International Journal of Research & Method in Education
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713727792

The issue of quality in qualitative research

Martyn Hammersley [a]
[a] The Open University, UK

Online Publication Date: 01 November 2007
To cite this Article: Hammersley, Martyn (2007) 'The issue of quality in qualitative research', International Journal of Research & Method in Education, 30:3, 287 - 305
To link to this article: DOI: 10.1080/17437270701614782
URL: http://dx.doi.org/10.1080/17437270701614782

International Journal of Research & Method in Education
Vol. 30, No. 3, November 2007, pp. 287–305

The issue of quality in qualitative research

Martyn Hammersley*
The Open University, UK

This article addresses the perennial issue of the criteria by which qualitative research should be evaluated. At the present time, there is a sharp conflict between demands for explicit criteria, for example in order to serve systematic reviewing and evidence-based practice, and arguments on the part of some qualitative researchers that such criteria are neither necessary nor desirable. At issue here, in part, is what the term 'criterion' means, and what role criteria could play in the context of qualitative enquiry. Equally important, though, is the question of whether a single set of criteria is possible across qualitative research, given the fundamental areas of disagreement within it. These reflect divergent paradigms framed by value assumptions about what is and is not worth investigation. In addition, there are differences in methodological orientation: over what counts as rigorous enquiry, realism versus constructionism, and whether the goal of research is to produce knowledge or to serve other goals.

The value of qualitative research has recently come to be questioned again, after many years during which it was widely accepted. In large part, this has been stimulated by external criticism of educational research, on the grounds that it does not serve evidence-based practice well.[1] There are those who argue that much qualitative research is of poor standard, but more usually the complaint is that there is no clearly defined set of quality criteria available for judging it, so that it is of *uncertain* quality. There are two assumptions underpinning these criticisms that need to be addressed. First, what is involved here is an at least implicitly *comparative* assessment: it is assumed that clearly defined criteria of quality are already available for quantitative research.[2] The other assumption is that explicit assessment criteria are needed, and two main reasons seem usually to be involved. First, it is believed that unless researchers operate on the basis of such criteria, their work will be of poor quality. Secondly, it is argued that users of research require some reliable means of judging its quality, and that a set of criteria would meet this need.

*Faculty of Education and Language Studies, Level 3 Stuart Hall Building, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK. Email: m.hammersley@open.ac.uk

ISSN 1743-727X (print)/ISSN 1743-7288 (online)/07/030287–19

DOI: 10.1080/17437270701614782

In this article, I will focus on the second set of assumptions, examining some of the arguments about whether criteria are necessary, before going on to consider whether a single set is possible given the present state of qualitative enquiry.

Are criteria necessary?

Whether there are criteria by which qualitative research can be judged, and if so what character these should have, are issues about which there has been much debate but little agreement. There have been many attempts to identify criteria.[3] Some writers have tried to apply what they see as traditional quantitative criteria, such as validity and reliability, to qualitative work. Others have reformulated these epistemic criteria and, very often, added non-epistemic ones, whether in terms of 'giving voice' to the marginalised or bringing about practical or political effects of some kind—improving education, challenging injustice or oppression, and so on. At the same time, there are some writers who appear to reject the very possibility of criteria, at least as conventionally understood. For example, many years ago, John Smith argued that 'there can be no criteria to reconcile discourse (to sort out the trustworthy from the untrustworthy results)' (Smith, 1984, p. 384; see also Smith, 1989b and Schwandt, 1996). Smith claimed that attempts to apply criteria to qualitative research would inevitably result in confusion and inconsistency, because criteria are incompatible with the basic philosophical assumptions of this type of enquiry. More recently, he has insisted that 'criteria should not be thought of in abstraction, but as a list of features that we think, or more or less agree at any given time, and place, characterize good versus bad inquiry'. And, in addition: 'This is a list that can be challenged, added to, subtracted from, modified, and so on, as it is applied in actual practice—in its actual application to actual inquiries' (Smith & Deemer, 2000, p. 894). I suspect that, despite appearances, Smith's views have not changed much over the period concerned; the difference between the two formulations hinges on what is meant by the term 'criterion'. And this is the nub of the matter.

We can provide two contrasting definitions which illustrate the scope for disagreement about this. At one end of the spectrum, 'criterion' means an observable indicator that can tell us (along with other criteria) whether or not the findings of a study are valid, or are of value in more general terms. In other words, a set of criteria here amounts to a checklist that is sufficiently explicit and concrete in its reference for it to be filled in with little error; and also comprehensive enough to cover everything that needs to be taken into account in judging quality. Very often such criteria are seen as relating to the procedures employed in carrying out the research, on the implicit assumption that these determine the validity of findings. Towards the other end of the spectrum are the kinds of criteria outlined by Smith and Deemer: a list of considerations, never fully explicit, that it is agreed in local circumstances should be taken into account in judging qualitative work, a list that can serve as no more than a reminder and that is always open to revision in the process of being used—indeed, which only gains meaning in particular contexts.[4] It seems to me that the second definition comes closer to what is possible and desirable than the first, though it is not entirely satisfactory.

The best place to start in thinking about the nature and role of assessment criteria is with how researchers actually go about judging quality in doing their work. It is probably beyond dispute that this does not involve simply applying a set of explicit and concrete rules of assessment. Instead, they make specific, and often largely implicit, judgements. However, to some extent they can, and will often, explicate these judgements in terms of methodological principles. This picture is as true of quantitative as of qualitative researchers, whatever other differences there are between them. Moreover, I suggest that this reliance on judgement is inevitable. The task of judging quality in the context of a relatively complex activity like research cannot be sensibly reduced to the application of explicit, concrete and exhaustive indicators. Instead, formulations of criteria, in terms of considerations that might need to be taken into account, *come out of* the process of judgement and are modified by it; to one degree or another. Of course, they also feed back into it, in that they are used subsequently by researchers in judging quality in other situations, but in this process they always have to be interpreted and may be *re*-interpreted. Furthermore, they will be applied selectively, depending upon the nature of the knowledge claims and research involved.

While I am rejecting the idea of a finite set of explicit and exhaustive criteria that can substitute for judgement, or render its role minimal, the above discussion indicates my belief that criteria, in the form of guidelines, can play an important role in the work of researchers. They may facilitate reflection on previous judgements that enables us to learn from our own experience, and from one another. Such learning may also involve exploring the implications of applying locally used criteria in new contexts, and considering the extent to which there are commonalities across fields. This is important because such reflection can lead us to conclude that locally used criteria need to be changed or developed, even as regards their original context. In other words, it seems to me that changes in criteria arise not just, as Smith and Deemer suggest, from the practical work of particular researchers in particular situations but also from attempts to universalise and systematise current criteria. The aim here is not to produce a single, universal system of assessment rules that will eliminate judgement; it is rather that the process of thinking about how the criteria would apply in other contexts and how they relate to one another is a productive force in the development of research practice.

In a much more mundane way, but no less importantly, criteria, in the form of guidelines, can remind researchers of what they ought to take into account in assessing their own and others' research. This is not a small matter. When engaged in any complex activity, it is easy to overlook what one would, in other circumstances, routinely take into consideration. Guidelines may also facilitate students' learning how to assess research, though here especially care must be taken to avoid the implication that stated criteria can in some sense be a substitute for acquiring the practical capacity to make sound judgements.

However, as noted earlier, much of the pressure for qualitative criteria comes not so much from the context of *researchers* judging research, or even students learning to do this, but rather from that of lay 'users' of research (notably policymakers and practitioners) assessing its quality. It might seem, on the basis of the above discussion, that

the problem here is that, by contrast with researchers, lay people will not have the background knowledge, skill and experience necessary to judge quality: so that they will be overly reliant upon the guidelines. This is true in one sense and false in another. And this ambiguity points to the crucial problem of the interface between research and other sorts of specialised activity. It is true that lay users will not have equivalent knowledge of other research, or the experience of and practical capacity for making judgements about research that comes from practising it. As I have indicated, no set of criteria can substitute for this. On the other hand, lay users of research are not simply ignorant or lacking in all skill in assessing knowledge claims for relevance and validity. Indeed, this capacity is an essential element of many forms of activity, indeed of everyday life generally. The problem faced by lay users, then, is twofold: how to assess research findings in the terms in which these present themselves; and how to relate these findings to what they already know and how they normally judge validity. There seems to be a built-in tendency here for extreme reactions: for lay users either to adopt too reverent an attitude towards the products of research or simply to dismiss these if they conflict with what is taken to be existing knowledge. Oscillation between these extreme positions is not uncommon, and can be difficult to avoid.

Any set of criteria offered by researchers could only deal with the first of these tasks—how to assess findings in research terms—it could not help with the problem of relating them to practical, experiential knowledge; not least because the latter will be highly variegated. And it could only help to some extent even in relation to this first task: to repeat the point, no set of guidelines can remedy lay users' lack of practical experience of research. The implication of this is that, to a considerable extent, they must trust researchers: they must assume, unless there is evidence to the contrary, that the latter have carried out their work effectively, that they have presented their findings honestly, that the normal processes of criticism within the research community have reduced the danger of error, and so on. Whether researchers *deserve* such trust depends upon how they operate, but there is no mechanism that can check independently that the knowledge they produce is sound, or even that they have pursued this knowledge in the most effective way possible.

The need for trust cannot be eradicated in this context, any more than it can be in other forms of contact between specialised practitioners and lay people. As already hinted, trust does not have to be blind trust, it may be possible for lay users to make some judgement about the coherence and reasonableness of what researchers say, but they must take care not to dismiss what does not fit their current frame of reference for that reason alone. There is clearly a dilemma here. On the one hand, the whole point of research is that it may produce un-commonsensical conclusions, and therefore its findings must not be dismissed because they are counter-intuitive. On the other hand, people's experience and background knowledge will often give them important resources with which to interrogate, modify, or reject research findings; especially in terms of how these relate to particular contexts of action with which they are familiar, and what their implications are for practical action. So, good judgement is required by lay users, just as much as it is on the part of researchers, and it is even more difficult to exercise in their case.

In short, I am arguing that it is not possible for researchers to make their judgements transparent, in the sense of fully intelligible to *anyone*, irrespective of background knowledge and experience. Indeed, there are limits to the extent to which these judgements can be made intelligible even to fellow researchers, because of the situated nature of judgement. Certainly, it is the case that such intelligibility is an *achievement*, it is not automatic: speakers need to be able to formulate the situation, the reasons for making the judgements that they did, and so on, in ways that facilitate understanding; and, equally importantly, the audience must be able to draw the right inferences from what is said, on the basis of the background resources they have. The greater the experiential distance between speaker and hearer, the larger this problem of communication will be. And because the use of guidelines always depends upon background knowledge and judgement, they cannot solve this problem even if they can serve as a useful resource in dealing with it.[5]

Indeed, guidelines can exacerbate the problem where lay audiences interpret them in ways that are at odds with the practical expertise of researchers, or treat them as if they could substitute for that expertise. And there are current circumstances that encourage this. The demand for qualitative research criteria that lay users can employ has arisen, to a large extent, in the context of the evidence-based practice movement (Trinder with Reynolds, 2000; Thomas & Pring, 2004) and what has been termed 'the new public management' (Pollitt, 1990; Ferlie *et al.*, 1996; Ridley, 1996; Clarke & Newman, 1997). A central theme here is 'transparency', for example it is demanded that the basis on which professionals work should be made explicit, so that the lay people who use their services can judge the quality of what is provided.

There are two ways in which social and educational research is caught up in this. First, it is seen as capable of supplying the evidence on which more effective policy and practice can be based: it is regarded as offering the impersonal evidence that is essential in order to hold professional practice to account. But, and this is the crucial point in this context, in order for it to do this it must itself be transparent, so that lay users can determine which research findings can and cannot be relied upon. Fortunately or unfortunately, depending upon one's point of view, the idea that research can serve this function is a mirage. For reasons already explained, and they parallel those that apply to other forms of professional practice, there is no possibility of transparency; in the sense of lay people being in a position *consistently* to make judgements, about the quality of particular research studies, that are equally as good as those of researchers who work in the relevant field.

To a large extent, then, I think that qualitative researchers are right to resist demands for a fully explicit set of evaluation criteria: the nature of research (in general, not just of qualitative work) is such that these cannot operate in the way that is frequently assumed by those who demand them. And the same is true as regards lay use of research findings, here too good judgement is required and this cannot be reduced to explicit rules.

Even so, I believe that qualitative researchers need to give much more attention than is currently done to thinking about the considerations that must be taken into account in assessing the likely validity of knowledge claims, exploring the consistency

of these with one another, and considering how they apply in other situations from those in which they were generated. This is not so much a matter of producing lists of criteria, in the sense of considerations that might need to be taken into account, useful though that undoubtedly is, but rather of collectively developing and talking about what is involved in assessing the likely validity of findings from particular studies. Currently, however, there is a major obstacle to this process: the plurality of approaches to qualitative research that now exists, and the way that these are often regarded as incommensurable paradigms. It is to this issue that I now turn.

Is a single set of qualitative criteria possible?

It is an obvious fact about qualitative research today that it is divided not just in terms of substantive focus, or even according to the use of particular methods, but by divergent theoretical, methodological and value assumptions: about the nature of the phenomena being investigated and about how they can and should be researched.[6] In other words, there are different, and to some extent competing, 'paradigms'. The key questions that this raises are, of course, whether it is possible and desirable to overcome these differences, and if so how this can be done.

The term 'paradigm', in the sense being used here, derives, of course, from the work of the historian and philosopher of science Thomas Kuhn. There has been much discussion of what Kuhn meant by 'paradigm', and it is clear that he used the word in a variety of ways.[7] However, his core argument was that natural scientists working within a mature field operate with a set of assumptions, about the relevant region of the world and how to understand it, that are embodied in their use of certain studies or theoretical models as exemplars; in other words, they take these for granted as the basis for further work. Of course, there are always things that do not seem to fit into their current understanding, and the task of what he called 'normal science' is to treat these as puzzles that can be solved *without modification to paradigmatic assumptions*. However, Kuhn famously argued that, over time, some puzzles turn out to be anomalies that cannot be resolved within these limits; and that, at some point, an alternative paradigm will arise that takes over the field, because it is able to explain all of what the previous paradigm covered and also resolves the anomalies that had arisen within it. This is the process of scientific revolution.

Now there have been many disputes about the accuracy of Kuhn's account of natural science, but the task of applying it to social science, including to educational research, is even more complex and uncertain. Kuhn regarded the social sciences as pre-paradigmatic, at best, and therefore as not characterised by competing paradigms, in his sense of the word. And, even were we to ignore this, there is very little social and educational research that approximates to his notion of normal science; instead much of it, over the past three decades, seems to have been in a state of continual 'revolution', albeit of a kind that is not recognisably scientific in Kuhnian terms.

At the same time, it probably *is* the case that social scientists, including educational researchers, operate on the basis of exemplar studies and models, and that this is part of the reason why they often adopt rather different approaches to the same topic, and

why they sometimes disagree so sharply about the nature of research. Read correctly, Kuhn's work is an important counter to any tendency to adopt too cognitivist a conception of conflicting approaches amongst researchers. It is not that people first acquire epistemological and ontological assumptions and then decide how they are going to investigate the social world. Rather, they acquire particular research practices and various methodological and philosophical assumptions, consciously and unconsciously, more or less simultaneously, and each shapes the other. This means that the differences among qualitative researchers are embedded in diverse forms of situated practice that incorporate characteristic ways of thinking about the research process. Furthermore, we are not so much faced with a set of clearly differentiated qualitative approaches as with a complex landscape of variable practice in which the inhabitants use a range of labels ('ethnography', 'discourse analysis', 'life history work', 'narrative study', 'activity theory', 'interpretivism', 'feminist standpoint epistemology', 'postmodernism', and so on) in diverse and open-ended ways in order to characterise their orientation, and probably do this somewhat differently across audiences and occasions.

So what is involved in paradigm conflict is not simply a clash of ideas about what form social and educational research ought to take; but, rather, a divergence in *practices*. I can develop one aspect of this by drawing on the methodological arguments of the sociologist Max Weber. He inherited the view, from neo-Kantianism, that social science, by contrast with natural science, must be idiographic rather than nomothetic in orientation. What this means is that social scientists are primarily interested in understanding particular social phenomena in their socio-historical contexts, rather than in discovering universal scientific laws. Moreover, by contrast with the neo-Kantians, Weber believed that in identifying and conceptualising particular phenomena for study, social scientists could not draw on objective or eternal values to define what is worth investigation. Rather, value-relevance is perspectival: there are discrepant value perspectives, which give different aspects of the social world more or less importance, and among which there is no rational basis that can determine choice. So, social scientists can adopt a range of different value perspectives in order to identify relevant phenomena for investigation, though they need not believe, and certainly should not claim to have established through research or on any other basis, that the one they have adopted is the only legitimate or true one.

The point of this is that if, in order to identify phenomena that are worth investigating, we necessarily draw on some set of values, to which there are always alternatives—and if, as seems likely, what are involved here are not only values but also factual assumptions about the nature of the phenomena concerned—then there will inevitably be significant lines of division within educational research. What I mean by this is that there will be forms of research within this field that work within discrepant frameworks of assumption.[8]

Now, the reason why I employed the work of Weber to make this point is because he did not believe that social science research is necessarily, even less that it *should* be, political; in the sense of being geared towards political goals. There are quite a lot of qualitative researchers today who insist that research ought to be political in this sense: that it should aim at the eradication of social inequality of various kinds, that

it should improve the lives of some group of people (children, for example), that it should serve the goals of education or of some political cause.[9] It is not hard to see that this is likely to lead to incommensurable paradigms reflecting sharply divergent orientations; unless one holds to the enlightenment idea that there is a single all-embracing conception of 'the good' that will be recognised (at least eventually) by everyone. Weber certainly did not believe this, and this was why he thought social research should aim to be value neutral: that practical values (values other than truth) should not constitute the goal of social science enquiry, *even though they are needed to provide the value-relevant framework in terms of which the phenomena to be studied are identified*. My point, then, is that if—even from Weber's point of view, rather than from a position which assumes that all social science is inevitably political (in the broadest sense of that word) and therefore necessarily divided by allegiances to discrepant worldviews—we can see how educational research could be characterised by competing paradigms (in a non-Kuhnian sense), then we perhaps should resist any inclination to dismiss paradigm differentiation in social science as entirely the product of bias, theoretical or methodological fashion, career-building, etc., as is sometimes done.

Weber thought that value relevance is not a matter of partisan definition; he believed that those who fundamentally disagree about an issue could nevertheless agree about its importance and what is relevant to it. Indeed, he argued that research could play a role in facilitating the resolution of disagreements through both providing relevant factual evidence and clarifying the implications of different value positions; though he recognised that there was no guarantee that it could completely resolve any disagreement, since there are ineradicable conflicts among fundamental values. It is equally important, though, that these conflicts can also generate differences in view about what are and are not important issues, or about the priority of different issues. As a consequence, there may be research topics that are believed to be important by some researchers but are regarded as pointless or as positively harmful, or at least as much less important, by others. Let me try to illustrate this with an example from the field of education.

There is a body of research on gender inequalities in children's participation in school classrooms, both in terms of level and kind of participation.[10] In order to accept that this is a worthwhile topic of enquiry, one must, first of all, believe that there should be equality between the sexes in some sense of that term (and this might not be accepted by the radical religious right or even by some radical feminists, both of whom believe that there are fundamental differences in orientation between males and females, though of course they conceive of these in discrepant ways). In order to see this topic as important, one probably also has to assume that particular kinds of classroom inequality are consequential for educational achievement; and, in addition, one must value highly educational achievement, defined for example in terms of examination success. Yet, there are those who question whether gender differences, or those of other kinds, in classroom participation have determinate and consequential effects on outcomes; and there are others who deny that examinations measure education and/or who insist that examinations are at odds with true education. What

I am trying to show, then, is that research in this field is framed by a range of both value and factual assumptions that are open to reasonable disagreement. What is involved here is not an instance of absolute 'incommensurability' (a problematic term even in the context of Kuhn's work) between points of view; it is rather that the fewer of the assumptions underpinning a field of research people share the more difficult it will be for them to understand the point of the research, and the less inclined they will be to see it as of value. Moreover, while there is scope for persuasion, so that one might come to regard the significance of a particular form of research one did not previously value, this is likely to require arguments about values not just about facts. And, as a result, there is little guarantee that even lengthy discussion will produce a consensus.

So I am arguing that some of the variation to be found within educational research derives from the fact that it operates in diverse value-relevance niches. And I believe that these niches have consequences not just for judgements about what sorts of work are and are not worthwhile in terms of relevance, but even for judgements of validity. This is because the threshold that knowledge claims must reach, in terms of likely validity, if they are to be accepted, is determined by relevant research communities—it is not something that can be laid down by the philosophy of science or by any central authority. And their judgements about this may differ sharply because of their varying background assumptions (Hammersley, 2000).

Besides this variation in orientation arising from researchers operating in different value-relevance niches, there are also some more abstract sources of paradigm differentiation. Moreover, these two aspects of methodological pluralism are not, of course, completely independent: those working within particular niches draw on, modify, develop, and perhaps even misuse, more general ideas for their own purposes. They also sometimes portray educational research in general terms, occasionally as if it was simply their own work writ large, thereby contributing to the intellectual resources that others may draw on in facilitating and justifying work in other niches.

What I want to do in the rest of this article is to discuss three issues, about which there are sharp disagreements within the literature of qualitative research methodology, that have important implications for judgements about quality. These relate to:

- different conceptions of scientific method and rigour;
- the conflict between constructionism and realism;
- the relationship between research and various other kinds of practical activity, including politics.

Scientific method and rigour

Natural science was the methodological model for most social research, including that in education, for much of the twentieth century. Conceptions of scientific method determined what counted as rigorous investigation. However, this did not prevent considerable disagreement about how social and educational research ought to be pursued. There are several, interrelated, reasons for this.

First, different natural sciences were sometimes taken as exemplifying scientific method: for example, some researchers treated physics as the premier science; others took nineteenth-century biology (in particular, botany) as a more appropriate guide. Furthermore, there were also disciplines broadly within the social sciences that were treated by those working in other fields as exemplifying scientific method. Key examples here are behaviourist psychology, neo-classical economics, and structuralist linguistics. The point is, of course, that different exemplars of science or rigour can lead to very different prescriptions and forms of assessment.

Second, there were divergent philosophical interpretations of the 'method' employed by natural scientists. Broadly speaking, moving from the nineteenth into the twentieth century there was a shift from an inductivist conception of enquiry, in which scientific laws were logically derived from observation of repeated patterns of occurrence, towards one which stressed the testing of hypotheses deduced from theories that were necessarily a product of speculative thought and that perhaps could never be proven to be true, only falsified at best, through being tested against evidence. Quite a lot of the conflict between social scientists promoting quantitative and those adopting qualitative methods in the first 70 years of the twentieth century stemmed from commitment to different notions of scientific method along these lines, with qualitative researchers tending to adopt an inductivist approach, loosely defined.

Finally, there was variation in views about the degree and character of the differences between physical and social phenomena, and what the implications of these were for how far the methods of natural science needed to be modified if they were to be applied in the social field. At one extreme, there were those who took physics as exemplifying scientific method, and who believed that there were no distinctive features of social phenomena, or that if there were these did not stand in the way of rigorous measurement and control of variables. By contrast, others insisted that social phenomena had to be approached quite differently from physical phenomena. Indeed, it was sometimes argued that whereas the latter could only be studied from outside, social phenomena could and should be understood from within, so that a deeper form of knowledge was available. And it was claimed that this inner understanding required the researcher to draw upon his or her psychological and/or cultural resources to grasp the meanings that informed the actions of the people being studied, since these meanings are crucial for what it is that people set out to do, and why. In particular, it was insisted that these meanings cannot be inferred from external behaviour. To use Clifford Geertz's example: there are significant differences between a facial tic that forces the closure of one eye, a wink, someone pretending to wink, and someone practising winking; and it is not possible to infer from physical behaviour alone which of these is taking place (Geertz, 1973, p. 6). Given this, some other, or additional, means of access must be secured to the cultural meanings that inform people's behaviour if we are to be able even to describe it accurately, let alone to explain it. All this did not necessarily lead to a rejection of science, but rather to the construction of a distinctive conception of science, one that was particularly suited to studying the social world.

Now qualitative researchers today vary in whether they see their work as scientific and, if they do, in what they take this to imply. And, aside from this, there are significant differences concerning what is regarded as possible or legitimate in epistemic terms. A recent illustration is what has been called the 'radical critique of interviews' (Murphy *et al.*, 1998, pp. 120–123). This does not just raise questions about over-reliance on interview data—a common complaint, for example, on the part of ethnographers who, in the past at least, tended to stress the centrality of participant observation. What makes recent criticisms of interview research radical is that they challenge the use of this type of data *as a window into the minds of informants* and/or *as throwing light on the social worlds in which informants live* (Dingwall, 1997; Silverman, 1997; Atkinson & Coffey, 2002). In other words, the critics deny that interviews can tap stable attitudes or perspectives that govern people's behaviour beyond the interview situation; and/or that they can be a sound source of witness information about what happened, or what happens, in particular settings or in the world more generally. In effect, the critics reject the two main uses to which social scientists have put interview data in the past, and the ways many qualitative researchers continue to use them today (Hammersley, 2003a).

It is worth pointing out that the radical critique of interviews challenges a conception of rigour that is characteristic of some forms of ethnography—for example that exemplified by interpretive anthropology—and replaces it with a view of rigour that is characteristic of many forms of discourse analysis. Discourse analysts often insist that the data must be presented to readers so that the latter can assess directly the validity of the inferences made, and perhaps also that inference must not range beyond what is 'observable' in the data. By contrast, ethnographers often argue that it is not possible for them to make all their data available to readers, that the validity of their inferences depends upon the success with which they have learned the culture of the people they are studying, and thereby become able to interpret accurately what meanings various phenomena have for them. Moreover, some would question whether discourse analysts are actually presenting, or can present, 'the data' to readers; given that transcriptions involve theoretical assumptions about language use, that how these are read will depend upon cultural background, and so on (see Hammersley, 2003b).

So, without even touching on post-structuralism and postmodernism, the usual bogeymen in this context, I have tried to show that there are divergent conceptions of the requirements of rigorous enquiry to be found amongst qualitative researchers. And it is not difficult to see how these can lead to disagreement in judgements about what is good quality work.

Realism versus constructionism

Constructionism is one of the sources from which the radical critique of interview analysis arose. The generic move of constructionism against realism is to insist that social phenomena do not exist independently of people's understandings of them, that those understandings play a crucial generative role. Of course, many realists

would accept, indeed insist, that social phenomena are the product of 'people acting together' rather than entirely a result of social forces operating beyond their control. What is distinctive about constructionism, as I am using that term here, is that it takes the argument that social phenomena are culturally constituted and draws from it the conclusion that these phenomena can only be understood by describing *the processes by which* they are culturally constituted as the things they are. In other words, a fundamental re-specification of the goal of enquiry is required. The focus thus becomes, not the phenomena themselves, and certainly not what might have caused them or what effects they have, but rather the processes by which they are identified and in effect created through identification by culture members. For example, in focusing on home–school relations, rather than studying families as groups interacting within and beyond the context of their homes, the focus may become how people talk about family relationships, both explicitly and perhaps even more importantly implicitly: how they use notions of family form, ideas about kin obligations, etc. in the course of their interactions with one another; thereby creating what they 'find' to be, for example, functional or dysfunctional families. So, for constructionists, notions like 'family' are not descriptions of patterns of social relations that exist independently of them; the social significance of such concepts lies in their functional use rather than in any representational capacity.

It is perhaps worth emphasising that it is not just the realist/constructionist divide that generates divergences here in how research is assessed. There are important differences in orientation *within* constructionism. One version is modelled on ethnomethodology (see Zimmerman & Pollner, 1971).[11] It suggests that there are means whereby the constitutive processes by which social phenomena are ongoingly produced can be uncovered or displayed in a manner that does not involve any cultural interpretation or inference on the part of the researcher. All that is involved, it is claimed, is description, in the sense of explication. The terms of this explication are entirely those of the culture embodied in the actions being explicated. This position does not imply any form of epistemological scepticism; in other words, it need not be denied that the actions and institutions people produce through their actions have a real existence as particulars. What *is* denied, though, is that these can be grouped into natural kinds in any other way than in terms of the processes that constituted them. Furthermore, we must remember that there is no one-to-one correspondence between some constitutive cultural notion, like family form, and the actions it is used to produce and how they interrelate with other actions. This is because such notions may be 'honoured in the breach', 'stretched' in various ways, joked about, and so on.

A quite different way of operating, characteristic of radical forms of constructionism, is to treat researchers as themselves necessarily engaged in constituting the social world, or particular social phenomena, through the writing process; rather than simply describing or displaying how others construct the social world through their discourse or actions.[12] This often leads to a blurring of the boundary between social research and imaginative literature. In fact, from this point of view, the whole of conventional educational research cannot but be, in effect, simply a form of fiction which is falsely conscious of its own character.[13]

These two approaches, which by no means exhaust the kinds of work constructionism has generated, imply very different modes of assessing research from one another, as well as from realism. For the first, the concern, presumably, is with whether the descriptions of the constitutive processes of social life are accurate, in other words whether what is displayed is indeed these processes. By contrast, for the second position, any idea of validity in this sense is rejected, and the accounts produced by researchers must be judged in non-epistemic terms, according to aesthetic, ethical and/or political criteria. One way of thinking about the shift from realism to this latter kind of constructionism is in terms of the replacement of Kuhn by Rorty as the patron philosopher of many qualitative researchers. Whereas Kuhn still sees natural science as engaged in a process of enquiry, in which knowledge is accumulated, albeit in a discontinuous rather than a continuous way, Rorty abandons the residual realism to be found in Kuhn's account. In effect, he erases the distinction between enquiry, which is concerned with gaining knowledge, and conversation, conceived as guided by an interest in 'edification'. Indeed, he treats enquiry, understood in its conventional way, as labouring under a mistaken conception of itself, one that assumes that it is possible (and desirable) to claim superior knowledge of reality.[14]

There are many varieties of both constructionism and realism, but I hope this discussion gives some sense of the implications that such a contrast in orientation can have for judgements about the quality of particular pieces of research. The point is that these amount to fundamental differences that it may be impossible to bridge.

Activism: the relationship of research to politics, policymaking and practice

What comes under the heading of 'activism', like that of 'constructionism', is quite diverse in character. But my point is that there are various qualitative approaches that explicitly reject the idea that the production of knowledge should be the only immediate goal of enquiry—in place of what they see as mere contemplation they insist on action.[15] Advocates of some of these approaches believe that research should form an integral part of educational practice, and argue that it is rendered useless, or at least debased in value, when separated out from this.[16] Indeed, there are those who claim that any process of institutionalisation, whereby knowledge production is concentrated in special institutions, such as universities, is now being reversed through the rise of a new mode of knowledge production which occurs in the context of and directly addresses practical problems of various kinds (Gibbons *et al.*, 1994; Gibbons, 2000). Other approaches insist on the *political* character of research, requiring that it must be explicitly directed towards bringing about change of one sort or another: challenging capitalism, patriarchy, racism, the social conditions that generate disability, homophobia and so on.

Now, what is relevant here is that all these approaches introduce extra or alternative considerations in judging the quality of research, additional to the traditional epistemic ones concerned with the production of knowledge. The nature of these criteria varies, of course, depending upon the form of practice to which research is to be tied, or with which it is to be integrated. In some cases the criteria are educational

in character, but they may also be political, ethical, aesthetic or even economic—for example concerned with whether there is demand for the knowledge being produced, whether the research offers value for money, etc.

In their early forms, these activist conceptions of research treated the goal of producing knowledge as compatible with, or even conducive to, the pursuit of practical goals. For example, in its original form, Marxism conflated epistemic with other considerations, on the basis of the Hegelian assumption that the development of knowledge and the realisation of other ideals are strongly interrelated in a historical dialectic. More recently, that assumption has been abandoned by 'critical' researchers, and as a result practical or political goals have often been raised above, or have replaced, epistemic ones. This has usually been done on the argument, associated with some kinds of constructionism, that the pursuit of knowledge, in the conventional sense of that term, is futile. Alternatively, there are arguments for interventionism that dismiss traditional kinds of qualitative research on ethical grounds as a form of voyeurism.[17] There are also arguments for intervention that are aesthetic rather than either scientific or ethical in character. Here research is to be a form of performance art, indeed it may be argued that it cannot but be this, and the requirement is that it be consciously shaped aesthetically, at most only hiding behind the masks of science or ethics in an ironic, knowing way.[18] It is not difficult to see that the criteria relevant here would be significantly different again.

Concluding thoughts

In this article, I have explored some of the issues surrounding the idea that a single set of assessment criteria for qualitative research is possible and desirable. I raised questions about what the term 'criteria' refers to, but suggested that guidelines can be desirable, so long as they are not seen as a substitute for the practical capacity to assess research. In the second half of the article, I argued that some fundamental differences generate the methodological pluralism that is characteristic of qualitative research today, and that this has profound consequences for any attempt to identify a single set of guidelines for qualitative enquiry.

There have been various responses to this methodological pluralism, including the following:

1. The argument that the apparent differences are spurious, perhaps amounting simply to rhetoric: that when it comes to actually doing research there is much less variation than the methodological debates suggest. From this pragmatist position, the wilder flights of methodological fancy may be regarded as of value as antidotes to traditional prejudices, but it will usually be insisted that one should not take methodology 'straight' or in large doses.[19]
2. There are deep-seated incompatibilities in approach, and we must simply recognise that diverse forms of social research will develop, each having its own distinctive conception of what is good quality work, and its own ways of applying this. These differences must be respected.[20]

3. There are serious differences in perspective, but some means need to be found to at least reduce them, so as to increase the level of agreement across educational researchers' judgements about what is and is not good quality work.[21] Developing guidelines may serve a useful function in this.

My position is, broadly speaking, the third. However, the sources of division I have identified in this article vary in character and in their implications for assessing quality. The first—diversity in value relevancies—is probably not a serious barrier to generating a common set of guidelines, so long as it is recognised that these cannot be used as a fully explicit and exhaustive checklist, and that somewhat divergent guidelines may be required in different substantive fields. As regards conflicting views of what counts as rigour, I suspect that (at least in the longer run) these are open to resolution through reflection on the research process, since different conceptions of rigour must be judged in practical terms according to their productivity. A rather different strategy may be required as regards the realism–constructionism divide: this may be resolvable by treating it as, to a large extent, a disciplinary difference; in other words as resulting from a fundamental difference in focus. In these terms, there is no need for competition or conflict, so long as neither side formulates its position in such a way as to render the alternative approach illegitimate, and that each side is suitably modest in the claims it makes for itself. After all, one would not accept the same assessment criteria to hold across disciplines.

Much more difficult to deal with, it seems to me, are the differences generated by what I have referred to as activism. As we saw, this either requires research to supplement its goal of producing knowledge with other goals, or proposes the substitution of these for the pursuit of knowledge. The former position is likely to render research incoherent, unable to pursue any goal effectively, least of all the production of knowledge; while the latter amounts to the abandonment of research in favour of some other activity (Hammersley, 2000). Furthermore, aside from the problem of reconciling activist and non-activist conceptions of research, we should note that there is likely to be little agreement as regards assessment criteria even *among* activist researchers, given that they can be motivated by very different political or practical enterprises. Here, it seems to me, we need collectively to define exactly what can and cannot count as research; but how one could persuade those whose main commitments lie elsewhere that its only immediate goal should be the production of knowledge is far from clear.

My conclusion is that guidelines for qualitative research are desirable, for the reasons I outlined at the beginning of the article. However, the barriers to our being able to produce any set of common guidelines, even among qualitative researchers, are formidable. At the same time, we should not simply accept methodological pluralism at face value, reinforcing it by treating each qualitative approach as having its own unique set of quality criteria. Dialogue on this issue across different approaches, and indeed across the qualitative–quantitative divide, is essential for the future of social and educational research.

Notes on contributor

Martyn Hammersley is Professor of Educational and Social Research at the Open University. His early research was in the sociology of education. Later work has been concerned with the methodological issues surrounding social and educational enquiry. His most recent books are: *Taking sides in social research* (Routledge, 2000); *Educational research, policymaking and practice* (Paul Chapman, 2002); and *Media bias in reporting social research? The case of reviewing ethnic inequalities in education* (Routledge, 2006).

Notes

1. For accounts of the history of these challenges to educational research, in the UK and the USA, see, for example, Hammersley (2002) and Biesta (2007), pp. 1–5.
2. I have suggested elsewhere that this is open to doubt (Hammersley, 2007).
3. These are lucidly reviewed and consolidated by Spencer *et al.* (2003). See also Altheide and Johnson (1994); and the recent discussion by Furlong and Oancea (2005) of criteria for assessing applied and practice-based research. An assessment of the latter is provided in Hammersley (forthcoming). There has also been much discussion of the issue within the context of the ESRC's Teaching and Learning Research Programme, focusing in particular on the issue of 'warrant': see, for example, James *et al.* (2005).
4. There could be a position that is beyond that of Smith and Deemer (2000), a position where the whole point of criteria is to transgress them! Also near this far end of the spectrum would be Feyerabend's rejection of method (Feyerabend, 1975).
5. Here I am of course abstracting away from the, very real, possibility that the judgements of both researchers and 'users' will be biased by their values, preferences, and interests.
6. I am not suggesting that quantitative research is entirely homogeneous; it is not. Experimental and survey research have conflicting orientations in some key respects. However, the difference does not lie at as deep a level as the differences in orientation among qualitative researchers.
7. For excellent recent discussions of Kuhn's work, see Hoyningen-Huene (1993), Bird (2000) and Sharrock and Read (2002).
8. For an account of some of this diversity in the field of educational research, see Hammersley (2007).
9. For discussion of these orientations, see Hammersley (2000, 2004a).
10. For a more detailed discussion of work in this area, see Hammersley (1990). I have specifically avoided the other fields that have become almost standard examples in demonstrating that research can be motivated by controversial values: the case of research on racial differences in intelligence and that of research on 'effective schooling'. For a recent discussion of the latter that illustrates my point, see Clark (2005).
11. Empirical work inspired by this programme in the field of education has, for example, involved detailed analysis of classroom processes: see Baker (1997).
12. To a large extent, this arose from structuralism and the various moves beyond or behind it that have come to be labelled 'post-structuralism'. There are significant parallels, as well as important differences, between ethnomethodology, inspired by phenomenology, and structuralism.
13. There are few examples of this approach within education. For what is probably the most entertaining exemplar in social science generally, see Ashmore (1989).
14. The issues involved here are complex ones. See Putnam's (2002, pp. 99–100) criticisms of Rorty's position.
15. The term 'activism' is not entirely satisfactory, not least because it might seem to imply an acceptance of the arguments of activists: that conventional kinds of educational research are

inactive, being the pastime of those closed off from the world in their ivory towers. Needless to say, I do not accept that implication.

16. Some advocates of action research seem to come close to this position, see Elliott (1988). For a critique of this, see Hammersley (2003c, 2004b).
17. See, for example, Denzin (1992), p. 131.
18. I have not been able to find any clear presentation of this argument for intervention, but an example of the approach is Miller and Whalley (2005). I am not suggesting that these authors would accept my description as an accurate account of their motivation.
19. Seale (1999) provides an example of this sort of position.
20. Hodkinson (2004) illustrates this position, see Hammersley (2005). It is worth noting that this does not have to be taken as implying that there are incommensurable paradigms, the exponents of which simply cannot understand one another. An alternative metaphor is different language communities, where while true translation may not be possible learning the other language is. This is an analogy that Kuhn uses in his later work, as against the perceptual analogy on which he relies in *The Structure of Scientific Revolutions* (See Kuhn, 1970).
21. Feuer *et al.* (2002) exemplify this position.

References

Altheide, D. & Johnson, J. (1994) Criteria for assessing interpretive validity in qualitative research, in: N. Denzin & Y. Lincoln (Eds) *Handbook of qualitative research* (Thousand Oaks, CA, Sage), 485–499.

Ashmore, M. (1989) *The reflexive thesis* (Chicago, IL, University of Chicago Press).

Atkinson, P. & Coffey, A. (2002) Revisiting the relationship between participant observation and interviewing, in: J. F. Gubrium & J. A. Holstein (Eds) *Handbook of interview research* (Thousand Oaks, CA, Sage), 801–814.

Baker, C. D. (1997) Ethnomethodological studies of talk in educational settings, in: B. Davies & D. Corson (Eds) *Oral discourse and education* (vol. 3) Encyclopedia of Language and Education (Dordrecht, Kluwer Academic), 43–52.

Biesta, G. (2007) Why 'what works' won't work: evidence-based practice and the democratic deficit in educational research, *Educational Theory,* 57(1), 1–22.

Bird, A. (2000) *Thomas Kuhn* (Princeton, NJ, Princeton University Press).

Clark, C. (2005) The structure of educational research, *British Educational Research Journal,* 31(3), 289–308.

Clarke, J. & Newman, J. (1997) *The managerial state* (London, Sage).

Denzin, N. K. (1992) Whose Cornerville is it anyway?, *Journal of Contemporary Ethnography,* 21, 120–132.

Dingwall, R. (1997) Accounts, interviews and observations, in: G. Miller & R. Dingwall (Eds) *Context and method in qualitative research* (London, Sage), 51–65.

Elliott, J. (1988) Response to Patricia Broadfoot's presidential address, *British Educational Research Journal,* 14(2), 191–194.

Ferlie, E., Pettigrew, A., Ashburner, L. & Fitzgerald, L. (1996) *The new public management in action* (Oxford, Oxford University Press).

Feuer, M. J., Towne, L. & Shavelson, R. J. (2002) Scientific culture and educational research, *Educational Researcher,* 31(8), 4–14.

Feyerabend, P. (1975) *Against method* (London, Verso).

Furlong, J. & Oancea, A. (2005) *Assessing quality in applied and practice-based educational research: a framework for discussion* (Oxford, Oxford Department of Educational Studies). Available online at: http://www.bera.ac.uk/pdfs/Qualitycriteria.pdf (accessed 26 February 2006).

Geertz, C. (1973) *The interpretation of cultures* (New York, Basic Books).

Gibbons, M. (2000) Mode 2 society and the emergence of context-sensitive science, *Science and Public Policy,* 26(5), 159–163.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P. & Trow, M. (1994) *The new production of knowledge; the dynamics of science and research in contemporary societies* (London, Sage).

Hammersley, M. (1990) An evaluation of two studies of gender imbalance in classroom interaction, *British Educational Research Journal,* 16(2), 125–143.

Hammersley, M. (2000) *Taking sides in social research* (London, Routledge).

Hammersley, M. (2002) *Educational research, policymaking and practice* (London, Paul Chapman/Sage).

Hammersley, M. (2003a) Recent radical criticism of interview studies: any implications for the sociology of education?, *British Journal of Sociology of Education,* 24(1), 119–126.

Hammersley, M. (2003b) Conversation analysis and discourse analysis: methods or paradigms?, *Discourse and Society,* 14(6), 751–781.

Hammersley, M. (2003c) Can and should educational research be educative?, *Oxford Review of Education,* 29(1), 3–25.

Hammersley, M. (2004a) Should ethnographers be against inequality? On Becker, value neutrality and researcher partisanship, in: B. Jeffrey & G. Walford (Eds) *Ethnographies of educational and cultural conflicts: strategies and resolutions* (Oxford, Elsevier), 25–43.

Hammersley, M. (2004b) Action research: a contradiction in terms?, *Oxford Review of Education,* 30(2), 165–181.

Hammersley, M. (2005) Countering the 'new orthodoxy' in educational research: a response to Phil Hodkinson, *British Educational Research Journal,* 31(2), 139–155.

Hammersley, M. (2007) Assessing validity in social research, in: P. Alasuutari (Ed.) *Handbook of Social Research* (London, Sage) (forthcoming).

Hammersley, M. (2008) Educational research, in: G. McCulloch (Ed.) *Encyclopaedia of Education* (London, Routledge) (forthcoming).

Hammersley, M. (forthcoming) Troubling criteria: a critical commentary on furlong and Oancea's framework for assessing educational research, *British Educational Research Journal.*

Hodkinson, P. (2004) Research as a form of work: expertise, community and methodological objectivity, *British Educational Research Journal,* 30(1), 9–26.

Hoyningen-Huene, P. (1993) *Reconstructing scientific revolutions: Thomas S. Kuhn's philosophy of science* (Chicago, IL, University of Chicago Press) (First published in German in 1989).

James, M., Pollard, A., Rees, G. & Taylor, C. (2005) Researching learning outcomes: building confidence in our conclusions, *Curriculum Journal,* 16(1), 109–122.

Kuhn, T. S. (1970) *The structure of scientific revolutions* (2nd edn) (Chicago, IL, University of Chicago Press).

Miller, L. & Whalley, J. B. (2005) Stories from the field—'taking the piss': notes on collaborative practice as research, in: B. Somekh & C. Lewin (Eds) *Research methods in the social sciences* (London, Sage), 313–317.

Murphy, E., Dingwall, R., Greatbatch, D., Parker, S. & Watson, P. (1998) Qualitative research methods in health technology assessment: a review of the literature, *Health Technology Assessment,* 2(16), 1–260. Available online at: http://www.hta.nhsweb.nhs.uk/execsumm/summ216.htm (accessed 14 August 2002).

Pollitt, C. (1990) *Managerialism and the public services* (Oxford, Blackwell).

Putnam, H. (2002) *The collapse of the fact-value dichotomy and other* essays (Cambridge, MA, Harvard University Press).

Ridley, F. (1996) The new public management in Europe: comparative perspectives, *Public Policy and Administration,* 11(1), 10–29.

Schwandt, T. (1996) Farewell to criteriology, *Qualitative Inquiry,* 2(1), 58–72.

Seale, C. (1999) *The quality of qualitative research* (London, Sage).

Sharrock, R. & Read, R. (2002) *Kuhn: philosopher of scientific revolution* (Cambridge, Polity Press).

Silverman, D. (1997) Towards an aesthetics of research, in: D. Silverman (Ed) *Qualitative research: theory, method and practice* (London, Sage).

Smith, J. K. (1984) The problem of criteria for judging interpretive inquiry, *Educational Evaluation and Policy Analysis,* 6(4), 379–391.

Smith, J. K. & Deemer, D. K. (2000) The problem of criteria in the age of relativism, in: N. K. Denzin & Y. S. Lincoln (Eds) *Handbook of qualitative research,* (2nd edn) (Thousand Oaks, CA, Sage), 877–896.

Spencer, L., Richie, J., Lewis, J. & Dillon, L. (2003) *Quality in qualitative evaluation: a framework for assessing research evidence,* prepared by the National Centre for Social Research on behalf of the Cabinet Office.

Thomas, G. & Pring, R. (Eds) (2004) *Evidence-based practice in education* (Buckingham, Open University Press).

Trinder, L. with Reynolds, S. (Eds) (2000) *Evidence-based practice: a critical appraisal* (Oxford, Blackwell Science).

Zimmerman, D. H. & Pollner, M. (1971) The everyday world as a phenomenon, in: J. D. Douglas (Ed.) *Understanding everyday life* (London, Routledge), 80–103.