



# *Guidelines for Reporting and Evaluating Qualitative Research: what are the alternatives?*

**ALAN REID & STEPHEN GOUGH** *University of Bath, UK*

**SUMMARY** *There is more to guidelines for qualitative research reports than that proposed in Smith-Sebasto (2000). Our response to the article illustrates a range of alternatives to the position and criteria outlined by Smith-Sebasto. We draw on examples from across the wider literature to explore themes and issues regarding the purpose(s) of guidelines, their content and how they are structured. Our core argument is that deliberations on the process and product of judging the quality of qualitative research in environmental education research require the recognition of two points regularly contested within the literature on guidelines. Firstly, the existence of a wide variety of types, genres and forms of qualitative research; and secondly, the proposition that criteria for judging research quality contain within them, implicitly or explicitly, a defining view of what research is, and perhaps more contentiously, what it should be.*

## **Introduction**

The viewpoints expressed in this collection are not unique. Nor are they the first time researchers and commentators have deliberated about evaluative guidelines in qualitative research. Put more positively, they form part of a broader discussion and assessment of guidelines and criteria, which in this instance, takes place on the 'broad plains' of environmental education research. We use this metaphor for two reasons. Firstly, we believe that no research model has a monopoly on research in education—let alone environmental education (Peshkin, 1993). Secondly, we believe that drawing on a diversity of approaches, techniques and traditions is essential to the vibrancy, openness and continuance of environmental education research (Gough & Reid, 2000). Both positions are illustrated throughout the article.

Throughout the following sections, we introduce a plurality of proposals and responses regarding the purpose, content and structure of guidelines (cf. Gough

& Reid, 2000). The recurring theme, as indicated by the title of this article, is alternatives to the position and criteria outlined in Smith-Sebasto (2000). When the diverse aspects of guidelines are placed side by side, it is possible to consider how each offers comment on the others, including that of Smith-Sebasto (2000). In this way, the rich diversity of meaning of the 'quality of qualitative research' becomes apparent, and the sometimes contradictory messages the reader of this debate receives can be evaluated. Our core argument is that deliberations on the process and product of judging the quality of qualitative research in environmental education research require the recognition of two points regularly contested within the literature on guidelines. Firstly, the existence of a wide variety of types, genres and forms of qualitative research; and secondly, the proposition that criteria for judging research quality contain within them, implicitly or explicitly, a defining view of what research is, and perhaps more contentiously, what it should be.

We note, as Seale (1999b, p. 50) argues, the 'quality of research is not automatically determined by the imposition of generalised quality criteria, but such schemes can help sensitise researchers to the issues that a particular project may need to address'. Thus, we do not intend simply to confirm, hone, or challenge guidelines assembled already in this collection. In short, as our discussion shows, we recommend those stances within the wider research community that seek to advance a more *transparent*, rather than just a *technical*, appreciation of the methodological issues in judging the worth of qualitative research.

To signal our general approach to these matters, we start by probing remarks that frame this discussion in environmental education research. We also introduce our position on some of the wider issues regarding guidelines in this field. The main part of our article explores a range of alternatives to Smith-Sebasto (2000), while the discussion develops our themes, raising some of the subtleties and implications of these alternatives within the processes and products of guideline identification and application.

## Opening Remarks

We start by noting that it is now increasingly difficult to sustain the argument that the terminology and criteria for defining or evaluating qualitative research are either irrevocably idiosyncratic, ambiguous, or conceptually undeveloped (see Seale, 1999b). Nor is it that qualitative research is a marginal pursuit either within or outwith research in environmental education. In many cases, justifying the substance and relevance of qualitative research within the panoply of research practices available to 'the environmental education researcher' is simply no longer an imperative. Clearly, as elsewhere, there are variations in the quantity, quality, and status accorded to qualitative inquiries that are reported in environmental education research. Concerns have also been expressed about publishing reports of qualitative research in a field that can still be dominated by 'positivist traditions' (see Mrazek, 1993), as well as that of the scholarly standards of the work that is published (Smith-Sebasto, 2000). What we challenge is both the persistence of a narrowness in the range of examples, and the insufficiencies of broader argumentation, of qualitative research in journals that are marketed solely for an environmental education research audience [1].

Although recognising their continued importance to these discussions, the case for and against the (in)adequacies of criteria like *validity*, *reliability* and *generalisability* has been made at length elsewhere so will not be repeated here [2]. Emden and Sandelowski (1998, p. 207) liken their 'trajectories' within qualitative inquiry to that of being at one of five successive stages: that of being 'championed, translated, exiled, redeemed and surpassed'. We agree with those who argue that these criteria are appropriate to informing guidelines for the evaluation of quantitative research, and that quantitative and qualitative research methodologies may complement each other. What we disagree with, like Emden and Sandelowski, is the value of applying these criteria to qualitative research by little more than direct analogy, that is, their commensurability is and should be contested, whether or not they are argued to be of 'parallel form' [3].

Neither shall we repeat the debates over the merits of the various interpretive, critical and other approaches within what amounts to a 'broad church' of qualitative inquiry. In our view, this debate has, on occasion, been unnecessarily reduced to one that focuses on the forcing of 'paradigm shifts' in environmental education research. Such a focus may reflect a desire to supplant 'traditional' and 'quantitative' approaches, or, by default or design, to give voice, power and status to a growing cadre of 'new paradigm' researchers who espouse alternative research programmes within environmental education research. Whatever the merits or weaknesses of this focus, our caution is that by neglecting the limitations of 'alternative approaches' and concentrating exclusively on their strengths, the outcome may exhibit the same limiting lack of reflexivity of which the more traditional ('dominant' or 'hegemonic') approaches in this area are accused (cf. Robottom & Hart, 1993; Carr, 1994; Stronach & MacLure, 1997). As Seale (1999b) cautions, it is not that reflection on methodological considerations is not necessary in research. Rather, it is the extent to which decisions regarding attention to matters like 'quality' hinder or assist the craft and passage of the research project which invoke more pragmatic and sceptical considerations into the debate. We will illustrate these opening remarks more fully throughout this article, but before we do so, we introduce our own attempt to contextualise Smith-Sebasto's proposed guidelines with reference to wider debates.

## Raising the Debate

The debate about guidelines is neither new, in and of itself, nor exclusive to environmental education research. Engel and Kuzel's (1992) article on health-related research is one of many [4] which discuss trends across the social science literature with that of other areas using qualitative research, each of which share a concern with questions of whether and/or how to judge its quality. Thus in discussing what constitutes goodness in qualitative research in nursing, and what kinds of value judgements we might make, Emden and Sandelowski (1999, p. 2) cite Smith (1987) as suggesting that one has not seriously thought about the nature of inquiry 'until one has confronted the criteria issue'. Engel and Kuzel (1992), like Emden and Sandelowski (1998, 1999), map the conceptual resources for these confrontations to 'goodness criteria' developed in sociology, anthropology and psychology. If we consider similar analyses in the work of Elder and Miller (1995), Hasselkus (1995) and Chapple and Rogers (1998) among others, it

is clear that boundaries between disciplines are neither permanent nor impermeable: in each case, insights from sociology, anthropology and psychology—along with those from, say, literary theory, the philosophy of science, and axiology, to name but three—have found their way into the discussion.

In the case of Smith-Sebasto (2000), we observe a relatively small sample of documents on how we might conduct and report qualitative research. Nevertheless, the circle of citations does extend beyond that limited to environmental education research journal guidelines and education research textbooks to include articles from other fields like family research (Ambert *et al.*, 1995) and health research (Knafl & Howard, 1984). At first glance then, there might be considerable advantage in considering the methods for their selection and mobilisation in Smith-Sebasto (2000) as a means to identifying alternatives within this article. One way of achieving this might be to apply the proposed guidelines to a qualitative report, say Smith-Sebasto's, and by way of identifying the strengths and shortcomings of the report, reflect on those of the guidelines as well. An alternative strategy might be to repeat and extend Smith-Sebasto's initial search criteria and database search. This bodes well with advocates of triangulation and replication of research, but, like our first strategy, is not without its problems. One of the tensions it creates relates to the ironies that such a strategy might raise. For example, of what value are these searches to the debate if they are inconsistent or do not cohere with each other—criteria themselves which have a particular 'modernist ring' about them (Seale, 1999a)? Another is that the term 'alternatives' now relates more to that of substance than strategy, that is, we are bypassing debates on whether to judge the quality of research to focus more explicitly on *how*. The latter strategy was, of course, the one that took place at Bath in preparing the substance of the article. What we hope to show is that in spite of its weaknesses, our methods, findings and reflections on our experience with these searches raises a variety of questions and issues about guidelines and criteria and their framing, matters we return to at greater length in our discussion.

### Identifying Alternatives

At Bath then, additional sources on reporting and evaluating qualitative research were identified through the BIDS information services, including its links, indices and databases on the internet at: <http://www.bids.ac.uk>. In our case, they include access to ERIC, BIDS-ISI and BIDS-EDUCATION and MEDLINE. We employed additional search terms like variants of *qualitative research*, *qualitative method/s*, *qualitative methodology*, along with those for *guidelines*, *criteria*, *criteriology*, *principles*, *standards*, *evaluation*, *validity*, *reliability*, *rigor/rigour*, *goodness*, *legitimation* and *plausability*, either in isolation, combination, or to 'scope' the electronic retrieval of findings. In total, our electronic searches produced approximately 40 further references within a variety of cognate areas, with each 'result' addressing either specifically or loosely the key terms within Smith-Sebasto's article [5].

We do not claim that this search was extensive or exhaustive. It reflects a variety of constraints on the search process which relate in part to our aptitudes and inclinations regarding the task of preparing this response and to its 'technologies', like the availability and scope of citation indices, search facilities,

and our search strategies. Yet, as it has already been stated, there are many diverse ideas of quality in qualitative research. How then can any one set or search be given priority? There is the distinct possibility of total fragmentation here, particularly in representing the divergent positionings of guidelines as, for example, 'unassailable criteria', 'rules of thumb' or 'guiding ideals' (cf. Schwandt, 1996). In response, even in comparing the citations and core arguments on qualitative research as made available via the abstracts and citation lists, it is clear that particular key texts, issues and dilemmas in reporting and evaluation are repeatedly cited and counter-cited [6,7]. More importantly, the examples we use are not arbitrary, they represent the exercise of 'authorial presence' in assembling a range of contrasting themes, voices and issues that might be read, side by side, as commenting on each other and Smith-Sebasto (2000) as alternatives in conducting and reporting qualitative research. The sections include elements of the substance, argumentation, and issues raised by these documents to illustrate how one might—or even, should—judge the worth of qualitative research. Our examples are drawn from the literature in communications, nursing and health, human geography, and the social sciences more generally. But those are not the only sources, and this is not the only way of constructing a response. Thus, repeat, further and alternative searches are welcome to further the discussion and range of views, and critiques of our strategy and the value of what we present below.

### **What is it that Qualitative Researchers Aspire to When Conducting and Reporting their Research?**

Our introductory examples are drawn from the field of communications research. Here, issues of publishing, reporting and evaluating research remain vibrant. To illustrate, we refer readers to two special editions of the *Western Journal of Communication*, 41(1) from 1977, and 58(1) from 1994, which discuss the admissibility of evidence, criteria to judge evidence, and the various implications of methodology for our conceptions of evidence in reporting research (e.g. Fitch, 1994; Tompkins, 1994). More specifically, in raising the question of what it is that qualitative researchers aspire to when conducting and reporting their research, we note the work of Avery and McCain (1984). The authors reflect on their experience as editors of communications research journals at the beginning of the 1980s, a time which witnessed the growth of qualitative research in a field traditionally dominated by quantitative research, like others more closer to home.

Avery and McCain (1984) reviewed a variety of journal submissions and publications, on the basis of which they make two main observations about evaluating and reporting qualitative research. Their initial observation, which resonates with Smith-Sebasto's more recent concerns, is that 'qualitative research' is a relatively ill-defined descriptor which embodies a wide range of research activity (cf. Denzin & Lincoln, 1994a). As the term varies in meaning, they argue that a diversity of interpretations ensures that evaluators of qualitative studies bring to bear radically different expectations, sets of evaluative criteria, and predispositions regarding format and style of presentation. In ways that also echo both arguments about the quality of environmental education research cited by Smith-Sebasto [8] and the earlier stages of the trajectories

described by Emden and Sandelowski (1998), Avery and McCain then observe the uneasiness they perceive within their research community over appropriate standards for the execution and reporting of qualitative studies, adding that this might well have led certain reviewers to 'be more conservative in their estimation of a manuscript's worth' as well as be 'less willing to give specific guidance for revising the manuscript'.

A similar study was carried out by Zaruba *et al.* (1996) in education. They reviewed critiques of 57 manuscripts that had used qualitative research methods and were submitted to the *Review of Higher Education*. In developing their case that 'standards in qualitative inquiry should be distinct from those within the qualitative tradition' where appropriate, the authors refer to the trustworthiness criteria developed by Lincoln and Guba (1985) and related works to generate a set of standards for determining good qualitative research. Although generating guidelines fell beyond the remit of their particular article, Avery and McCain recommend the recognition of personal judgements and limitations in evaluating such research, as they recognised that much qualitative research was then 'new' to its audience. The substance of the guidelines by Zaruba *et al.* remains little more than general too: 'good qualitative research is thorough; informed; well written; balanced; useful; and educative' (cf. Emden & Sandelowski, 1998, p. 208).

To us, the observations and recommendations of the two surveys imply two points about the development of guidelines. First, in a realm that claims to be governed by peer review rather than reputation, we would argue that these examples may also link concerns about guidelines with those about how the impact and dynamics of the diversity, operations and health of a single discipline amongst many, might structure institutional academic life (Bourdieu, 1988). However, if the links are to be made, we would argue that they should not be reduced to that of expounding a 'theory of gatekeeping', where the focus is that of identifying how 'power' and 'control' determine what comes to be inside or outside the common property of the 'disciplinary circle', as if the observer were merely tracking intellectual currency or fashion. Our second point follows from this: within our search, for us the informative and interesting aspects of such debates are those that do not stop at considerations of the culture and authority by which status is conferred on a research report, but those which also examine the reasons and intentions they explicate about how we come to value some reports above others.

Thus when Engel and Kuzel (1992) locate these forms of discussion of quality as part of a larger debate about the efficacy of alternative paradigms in the human sciences, including that of who 'controls the terms of the debate', we would suggest that those who would deny the desirability of lists of guidelines should not unwittingly extend this to exclude an evaluation of those which exist or are proposed (see Garratt & Hodgkinson, 1998). In the case of Engel and Kuzel, they recount the shift from 'realist' to 'non-' and 'anti-realist' camps. By contrasting empirical-analytical assumptions with hermeneutic or interpretive perspectives along with their concomitant 'within-group' debates, Engel and Kuzel suggest that 'there can be no fixed or certain standards whose applications are algorithmic. In this view, standards must be heuristic at best, relying on tacit assumptions in their application' (p. 505). Their analysis in a journal aimed at qualitative health researchers is reminiscent of that about 'paradigm wars' and

'paradigm dialogues' in environmental education research (as, for example, well documented in the work of Guba (1990) more generally, and Mrazek (1993) in particular for environmental education research) [9]. Their themes are extended with reference to Smith's (1987) oft-cited article on publishing qualitative research. Firstly, Engel and Kuzel note the variety of approaches and traditions to qualitative research, and secondly, that the criteria for judging the quality of the work across this range may be particular to the approach and markedly different across approaches. To illustrate, Engel and Kuzel contrast case studies in symbolic interactionism and critical ethnography as a means to identifying the commonalities and differences within the 'hermeneutical tradition'. Both approaches operate as if knowledge is socially constructed, contextual, and interpretive, but they vary in the role, meaning and status accorded to: (a) 'empowerment', and (b) evidence of 'conflicting discourses', as criteria for judging the quality of qualitative research [10] (cf. Guba, 1990; Sandelowski, 1993; Guba & Lincoln, 1994; Schwandt, 1994; Bailey, 1997).

The reason we include this example from Engel and Kuzel is that although it is indicative of the significance of meta-level arguments in disciplines not necessarily directly connected with the focus of environmental education research, others, including environmental education researchers like Smith-Sebasto, still occupy themselves with the form, content and extent of guidelines in ways more specific than Avery and McCain and Zaruba *et al.* In this case, we might also suggest that discourses on guidelines are not just about establishing what we might recognise as either the 'private languages' or 'public discourses' that seek to establish and contend family resemblances, connections and aspirations within 'qualitative research' and research more generally [11]. They also return us to those matters of the institutional politics of what counts as research, that is, the conducting and legitimating of research, and those that relate to reflecting on how this state of affairs might be evaluated. Thus one possibility for combining these discussions in environmental education research might draw on Britzman's (1995, p. 231) remarks concerning the value of poststructuralist theories in problematising contemporary reporting practices:

Poststructuralist theories raise critical concerns about what it is that structures meanings, practices, and bodies, about why certain practices become intelligible, valorised, or deemed as traditions while other practices become discounted, impossible, or unimaginable. For poststructuralists, representation is always in crisis, knowledge is constitutive of power, and agency is the constitutive effect, and not the originator, of situated practices and histories.

On the one hand, researchers might investigate the 'archaeology' and 'genealogy' of constructions of what counts in research through its 'disciplinary artefacts'. On the other, researchers might bear in mind the 'likelihoods' of how contemporary praxis and its possible developments might continue, or disrupt, the 'sedimentation' of these constructions in the future. We return to this thread later in our discussion section. Before then, though, we consider an example of the role of contesting interests in human geography as another possibility for both extending the debate and introducing alternatives about what counts as good quality research in environmental education research.

## Contesting a Coalition of Interests

A contemporary example of the contestation of interests and practices can be identified in the exchanges between Bailey *et al.* (1999a,b) and Baxter and Eyles (1997, 1999). This follows the latter's (1997) call for detailed evaluative criteria to enhance the rigour of qualitative research in human geography, a subject area which has recently experienced its own literary, cultural and postmodern 'turns' (Pile, 1991). As part of a broader 'quest for quality', we suggest that the proposing of guidelines might also be read as signifying a 'request for recognisability' of specific aspects of qualitative research, and in so doing, the desire to open a platform for dialogue between researchers of differing practices (Moore *et al.*, 1986). One observation is that in identifying, selecting and displaying their preferences within qualitative research and research methodology, each article in the exchange seeks to influence the others' and the readers' conceptions of standards of quality. Another is that if such discourses are understood as illustrative, evaluative and socially constructive in import, they might also provide models which socialise and/or are constitutive of other practices in qualitative research in their field, whether the practitioner is 'quantitatively schooled', 'traditionally educated', 'a qualitative novice' or 'expert' (Bailey, 1997).

Briefly, the exchanges between the authors followed Baxter and Eyles' review of nearly 50 papers by qualitative social geographers who were mainly using in-depth interviews. Baxter and Eyles (1997, p. 505) 'found little explicit reference to the principle(s) adopted to enhance "rigor" and to ensure meaningful inference'—principles which are set out here in Table 1. Reviewing what they describe as 'the modest explicit discussion of evaluative criteria in these papers' (p. 505), they then inspect a scheme which drew primarily on the work of Lincoln, Denzin, Rose and Guba in evaluation research, viz. general principles of evaluation like *credibility*, *transferability* (traceability), *dependability* (honesty) and *confirmability* (see Table 2 for definitions, assumptions and strategies/practices to satisfy each criterion). A set of questions as guidelines to evaluation were then developed from the core of their discussion of these principles (see Table 3), and these were then also applied to examples of either meeting or failing to meet the criteria within their own research (see Baxter & Eyles, 1997, pp. 517–520). That their guidelines for evaluation might be applied 'as a general set of criteria rather than rigid rules for assessing qualitative work' (p. 505) neatly summarises their intention of promoting and illustrating systematic methods of evaluation of qualitative research in human geography.

By way of response, while maintaining that systematic methods of evaluation are required, Bailey *et al.* (1999, pp. 169–170) repeat the wider argument that standardised evaluation is inappropriate on the grounds that qualitative research is both dynamic in nature and varied in epistemology and method. They then proceed to argue that 'the criteria for the evaluation of individual projects must arise from the research process itself' (p. 170), that is, they contend that evaluative criteria are developmental and not *a priori* (see also Baxter & Eyles, 1999). Nonetheless, examples of how general evaluative principles might be put into practice are suggested by Bailey *et al.*, and are reproduced here in Table 4. More specifically, in advocating the 'reflexive management' of the research process as part of the pursuit of the 'qualitative validity' of the study under



TABLE 1. Strategies for establishing qualitative 'rigour' in geographic work

| Strategy                          | Comments  |
|-----------------------------------|---|
| Rationale for methodology         | <ul style="list-style-type: none"> <li>• qualitative methods are argued to be the most (or only) appropriate way to address the research question</li> </ul>  |
| Multiple methods                  | <ul style="list-style-type: none"> <li>• more than one method used for studying the problem</li> </ul>  |
| Respondent                        | <ul style="list-style-type: none"> <li>• a description of the group(s) of respondents (e.g. number and gender ratio is given)</li> </ul>  |
| Interview quotations              | <ul style="list-style-type: none"> <li>• the words of the respondent may be read</li> </ul>   |
| Interview practices               | <ul style="list-style-type: none"> <li>• details of how interviews were conducted (e.g. use of interview schedules and autobiography are provided)</li> </ul>   |
| Procedures for analysis           | <ul style="list-style-type: none"> <li>• a description of how data were converted/condensed into theoretical constructs is given</li> </ul>   |
| Immersion/lengthy fieldwork       | <ul style="list-style-type: none"> <li>• it is argued that long field seasons (self-proclaimed or stated to enable exceed one year) develop rapport with respondents and/or enable deep understandings of the research situation</li> </ul> |
| Revisits                          | <ul style="list-style-type: none"> <li>• revisits to respondents are made usually (but sometimes unstated) to clarify meanings and build rapport</li> </ul>   |
| Verification by respondents       | <ul style="list-style-type: none"> <li>• respondents were contacted to verify interpretations/meanings</li> </ul>   |
| Appeals to interpretive community | <ul style="list-style-type: none"> <li>• an existing theory is supported (or refuted) by the findings, i.e. there is more than reference to the literature</li> </ul>   |
| Rationale for verification        | <ul style="list-style-type: none"> <li>• rationale for showing that there is agreement between constructs/interpretations and the meanings held by respondents is provided.</li> </ul>  |

Source: adapted from Baxter and Eyles (1997, p. 507).

scrutiny, Bailey *et al.* return to examples from the literature on grounded theory approaches (notably, that of Glaser, Strauss, Corbin and Spradley) to locate and illustrate their recommendations about procedures for qualitative reporting. These have two dimensions. Firstly, Bailey *et al.* argue that they be open, representative and valid in terms of analysis rather than just research design. Secondly, the authors revisit some of the concepts usually raised in reference to the efficacy of grounded theory techniques, but which now also echo the debates surrounding the quality and acceptance of computer-aided qualitative data analysis software as alternatives to 'manual technologies' of analysis within sociological and health research fields. These broader issues include the value of procedures that facilitate the production of quality indicators like 'audit trails' of the research process, negative case analysis, and 'analytical closure' (see also, for example, Appleton, 1995; Chapple & Rogers, 1998; Richards, 1999) [12].

What we would note about these examples is that both Baxter and Eyles (1997) and Bailey *et al.* (1999a) focus their proposed criteria on the *process* of qualitative research rather than the *product*, i.e. the quality of the inquiry process itself rather than the 'worth' of the narrative composed and presented in the research report. They are not alone in focusing debates in this way. Differences between criteria for each aspect can be identified in discussions of the production of both *authority* and *authenticity* in writing research reports, as in ethnography (Coffey, 1996); or *fidelity* and *believability* as criteria in narrative inquiry (Blumenfeld-Jones, 1995). They are also illustrated in environmental education research within Hart's (1996) discussion of the researching and reporting of teachers' thinking and practice. By way of contrast, Edge and Richards (1998, p. 334) can be

TABLE 2. Criteria for evaluating qualitative research

| Criterion       | Definition   | Assumptions   | Strategies/practices to satisfy criteria  |
|-----------------|--|---|---|
| Credibility     | Authentic representations of experience  | Multiple realities<br>Causes not distinguishable from effects<br>Empathetic researcher<br>Researcher as instrument<br>Emphasis of the research endeavour                          | Purposeful sampling<br>Disciplined subjectivity / bracketing<br>Prolonged engagement<br>Persistent observation<br>Triangulation<br>Peer debriefing<br>Negative case analysis<br>Referential adequacy<br>Member checking |
| Transferability | Fit within contexts outside the study situation  | Time and context-bound experiences<br>Not responsibility of 'sending' researcher<br>Provision of information for 'receiving' researcher   | Purposeful sampling<br>Thick description  |
| Dependability   | Minimization of idiosyncrasies in interpretation<br>Variability tracked to identifiable sources          | Researcher as instrument<br>Consistency in interpretation (same phenomena always matched with the same constructs)<br>Multiple realities<br>Idiosyncrasy of behaviour and context | Low-inference descriptors, mechanically recorded data<br>Multiple researchers<br>Participant researchers<br>Peer examination<br>Triangulation, inquiry audit  |
| Confirmability  | Extent to which biases, motivations, interests or perspectives of the inquirer influence interpretations | Biases, motivations, interests or perspectives of the inquirer can influence interpretation<br>Focus on investigator and Interpretations  | Audit trail products<br>Thick description of the audit process<br>Autobiography<br>Journal/notebook   |

Source: Baxter and Eyles (1997, p. 512), based on Lincoln and Guba (1985).

taken as broadening these examples by identifying three possible avenues as a means to developing a research programme in the face of guidelines: 'to extend conventional research parameters; to align one's work with a recognised radical alternative; to develop an individual stance', such that:

Any effective response must address questions of *position* (the research tradition in terms of which one wishes to be evaluated), *voice* (in terms of who is given space to speak) and *representation* (the form of discourse chosen). Taken together, these form the basis of *legitimization* (in terms of disputable argument) and *authentication* (in terms of contextualized credibility). [p. 334; italics added]

Similar points about research product are made by Flaherty (1996) in his discussion of truth, validity and method in relation to methodological reflexivity, and by Koch (1994, 1996, 1998) and Koch and Harrington (1998), in

TABLE 3. Checklist for evaluating qualitative interview research

| Question   | Elaboration examples   |
|--|--|
| 1. What was the natural history of the research?             | Original purpose(s) of the research<br>Rationale for methodology<br>How research developed over time<br>Fieldwork relations  |
| 2. What data were collected and by what methods?             | Method of note-keeping<br>Method of tape-recording   |
| 3. How was the sampling done?                                | Explicit delineation of sample frame (working universe)<br>Random or purposeful?<br>Rationale for type of sampling used  |
| 4. How was the data analysis done?                           | Procedures for summarising and presenting data<br>How data were selected for presentation  |
| 5. What results are presented?                               | Description of researcher's objective for results presentation (e.g. theory-building or description)<br>Differentiation of data-derived as opposed to pre-existing constructs<br>Differentiation of participant concepts as opposed to theoretical (researcher-derived) constructs |
| 6. How credible and dependable are the data-construct links? | Details of the relationship(s) between the data and constructs/concepts derived from data (e.g. member checking)   |
| 7. How credible is the theory/hypothesis?                    | Specification of the relationship between constructs/concepts and theory/hypotheses  |
| 8. How transferable are the findings?                        | Recognition of the limits imposed by the sampling strategy   |

Source: adapted from Baxter and Eyles (1997, p. 518), originally based on Rose (1982) and Lincoln and Guba (1985).

recounting the complexities raised by Gadamer's hermeneutics for both reflexivity and rigour. Reflecting on his ethnographic inquiries, Flaherty identifies intellectual curiosity, systematic empiricism, analytical induction, pragmatic orientation, and commitment to progress, as evaluative principles for his and similar research programmes. In contrast, Koch (1998, p. 1189) has repeatedly argued that reflexive research should be characterised by ongoing self-critique and self-appraisal, including the comment that if the research product is 'well sign posted, the readers will be able to travel easily through the worlds of the participants and makers of the text (the researchers) and decide for themselves whether the text is believable or plausible (our terms for rigor)' (Koch & Harrington, 1998, p. 887).

Thus despite their debt to Lincoln and Guba (particularly 1985—see note 4), neither set of authors discuss Lincoln and Guba's (1990) proposals for judging the 'typical product of alternative paradigm inquiry', that of the case report. Richardson (1994, p. 521), for example, argues that literary criteria like coherence, verisimilitude and interest be applied to the report. However, neither are such guidelines solely about good writing and the clear articulation of all elements of the study and its execution. Other criteria soon come into play. Lincoln and Guba (1990) discuss four additional criteria for judging the quality of qualitative research, but this time *resonance, rhetoric, empowerment, and appli-*

TABLE 4. Some principles for the evaluation of qualitative research in human geography

| Principle  | Examples of practice   |
|--|--|
| The need for theoretical sensitivity   | Being clear about the paradigm of inquiry. Being well grounded in academic literature and professional and personal experience. Continual interaction with the data.   |
| Reflexive management that strengthens qualitative validity   | Keeping a research diary that develops and documents reflexive management. This helps to make links between scientific and rhetorical research processes, so entries may range from 'logical' data links to 'blind leaps of faith' (which are then 'tested' by extending the data).  |
| Constant comparison by continued questioning   | Comparing categories developed from coded data with other data, and challenging with new data for verificational purposes, similar to qualitative triangulation. However, how, where and when to extend the emerging data is data-led, rather than pre-planned.  |
| Thorough documentation of procedures to leave a 'paper trail' audit that strengthens qualitative validity  | Continuously familiarising with, reflecting upon and linking and evaluating the data. This documentation centres on techniques demanded by grounded theory and outlined in this article, such as coding procedures, diaries, memos and diagrams.   |
| Clear and open reporting of procedures   | Including concise precis of procedures within public write-ups (journal articles, etc), but also making public (and indicating where others can access) longer description of methodology.   |
| Clear discussion of how theory 'fits' the reality of the respondents' lives, with rationale offered for 'negative' cases (not every incidence can fit) | Including concise summary within public write-ups, exploring alternative explanations where appropriate and acknowledging cases that do not appear to 'fit'; more detailed discussion documented elsewhere, as above. The research categories that emerge from the inquiry must be relevant to the respondents' lives. They should be able to explain the reality (for the respondents) of what is happening or what has happened. |
| Generation of criteria for evaluation of particular research   | These include details of evolving sampling procedures (as outlined in this article), a consideration of how the generated research categories explain the main concerns or problems of the subjects studied, and an ability to show how these research categories can account for most of the different ways in which the subjects deal with the identified concerns or problems.  |
| Recognition of the researcher(s) influence on the research findings (interpretation is always partial)   | Providing a clear statement of the background and position of researcher/research team at start of write-up; some reflection on the effect it has had at each stage.   |
| Use of archives for data and documentation relating to research procedures   | Using central archives where existent (e.g. ESRC), and establishing local archives where possible, to allow data and documentation on methodology, etc to be accessed.   |

Source: Bailey *et al.* (1999a, p. 175).

*capability* are positioned as product-related terms. This, like Edge and Richards, Flaherty, and Koch, can be read as a move to prompt a shift in the theoretical underpinnings of criteria for qualitative research away from such notions as 'artificial neutrality' in its process, to those which embrace concerns regarding reflexivity, literary worth, transformative and responsive potential, and vicarious experience in the product. In a style later redolent within Robottom and Hart's (1993) work on competing paradigms in environmental education re-

search, Lincoln and Guba speak of qualitative inquiry leading to a strategic transformation of social science and the visibility of researchers' political and social agenda, where (pp. 58–59):

... studies will resonate with the belief system of the alternative paradigm. That is, their basic assumptions will be exemplified in every way by the means and manner of the study's methodology or design strategy, and the final product of the inquiry will be composed with the emergent paradigm's world view in mind.

## The Health of the Debate

Proponents of these kinds of shifts towards diversity of inquiry and alternatives for quality judgements can also be found in health-related research. That the various fields of research which unite around health, nursing and medical sociology contain such recommendations is discussed at length by Seale (1999b) and Silverman (1998a). That the links between environment and health repay consideration in broadening the discussion in environmental education research have been highlighted previously in this journal (e.g. Sørensen, 1997 and the Health Promoting Schools movement), and, as such, might well be more instructive to the debate at hand than previously considered [13].

Before exploring some examples from this wide-ranging literature, we return to an earlier point about the institutional and positioning roles of discussions of criteria to suggest two further points that are raised by our reading of the health-related literature. The first concerns arguments about the 'death of the author': here, the emphasis is on the reader to interpret the qualitative research report as the author's viewpoint can no longer be regarded as in a privileged position for its interpretation. As the evaluative criteria brought to bear on a report can be those of an audience who are not necessarily the one that were 'hoped for' by the researcher, or, when qualitative researchers are writing their reports aware that their audience is still probably more familiar with quantitative research than qualitative kinds, clearer the 'mental sets' of the reader—their knowledge, expectations and traditions—may not be those that coincide with those of the author. Unfortunately, despite the best intentions of the author, this might result in little more than a hostile reception; a situation we observe rings particularly true of our next set of examples. And these examples illustrate our second point: the spectrum of responses to this state of affairs suggests that apologetic and defensive discourses might well be more the order of the day for some researchers, particularly for those who are having to identify their activities as qualitative research to an audience that appears even less sympathetic to that within environmental education research, or so we are led to believe.

Firstly then, from the field of nursing research, Bailey (1996) records how many nurse-researchers using qualitative strategies have been concerned with assuring the quality in their work to an audience who are more familiar with, and often prize the expectations and norms of, *quantitative, predictive, empiricist, positivistic or scientific* research [14]. Bailey then repeats another of the literature's widely used arguments: that recourse to the early literature (e.g. Campbell & Stanley, 1966; Cook, & Campbell, 1979) reveals that the concepts of validity and reliability as understood from 'the positivist perspective' are somehow inappro-

priate and inadequate when applied to qualitative research. In this case, for nursing, 'more recent literature suggests that because the positivist and interpretive paradigms are epistemologically divergent, the transfer of quality criteria from one perspective to the other is not automatic or even reasonable' (p. 186). Again, as with others who might advocate postpositivist, interpretive or qualitative method/ologies, Bailey then draws on Lincoln and Guba's framework to clarify guidelines concerning dimensions of broader notions of quality like *trustworthiness, credibility, authenticity, and goodness* for qualitative nursing research findings. We also note one other feature: the language of 'meta-paradigms' and 'ontological and epistemological perspectives' to discuss basic methodological divisions between the purposes, techniques and approaches of qualitative and quantitative research (cf. Emden & Sandelowski, 1998, 1999).

Bailey's is not the only example aimed at assisting health-related researchers in 'navigating the qualitative research literature with understanding'. It does, however, connect with the analysis of Edge and Richards and represents what might be viewed as the first of three types of responses by opponents of an uncritical acceptance of guidelines. These can, broadly speaking, be categorised as: *moderate rejection, radical rejection, and the redefinition*, of guidelines. That these types can be combined is well illustrated by Hall and Stevens (1991) when discussing rigour in feminist research in nursing from a 'postempiricist perspective'. Although stating that feminist inquiry is not confined to the use of a particular method and has no need for guidelines, i.e. there are grounds for a radical rejection of guidelines (cf. Lather, 1991), Hall and Stevens have made repeated use of the processes of qualitative methodology to argue that: 'The history of oppression, invisibility, and objectification of women underscores the need for a more relevant, just, and complete framework for evaluating research about and for women' (p. 27).

The task involves the critique of 'conventional empiricist criteria of reliability and validity', and the advancement of an alternative set of criteria and approaches. Their terms of reference are the various dimensions of *adequacy*, where adequacy implies that research processes and outcomes are well grounded, cogent, justifiable, relevant, and meaningful (p. 20), and its dimensions include *reflexivity, credibility, rapport, coherence, complexity, consensus, relevance, honesty and mutuality, naming and relationality*. A similar position is developed by Avis (1995, p. 1204), who argues that validity is an important concept in establishing the credibility of research findings, but who also maintains that 'the current debate about the criteria used to substantiate claims for the validity of research evidence is largely based on a set of distinctions between qualitative and quantitative methodology which are outdated and misleading' (p. 1203).

Positions like that of Avis have been developed around the argument that validity should be recognised as an epistemological concept (cf. Bailey, 1996), for example, 'whose application depends upon some fundamental positions taken about the nature of truth, representation and scientific methodology'. Similar concerns are evident in environmental education research when it is categorised with reference to ontology, epistemology and methodology (Robottom & Hart, 1993). With Avis, the rejection of 'traditional notions of criteria' suggests that there remain important, unresolved questions about 'the meaning and application of the concept of validity which have a crucial bearing on what criteria are used to establish the credibility of qualitative research evidence', (p. 1203). In a

slightly different vein, but to continue the thread of the previous section, a similar position to this, along with that advocated by Lincoln and Guba (1990), is the one developed by Meleis (1996). Meleis seeks to challenge and reinterpret criteria for rigour and quality in nursing research by broadening the debate to include notions of 'culturally competent scholarship' both for the processes by which researchers develop knowledge, and for reviewers in evaluating the rigour and credibility of scholarship. The criteria are those of: *contextuality, relevance, communication styles, awareness of identity and power differentials, disclosure, reciprocity, empowerment and time*. With such a proliferation of terms and recommendations amidst this widespread contestation of guidelines, what then of researchers who appear more content with the case for their development and application?

### **Guidelines: just following rules or modelling behaviour?**

Whatever the contributions of qualitative research to a particular field, in the context of environmental education research, as in health-related research, it has sometimes been viewed as what Chapple and Rogers (1998, p. 556) refer to as 'a "soft option", lacking scientific rigor, open to possible bias or even fraud, and a form of research unsuited to evidence based practice' (cf. Boulton *et al.*, 1996). In response to this line of criticism, Burns (1989) like Smith-Sebasto, responds by stating that one of the greatest obstacles to the identification of excellence in qualitative studies is 'the lack of generally accepted criteria', p. 44. An immediate question to ask is, in their provision, is research activity condemned to stagnate—i.e. remaining well within its usual frames? How do, for example, the sets of standards recorded in Tables 5 to 8 further the reflexive and dialogical nature of qualitative research, or the candour and critical attitude of researchers (Cermak & Stepanikova, 1997, 1998)? The following examples illustrate the extent to which this might be the case, before we introduce some more general reflections on the framing of the examples in our concluding discussion.

Like Sandelowski (1986, 1993) and Secker *et al.* (1995), Clarke (1992) discusses a variety of deficiencies within qualitative inquiries reported in nursing research. Issues for some readers identified by Clarke include sampling, reliability and attribution of meaning given to statements by interviewees, while recommendations for the provision for replication and standardisation, the use of counts of events, and evidence of independent checks (e.g. figures for intercoder reliability) might be viewed as more problematic by some qualitative researchers than attention to theoretical sensitivity within the interpretation of data, triangulation, negative case analysis, and testing for rival hypotheses (cf. Sandelowski, 1993). More importantly, what these examples illustrate are just some of the various methods and techniques emphasised within guidelines discourses that might 'improve the quality' of qualitative research, many of which are nested within a recognition of the resistance to qualitative inquiry in their field [15].

Accordingly, Elder and Miller (1995) provide a series of questions to aid the reading and evaluating of qualitative research studies for family physicians (Table 5). Similarly, Beck (1993, 1994) proposes a series of guidelines for evaluating the credibility, fittingness and auditability of qualitative research, where again the focus is one which highlights the way that a study is

conducted more than the way it is reported (Table 6). Burns (1989) too, after suggesting the skills necessary for critiquing qualitative studies (see Gough & Reid, 2000), proposes two detailed sets of general standards for constructing and evaluating qualitative research reports, each of which implies something of the author's expectations regarding the form and range of research that might be acceptable within the field of nursing, and what might be lacking in either aspect. In this case, equivalents of Burn's first set of expectations are found throughout the literature on guidelines and relate to the various elements of a qualitative research report. For Burns, the elements refer to the gathering of information on the investigated phenomenon and to the analysis and interpretation of the collected material, such as: the phenomenon under investigation; the purpose of the research along with the research questions and the significance of the topic; statements about assumptions and meta-theoretical perspectives and the literature review; the researcher's credentials, context, and role; discussions of ethical implications, sampling and subjects, and data analysis strategies; and, the researcher's conclusions, the study's implications, and any suggestions for further study (see Burns, 1989, pp. 46–48) [16]. The second set of standards relates to those required in critiquing qualitative studies, along with threats to those standards. To illustrate, Burns lists the following nine threats to the standard [17] of 'descriptive vividness' (1989, p. 48):

1. Failure to include essential descriptive information.
2. Lack of clarity in description.
3. Lack of credibility of description.
4. Inadequate length of time at site to gain familiarity necessary for vivid description.
5. Insufficient depth to description.
6. Insufficient skills in writing descriptive narrative.
7. Reluctance to reveal self in written material.
8. Inadequate self-awareness.
9. Poor observational skills.

Similar sets of criteria for the evaluation of qualitative research articles include those of the British Sociological Association Medical Sociology Group (1996), as cited by Seale (1999b), and listed in Table 7, and more briefly, Huff (1999, p. 159).

The fact that those working within a qualitative tradition in health-related research place emphasis on different criteria or standards when assessing the quality of a piece of qualitative research has more recently been illustrated by Chapple and Rogers (1998, p. 558). Their comparative table of criterial emphasis in the work of Popay *et al.* (1998), Seale and Silverman (1997) and Hoddinott and Pill (1997) for health service research is reproduced in Table 8.

Nevertheless, there are also many similarities to guidelines and criteria, some of which are used by Smith-Sebasto, others which represent distinct alternatives. More broadly, Popay *et al.* (1998, pp. 343–344) recognise that although there is no absolute list of criteria as to what constitutes good qualitative research, we may categorise the kinds of responses to what makes good qualitative research in three ways:



TABLE 5. Reading and evaluating qualitative research studies

**Part A: Questions to ask in reading qualitative research**

1. What is the study question?
2. Is an appropriate research approach taken to the initial study question?
3. What is the specific study design?
  - Who are the participants and how are they chosen?
  - How are the data collected?
  - How are the data analysed?
4. Is the final study design appropriate for the study question?

**Part B: Questions to ask in assessing qualitative research***Trustworthiness*

1. Who are the investigators and what is happening to them?
2. Do the investigators keep following up?
3. Is there sufficient contact between the investigators and the participants?
4. Did the investigators become too close to their participants?
5. Did the investigators cross-check their data?
6. Did the investigators "rule out" other theories?
7. Can the reader follow how the investigators got from the problem to the plan?
8. Were there outside checks?
9. Does the study provide sufficient information for readers to determine whether the study applies to them?

*Believability*

Does it all come together coherently?

- Is it parsimonious?
- Is it consistent?
- Is it fertile?
- Is it clear?

*Contribution*

1. Is the study clinically convincing?
2. Does the study make a contribution to the discipline of family medicine?

---

Source: Elder and Miller (1995).

On one side, there are those who argue that there is nothing unique about qualitative research and that traditional definitions of reliability, validity, objectivity, and generalizability apply across both qualitative and quantitative approaches. On the other side, there are those post-modernists who contend that there can be no criteria for judging qualitative research outcomes (Fuchs, 1993). In this radical relativist position, all criteria are doubtful and none can be privileged. However, both of these positions are unsatisfactory. The second is nihilistic and precludes any distinction based on systematic or other criteria. If the first is adopted, then, at best, qualitative research will always be seen as inferior to quantitative research. At worst, there is a danger that poor-quality qualitative research, which meets criteria inappropriate for the assessment of such evidence, will be privileged.

Popay *et al.* (p. 344) favour a third approach, as based on the work of the likes of Denzin and Lincoln (1994a) and Hammersley (1987):

Here, it is accepted that some criteria may be equally applicable to the evaluation of any research product, regardless of the methods. However, this third approach also acknowledges differences. In giving

TABLE 6. Guidelines for the evaluation of credibility, fittingness and auditability

---

|  |
|--|
| <b>Part A: Evaluation of the credibility of qualitative research</b>   |
| 1. Did the researchers keep in-depth field notes regarding the researcher–informant relationships?   |
| 2. Were the effects of the researcher’s presence on the nature of the data collected considered?   |
| 3. Did the researchers keep field notes of their actions, interactions, and subjective states during the investigation?  |
| 4. Did the researchers discuss their own behaviour and experiences in relation to the informant’s experiences?   |
| 5. Did the researchers become so involved with the informants that they “went native,” that is, had difficulty separating their own experiences from those of informant? |
| 6. Were multiple methods of data collection (triangulation) used to determine the congruence of the results among them?  |
| 7. Were the readers provided with rich excerpts from the transcripts or field notes?   |
| 8. Did the researchers validate the findings with the informants?  |
| 9. Did the researchers search for negative instances of categories or discounting evidence for tentative constructs (Field & Morse, 1985)?                               |
| 10. Were data analysis procedures reviewed by a judge panel to prevent researcher bias and selective inattention (Morse, 1989)?  |
| 11. Do the readers view the findings as meaningful and applicable in terms of their own experiences?   |
| <b>Part B: Assessment of the fittingness of qualitative research</b>   |
| 1. Did the researchers establish the typicality of the informants and their responses?   |
| 2. Did the researchers check for the representativeness of the data as a whole?  |
| 3. Did the theoretical sampling result in a range of informants experiencing the phenomenon under study?   |
| 4. Were the data made to appear more similar or congruent than they really were?   |
| 5. Did the study results fit the data from which they were generated?  |
| <b>Part C: Critiquing the auditability of qualitative research</b>   |
| 1. Was a tape recorder or other mechanical device used to record the interviews?   |
| 2. If a tape recorder was not used, did the researchers write their field notes immediately after the interview or observation to increase accurate recall?              |
| 3. Was an in-depth description of the strategies used to collect and analyse the data provided to the readers?   |
| 4. Were the characteristics of the informants described and the process used to choose the informants?   |
| 5. Were low inference descriptors, informants’ verbatim accounts, included to substantiate the categories developed during data analysis (LeCompte & Goetz, 1982)?       |
| 6. Were the social, physical, and interpersonal contexts within which the data had been collected discussed by the researchers?  |
| 7. Did the researchers specifically define the categories developed and also identify their theoretical antecedents?   |
| 8. Did more than one researcher perform the theoretical coding?  |
| 9. Did colleagues review the data to determine if they had identified the same categories and constructs as the researchers had?   |
| 10. Could another investigator clearly follow the decision trail used by the researchers in the study?   |

---

Source: Beck (1993).

greater recognition to the similarities between research traditions, we must not lose sight of two fundamental differences: the type of knowledge that different methods can generate (i.e., the epistemological difference) and the type of reality or object to which different methods are relevant (i.e., the ontological difference).

TABLE 7. Criteria for the evaluation of qualitative research papers

- 
1. Are the methods of the research appropriate to the nature of the question being asked?
    - Does the research seek to understand processes or structures, or illuminate subjective experiences or meanings?
    - Are the categories or groups being examined of a type which cannot be preselected, or the possible outcomes cannot be specified in advance?
    - Could a quantitative approach have addressed the issue better?
  2. Is the connection to an existing body of knowledge or theory clear?
    - Is there adequate reference to the literature?
    - Does the work cohere with, or critically address, existing theory?

### Methods

3. Are there clear accounts of the criteria used for the selection of subjects for study, and of the data collection and analysis?
4. Is the selection of cases theoretically justified?
  - The unit of research may be people, or events, institutions, samples of natural behaviour, conversations, written material, etc. In any case, while random sampling may not be appropriate, is it nevertheless clear what population the sample refers to?
  - Is consideration given to whether the units chosen were unusual in some important way?
5. Does the sensitivity of the methods match the needs of the research questions?
  - Does the method accept the implications of an approach which respects the perceptions of those studied?
  - To what extent are any definitions or agendas taken for granted, rather than being critically examined or left open?
  - Are the limitations of any structured interview method considered?
6. Has the relationship between fieldworkers and subjects been considered, and is there evidence that the research was presented and explained to its subjects?
  - If more than one worker was involved, has comparability been considered?
  - Is there evidence about how the subjects perceived the research?
  - Is there evidence about how any group processes were conducted?
7. Was the data collection and record keeping systematic?
  - Were careful records kept?
  - Is the evidence available for independent examination?
  - Were full records or transcripts of conversations used if appropriate?

### Analysis

8. Is reference made to accepted procedures for analysis?
    - Is it clear how the analysis was done? (Detailed repetition of how to perform standard procedures ought not to be expected.)
    - Has its reliability been considered, ideally by independent repetition?
  9. How systematic is the analysis?
    - What steps were taken to guard against selectivity in the use of data?
    - In research with individuals, is it clear that there has not been selection of some cases and ignoring of less interesting ones? In group research, are all categories of opinion taken into account?
  10. Is there adequate discussion of how themes, concepts and categories were derived from the data?
    - It is sometimes inevitable that externally given or predetermined descriptive categories are used, but have they been examined for their real meaning or any possible ambiguities?
  11. Is there adequate discussion of the evidence both for and against the researcher's arguments?
    - Is negative data given? Has there been any search for cases which might refute the conclusions?
-

TABLE 7. Criteria for the evaluation of qualitative research papers—*continued*

- 
12. Have measures been taken to test the validity of the findings?
- For instance, have methods such as feeding them back to the respondents, triangulation, or procedures such as grounded theory been used?
13. Have any steps been taken to see whether the analysis would be comprehensible to the participants, if this is possible and relevant?
- Has the meaning of their accounts been explored with respondents? Have apparent anomalies and contradictions been discussed with them, rather than assumptions being made?

**Presentation**

14. Is the research clearly contextualised?
- Is all the relevant information about the setting and subjects been supplied?
  - Are the variables being studied integrated in their social context, rather than being abstracted and decontextualised?
15. Are the data presented systematically?
- Are quotations, fieldnotes, etc. identified in a way which enables the reader to judge the range of evidence used?
16. Is a clear distinction made between the data and their interpretation?
- Do the conclusions follow from the data? (It should be noted that the phases of research—data collection, analysis, discussion—are not usually separate and articles do not necessarily follow the quantitative pattern of methods, results, discussion.)
17. Is sufficient of the original evidence presented to satisfy the reader of the relationship between the evidence and the conclusions?
- Though the presentation of discursive data is always going to require more space than numerical data, is the article as concise as possible?
18. Is the author's own position clearly stated?
- Is the researcher's perspective described?
  - Has the researcher examined their own role, possible bias, and influence on the research?
19. Are the results credible and appropriate?
- Do they address the research question(s)?
  - Are they plausible and coherent?
  - Are they important, either theoretically or practically, or trivial?

**Ethics**

20. Have ethical issues been adequately considered?
- Has the issue of confidentiality (often particularly difficult in qualitative work) been adequately dealt with?
  - Have the consequences of the research—including establishing relationships with the subjects, raising expectations, changing behaviour etc.—been considered?
- 

*Source:* British Sociological Association Medical Sociology Group (1996), cited in Seale (1999b, pp. 189–192).

Thus when Hammersley (1987) suggests that qualitative research needs to be assessed in terms of its capability to generate theory, be empirically grounded and scientifically credible, produce findings that can be transferable to other settings, and be internally reflexive with regard to the roles played by the researcher and researched, Popay *et al.* (1998) note that only the last of these marks a significant departure from research undertaken from within a quantitative tradition. Does this mean all that is required is to adjudicate which of the various lists of criteria advanced in the literature we should use in order to do the job well, or does attention to their broader framing repay attention?

TABLE 8. Variations in emphasis on criteria within health services research

| Emphasis in Popay, Rogers and Williams  | Emphasis in Seale and Silverman  | Emphasis in Hoddinott and Pill   |
|---|--|--|
| The researcher should use methodology that is responsive to particularities of situations—flexibility rather than standardisation. Show evidence that the research illuminates the subjective meanings, actions and social context of those researched. | The researcher should examine relevant historical evidence, and display cultural, political and contextual sensitivity. Supports generalisations by counts of events (quasi-statistics).                       | The researcher should make his or her roles and qualifications clear.  |
| Use a sample that produces the type of knowledge necessary to understand the structures and processes within which individuals or situations are located.   | Ensure representativeness of cases, including the use of combined qualitative and quantitative methods to support generalisations.   | Make it clear who conducted the interview and where the interview took place.  |
| Include enough detail to allow the reader to interpret the meaning and context of what is being researched.   | Test hypotheses in data analysis and include consideration of deviant case analysis, and grounded theory.  | Be explicit about how the respondents were recruited.  |
| Recognise that computer programs such as NUDIST or ETHNOGRAPH are neither necessary nor sufficient for rigorous qualitative analysis.   | Use computer programs to assist qualitative data analysis, thus ensuring systematic analysis of representative instances of data.  | Be explicit about who recruited the respondents.   |
| Show evidence that people's perceptions and experiences have been treated as knowledge in their own right. Show evidence that research moves from a description of data, through examples, to an analysis of the meaning and significance of it.        | Recognise that 'it simply makes sense to argue that all knowledge and feelings are of equal weight and value'. Record data objectively and comprehensively, preferably with the help of conversation analysis. | Be explicit about how the research was explained to the respondents.   |
|   |  | Be explicit about whether the interviewer was known to the respondents and how he/she was introduced. Show whether the methodological issues about the influences of the interviewer on the data were addressed. |

Source: adapted from Chapple and Rogers (1998, p. 558).

## Discussion

Quality in social science research is not a straightforward issue. Quality could refer to the relevance of a study, to the degree in which it yields useful and applicable information, to the degree it enhances values such as democracy and social justice, to the degree it empowers powerless people. Finally, it could also refer to the technical quality of a piece of work, that is, to the degree in which it conforms to the methodological expectations of a community of scientists.

This quotation from Meulenbergh-Buskens (1997, p. 111) signals a shift in the emphasis of this article. Our response to Smith-Sebasto (2000) has involved a

number of excursions into the literature [18] on qualitative research. We have illustrated a variety of general themes and issues regarding evaluative guidelines in qualitative research, and remain aware that many more could have been raised, including those of Gough and Reid (2000), Hart (2000) and Marcinkowski (2000). More generally, we note that the alternatives mentioned have been mainly those of content rather than frame of reference: the process of critiquing 'is not one of making judgements about whether or not certain rules have been followed or counting the number of criteria that have been met' (Hasselkus, 1995, p. 81), as if ideology-free readings can be achieved. In response, looking across the complexity and multiple valences of our examples, this final section uses a series of questions and issues to further our appreciation of how we might frame what an alternative might be—whether they apply to guidelines and criteria for qualitative research in general, our search strategy, or the relevance of the examples to the wider terms of the debate.

### *Some Questions*

We start by suggesting that bringing our notions of resemblance and connection in qualitative research into question is a salutary step in debates about guidelines. In particular, we begin by asking: where and how do those activities signified by 'qualitative research' situate themselves as such? For example, we might ask:

- Are they *with* or *against* conceptions of research activity as either a science, art, craft, professional activity, a hybrid of these categories, or something else (Gough & Reid, 2000)?
- Are its assumptions, terminology, language, and value systems those of 'naturalistic inquiry' (Lincoln & Guba, 1985) or something else?
- Which and whose practices do we model, seek to emulate, or improve on?

We are not suggesting that these are the only ways to investigate this broader question. For example, some might ask how resemblance and connection influence our perception of the status, persuasiveness and plausibility of the research. Likewise:

- Does coherence mean they are in some sense part of a paradigm (Atkinson, 1995)?
- What is essential? Ambiguous? Where is there lack of consensus on qualitative research (Beck, 1994)?

Such questions might suggest a need to focus on the purpose of a research project, along with its process, ends, and the extent of prior planning evident in the report (e.g. Burns, 1989; Marshall, 1989; Miles & Huberman, 1994). Alternatively, and what has interested us, is our justifications for the inevitable limitations and narrowness to our experience and knowledge of research from the past, in the present, and for the future—what Britzman (1995) refers to as investigating how our researching and reporting remain partial and governed by discourses of our time(s) and place(s).

Another issue we would raise to further our understanding of the framing of

alternatives concerns that of *whose* and *which* orthodoxy, position and logic researchers accept, contest and extend in examining and justifying the worth, standpoint and truth of their research endeavours. Sub-questions of why we should, rather than how we might, ask questions of quality might include:

- Are we seeking indicators of quality that are, say, definitive, provisional, partial, impartial, tailored, general, minimalist, extensive, summative, formative, normative, absolute, immutable, literal, self-certifying?
- Do we—how do we—invite the checks of others?
- Do we act as though evaluating research were a private rather than public matter?
- Is there—can there be—a shared (although not necessarily stable) language of concepts and practices?
- Do we consider discussing how might our research be refuted, falsified, negated, even wrong, as part of its evaluation?

We would also ask that in responding to such questions, are we more or less tentative in the authority of our theories, practices and findings?

Such questions and issues can be explored further in a variety of ways. The one that we have become most familiar with in preparing this response is reflections on what Schwandt (1996), Emden and Sandelowski (1998, 1999), Garratt and Hodgkinson (1998) and Seale (1999a,b) refer to as the project of *criteriology*.

### *Criteriology*

Seale's (1999a) review of modernist and postmodern phases of qualitative inquiry identifies the proliferation of difference and the change in emphasis in the work of a variety of researchers and commentators as they discuss, argue and advocate particular sets of criteria and guidelines for judging good quality qualitative research. Our article replicates aspects of this, but for Seale, the work of LeCompte and Goetz (1982), Lincoln and Guba (1985), Kirk and Miller (1986), Guba and Lincoln (1989, 1994), and Altheide and Johnson (1994) in particular become ciphers for groupings within the qualitative research community and their deliberations over appropriate and worthwhile criteria (see also Seale, 1999b; Gough & Reid, 2000). In terms of strategy, whether it is better not to pursue commensurability but rather to focus evaluation on understanding and how understanding is possible has been raised by Imre (1994). To a poststructuralist inquiry of *criteriology*, this might be taken as an invocation to investigate those modes of intelligibility which constitute textual identities and circulate competing regimes of truth, as might be illustrated in the examples in our previous section.

In Seale's case, rather than focusing on how the generation of, and competitions between, sets of criteria might provide evaluative practices that prize desirables that include truth status, factual accuracy, style of representation, authority, personal experience, ethical stance, political commitment or methodological innovation, a different route is taken. Seale (1999a), like Imre, avoids the inconclusiveness of these debates about what is and why something is desirable [19], by reframing the discussion to introduce a range of socio-cultural perspectives on *criteriological* projects, including the observation that:

The urge to generate criteria for judging good quality studies seems irrepressible in these methodological writings, perhaps in part due to the requirement that social researchers impress the worth of their efforts on sceptical audiences, such as research funding bodies. ... In qualitative research the project of criteriology experiences particular contradictions because of the difficulty in regulating and constraining an endeavour whose guiding philosophy often stresses creativity, exploration, conceptual flexibility and a freedom of spirit. Additionally, though, conceptual proliferation is a marker of the paradigm shifts and crises of legitimation and representation, that have characterised 'moments' in the recent history of qualitative methods (Denzin & Lincoln, 1994b).

The desire to initiate searches for 'permanent or universal criteria for judging qualitative research' (cf. Schwandt, 1996) has been swiftly dismissed by the likes of Garratt and Hodgkinson (1998, p. 515) via an anti-foundational analysis of the criteria for criteria. However, it does remain that what replaces or substitutes these desires might then be problematised and deconstructed during the next 'moments' of qualitative research, such as investigating how it is itself constituted by its own theory, policy *and* practice. Similarly, despite the anxieties surrounding the postmodern twists to a number of academic projects, we would note that the cynicism at a belief in shared values may be more of an obstacle to common sets of guidelines than a relativism of knowledge, abbreviating criteria to little more than shorthands that signify differences of belief, opinion or taste. Perhaps more positively, drawing on Britzman (1995, pp. 230–231), we might ask how this debate about qualitative inquiries in environmental education research is made intelligible by questions like:

- What is it that structures the methodological imperatives for environmental education research, the researcher's stances and their voices?
- What are the kinds of theoretical traditions through which data are constructed, represented, and narrated?
- What are taken to be problems suited for environmental education research?
- What are the problems of how one might read against the environmental education research grain?

### *Guidelines and Theory, Policy and Practice*

By referring to the 'triumvirate' of theory, policy and practice, we attest to our view that these dimensions to research are interdependent and not mutually exclusive. Judging the relationships between, and priority given to, theory, policy and practice within a research community's deliberations has been part of the broader evaluation debate we referred to earlier (cf. Denzin & Lincoln, 1994a). They are also illustrated in our final set of questions concerning the framing of possibilities and alternatives for guidelines projects.

Firstly, does identifying and applying criteria imply a desire for little more than increased methodological rigour? This emphasis may be taken to express the belief that via (often prespecified) guidelines for the application of methods, we may be able to ensure objective and universally valid judgements about research (cf. Maxwell, 1992). In this way, it might just be possible 'to establish



how things really are', enabling the research community to build a more 'complete' and 'accurate' picture of the world—a situation Johnson (1999) links with the pursuit of 'verificationism' [20]. This kind of perspective on guidelines is widely argued to be predicated on realist assumptions about the research task, as typified by the assumption that 'right method' is the best guarantor of truth and quality in research findings, unpolluted by subjective opinion or uncertainty. It is also often coupled with a commitment to faithful and accurate representation—i.e. researchers who say what they mean and mean what they say. But is pursuing and assembling correct criteria simply missing the point about the quality of qualitative research? For example, we do not focus on the mirror itself to find out about ourselves, we consider our reflection. Some opponents to a naïve realist position, particularly when drawing on the literature of the 'crisis of representation and legitimation' suggest that guidelines are little more than rules and prescriptions, amounting to a new orthodoxy for research practices which may even lead to 'methodolatry' (cf. Hasselkus, 1991). Ironically, considering the modifications and reactions of Denzin (e.g. 1997), Hammersley (e.g. 1992), Lather (e.g. 1993) and Lincoln (e.g. 1995) to those who seek refuge in method [21], it might also be suggested that even those like us who seek to engage critically with this widely held position will continue to reinscribe it merely by engaging the debate, to echo Derrida. A way out of this bind is suggested in Gough and Reid (2000).

Secondly, must identifying and applying guidelines be reduced to a 'policing practice' that divides good research from bad (see Scheurich, 1997, p. 87)? Garratt and Hodkinson (1998, p. 520) refer to guidelines as imposing a 'closed narratological framework': they tend to homogenise and fossilise research with what they term 'non-negotiable norms'. Whether guidelines apply to design, practice and/or accounts, this is a situation they cite Denzin (1997, p. 233) as arguing we should avoid. Besides, for Britzman (1995, p. 235), the dissatisfactions of writing ethnography within such a framework include the 'impossible moral imperatives of policing categories, insuring boundaries, and attempting to live the promises of a noncontradictory, transcendental self'.

Turning to matters raised by advocates of an 'ethically-informed' practical inquiry: is a criteriological project simply about constructing a conscious and/or discursively explicit account of the often tacit and intuitive criteria, commitments and dispositions for our interpretative judgements? Similarly, what are our unspoken rules and conventions? Further issues arise from the observation that generating a list of guidelines has become a common feature of their expression and contestation, as well-illustrated in Tables 1–8. One issue relates to their apparent openness to unending extension, reduction, selection and translation (Smith, 1990) [22]. One might also be suspicious of claims that lists of criteria represent no more than attributes of an openly descriptive, even reflexive, practice. Hence the question: How do guidelines represent a normative intention on the part of the researcher, reviewer or evaluator, that, as in the very nature of being constitutive of a public discourse, are implicitly inscribed to be applicable to others, i.e. can guidelines be anything other than inevitably and irredeemably exclusionary (Gough & Reid, 2000)? And, finally, what of the readers' reactions and their reinterpretation of research: 'most writings about the ways in which research should be judged are concerned almost exclusively with the ways in which the research was done, and not with *the ways in which the*

*standpoint of the reader will influence their judgement of that research'* (Garratt & Hodkinson, 1998, p. 524; italics added)?

In raising these and other critical matters about the framing of guidelines and the projects of criteriology, some responses may be characterised by what Geertz (1988, p. 131) calls 'a pervasive nervousness about the whole business'. In our view, similarly unsuccessful responses are those whose tone and substance either mystify or obscure explanations of approach, guideline and quality. In this article we have tried to avoid these pitfalls by shifting the focus of the article towards one which has introduced a selection of responses to these kinds of questions, but also raised them in and of themselves. Despite its limitations, our strategy has enabled us to demonstrate our argument that judging the quality of qualitative research reports requires the recognition of two issues regularly contested within the literature on guidelines. Firstly, the wide variety of types, genres and forms of qualitative research; and secondly, the proposition that the criteria for judging research quality contain within them, implicitly or explicitly, a defining view of what research is, and perhaps more contentiously, should be (Gough & Reid, 2000). Emden and Sandelowski (1999, p. 7) remind us that it is never too late to ask: 'Whose criteria?', 'Criteria for what?' and, 'Why criteria at all?' In closing this article and inviting the response of others, we quote Garratt and Hodkinson (1998, p. 535):

None of this means that researchers and others should stop writing about the ways in which they work and the ways in which their research and that of others should be judged. Such writing helps refine and develop our thinking about what doing and judging research entails, acts as an heuristic device for teaching others about these things, and represents a key part of what Gadamer might call the research tradition of which we are a part. Established approaches to doing and judging research are our collective prejudices, neither to be slavishly accepted nor wilfully rejected, but which should be placed continually at risk.

## Postscript

This response to the proposed guidelines in Smith-Sebasto (2000) set out to examine a variety of issues by which quality, guidelines and criteria are conveyed. We have continually offered the reader yet another response to set alongside the previous one. And this process has lasted to the final page of the main text. We have preferred to leave the reader with a patchwork of guideline figurations rather than to end by subordinating the variety to one common theme. It may be that the reader is unsure, finally, of the meaning to be placed on this material. But that, in part, is because there is no one fixed meaning to be found. Each reader may have a favourite alternative among those presented here, or the reader may prefer an alternative which the article has not explored. Also, each reader will probably put a different evaluation on the individual issues current in scholarship. The article, then, presents an opportunity for the reader of qualitative research to explore a range of possible alternatives in judging the quality of qualitative research. But hopefully it is more than this. We believe it provides a tool, and it is hoped, a rationale, for further research and

development for the qualitative researcher engaged with environmental education beyond that of Smith-Sebasto (2000).

## Notes on Contributors

STEPHEN GOUGH is a member of the Centre for Research in Education and the Environment at the University of Bath. *Correspondence:* Department of Education, University of Bath, Bath BA2 7AY, UK.

ALAN REID is a member of the Centre for Research in Education and the Environment at the University of Bath.

## NOTES

- [1] The range of issues raised in Mrazek (1993) on the remit and scope for work in alternative paradigms is taken to be illustrative of this point, as is that of a comparison of the submissions criteria assembled in the appendices of Smith-Sebasto (2000) with those used in mainstream qualitative research journals.
- [2] Examples can be found throughout Hammersley (1987, 1991, 1992), Stiles (1993), Seale (1999a), and the articles edited by Guba (1990) and Denzin and Lincoln (1994a).
- [3] These matters are explored in Gough and Reid (2000), but see Swanborn (1996) and Lincoln (1995) for a more detailed discussion of the implication of a 'gold standards' model for guidelines, particularly when used as a means to 'safeguarding' quality within 'alternative approaches'.
- [4] They include Hammersley (1987), Marshall (1989), Denzin and Lincoln (1994a), Richardson (1994), Elder and Miller (1995) and Chapple and Rogers (1998) among others cited in the references to this article.
- [5] The majority of sources are cited in the references to this article.
- [6] By way of illustration, *rigour*, *trustworthiness* and *authenticity* as developed by Lincoln and Guba (1985) are persistently cited in the fields of nursing, occupational therapy, education, family medicine and human geography. More specifically, *rigor* is often referred to via the recitation of four sub-criteria:
  - *Truth value*: how can one establish confidence in the 'truth' of the findings of a particular inquiry for the subjects (respondents) with which and the context in which the inquiry was carried out?
  - *Applicability*: how can one determine the extent to which the findings of a particular inquiry have applicability in other contexts or other subjects (respondents)?
  - *Consistency*: how can one determine the extent to which the findings of an inquiry would be repeated if the inquiry were replicated with the same (or similar) subjects (respondents) in the same (or similar) contexts?
  - *Neutrality*: how can one establish the degree to which the findings of an inquiry are determined by the subjects (respondents) and conditions of the inquiry and not by the biases, motivations, interests or perspectives of the inquirer?

For a more detailed discussion of the 'sources' for debating criteria and guidelines, see Seale (1999a).
- [7] In addition, it might well be argued that 'disciplinary myopia' remains much in evidence across subject areas, as do, wittingly or not, preferences for particular textual formats, e.g. those authors who tend to rely on anglophonic textbooks by high-profile 'authorities', as opposed to, say, journal articles by less well-known authors in research fields that might well be deemed 'unsexy', irrelevant or unsuitable to a research community or its citation managers. *Qualitative Health Research*, for example, is indexed in a variety of psychological, health and sociological abstracting services, but as of yet, has not been included in MEDLINE, much to the chagrin of its editor, Janice Morse.

- [8] That is, those of Lewis (1981–82), Stewart (1982), Iozzi (1989), Marcinkowski (1988/89), Gigliotti (1990), Leeming *et al.* (1993), Naish (1993), Adler (1993a,b) and Kwong (1995).
- [9] See also the contributions and responses to the theme of 'Beyond paradigms: coherentism and holism in educational research' edited by Lakomski in the *International Journal of Educational Research* (1991), 5(6). <sup>[10]</sup>As an aside, in following Smith (1987) by arguing that journal editors and their boards have a strategic role to play in focusing the substance and forms of such discussions (cf. Morse, 1996, 1997, 1998, 1999), the inclusion of sets of review and refereeing criteria by Smith-Sebasto makes an interesting case in point for the broader literature. Chapple and Rogers (1998, p. 560) highlight how a submission's specifications (particularly those that result in insufficient information for evaluation against criteria), a journal's current and future audience, restrictions against multiple and lengthy publications, and issues relating to dissemination if a qualitative report is published elsewhere, might each affect the response of the reader to qualitative research in their field. A particular dilemma for the report writer it is argued is that, on the one hand, the mechanisms and shorthands for conveying rigorous practice to readers may risk reification and a lack of scrutiny of practices, and on the other hand, research may be misunderstood and judged inferior to other methods that represent the 'gold standard' in that field. Underlying questions then for editors concern what it is that is appropriate and possible, never mind the constraints of the research budget from which the report derives, or what might happen if resubmitted (Peters & Ceci, 1982).
- [11] For example, Hasselkus (1995) read in conjunction with Koch and Harrington (1998), questions whether guidelines for qualitative research in general as derived from ethnography are in some sense superior to those from either phenomenology, grounded theory or hermeneutics.
- [12] See, for example, Buston (1997), Coffey *et al.* (1996), Weitzman and Miles (1995), and Richards and Richards (1994).
- [13] We also note that the results of our search located the majority of the literature on guidelines in these areas, as illustrated in the references.
- [14] These terms are unfortunately used quite interchangeably throughout the literature, and we also note, in some cases, such terms have become little more than pejorative to their opponents, a point which resonates with some of our opening remarks.
- [15] See Gliner (1994); Mays and Pope (1995); Peck and Secker (1999); Rodgers and Cowles (1993); Silverman (1998a); and Table VII. Silverman (1998b) has highlighted four criteria of quality (conceptual depth, rigour, thoughtful research design and practical relevance) to counter his colleagues' reticence:
- (1) How far can we demonstrate that our research has mobilised the conceptual apparatus of our social science disciplines and, thereby, helped to build useful social theories?
  - (2) How far can our data, methods and findings satisfy the criteria of reliability and validity or, put more crudely, counter the cynic who comments 'Sez you'?
  - (3) To what extent do our preferred research methods reflect careful weighting of the alternatives or simple responses to time and resource constraints or even an unthinking adoption of the current fashions?
  - (4) How can valid, reliable and conceptually defined studies of health care processes contribute to practice and policy by revealing something new to practitioners, clients and/or policy makers?
- [16] Drisko (1997) compiles a similar list of criteria for conducting qualitative research and evaluating qualitative research reports in social work: (1) identification of the chosen philosophy/epistemology, (2) identification of audience and objectives, (3) specification of the study method, (4) identification of biases, (5) maintenance of social work ethics, and (6) assurance of consistency between conclusions and study philosophy, objectives, and presented data. Like Seale and Silverman (1997), the author emphasises the need to examine potentially disconfirming data in all qualitative research. Like Burns (1989), producing guidelines will direct qualitative researchers, inform the teaching of qualitative research methods, and aid review of qualitative research reports. In many cases, these lists draw, *inter alia*, on procedures for 'self-documentation' as described in Huberman and Miles (1983) and Miles and Huberman (1984, 1994), and critiqued by Marshall (1985, pp. 354–357).
- [17] Other standards include: methodological congruence; analytic preciseness; theoretical

connectedness; and heuristic relevance. Sub-elements for heuristic relevance, for example, include: rigour in documentation, procedural rigour, ethical rigour, and auditability for methodological congruence; intuitive recognition; relationship to existing body of knowledge, and applicability (see Burns, 1989, pp. 48–51).

- [18] An obvious limitation to our excursions has been our reading of anglophonic works only. Put simply, the references to this article are indicative of what we have and have not accessed.
- [19] Seale suggests this can lead to the abandonment of the research enterprise altogether, particularly if /when postmodernism is taken too seriously.
- [20] See Garratt and Hodkinson (1998) for a more detailed critique of this position in qualitative inquiry.
- [21] See also, for example, the discussions in Heshusius (1990), Garratt and Hodkinson (1998), Silverman (1998b) and Seale (1999a).
- [22] And it appears, seldom are they met in full, the desirability of which anyway has been repeatedly called into question (see Zeller & Farmer, 1999).

## REFERENCES

- ADLER, J. (1993a) A child's garden of misinformation, *Consumer's Research*, 76(September), pp. 11–16.
- ADLER, J. (1993b) The greening of America's youth, *Wall Street Journal*, 14 April, p. A14.
- ALTHEIDE, D. & JOHNSON, J. (1994) Criteria for assessing interpretive validity in qualitative research, in: N. DENZIN & Y. LINCOLN (Eds) *Handbook of Qualitative Research* (London, Sage).
- AMBERT, A., ADLER, P. A., ADLER, P. & DETZNER, D. (1995) Understanding and evaluating qualitative research, *Journal of Marriage and the Family*, 57(4), pp. 879–893.
- APPLETON, J. (1995) Analysing qualitative interview data: addressing issues of validity and reliability, *Journal of Advanced Nursing*, 22(5), pp. 993–997.
- ATKINSON, P. (1995) Some perils of paradigms, *Qualitative Health Research*, 5(1), pp. 117–124.
- AVERY, R. & MCCAIN, T. (1984) Evaluating qualitative research: recognising our personal judgements, paper presented at the 34th Annual Meeting of the International Communication Association, San Francisco, CA, 24–28 May.
- AVIS, M. (1995) Valid arguments: a consideration of the concept of validity in establishing the credibility of research findings, *Journal of Advanced Nursing*, 22(6), pp. 1203–1209.
- BAILEY, P. (1996) Assuring quality in narrative analysis, *Western Journal of Nursing Research*, 18(2), pp. 186–194.
- BAILEY, P. (1997) Finding your way around qualitative methods in nursing research, *Journal of Advanced Nursing*, 25(1), pp. 18–22.
- BAILEY, C., WHITE, C. & PAIN, R. (1999a) Evaluating qualitative research: dealing with the tension between 'science' and 'creativity', *AREA*, 31(2), pp. 169–178.
- BAILEY, C., WHITE, C. & PAIN, R. (1999b) Response, *AREA*, 31(2), pp. 182–183.
- BAXTER, J. & EYLES, J. (1997) Evaluating qualitative research in social geography: establishing rigor in interview analysis, *Transactions of the Institute of British Geographers*, 22(4), pp. 505–525.
- BAXTER, J. & EYLES, J. (1999) Prescription in research practice? Grounded theory in qualitative evaluation, *AREA*, 31(2), pp. 179–181.
- BECK, C. (1993) Qualitative research: the evaluation of its credibility, fittingness, and auditability, *Western Journal of Nursing Research*, 15(2), pp. 263–266.
- BECK, C. (1994) Reliability and validity issues in phenomenological research, *Western Journal of Nursing Research*, 16(3), pp. 254–267.
- BLUMENFELD-JONES, D. (1995) Fidelity as a criterion for practising and evaluating narrative inquiry, *International Journal of Qualitative Studies in Education*, 8(1), pp. 25–35.
- BOULTON, M., FITZPATRICK, R. & SWINBURN, C. (1996) Qualitative research in health care. II: a structured review and evaluation of studies, *Journal of the Evaluation of Clinical Practice*, 2(3), pp. 171–179.
- BOURDIEU, P. (1988) *Homo Academicus* (Cambridge, MA, Polity Press).
- BRITISH SOCIOLOGICAL ASSOCIATION MEDICAL SOCIOLOGY GROUP (1996) Criteria for the evaluation of qualitative research papers, *Medical Sociology News*, 22(1), pp. 69–71.
- BRITZMAN, D. (1995) The question of belief: writing a poststructural ethnography, *International Journal of Qualitative Studies in Education*, 8(3), pp. 229–238.

- BURNS, N. (1989) Standards for qualitative research, *Nursing Science Quarterly*, 2(1), pp. 44–52.
- BUSTON, K. (1997) NUD\*IST in action: its use and its usefulness in a study of chronic illness in young people, *Sociological Research Online*, 2 (3), <<http://www.socresonline.org.uk/socresonline/2/3/6.html>>.
- CAMPBELL, D. & STANLEY, J. (1966) *Experimental and quasi-experimental designs for research* (Chicago, IL, Rand McNally).
- CARR, L. (1994) The strengths and weaknesses of quantitative and qualitative research: what method for nursing?, *Journal of Advanced Nursing*, 20(4), pp. 716–721.
- CERMAK, I. & STEPANIKOVA, I. (1997) Validity in qualitative research in psychology, *Ceskoslovenska Psychologie*, 41(6), pp. 503–512.
- CERMAK, I. & STEPANIKOVA, I. (1998) Data validity control in qualitative psychological research, *Ceskoslovenska Psychologie*, 42(1), 50–62.
- CHAPPLE, A. & ROGERS, A. (1998) Explicit guidelines for qualitative research: a step in the right direction, a defence of the 'soft' option, or a form of sociological imperialism?, *Family Practice*, 15(6), pp. 556–561.
- CLARKE, L. (1992) Qualitative research: meaning and language, *Journal of Advanced Nursing*, 17(2), pp. 243–252.
- COFFEY, A. (1996) The power of accounts: authority and authorship in ethnography, *International Journal of Qualitative Studies in Education*, 9(1), 61–74.
- COFFEY, A., HOLBROOK, B. & ATKINSON, P. (1996) Qualitative data analysis: technologies and representations, *Sociological Research Online*, 1 (1), <<http://www.socresonline.org.uk/socresonline/1/1/4.html>>.
- COOK, T. & CAMPBELL, D. (1979) *Quasi-experimentation: design and analysis issues for field settings* (Chicago, IL, Rand McNally).
- DENZIN, N. (1997) *Interpretive Ethnography: ethnographic practices for the 21st century* (London, Sage).
- DENZIN, N. & LINCOLN, Y. (Eds) (1994a) *Handbook of Qualitative Research* (London, Sage).
- DENZIN, N. & LINCOLN, Y. (1994b) Introduction: entering the field of qualitative research, in: N. DENZIN, & Y. LINCOLN (Eds) *Handbook of Qualitative Research* (London, Sage).
- DRISKO, J. (1997) Strengthening qualitative studies and reports: standards to promote academic integrity, *Journal of Social Work Education*, 33(1), pp. 185–197.
- EDGE, J. & RICHARDS, K. (1998) May I see your warrant, please? Justifying outcomes in qualitative research, *Applied Linguistics*, 19(3), pp. 334–356.
- ELDER, N. & MILLER, W. (1995) Reading and evaluating qualitative research studies, *Journal of Family Practice*, 41(3), pp. 279–285.
- EMDEN, C. & SANDELOWSKI, M. (1998) The good, the bad and the relative. Part one: conceptions of goodness in qualitative research, *International Journal of Nursing Practice*, 4(4), pp. 206–212.
- EMDEN, C. & SANDELOWSKI, M. (1999) The good, the bad and the relative. Part two: goodness and the criterion problem in qualitative research, *International Journal of Nursing Practice*, 5(1), pp. 2–7.
- ENGEL, J. & KUZEL, A. (1992) On the idea of what constitutes good qualitative inquiry, *Qualitative Health Research*, 2(4), pp. 504–510.
- FIELD, P. & MORSE, J. (1985) *Nursing Research: the application of qualitative approaches* (Rockville, MD, Aspen).
- FITCH, K. (1994) Criteria for evidence in qualitative research, *Western Journal of Communication*, 58(1), pp. 32–38.
- FLAHERTY, M. (1996) Some methodological principles from research in practice: validity, truth, and method in the study of lived time, *Qualitative Inquiry*, 2(3), pp. 285–299.
- FUCHS, M. (1993) The reversal of the ethnographical perspective: attempts at objectifying one's own cultural horizon. Dumont, Foucault, Bourdieu? *Thesis Eleven*, 34, pp. 104–125.
- GARRATT, D. & HODKINSON, P. (1998) Can there be criteria for selecting research criteria? A hermeneutical analysis of an inescapable dilemma, *Qualitative Inquiry*, 4(4), pp. 515–539.
- GEERTZ, C. (1988) *Works and Lives: the anthropologists author* (Stanford, CA, Stanford University Press).
- GIGLIOTTI, L. (1990) Environmental education: what went wrong? What can be done? *Journal of Environmental Education*, 22(1), pp. 9–12.
- GLINER, J. (1994) Reviewing qualitative research: proposed criteria for fairness and rigor, *Occupational Therapy Journal of Research*, 14(2), pp. 78–90.

- GOUGH, S. & REID, A. (2000) Environmental education research as profession, as science, as art and as craft: implications for guidelines in qualitative research, *Environmental Education Research*, 6(1), pp. 47–57.
- GUBA, E. (Ed.) (1990) *The Paradigm Dialog* (Newbury Park, CA, Sage).
- GUBA, E. & LINCOLN, Y. (1989) *Fourth Generation Evaluation* (London, Sage).
- GUBA, E. & LINCOLN, Y. (1994) Competing paradigms in qualitative research, in: N. DENZIN & Y. LINCOLN (Eds) *Handbook of Qualitative Research* (London, Sage).
- HALL, J. & STEVENS, P. (1991) Rigor in feminist research, *Advances in Nursing Science*, 13(3), pp. 16–29.
- HAMMERSLEY, M. (1987) Some notes on the terms ‘validity’ and ‘reliability’, *British Educational Research Journal*, 13(1), pp. 73–81.
- HAMMERSLEY, M. (1991) A note on Campbell’s distinction between internal and external validity, *Quality & Quantity*, 25(4), pp. 381–387.
- HAMMERSLEY, M. (1992) Some reflections on ethnography and validity, *International Journal of Qualitative Studies in Education*, 5(3), pp. 195–203.
- HART, P. (1996) Problematising enquiry in environmental education: issues of method in a study of teacher thinking and practice, *Canadian Journal of Environmental Education*, 1(1), pp. 56–88.
- HART, P. (2000) Requisite variety: the problem with generic guidelines for diverse genres of inquiry, *Environmental Education Research*, pp. 37–46.
- HASSELKUS, B. (1991) Qualitative research: not another orthodoxy (Editorial), *Occupational Therapy Journal of Research*, 11(1), pp. 3–7.
- HASSELKUS, B. (1995) Beyond ethnography: expanding our understanding and criteria for qualitative research (Editorial), *Occupational Therapy Journal of Research*, 15(2), pp. 75–84.
- HESHUSIUS, L. (1990) Discussion on criteria, in: E. GUBA (Ed.) *The Paradigm Dialog* (Newbury Park, CA, Sage).
- HODDINOTT, P. & PILL, R. (1997) A review of recently published qualitative research in general practice. More methodological questions than answers? *Family Practice*, 14(4), 313–320.
- HUBERMAN, A. & MILES, M. (1983) Drawing valid meaning from qualitative data: some techniques for data reduction and display, *Quality and Quantity*, 17(3), pp. 281–339.
- HUFF, A. (1994) *Writing for Scholarly Publication* (Thousand Oaks, CA, Sage).
- IMRE, R. (1994) Commentary: epistemological disagreements and the search for understanding, in: E. SHERMAN & W. REID (Eds) *Qualitative Research in Social Work* (New York, Columbia University Press).
- IOZZI, L. (1989) What research says to the educator. Part one: environmental education and the affective domain, *Journal of Environmental Education*, 20(3), pp. 3–9.
- JOHNSON, M. (1999) Observations on positivism and pseudoscience in qualitative nursing research, *Journal of Advanced Nursing*, 30(1), pp. 67–73.
- KIRK, J. & MILLER, M. (1986) *Reliability and Validity in Qualitative Research* (Newbury Park, CA, Sage).
- KNAFL, K. & HOWARD, M. (1984) Interpreting and reporting qualitative research, *Research in Nursing and Health*, 7(1), pp. 17–24.
- KOCH, T. (1994) Establishing rigor in qualitative research: the decision trail, *Journal of Advanced Nursing*, 19(5), pp. 976–986.
- KOCH, T. (1996) Implementation of a hermeneutic inquiry in nursing: philosophy, rigor and representation, *Journal of Advanced Nursing*, 24(1), pp. 174–184.
- KOCH, T. (1998) Story telling: is it really research? *Journal of Advanced Nursing*, 28(6), 1182–1190.
- KOCH, T. & HARRINGTON, A. (1998) Reconceptualising rigor: the case for reflexivity, *Journal of Advanced Nursing*, 28(4), pp. 882–890.
- KVALE, S. (1996) *InterViews: an introduction to qualitative research interviewing* (London, Sage).
- KWONG, J. (1995) On coercive environmental education, *Religion & Liberty*, 5(2), pp. 7–10.
- LAKOMSKI, G. (1991) Beyond paradigms: coherentism and holism in educational research, *International Journal of Educational Research*, 5(6), pp. 501–597.
- LATHER, P. (1991) *Getting Smart: feminist research and pedagogy with/in the postmodern* (New York, Routledge).
- LATHER, P. (1993) Fertile obsession: validity after poststructuralism, *Sociological Quarterly*, 34(4), pp. 673–693.
- LECOMPTE, M. & GOETZ, J. (1982) Problems of reliability and validity in ethnographic research, *Review of Educational Research*, 52(1), pp. 31–60.



- LEEMING, F., DWYER, W., PORTER, B. & COBERN, M. (1993) Outcome research in environmental education: a critical review, *Journal of Environmental Education*, 24(4), pp. 8–21.
- LEWIS, G. (1981–82) A review of classroom methodologies for environmental education, *Journal of Environmental Education*, 13(2), pp. 12–15.
- LINCOLN, Y. (1995) Emerging criteria for quality in qualitative and interpretive research, *Qualitative Inquiry*, 1(3), pp. 275–289.
- LINCOLN, Y. & GUBA, E. (1985) *Naturalistic Enquiry* (Beverly Hills, CA, Sage).
- LINCOLN, Y. & GUBA, E. (1990) Judging the quality of case study reports, *International Journal of Qualitative Studies in Education*, 3(1), pp. 53–59.
- MARCINKOWSKI, T. (1988/89) Commentary on ‘when words speak louder than actions’, *Journal of Environmental Education*, 20(2), pp. 3–5.
- MARCINKOWSKI, T. (2000) The NAAEE workshop on developing guidelines for qualitative research in environmental education: an analysis of premises, processes and products, *Environmental Education Research*, 6(1), pp. 27–35.
- MARSHALL, C. (1985) Appropriate criteria of trustworthiness and goodness for qualitative research on education organisations, *Quality and Quantity*, 19(4), pp. 353–373.
- MARSHALL, C. (1989) Goodness criteria: are they objective criteria or judgement calls? in: E. GUBA (Ed.) *The Paradigm Dialog* (Newbury Park, CA, Sage).
- MAXWELL, J. (1992) Understanding and validity in qualitative research, *Harvard Educational Review*, 62(3), pp. 279–300.
- MAYS, N. & POPE, C. (1995) Rigor and qualitative research, *British Medical Journal*, 311(6997), pp. 109–112.
- MELEIS, A. (1996) Culturally competent scholarship: substance and rigor, *Advances in Nursing Science*, 19(2), pp. 1–16.
- MEULENBERG-BUSKENS, I. (1997) Turtles all the way down? On a quest for quality in qualitative research, *South African Journal of Psychology*, 27(2), pp. 111–115.
- MILES, M. & HUBERMAN, A. (1984) *Qualitative Data Analysis* (Beverly Hills, CA, Sage).
- MILES, M. & HUBERMAN, A. (1994) *Qualitative Data Analysis. An expanded sourcebook*, 2nd edn (London, Sage).
- MOORE, G., BURNETT, M. & MOORE, B. (1986) Approved practices in reporting quantitative research, *Journal of Vocational Education Research*, 11(4), pp. 1–24.
- MORSE, J. (1989) *Qualitative Nursing Research: a contemporary dialogue* (Rockville, MD, Aspen).
- MORSE, J. (1996) What is method? (Editorial), *Qualitative Health Research*, 6(4), pp. 467–468.
- MORSE, J. (1997) ‘Perfectly health, but dead’: the myth of inter-rater reliability (Editorial), *Qualitative Health Research*, 7(4), pp. 445–447.
- MORSE, J. (1998) Validity by committee (Editorial), *Qualitative Health Research*, 8(4), pp. 443–445.
- MORSE, J. (1999) Silent debates in qualitative inquiry (Editorial), *Qualitative Health Research*, 9(2), pp. 163–165.
- MRAZEK, R. (Ed.) (1993) *Alternative Paradigms in Environmental Education Research*, Monographs in Environmental Education and Environmental Studies, Vol. VIII (Troy, OH, The North American Association for Environmental Education).
- NAISH, M. (1993) Forum introduction: ‘Never mind the quality—feel the width’—How shall we judge the quality of research in geographical and environmental education?, *International Research in Geographical and Environmental Education*, 2(1), pp. 64–65.
- PECK, E. & SECKER, J. (1999) Quality criteria for qualitative research: does context make a difference?, *Qualitative Health Research*, 9(4), pp. 552–558.
- PESHKIN, A. (1993) The goodness of qualitative research, *Educational Researcher*, 22(2), pp. 23–29.
- PETERS, D. & CECI, S. (1982) Peer review practices of psychological journals: the fate of published articles, submitted again, *The Behavioural and Brain Sciences*, 5, pp. 187–255.
- PILE, S. (1991) Practising interpretative geography, *Transactions of the Institute of British Geographers*, 16, pp. 458–469.
- POPAY, J., ROGERS, A. & WILLIAMS, G. (1998) Rationale and standards for the systematic review of qualitative literature in health services research, *Qualitative Health Research*, 8(3), pp. 341–351.
- RICHARDS, L. (1999) Data alive! The thinking behind NVivo, *Qualitative Health Research*, 9(3), pp. 412–428.
- RICHARDS, L. & RICHARDS, T. (1994) Using computers in qualitative analysis, in: N. DENZIN & Y. LINCOLN (Eds) *Handbook of Qualitative Research* (Thousand Oaks, CA, Sage).



- RICHARDSON, L. (1994) Writing: a method of inquiry, in: N. DENZIN & Y. LINCOLN (Eds) *Handbook of Qualitative Research* (London, Sage).
- ROBOTTOM, I. & HART, P. (1993) *Research in Environmental Education: engaging the debate* (Geelong: Deakin University Press).
- RODGERS, B. & COWLES, K. (1993) The qualitative research audit trail: a complex collection of documentation, *Research in Nursing and Health*, 16(3), pp. 219–226.
- ROSE, G. (1982) *Deciphering Sociological Research* (London, Macmillan).
- SANDELOWSKI, M. (1986) The problem of rigor in qualitative research, *Advances in Nursing Science*, 8(3), pp. 27–37.
- SANDELOWSKI, M. (1993) Rigor or rigor-mortis: the problem of rigor in qualitative research revisited, *Advances in Nursing Science*, 16(2), pp. 1–8.
- SCHEURICH, J. (1997) *Research Method in the Postmodern* (London, Falmer).
- SCHWANDT, T. (1994) Constructivist, interpretivist approaches to human inquiry, in: N. DENZIN & Y. LINCOLN (Eds) *Handbook of Qualitative Research* (London, Sage).
- SCHWANDT, T. (1996) Farewell to criteriology, *Qualitative Inquiry*, 2(1), pp. 58–72.
- SEALE, C. (1999a) Quality in qualitative research, *Qualitative Inquiry*, 5(4), pp. 465–478.
- SEALE, C. (1999b) *The Quality of Qualitative Research* (London, Sage).
- SEALE, C. & SILVERMAN, D. (1997) Ensuring rigor in qualitative research, *European Journal of Public Health*, 7(4), pp. 379–384.
- SECKER, J., WIMBURNE, E., WATSON, J. & MILBURN, K. (1995) Qualitative methods in health promotion research: some criteria for quality, *Health Education Journal*, 54(1), pp. 74–87.
- SILVERMAN, D. (1998a) Qualitative research: meanings or practices?, *Information Systems Journal*, 8(1), pp. 3–20.
- SILVERMAN, D. (1998b) The quality of qualitative health research: the open-ended interview and its alternatives, *Social Sciences in Health*, 4(2), pp. 104–118.
- SMITH, J. (1990) Alternative research paradigms and the problem of criteria, in: E. GUBA (Ed.) *The Paradigm Dialog* (Newbury Park, CA, Sage).
- SMITH, M. (1987) Publishing qualitative research, *American Educational Research Journal*, 24, pp. 173–184.
- SMITH-SEBASTO, N. (2000) An exploration of potential guidelines for conducting and reporting environmental education research: qualitative methods of inquiry, *Environmental Education Research*, pp. 9–26.
- SØRENSEN, N. (1997) The problem of parallelism: a problem for pedagogic research and development seen from the perspective of environmental and health education, *Environmental Education Research*, 3(2), pp. 179–187.
- STEWART, J. (1982) Empirical research reported in the Journal of Environmental Education: a critique, *Journal of Environmental Education*, 14(1), pp. 42–44.
- STILES, W. (1993) Quality-control in qualitative research, *Clinical Psychology Review*, 13(6), pp. 593–618.
- STRONACH, I. & MACLURE, M. (1997) *Educational Research Undone: the postmodern embrace* (Buckingham, Open University Press).
- SWANBORN, P. (1996) A common base for quality control criteria in quantitative and qualitative research, *Quality and Quantity*, 30(1), pp. 19–35.
- TOMPKINS, P. (1994) Principles of rigor for assessing evidence in 'qualitative' communication research, *Western Journal of Communication*, 58(1), pp. 44–50.
- WEITZMAN, E. & MILES, M. (1995) *Computer Programs for Qualitative Data Analysis* (Thousand Oaks, CA, Sage).
- ZARUBA, K., TOMA, J. & STARK, J. (1996) Criteria used for qualitative research in the refereeing process, *Review of Higher Education*, 19(4), pp. 435–460.
- ZELLER, N. & FARMER, F. (1999) 'Catchy, clever titles are not acceptable': style, APA, and qualitative reporting, *International Journal of Qualitative Studies in Education*, 12(1), pp. 3–19.