in the field, where there is little evidence of consensus on many features either of study design or methodological tradition.

To me, however, the essentialist approach is too limited, and confines the appraisal to generic issues rather without allowing an assessment of vitally important problems of execution or interpretation. I think it really is necessary to define what the expectations are of particular designs and traditions. I agree that there will be many difficulties about determining good practice across study designs and methodological

approaches, many of them identified in the report. It often seems to me that the distinctions between the various approaches are often overstated and that in practice most studies one sees seem to revert to some form of thematic analysis as their central analytic strategy and very often some form of opportunistic or guided opportunistic sampling for their sample. I think we need to see some consolidation going on, rather than the continual invention of new approaches, and greater clarity about what precisely might constitute a particular study design or a particular methodological approach. One way of building on the work of the Framework would be to see it as part of a modular structure. The current Framework could form the core, and add-on Frameworks could be developed to address the study designs and methodology issues.

There is one other major area that I feel is not well addressed by the Framework as it stands. I have referred already to what are to some extent issues of procedure, to do with the techniques associated with particular designs and methodological traditions, but perhaps even more important are issues of interpretation. Someone may follow the procedures of a particular study design or methodological tradition to the letter, but, because of failings in the quality of interpretation, end up with a poor piece of work that lacks insight or vividness. I don't feel that the current question on richness of data adequately addresses this problem.

Finally, a few minor things. I found the order of the questions a little quirky. It seems a little odd to start with "how credible are the findings?" – to me this should have been nearer the end. I accept, as the authors point out, that you can do appraisals using whatever order of questions seems best suited to what you want to do, but still!

I also found the phrasing of some of the questions and their labelling somewhat idiosyncratic. For example, question 10 – *how well has the approach to and formulation of the analysis been conveyed?* – reads like a "reporting" question but is labelled "analysis". It may be that the analysis has been conducted very well – as in Goffman's analysis of asylums – but "conveyed" very poorly – as in Goffman's analysis of asylums. I think I would also query the use of the label of "reflexivity and neutrality" – to me these terms are value-laden and open to misinterpretation.

Notwithstanding these concerns, I think that this framework and its associated report have achieved a great deal. The authors have provided a scholarly and insightful analysis of issues in appraising qualitative research, and have made an excellent contribution to methods for appraisal. I would like to congratulate the authors and wish them very well in their future work in this area.