The demand for criteria of assessment for qualitative research arises in two contexts that are significantly different: for use within the research community; and for use by sponsors, funders, or users of research, many of whom will not be active researchers. I will concentrate on the first of these contexts, because this more fundamental. However, I will touch on the second towards the end.

We might ask why criteria are needed by researchers to judge their own and one another’s work. There are at least two possible answers to that question. First, it might be argued that only if they operate on the basis of explicit procedures, both in carrying out research and in assessing its products, will qualitative researchers produce reliable findings. This, of course, is to adopt a conception of scientific method, reflecting one strand of positivism, which even those qualitative researchers who see their work as following a scientific ideal would not accept. Secondly, it might be argued that, given the diversity of approach within educational research, and specifically among qualitative researchers, there is a need to make explicit the criteria by which work within the various genres should be judged. While I have some sympathy with this second argument, it carries what in my view is the undesirable implication that we should simply accept the various paradigms that currently exist as ‘valid in their own terms’, only requiring of them that they spell out their distinctive criteria (which, in any case, some will deny is possible).

In the other paper, I think I managed to show that there are understandable (though not necessarily justifiable) reasons why there is considerable variation amongst qualitative researchers in their judgments about the quality of particular pieces of work, and in their views about how we should judge quality. To recap, these stem partly from what we might call niche-relevance, but also from significant disagreements about whether qualitative research can or should be scientific, and what this means; about whether it is possible to produce knowledge of social phenomena or only of the discursive practices through which they are constituted; and/or about whether the immediate task of research is simply to produce knowledge or also (or instead) to serve some other practical function in the world.

Not everyone believes that this diversity in orientation within educational research is a bad thing, indeed this would be denied by many qualitative researchers. However, even if one thinks that it is undesirable, there is still a problem concerning how increased agreement can be brought about. There are those who favour a Leninist solution to the question of what is to be done: in other words, the imposition of some particular methodological model on the research community. And this does seem to be what is currently being attempted in the United States. In my view, this would be unlikely to be successful even if the model being imposed were a sound

---

1 I have explained elsewhere (Hammersley 2001) why I believe this positivist idea to be false, drawing on the work of Polanyi. There is an important tension here, however. It is not the case, of course, that the judgments made by scientists are simply intuitive, in the sense of being idiosyncratic claims to knowledge that have to be accepted or rejected at face value. There is a requirement in science that judgments be explained to others to the extent that this can be done, and corroborated wherever possible by other research. And what cannot be sufficiently explained, assessed or corroborated cannot be accepted as knowledge. In my view this must be applied in educational research as much as in any other scientific field of inquiry. But there is no way in which to guarantee the validity of judgments by proceduralisation. Indeed, seeking to proceduralise the practice of research and the assessment of its findings is counterproductive: it involves distortion of the tacit knowledge that we cannot but rely upon.

2 Paper prepared for a seminar on Assessing Quality in Case Study and Qualitative Research, July 2005, forming part of the ESRC TLRP Seminar Series on Quality in Educational Research.

The boundary between these two audiences is not a sharp one, and this reflects the fact that neither audience is homogeneous. This arises partly from the fact that closeness to a particular research field is a matter of degree; social scientists working in quite different fields are in a position that is not dissimilar from that of lay people in judging work there. Equally important, those who are researchers play other roles and indeed may move on to playing those other roles to the exclusion of being engaged in research, and it does not take long to lose track of what is happening in a field.
one. But, in any event, I am not convinced that we know what is the right model, though as will be clear from what follows I have definite views about this. Given the uncertainty, the only route to a solution, I suggest, is through dialogue; though I do not have great confidence that we will get to a destination any time soon; especially since the form of dialogue required is not simply a matter of the various parties coming to understand one another better but rather the kind of dialogue that Gadamer talks about, in which we participate so as to allow ourselves to come to understandings that none of us had before. So, what follows is my contribution to what I hope will be such a dialogue. Given the time and space available, of course, it can only be a sketch.

**Reducing diversity in orientation within educational research**

Even in principle, I do not believe that there is a single way in which the sharp differences in methodologies currently to be found amongst educational researchers can be reduced. At least two, complementary, strategies are required. First, we need clarification about the boundary around inquiry, and about different kinds of inquiry in terms of their relationship to practice and policymaking. Here, I take it that the distinguishing feature of inquiry, of any kind, is that its goal is the production of knowledge. As regards kinds of inquiry, personally, I do not find the available distinctions (pure/applied, strategic/non-strategic, basic/practice-driven, etc) helpful, and I have tried to develop a better typology (Hammersley 2002;ch6; 2003b). This involves, first of all, distinguishing between inquiry-subordinated-to-another-activity from specialised inquiry, in which the production of knowledge is the exclusive immediate goal, for which I will reserve the label ‘research’. Within research, we need to distinguish between that which is practical, being concerned to provide information needed by some set of practitioners, and that which is academic or scientific, which is aimed at contributing to a body of knowledge, whether about some value-relevant issue or about some theoretical problem. It seems to me that all these forms of inquiry are of value, but that they demand very different orientations. While I would not want completely to rule out combining them, or blurring the boundaries between them, I believe that this will often lead to serious conflicts in orientation. Recognising these different types of research is intended to resolve some of the methodological differences that currently exercise us: since they relate to genuine differences in purpose. However, it will not deal with them all.

The second strategy involves showing that some of the methodological arguments that are currently influential in the field of qualitative research are internally incoherent. I think this is true of constructionism, for example (Foster et al 1996:20-2; Hammersley 2003a). Furthermore, in some cases what is involved is a rejection of assumptions that are constitutive of any form of inquiry, notably the idea that knowledge (in the conventional sense of representation of fact) is both possible and desirable. There may be good reasons to abandon those assumptions in relation to some fields of inquiry (astrology, to use a hackneyed example), but it should be made clear that this is what is being done, not some redefinition or rebranding of inquiry. Furthermore, the application of this kind of scepticism across the board leaves the sceptic with nowhere to stand, this being the most fundamental kind of incoherence.

To spell this out a little more, regards constructionism, suspending belief in the independent existence of social phenomena in order to focus on discursive practices is a reasonable methodological strategy. It is analogous to psychologists who study attribution processes treating as irrelevant the question of whether or not the attributions are or are not correct. However, to extend constructionism into making ontological assumptions about what does and does not exist, or about what is most fundamental (say, discursive practices), is self-undermining. The problem is analogous to the consequences for behaviourist psychologists of applying behaviourism to themselves and to how they should go about doing research. Applied consistently as an ontological doctrine, constructionism amounts to an abandonment of inquiry in favour of writing propaganda or imaginative literature.

Finally, those conceptions of research which add practical goals to the epistemic one of producing knowledge increase the likelihood of bias, since there are likely to be incompatibilities between pursuing the two sorts of goal. As a result, doing this is irrational to the extent that one is committed to the activity of research. Doing good quality research is difficult enough as it is, without making the task even more demanding.

I do not expect these arguments about the goal of research, about the legitimate variation in its forms, or about the incoherence and ill-advised character of many versions of qualitative research, immediately to convince anyone who does not already share my commitments here. This outline is simply intended to indicate my starting position; and why I believe that it is possible to reach an understanding of

---

3 On Gadamer’s conception of dialogue, see Zuckert 2002.
4 This is, of course, at odds with the views of many educational researchers, who see their work as directly linked to practical or political goals. It is also at odds with those discussions of research assessment that include criteria relating to these practical goals, for example Furlong and Oancea 2005.
5 In these terms, some ‘action research’, ‘interventionist research’, ‘interactive social science’, etc. would come under the heading of practical research, some under the heading of inquiry-subordinated-to-another-activity.
6 If the discourse being studied cannot have a representational function, why should we allow the writings of the discourse analyst to have such a function?
what is involved in evaluating research that is common, in general terms, not just across qualitative work but also across the quantitative-qualitative divide.

Types of methodological evaluation

It is important to recognise that several different kinds of evaluative judgment can be involved in research, focused on different objects and serving different purposes. At least the following possibilities can be identified:

a) Assessing how well a study is presented in a research report, in terms of whether it is clear and provides all the information we need.
b) Assessing the findings of a study, or of a body of research, to determine whether they should be believed.
c) Assessing a piece of research or a body of research studies to decide whether they were carried out well.
d) Assessing a study or group of studies in order to judge whether the methods they employed are ones that seem likely to be fruitful in other contexts.
e) Assessing the expertise or competence of particular researchers on the basis of the work they have produced; for example, in order to decide whether they should be awarded a PhD.

While these different kinds of assessment share something in common, for instance in judging findings we will also need to take account of the methods used, there are nevertheless very important differences. This is signalled, for example, by the fact that research may not have been done in the best way possible but may still have produced convincing findings; and vice versa. To underline the differences, I have indicated in Tables 1, 2 and 3 what I think would need to be taken into account under the first three headings.

<table>
<thead>
<tr>
<th>TABLE 1 CONSIDERATIONS IN ASSESSING THE ADEQUACY OF RESEARCH REPORTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The clarity of writing:</td>
</tr>
<tr>
<td>a. Consistency in use of terms</td>
</tr>
<tr>
<td>b. Are definitions provided where necessary?</td>
</tr>
<tr>
<td>c. Are sentences well constructed</td>
</tr>
<tr>
<td>d. Is there use of excessive rhetoric?</td>
</tr>
<tr>
<td>2. The problem or question being addressed:</td>
</tr>
<tr>
<td>a. Is this clearly outlined?</td>
</tr>
<tr>
<td>b. Is sufficient rationale provided for its significance?</td>
</tr>
<tr>
<td>3. The formulation of the main claims:</td>
</tr>
<tr>
<td>a. Are these made clear?</td>
</tr>
<tr>
<td>b. Are the relations between subordinate and superordinate claims (including evidence) made sufficiently explicit?</td>
</tr>
<tr>
<td>c. Is the nature of each claim (as description, explanation, theory, evaluation, or prescription) indicated?</td>
</tr>
<tr>
<td>4. The formulation of the conclusions:</td>
</tr>
<tr>
<td>a. Is there a distinction between main claims about the cases studied and general conclusions?</td>
</tr>
<tr>
<td>b. Is the basis for the conclusions made signalled?</td>
</tr>
<tr>
<td>5. The account of the research process and of the researcher:</td>
</tr>
<tr>
<td>a. Is there sufficient, and not too much, information about the research process?</td>
</tr>
<tr>
<td>b. Is there sufficient, and not too much, information about the researcher? (In other words, is what is necessary and no more provided for assessing the validity of the findings, the value of the methods, the competence of the researcher, depending upon which is the focus.)</td>
</tr>
</tbody>
</table>
TABLE 2 CONSIDERATIONS IN ASSESSING THE VALIDITY OF FINDINGS

1. The main claims and evidence:
   a. Are the main claims plausible enough to be accepted at face value?
   b. If not, is evidence provided?
   c. If so, is the evidence sufficient, both in terms of strongly implying the validity of the main knowledge claim and in being sufficiently plausible or credible to be accepted?
   d. If not, is a further layer of evidence provided?
   e. If so, is this evidence sufficient? And so on.

2. The relationship between the findings about the cases studied and the conclusions drawn:
   a. Where these are empirical generalisations about some finite population, on the basis of whatever evidence is provided, are they sufficiently plausible or credible to be accepted?
   b. Where they are theoretical statements of a conditional causal kind, on the basis of the evidence provided, are they sufficiently plausible or credible to be accepted?

So, to start with, we need to be clear about which sort of assessment we are concerned with. Here, I will focus mainly on the second type, since in the context of research as an activity it is the core one, in the sense that it relates directly to the goal of inquiry: the production of knowledge. However, it is perhaps just worth underlining that my account of what is required in terms of presentation would be very controversial amongst qualitative researchers. It is not just postmodernists who would find it unacceptable, the same would be true of many others, probably including most ethnographers.

On the basis of work that has been done in the philosophy of science and in epistemology more generally – in the wake of the discrediting of all forms of foundationalism - it seems to me that assessing the validity of research findings involves the process outlined in Table 2. First, there must be a judgment about whether the knowledge claim concerned is sufficiently plausible to be accepted at face value, because it is strongly implied by whatever is taken to be already known. Secondly, where the claim is judged to be insufficiently plausible in this sense, as will usually be the case with research findings, otherwise their news value would be very small, then there is a need for empirical evidence to support the claim. This evidence must be judged, first, in terms of how strongly it supports the initial knowledge claim (how strong is the line of inference from one to the other?); and, secondly, in terms of whether it, the evidence itself, is likely to be valid. Assessing the validity of the evidence involves judging whether it is sufficiently plausible to be accepted at face value, because it is strongly implied by what we take to be already known, or alternatively according to whether it is credible, in the sense that it is based on a mode of production (a research method employed in particular circumstances) that is unlikely to generate significant error. Where this initial evidence is judged as not offering strong enough inferential support for the knowledge claim then parallel evidence is required, to provide multiple strands that offer the necessary strength of inference. Where the evidence is judged to be neither sufficiently plausible nor sufficiently credible to be accepted at face value, then further evidence in support of the validity of the initial evidence is required. And this process must continue until evidence has been produced that is, firstly, sufficiently plausible and/or credible in itself, and, secondly, that provides the necessary amount of strong inference in support of the knowledge claim.8

7 A good example of the kind of philosophical view I am depending on here is Haack 1993.
8 It is important to note that reliance on judgments of plausibility, in the sense used here, does not imply that we can simply hold on to beliefs that we find ‘sufficiently plausible’. As researchers we are required to try to justify to others what we take to be knowledge. If most other researchers in the relevant field do not find a knowledge claim sufficiently plausible (or credible) to be accepted, then we must provide further evidence. Only if most of the relevant research community at a particular time find a knowledge claim or assumption sufficiently plausible can it be accepted at face value, and even then only ‘until further notice’ - should doubt be cast upon it in the future, then it will need to be investigated.
To a large extent, I see what I have outlined here as simply an explication of the logic implicit in how most researchers go about assessing the validity of findings; even if they do not always follow that logic fully. Of course, at the same time it raises problems for some kinds of research. For example, as I noted in my other paper, many ethnographers argue that what they present in research reports cannot be supported with evidence of a kind that would convince readers without the latter trusting the ethnographer’s ability to interpret the relevant cultural phenomena soundly. Otherwise, readers would need themselves to have acquired the relevant cultural competence in order to make sense of and derive the correct inferences from particular items of evidence. I have some sympathy for this argument, but only in so far as I believe that full explicitness is never possible in any kind of research.

The functional relativity of evidence

So, fairly obviously, evidence is usually essential in assessing research findings. However, it is important to recognise that what counts as ‘evidence’ is always relative, as Richard and Gary both indicated in their talks at the last session. It is only in relation to some particular knowledge claim, or type of knowledge claim, and also in relation to the purpose that any evidence would be designed to serve, that we can determine what would count as evidence, and indeed as good evidence. In other words, the nature of evidence, and its quality, are not intrinsic: it is only in relation to some function that we can determine what is and is not evidence, and what would be stronger or weaker evidence.

One aspect of this relates to different phases or aspects of inquiry. Charles Peirce identified three different types of inferential process that take place within inquiry: abduction (or retroduction), deduction, and induction. Abduction is the development of an explanatory or theoretical idea, this often resulting from close examination of particular cases. The process of inference involved here is not a matter of strict logic, it is not deductive nor does it follow some ‘inductive logic’, it is ampliative, in the sense that the conclusion is not already present in the premises. According to Peirce, abduction leads the way into the next task: having developed a theoretical idea that could account for the phenomenon of interest, we must then deduce implications or hypotheses from this idea that can be tested. And multiple hypotheses should be generated not just one. This second process of inference is deductive, in other words strictly logical. Finally, there is the process of induction, this is the task of testing these hypotheses against data from particular cases, and on the basis of the evidence offered by these cases inferring back to a conclusion about the validity or invalidity of the theory. This, like abduction, cannot be a matter of strict logic. And it is a complicated, uncertain, and lengthy business. Often, a single study will not be able to cover all of it.

Now the point for my purposes here is that what is good evidence for abduction is different from what is good evidence for induction. Those who rush to the process of hypothesis-testing are underplaying what is involved in deduction and induction. Equally, qualitative researchers who state that their work is exploratory in character will then go on to draw apparently definitive theoretical conclusions from it are neglecting what is involved in deduction and induction. Indeed, it has sometimes been explicitly argued that qualitative research is simply concerned with abduction, as if this could stand on its own (see Coffey and Atkinson 1996:155, 162-3). I do not believe that this is true, but it does seem to me that Peirce is somewhat misleading in his account of inquiry, in that he focuses almost entirely on research that is concerned with developing and testing theories. Yet, in the case of social and educational research at least, theories are not the only legitimate goal of inquiry. Instead we may aim at producing descriptions or explanations. And what is required as regards evidence varies depending upon the intended product. So this is another aspect of the functional relativity of evidence.

For the most part, descriptions as main findings of a study require only descriptive evidence, although sometimes explanations can be used in a subordinate way to indicate that we might expect a particular description to hold in the case concerned. As regards explanations, we need descriptive evidence about both the explanandum and the explanans, both about what is being explained and about what is held to explain it. And it will also be necessary to show that this theory applies to the case concerned. In seeking to produce theories, we need descriptive evidence about the key variables in various relevant cases, showing that the associations the theory predicts do appear. And we also need evidence that minimises the chances that some other factor than the one at the centre of our theory was what generated the association.

Finally, of course, what is and is not evidence depends upon the particular knowledge claim concerned: what will be evidence for or against one knowledge claim will not be evidence for or against another. In other words, everything depends upon the question we are seeking to answer. And it should be said that educational, and other, researchers are not always as careful as they should be in formulating either the questions they are addressing or the knowledge claims they are making. These are sometimes ambiguous, with the result that what would count as evidence is uncertain. Equally misleading are references to building up an ‘evidence-base’, as if evidence could be all-purpose.

---

9 On Peirce’s conception of inquiry see, for example, Misak 991.
10 I am not entirely sure that this is true, but I will leave this on one side.
Up to now in this paper, then, first of all, I have explained why I do not believe that we should simply accept the wide range of genres of educational, and particularly qualitative, research that currently exist, and by implication why we should not accept some of the assessment criteria associated with them. I then tried to indicate the different kinds of assessment that are involved in research, and to make clear that which one is being pursued makes a difference to what is relevant. And I outlined the process that I think is involved in assessing the validity of findings. Finally, I emphasised the functional relativity of evidence. In the final part of the paper I want to address the issue of whether or not there can be criteria of assessment.

The problem of criteria

As I indicated earlier, the demand for criteria comes in part from positivism. However, even if we reject positivism this does not necessarily mean that criteria are impossible or undesirable. Crucial here, though, is what we mean by the term ‘criterion’. Earlier, in Tables 1 to 3, I provided what might be taken to be lists of criteria. Even if these are accepted as of value, and I have already indicated why they may well not be by many qualitative researchers, what should be clear is that these are lists of what we might call issues for threshold judgment: they remind the reader what to take account of, what they need to make a judgment about. However, they are not operationalised in such a way that anyone who did not already know how to make the judgments concerned would be able to apply them, nor do they tell the reader what would be sufficient in each case. And they do not tell us whether all of the thresholds have to be met for a positive conclusion to be reached, whether high scoring on one can counterbalance a lower ‘score’ on another, and so on. Moreover, I do not believe that they can be operationalised in this way, or that attempting to do this would be desirable.

Three things follow about the issue of criteria from what I have already said. First, there would need to be different sets of criteria depending upon what was being assessed (adequacy of research report, validity of findings, effectiveness of research methods, and so on). Secondly, less obvious but equally important, it is not possible to have purely formal criteria for assessing either research findings or research methods. What I mean by this is that there will necessarily be reliance on background knowledge, including tacit knowledge, and judgment. This is true in the case of assessing the validity of research findings, firstly, because both they and any evidence offered in support of them must be judged in terms of plausibility, which requires knowledge of what is already established as known in the relevant research field. Secondly, assessing the credibility of knowledge claims and evidence requires

---

11 Much the same is true of the framework for assessing qualitative research evidence produced by Spencer et al 2003.

knowledge, including tacit knowledge arising from experience, sufficient to make judgments about the likely level of error in the use of particular sources of data for particular purposes. Finally, in assessing research findings, criteria, or at least their application, must vary depending upon the form of inference being assessed and the intended research product.

It is not that either the substantive or the practical knowledge involved in making a validity assessment cannot be explicated, or that it does not need to be. It will have to be explicated within the research community, on some occasions, to some degree, because members of that community will not all share exactly the same expertise and experience. However, the explication will be for this audience, an audience where the chances of mutual understanding are, in principle, relatively high because everyone is engaged in similar kinds of research. Explication is likely to be much more difficult, though not entirely impossible, in relation to lay audiences. The situation will parallel a doctor’s attempts to explain to a patient why it is that she believes him to be suffering from a particular illness, and why she thinks the course of treatment she recommends is likely to be the most effective. Up to a point this sort of explication can be done, but there is always a substantial element of trust in professional judgment involved, whether we like it or not.

Now, one response to all this may be to ask why these are problems for qualitative researchers when quantitative researchers have been able to provide a limited number of formal criteria? My answer to this question is that the idea that there are formal criteria available for assessing either the findings or the methods of quantitative research is a myth. There is not even agreement about the meaning of the standard concepts usually appealed to in relation to quantitative work - validity (internal and external) and reliability - and furthermore they involve some important problems. Moreover, these standards are not transparent criteria. While reliability can be specified in terms of the results of reliability tests, these can take different forms, and there are no equivalent tests in the case of validity.

Furthermore, if we look at the criteria that are actually used, for example in systematic reviews, what we find is that these are exclusionary criteria. What I mean by this is that rather than being designed to include only those studies whose findings are likely to be valid and reliable, they are actually designed to exclude those which it is believed cannot be valid or reliable, for example because they did not involve random allocation to treatment and control groups, did not have a control group, had

12 For this argument, see Hammersley 1987 and 1991. There are also serious disagreements over the value of particular techniques - one example we might use here is the long-running debate over significance testing (see, for example, Oakes 1986).
too small a sample, and so on. Not only is it the case that what is allowed in by use of these criteria may still be defective, for reasons that are not covered by the criteria, but also the exclusionary criteria may even be wrong about what they exclude since they fail to take account of the functional relativity of evidence. And this neglect stems from an influential assumption built into much thinking about systematic reviews, to the effect that in reviewing a field of inquiry one is dealing with studies all aimed at answering the same question. Yet this kind of aggregation is a very impoverished view of the idea of the cumulation of knowledge.

It seems to me that these arguments raise serious problems for the specification of criteria for use both by researchers themselves and by lay people, such as funders and users. The problems are, first, that any comprehensive list is likely to be very long (given that in assessing the validity of findings, we will also need to evaluate the methods used, and to some degree even the competence of the researcher); and, secondly, that the criteria can serve as little more than reminders, they cannot be transparent, in the sense of being applicable with equal effectiveness by anyone, because of necessary reliance on background knowledge and expert judgment.

There is another issue worth mentioning in relation to the assessment of research findings by lay audiences, and in particular practitioner users. There is a tendency sometimes to assume that they should do this in the same way as researchers. In fairly abstract terms this is perhaps true. They too will need to make judgements on the basis of plausibility and credibility. However, there are likely to be at least two differences. First, in relation to plausibility, very often, what they take to be established knowledge will be different from that which is taken to be well established within the relevant research community. This is because they will have practical knowledge deriving from individual and collective experience. Secondly, in my view there ought to be a difference in how what is sufficiently likely to be valid should be judged. In the context of research this must be done in such a way as to err on the side of avoiding false positives rather than false negatives. In the context of practical action, by contrast, it seems to me that the threshold of acceptance for a belief will vary depending upon the likely cost of error (in other words its consequentiality) as well as according to the likely cost of errors in particular directions. Where error in a particular direction would be very costly, the danger of this will be minimised. Where it is not likely to be very costly, or is easily remediable, there will often be less concern to avoid this sort of error. This does seem to me to be a feature of how we all operate in everyday life, and it is rational. Indeed, it is essential to our ability to learn by trial and error.

This takes us into the area that, in my view, is most problematic as regards arguments for evidence-based practice (Thomas and Pring 2004). The difficult question is: how should practitioners weigh research evidence against the evidence from their own and others’ practical experience? Advocates of evidence-based practice are right to point out that sometimes practical knowledge is defective, perhaps even irrational. However, their critics are also correct in insisting that research evidence cannot substitute entirely for, or automatically override, practical knowledge, and more sophisticated advocates of evidence-based practice recognise this. The dangers here are obvious: either too great a deference to research or too much reliance on what is ‘obvious’ to any professional practitioner in the field. There is a genuine problem concerning how to integrate these different kinds of knowledge when we accept that neither can trump the other.

Aside from these problems, I wonder whether we should not posit a two-stage process, in which it is the responsibility of researchers, through the research community, to determine the validity of research findings in research terms. It is then the responsibility of practitioners to judge their validity and relevance for the purpose of using them in dealing with various practical problems. Of course, this suggestion does not address one of the issues that, I suspect, motivates the demand for formal criteria to assess research: the wish to make research publicly accountable, to ensure that funds are being efficiently and effectively deployed, and are seen to be so deployed. That is a very difficult issue. Personally, I do not believe that so-called transparent accountability is possible in relation to any occupation, nor that attempts to achieve it are desirable, since one of their effects is to undermine trust on the side of clients and personal responsibility on the side of practitioners. That does not, of course, mean that efficiency and effectiveness are unimportant, only that our judgments about these matters can only be rather uncertain, and especially in relation to research.

The demand for clear and definitive assessments of the value likely to be added by particular pieces or forms of research stems from a shift in orientation from a conception of research funding as sponsorship to one that assumes it must be judged

---

13 Feinstein and Horwitz 1997:531 note the ‘dismay’ of proponents of meta-analysis at realising that many randomised controlled trials in medicine are scientifically unsatisfactory despite a ‘gold standard’ status. This article provides illuminating insights into the problems of relying exclusively on data from randomised controlled trials even in medicine.

14 On this see Hammersley 2005 forthcoming.

as investment: investors must be able to judge the likelihood that they will get sufficient return. All I will say here about this is that the investment model simply cannot be applied to research without the latter being distorted into something else, such as the production of commercial or political propaganda.

Summary

In this paper I have outlined what I think is involved in assessing qualitative, and other kinds of, research, especially the validity of research findings. I argued that we should not simply accept the existing array of qualitative approaches, that there is a need to develop common ground. And I indicated, very briefly, how I think this should be pursued. More specifically in relation to assessing research I identified some quite different forms of assessment and argued that they involve diverse requirements. I also stressed the functional relativity of evidence. In the process I outlined what I see as the implications of the nature of the assessment process for demands that transparent quality criteria be specified for qualitative research.

References


