Are There Assessment Criteria for Qualitative Findings?
A Challenge Facing Mixed Methods Research

Martyn Hammersley

Abstract: If findings from qualitative and quantitative components in mixed methods research are to be synthesised, the quality of each must be assessed. But an obvious problem is that there are no generally agreed criteria for assessing qualitative findings. The question of criteria has long been debated in the methodological literature. I argue that some important distinctions need to be made if progress is to be achieved on this issue. Perhaps the most important one is between the standards in terms of which assessment is carried out and the indicators used to evaluate findings in relation to those standards. I go on to outline what I believe is involved in such evaluations, rejecting the possibility of a detailed and explicit set of indicators that can immediately be used to determine the validity of knowledge claims. My approach broadly fits the framework of mixed methods research, since I deny that there is any fundamental philosophical difference between quantitative and qualitative methods. But it is at odds with widespread views, even within the realm of mixed methods, whose advocates seek radically to redefine the ontological, epistemological, and/or axiological assumptions of social scientific research, for example in the name of a *transformative* approach.

Table of Contents

1. Introduction

Assessment of the quality of findings from different sources is clearly central to mixed methods research (BRYMAN, BECKER & SEMPIK, 2008; DELLINGER & LEECH, 2007). This is an issue that has often been regarded as peculiarly problematic in the case of qualitative findings, since it is frequently assumed that there are well-established criteria for quantitative results. In response, efforts have been made to show how quantitative criteria could be applied to qualitative work, or to develop distinctive qualitative criteria.[1] However, conflicting sets of

---

1 For accounts of the various proposals, see HAMMERSLEY (2018a [1992]) and SPENCER, RITCHIE, LEWIS and DILLON (2003). More generally on the issue of criteria, see: ATTREE and MILTON (2006), BARBOUR and BARBOUR (2002), BERGMAN and COXON (2005), BREUER and REICHERTZ (2001), CALDERÓN GÓMEZ (2009), ELLIOTT, FISHER and RENNIE (1999), EMDEN and SANDELOWSKI (1999), FLICK (2007, 2018), HAMMERSLEY (2007, 2008a,

criteria have been developed, reflecting the deep divisions among qualitative researchers, these involving discrepant ontological, epistemological, and/or axiological assumptions. Indeed, some have raised doubts about the very possibility or desirability of criteria (SMITH, 2004). Thus, BOCHNER (2000, p.266) wrote that: "[t]he demand for criteria reflects the desire to contain freedom, limit possibilities, and resist change". The issue is further complicated by the promotion of "arts-based research" (LEAVY, 2019), as well as declarations about the *post-qualitative* inspired by *new materialisms* (LATHER, 2013; LATHER & ST PIERRE, 2013).[2] [1]

So how are we to respond to this impasse? It seems to me that there are some prior questions and assumptions that need attention before we can even attempt to answer the question of whether there can be criteria for assessing qualitative findings, or what the criteria should be. One of these is what the word *criterion* means in this context. [2]

2. What Are Criteria?

To answer this question, we need to step back in order to examine the nature of the evaluation process involved when researchers assess the results of one another's, and of their own, work. I think we can draw some lessons from the field of evaluation studies here (my source is SHADISH, COOK & LEVITON, 1991). We need to distinguish among the following, all of which have sometimes been treated as *criteria*:

- the *standards* or *dimensions* in terms of which research is judged;
- the *benchmarks* on these dimensions used to produce the evaluation (sound/unsound, good/bad, etc.): these could be an average, or a threshold (upper or lower), that allows research to be placed in one or other category according to differential quality;
- the *indicators* or *signs* used to decide where on each dimension particular pieces of research or findings lie. [3]

2008b, 2009), HATCH (2007), LEE (2014), LINCOLN (1995), MAYS and POPE (1995), REICHERTZ (2019), SANDELOWSKI (1978, 1993), SEALE (1999), SMITH (1990), SMITH and HODKINSON (2005, 2009).

2 The meaning of "post-qualitative" is by no means entirely clear (KUNTZ, 2020). However, advocates of post-qualitative approaches explicitly reject the *humanism* of previous qualitative work, with its representatives emphasising the need to understand human social life in a different way from how we explain the behaviour of animals and physical objects. Also rejected, frequently, is the secular orientation characteristic of much Western science, with social scientists denying the validity of non-naturalistic interpretations of social phenomena. What is proposed, instead, is ethical engagement with the world to enact new, more socially just possibilities. Advocates of this post-secular understanding of social life often draw upon indigenous peoples' ideas, according to which nature and culture are not separated but intertwined or even blended, so that animals, plants, and physical objects, as well as human beings, are treated as spiritual as well as material in character. As should be clear, the boundaries around research have been erased here, and this renders uncertain the question of what would be appropriate criteria of assessment for research accounts, and even whether criteria of assessment are appropriate.

Quite a common tendency, not least in the context of mixed methods research, is to focus on criteria as *indicators that can serve as direct and conclusive signs of quality* (BRYMAN et al., 2008; HEYVAERT, HANNES, MAES & ONGHENA, 2013). What is often demanded is a checklist of immediately intelligible, necessary and jointly sufficient, conditions for findings to be accepted or rejected.[3] There are at least two reasons why this interpretation of *criteria* has been popular. First, those judging research findings (such as policymakers or occupational practitioners using research) often want something of this kind because they thereby avoid having to make a judgment for themselves that would be laborious and/or that they do not have the necessary expertise to make. A second factor, reinforcing this, is the ideological demand for *transparency* of assessment, by which is meant that how research findings were produced should be laid out explicitly by researchers, and that there ought to be a set of indicators (of the determinate kind just mentioned) as to whether these findings are sound. This sort of transparency is sometimes regarded as essential to the intellectual authority of science. In other words, lay users of research should not have to rely on researchers' own judgments about the quality of their work, since these judgments are *subjective* and therefore open to bias. This argument is directly analogous to the demands for *transparency* applied to other professions, and the associated distrust of professional judgment (POWER, 1997). [4]

Part of the background to this is the common assumption that *transparent* and determinate criteria are already available for quantitative research. These criteria are frequently taken to include such matters as the following: *Was random allocation used?*, *Was random sampling employed?*, and *Were reliability tests applied that produced reasonably high scores?.* However, I suggest that if we were to examine how quantitative researchers actually evaluate quantitative studies we would find that this does not involve the application of a set of necessary and sufficient conditions, or even of a list of probabilistic indicators for counting up a score to determine whether some pre-set threshold has been reached. Rather, quantitative researchers necessarily engage in a process of interpretive judgment, weighing up various aspects of a study and/or of the findings produced.[4] [5]

We should also note that at least two of the examples of quantitative indicators I mentioned above cannot be applied across the board to all kinds of quantitative work. For instance, in survey research random allocation is rarely used, and researchers engaging in experimental work do not normally draw on random samples from a population. Moreover, this is not a matter of happenstance, it reflects the character of these methods. A second point is that these indicators do not cover all relevant aspects of studies that would need to be taken into account in assessing quality, and it is not clear that any coherent full set could be agreed upon. Equally important, none of the putative indicators would normally be accepted at face value by researchers. For instance, as regards random

---

3   On the emergence of what we might call *checklist culture*, see BOYLE (2020).

4   This was illustrated a long time ago in assessments of landmark studies: see CHRISTIE and JAHODA (1954), ELASHOFF and SNOW (1971), MERTON and LAZARSFELD (1950). It is increasingly recognised explicitly in textbooks dealing with quantitative method.

allocation it would usually be asked how effectively this had been implemented. And other considerations would also be taken into account, for example whether there was double-blinding, *and how effective this was likely to have been*. Meanwhile, in the case of random sampling, there will be questions about any stratification of the sample employed, the level of non-response, and so on. Multiple considerations are involved, often specific to the method employed, and these have to be interpreted in the context of the particular study. While there may be rules of thumb that are highly effective, they are no more than this. [6]

So, the differences between qualitative and quantitative approaches in this regard are not as great as often supposed. This illustrates, I suggest, that in social research, as with many other professional activities, full *transparency* is not possible: expert and experienced *judgment* is always required. While life would be simpler if direct and indubitable indicators of research quality were available, they are not and cannot be, even in relation to quantitative inquiry. And, given this, there is no good reason to expect them in the case of qualitative work. This raises difficult questions about the evaluation of research studies and findings in policy and practice contexts, but I will leave these on one side here (for a discussion, see HAMMERSLEY, 2013a, pp.74-76). [7]

So, in assessing their own work, as well as of that of colleagues in the same field, researchers rely upon a learned capacity for evaluation rather than on a process of *applying* explicit criteria that could be carried out by anyone, whatever their background experience and knowledge. Much the same is true of most other specialised occupations. As a result, any list of criteria in the form of indicators can only be an abstract and partial representation of what exercising this capacity involves. Of course, the degree of agreement among researchers in their judgments can vary. It may be true that among experimenters or survey researchers assessment takes place on the basis of a shared capacity for judgment that results in a large degree of consensus. Whereas, patently, given the divisions mentioned earlier, this is not true of qualitative research.[5] [8]

It might be concluded from my discussion so far that I believe that research assessment is *simply* a matter of judgment—that there is no role for explicitly stated *criteria* of any kind. While there are those who seem to believe this, I am not one of them (HAMMERSLEY, 2009). Among researchers, and others, stated criteria can be used as guides, not least in reminding us of what ought to be taken into account. Moreover, I do not believe that any sound learned capacity for assessment comes about *naturally*, in other words without reflection; nor that it is stable unless there is at least informal coordination amongst researchers within a field on the basis of such reflection. So, for example, accounts of useful signs, and of how these should be interpreted, are of value; even though they cannot serve as direct, infallible indicators. Indeed, I think they are essential to scientific work. Having sought to clarify what the term *criterion* means, I now want to turn back to the question of how qualitative findings are to be assessed. [9]

---

5   For evidence that this may not even be true of quantitative research, see for instance
    http://oxfordsociology.blogspot.com/ [Accessed: June 10, 2022].

3. Assessing the Validity of Qualitative Research Findings

If we assume, for the moment, that research is solely aimed at producing value-relevant knowledge, then the only *standards* for assessing findings concern their validity (in the sense of empirical truth) and their value-relevance (HAMMERSLEY, 2016, 2017a). I will restrict my discussion here to the first standard. In these terms, for research findings to be accepted they must meet some benchmark of likely validity, *one that is relatively high compared with those that operate in many everyday contexts* (HAMMERSLEY, 2011). However, what will need to be taken into account (i.e. what are appropriate *indicators*) in judging this will depend upon the form of knowledge claim involved. I suggest there are three legitimate types of knowledge claim made in social research (given the above specification of its task):

- *Descriptions*: These may be restricted to the phenomena studied, or be used as a basis for generalisation to a larger population.[6] An example would be*: In the workplace, bullying of women takes different (specified) forms from bullying of men (in the cases studied or more generally).* In order to evaluate such a description, we require indications of how well *bullying* was defined and instances of it identified, and how well the cases discussed were investigated. If generalisation from sample to population is involved then we need to know how members of the sample were selected, and/or how representative they are likely to be of the population.

- *Explanations*: Here, some feature of an object, or some event, is portrayed as occurring *because* a particular set of factors caused it, directly or indirectly. An example would be: *The increase in level of house burglaries and street theft is a consequence of changes in the state benefit system, leaving many poor people inadequate means of subsistence*. Here we would need not only descriptions of changes in the crime rate, and of the timing and character of changes in the benefits regime, but also evidence to the effect that the claimed causal relation was in operation.

- *Theories*: In putting forward a theory we are claiming that a particular *type* of feature or event (or variation in it) is *generally* brought about by the occurrence of, or variation in, some set of prior *types* of feature or event, this under specified conditions or *other things being equal.* An example would be: *smaller classes usually result in an increase in children's learning*. In order to obtain cogent evidence for the validity of this theory, examination of multiple cases would be required, ones where the causal variable being investigated (class size) takes different values and where other potential factors affecting the outcome (such as prior level of achievement on the part of members of the classes) were constant. Also required is accurate description/measurement (of class size and degree of learning) in each case; as well as careful description of putative causal processes linking the two. [10]

---

6   In my view, a distinction needs to be drawn between such empirical generalisation and theoretical inference, the latter involving inference from data to the truth or falsity of a theory that has universal application: see HAMMERSLEY (2018a).

As should be clear, each of these types of claim must be judged on somewhat different, albeit overlapping, grounds. Elsewhere, I have argued that in the process of assessing research findings judgments about their plausibility (this concerns their relationship to existing knowledge) and the credibility of the evidence offered in support of them will be involved (HAMMERSLEY, 2018a).[7] It is very unlikely that the main findings of a study will be sufficiently plausible to be accepted at face value, since if this were so they would have little news-value (one aspect of value-relevance). It is, of course, possible that what is judged to be sufficiently plausible is nevertheless false, but, in principle at least, a well-functioning research community will discover such errors over time. [11]

In assessing plausibility, research findings may be judged to be logically implied by existing knowledge, strongly suggested by it, compatible with it, in tension with it, or incompatible with it. Where a research finding is regarded as insufficiently plausible to be accepted at face value (and, as already suggested, this will almost always be the case), *it is not rejected but is subject to scrutiny in terms of the evidence offered or available to support It, and whether there is any that counts against it*. This evidence, too, must be assessed in terms of *its* plausibility, but will also need to be judged both in terms of how strongly it implies the truth of the finding and how *credible* it is—this concerns how great is any threat to its validity given how it was produced (for instance, we would generally give more credence to evidence coming from direct observation than to hearsay). Moreover, it may be necessary to have *further* evidence showing the reliability of the evidence offered for the main claims. This will at least partly take the form of information about how the research was carried out, but it could also involve evidence from other sources than that initially relied upon. This is a process that should continue until a conclusion is reached: either that there is sufficient evidential support to decide whether the knowledge claim being assessed is true or false (in terms of the appropriate benchmark) or that, in the absence of sufficient further evidence or information, judgment must be suspended (see Figure 1).

---

7 It seems obvious (to me, at least) that plausibility and credibility (in the senses of these terms used here) are the means by which, in everyday life, all of us judge knowledge claims in epistemic terms. However, there are two distinctive aspects of researchers' use of these *criteria*. The first concerns what is taken to be existing knowledge: researchers should rely predominantly on previous research findings, though some widely accepted common sense knowledge will also be employed, including the foundations of what HUSSERL referred to as the *natural attitude* (LUFT, 2002). The second difference is the adoption of a more deliberative approach in which, for example, the relationship between a prospective knowledge claim and what is taken to be already known is very carefully examined. And, in the case of credibility, there will be reliance upon expertise about the threats to validity associated with particular sources of data.
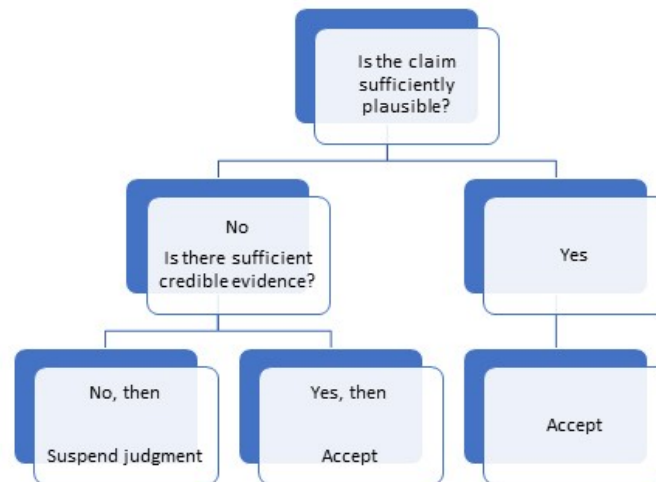
Figure 1: The process of evaluating research findings [12]

4. Are There Distinctive Criteria for Qualitative Research?

In my view, qualitative and quantitative findings should be judged by the same *standards*: validity and value-relevance. Moreover, in relation to both of these standards a relatively high threshold or benchmark will have to be met in order for the claim to social science expertise to be justified. Where any difference between assessment of quantitative and qualitative research arises is solely in the *indicators* used to determine the likely validity of knowledge claims, specifically as regards the credibility of evidence. [13]

Furthermore, such differences lie not at the level of quantitative versus qualitative *approaches* themselves, but are related instead to how one judges data and findings produced by *particular methods*. Different indicators are relevant, for example, in judging findings coming from participant observation versus those generated by unstructured interviewing, or those arising from theme analysis as opposed to those produced by discourse analysis. This parallels the fact that, as I noted earlier, the same indicators cannot be applied across all quantitative research: generally speaking, the question of whether random sampling was employed in a study only relates to surveys; that of whether random allocation was used mostly only applies to experiments. [14]

As an illustration of what is involved in assessing qualitative evidence, take the case of a descriptive finding derived from an interview with a school head teacher: That his relationship with his staff was autocratic, in the sense that he made decisions on his own and imposed them irrespective of whether the staff agreed with him.[8] I think we can judge this knowledge claim as by no means completely implausible at face value—some head teachers in British schools have been (and perhaps still are) autocratic, in a broad sense of that term. In the interview the head says *I do not believe that it is the job of a leader to be a democrat*, and reports that he appointed a head of drama against the opposition

---

8   For further discussion of data from this interview and its analysis see HAMMERSLEY (2017b).

of many of the staff. His statement of belief here certainly seems to indicate that his orientation may have been autocratic, but this is not sufficient on its own. Indeed, we might need to consider how it arose within the interview, and what he meant about not being a *democrat*: what sorts of behaviour was he ruling out? In the case of his own account of how he appointed a head of drama, we need to determine whether this account is accurate. We also must determine whether this indicates a stable pattern of action on his part. So, in assessing these interview data we may conclude that further evidence is necessary, for example of an observational kind and/or from interviews with other key members of the school staff. On top of this, though, there are some general questions we ought to ask in using interview data, about how the data were produced. These would include:

- What was the interviewer's relationship to the person/people interviewed?
- Where was the interview carried out: on whose territory, or was it in a neutral place?
- To what extent, and in what ways, were the data co-constructed, for example through the questions asked?
- Did the informant have any reason to lie about matters relating to the questions? [15]

These issues are not easy to address, but we must make the best assessment we can. In this case, the interviewer was a young PhD student, though the head teacher may have felt he was speaking to a wider audience through the student's research. The interview took place in the head's office, so very much in *his* territory. There was relatively little co-construction: much of the interview was a monologue. All this might lead us to suspect, for example, that the head exaggerated his own role, but we should not assume that this makes the knowledge claim about his autocratic orientation wholly inaccurate: it is a hypothesis that needs to be checked on the basis of further evidence. [16]

As this illustration makes clear, the assessment of qualitative findings is necessarily reliant on judgment, and this must be tailored to the particular knowledge claims and data involved. Some commentators have put forward certain techniques as short-cuts to conclusions about the validity of qualitative findings. Yet, neither triangulation (HAMMERSLEY, 2008c), nor respondent validation/member-checking (HAMMERSLEY & ATKINSON, 2019), nor the provision of an audit trail (HAMMERSLEY, 1997), can be taken as absolutely conclusive proof. While they can be used for generating valuable further evidence, there is no alternative to the process of assessment I outlined earlier. At the same time, it should be clear that this is not an entirely arbitrary process: by checking we can discover errors and revise what we take to be knowledge, increasing its likely validity. The contrast between *objective* criteria and *subjective* judgment is a false one. [17]

While I believe that the approach I have outlined is how the validity of knowledge claims should be assessed in social science, it fails to result in the *transparent* determinate proof that some seek in calling for criteria of assessment.

Furthermore, it involves assumptions that many qualitative researchers, and some mixed methods researchers, would reject. The latter point requires further attention. [18]

5. The Present Impasse

 As I noted earlier, there is little agreement about the issue of criteria among qualitative researchers, and there are those who question whether disagreement about the quality of qualitative work is actually to be deplored; indeed, some have suggested that it ought to be celebrated (BOCHNER, 2000; HAMMERSLEY, 2005; SMITH, 2004).[9] This reflects the fact that qualitative research is fractured into a considerable variety of approaches that differ in fundamental ways. They often vary even concerning the purpose or intended product of inquiry (in other words, in *axiological* terms), so we can ask whether the primary aim of research is taken to be:

1. solely to produce value-relevant knowledge? Or
2. to improve policymaking or practice? And/or
3. to challenge the socio-political status quo? And/or
4. to exemplify ethical or political ideals, for example those enshrined in versions of feminism or disability activism? [19]

Today, researchers frequently have more than one of these aims in mind, yet they result in conflicting implications about how best to pursue a particular piece of research, with the consequence that satisfying one of them means failing to satisfy others. In my view, the first of these goals is the only legitimate operational goal of any form of inquiry; though it is important to distinguish between the aim of research and the motives we may have for engaging in it—between what we set out to produce in doing research (knowledge), and the reasons why we think producing knowledge of the kind intended is worthwhile, what we hope we will achieve by doing this. Thus, 2, 3, and 4 are quite reasonable *motives* for engaging in inquiry, even though they are not appropriate operational goals: they should not govern how we carry out any social science inquiry.[10] [20]

There are also differences among qualitative researchers in their ontological and epistemological assumptions, these being concerned with the nature of social processes and how we can best understand them. A few key questions here include:

· Are they subjective phenomena that should be studied using *thick description* (GEERTZ, 1973)?
· Are they public forms of behaviour that can be objectively described and explained in qualitative terms? For example, the sequential organisation of talk-in-interaction studied by conversation analysts.

---

9 For responses to these arguments, see HAMMERSLEY (2008a, 2009).

10 In my view, ethical considerations are important *external* constraints on how research is pursued: see HAMMERSLEY and TRAIANOU (2012).

- Are they discursively constituted, so that the task of researchers is to document the process of constitution (POTTER, 1996)? [21]

The answers to these (and related) questions can have divergent implications for the nature of the data and evidence required in qualitative investigations, and for how findings should be evaluated. Disparities in axiological, ontological, and epistemological assumptions are reflected not just in differences between such approaches as ethnography, biographical/autobiographical work, discourse and narrative analysis, action research, and so on, but also in the existence of *diverse forms of each of these*. Furthermore, social researchers can make conflicting *theoretical* assumptions—from activity theory to phenomenology to symbolic interactionism to feminist standpoint theory, etc. To take just one example, as regards discourse analysis, REICHER (2000) highlighted a fundamental epistemological/ontological division that arises in relation to the status of interview or documentary data: are these data to be used for what information they convey about the world or for how they construct the phenomena to which they refer?[11] [22]

We must recognize that these fundamental differences in purpose and assumption can have very significant implications for how qualitative evidence should be assessed. And they represent a major obstacle blocking the way to agreement about this, whether in judgments about particular studies or in ideas about what are appropriate *criteria*. To follow up on REICHER's point, while some would treat data from unstructured or semi-structured interviews as evidence about the perspectives of informants, or even as information about events informants have experienced or witnessed, others reject such data as worthless for any purpose *other than investigating how talk is organised, or how reality is discursively constructed, within interviews* (HAMMERSLEY, 2017b). [23]

It seems to me that, in part at least, the axiological differences reflect a weakness in the boundary between research, on the one hand, and politics, policymaking, and other types of practice, on the other. Indeed, there are some social scientists who appear to want to erase such boundaries. Similarly, the differences in epistemology and ontology are derived from a weakened boundary between social research, on the one hand, and the humanities and arts, especially literary studies, philosophy, and imaginative literature, on the other. In my view, both these boundaries must be retained, because what is on each side of them varies in the nature of the goals pursued—for me, the only operational goal of social science is the production of descriptive, explanatory, or theoretical knowledge.[12] [24]

The key question is, of course, whether these disagreements can be resolved, and if so how. Some epistemological and ontological disagreements may be resolvable through comparing research studies based on different assumptions— in order to determine which are the most successful in answering research questions. But some of the disagreements, particularly axiological ones, are not

---

11 On what has been called the *radical critique of interviews*, see MURPHY, DINGWALL, GREATBATCH, PARKER and WATSON (1998), HAMMERSLEY (2008d, 2013b), and POTTER and HEPBURN (2005).

12 This relates to different kinds of expertise, on which see COLLINS (2014).

resolvable in this way because different criteria of success are associated with different goals. [25]

6. Qualitative Criteria and Mixed Methods Research

As I noted at the beginning of this paper, one of the main contexts where we encounter the issue of qualitative criteria is in mixed methods studies. And the account of the assessment process I have offered here is certainly compatible with the idea of mixing or combining quantitative and qualitative methods and findings. At the level of standards, I denied the distinction between the two approaches, and at the level of indicators I focused on specific methods. In these terms the problems surrounding how to assess qualitative evidence may appear to be largely displaced and therefore perhaps resolvable within mixed methods inquiry. [26]

However, there are aspects of the position I have taken that may be at odds with the views of some mixed methods researchers:

5. A first problem is that, in practice, mixed methods researchers retain the qualitative-quantitative distinction as the basis for mixing methods (SCHOONENBOOM & JOHNSON, 2017; SYMONDS & GORARD, 2010). In my view, that distinction needs to be abandoned in favour of a framework within which the choices of strategy that arise in different phases of the research process can be identified (HAMMERSLEY, 2018b). There are parallels here with the idea of *merging* methods (GOBO, FIELDING, LA ROCCA & VAN DER VAART, 2021). At the same time, by viewing research practices in terms of such a framework the process of *mixing* is complicated further, in that the forms this can take are multiplied.

6. A second problem is the tendency of some advocates of mixed methods to demand determinate or procedural criteria of assessment (see, for instance, O'CATHAIN, 2010). The aim of these researchers by demanding such proceduralisation is to minimize the role of judgment, often on the false assumption that it is a subjective source of error. As I have explained, this is not a fruitful approach; it is necessary to accept the ineradicable role of judgment in the process of assessment, whether the focus is qualitative or quantitative findings, and to recognise that this does not render it simply arbitrary, so long as precautions are taken to counter potential threats to validity.

7. A final problem is perhaps the most serious: the reproduction of axiological, ontological, and epistemological divisions *within* the mixed methods movement (GREENE, 2008; HARRITS, 2011; TASHAKKORI & TEDDLIE, 2010). Rather than tolerance for sharply different mixed methods paradigms, what is required, in my view, is a clear resolve to determine the boundaries of what counts as social scientific research, along with the adoption of an instrumental attitude towards ontological, and perhaps also towards epistemological, assumptions, with a view to reducing paradigm differences.

However, there may be as little appetite for this among mixed methods researchers as there is, currently, amongst those in the qualitative camp. [27]

7. Conclusion

There are no widely-accepted criteria for assessing qualitative findings, and there is such deep division among qualitative researchers that there is little prospect there for agreement; not just about what criteria would be appropriate but even about the possibility and desirability of criteria. Moreover, I have argued that if by *criteria* we mean *transparent and determinate indicators that anyone can use to assess the quality of research*, then there are, and can be, no assessment criteria for quantitative research either. I have argued that it is essential to recognise the distinction between standards and indicators, and that even indicators can be no more than probabilistic signs. I insisted that the same *standards* should apply to both quantitative and qualitative findings: truth and value-relevance. Furthermore, when it comes to indicators, these are specific to particular methods rather than being applicable across the board within quantitative or qualitative approaches. I also outlined the process by which the validity of research findings, quantitative or qualitative, ought to be assessed, relying on judgements about plausibility and credibility. [28]

So, while "research quality", in various senses, can be assessed in a rational and effective fashion, in my view this does not involve the kind of determinate indicators that are often assumed when people use the term *criteria*, and which are frequently felt to be needed. Furthermore, my account would not be accepted by many qualitative and some mixed-methods researchers, and perhaps not by all quantitative researchers either, because they will reject key assumptions on which it is based: about the purpose of research, about the nature of social phenomena, and/or regarding the very possibility of knowledge about these phenomena. In this respect there is a fundamental impasse over the question of criteria for assessing qualitative evidence; and, unfortunately, it is one that even a commitment to mixing methods does not automatically overcome. The fact that competing paradigms have emerged within the mixed methods community makes this clear. [29]

References

Attree, Pamela & Milton, Beth (2006). Critically appraising qualitative research for systematic reviews: Defusing the methodological cluster bombs. *Evidence & Policy*, *2*(1), 109-126.

Barbour, Rosaline & Barbour, Michael (2002). Evaluating and synthesizing qualitative research: The need to develop a distinctive approach. *Journal of Evaluation in Clinical Practice*, *9*(2), 179-186.

Bergman, Max & Coxon, Anthony (2005). The quality in qualitative methods. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *6*(2), Art. 34, https://doi.org/10.17169/fqs-6.2.457 [Accessed: March 11, 2022].

Bochner, Arthur (2000). Criteria against ourselves. *Qualitative Inquiry, 6*(2), 266-272.

Boyle, David (2020). *Tickbox*. London: Little, Brown.

Breuer, Franz & Reichertz, Jo (2001). Standards of social research. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *2*(3), Art. 24, https://doi.org/10.17169/fqs-2.3.919 [Accessed: March 11, 2022].

Bryman, Alan; Becker, Saul & Sempik, Joe (2008). Quality criteria for quantitative, qualitative and mixed methods research: A view from social policy. *International Journal of Social Research Methodology*, *11*(4), 261-276.

Calderón Gómez, Carlos (2009). Assessing the quality of qualitative health research: Criteria, process and writing. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *10*(2), Art. 17, https://doi.org/10.17169/fqs-10.2.1294 [Accessed: March 11, 2022].

Christie, Richard & Jahoda, Marie (Eds.) (1954). *Studies in the scope and method of "The authoritarian personality"*. Glencoe, IL: Free Press.

Collins, Harry (2014). *Are we all scientific experts now?.* Cambridge: Polity.

Dellinger, Amy & Leech, Nancy (2007). Toward a unified validation framework in mixed methods research. *Journal of Mixed Methods Research*, *1*(4), 309-332.

Elashoff, Janet & Snow, Richard (Eds.) (1971). *Pygmalion reconsidered*. Worthington, OH: Charles A. Jones.

Elliott, Robert; Fisher, Constance & Rennie, David (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology*, *38*(3), 215-229.

Emden, Carolyn & Sandelowski, Margarete (1999). The good, the bad, and the relative, Part 2: Goodness and the criterion problem in qualitative research. *International Journal of Nursing Practice*, *5*(1), 2-7.

Flick, Uwe (2007). *Managing quality in qualitative research*. London: Sage.

Flick, Uwe (2018). *An introduction to qualitative research* (8th ed.). London: Sage.

Geertz, Clifford (1973). *The interpretation of cultures*. New York, NY: Basic Books.

Gobo, Giampietro; Fielding, Nigel G.; La Rocca, Gevisa & van der Vaart, Wander (2021). *Merged methods*. London: Sage.

Greene, Judith (2008). Is mixed methods social inquiry a distinctive methodology?. *Journal of Mixed Methods Research, 2(*1), 7-22.

Hammersley, Martyn (1997). Qualitative data archiving: Some reflections on its prospects and problems. *Sociology*, *31*(1), 131-42.

Hammersley, Martyn (2005). Countering the "new orthodoxy" in educational research: A response to Phil Hodkinson. *British Educational Research Journal*, *31*(2), 139-155.

Hammersley, Martyn (2007). The issue of quality in qualitative research. *International Journal of Research and Method in Education*, *30*(3), 287-306.

Hammersley, Martyn (2008a). Troubling criteria: A critical commentary on Furlong and Oancea's framework for assessing educational research. *British Educational Research Journal*, *34*(6), 747-762.

Hammersley, Martyn (2008b). Assessing validity in social research. In Pertti Alasuutari, Leonard Bickman & Julia Brannen (Eds.), *The Sage handbook of social research methods* (pp.42-52). London: Sage.

Hammersley, Martyn (2008c). Troubles with triangulation. In Max Bergman (Ed.), *Advances in mixed methods research* (pp.22-36). London: Sage.

Hammersley, Martyn (2008d). *Questioning qualitative inquiry*. London: Sage.

Hammersley, Martyn (2009). Challenging relativism: The problem of assessment criteria. *Qualitative Inquiry*, *15*(1), 3-29.

Hammersley, Martyn (2011). *Methodology, who needs it?.* London: Sage.

Hammersley, Martyn (2013a). *The myth of research-based policy and practice*. London: Sage.

Hammersley, Martyn (2013b). *What is qualitative research?.* London: Bloomsbury.

Hammersley, Martyn (2016 [1998]). *Reading ethnographic research*. London: Longmans.

Hammersley, Martyn (2017a). On the role of values in social research: Weber vindicated?. *Sociological Research Online*, *22*(1), 7, https://www.socresonline.org.uk/22/1/7.html [Accessed: March 11, 2022].

Hammersley, Martyn (2017b). Interview data: A qualified defence against the radical critique. *Qualitative Research*, *17*(2), 173-186.

Hammersley, Martyn (2018a [1992]). *What's wrong with ethnography?*. London: Routledge.

Hammersley, Martyn (2018b). What is ethnography? Can it survive? Should it?. *Ethnography and Education*, *13*(1) 1-17.

Hammersley, Martyn & Atkinson, Paul (2019). *Ethnography: Principles in practice* (4th ed.). London: Routledge.

Hammersley, Martyn & Traianou, Anna (2012). *Ethics in qualitative research*. London: Sage.

Harrits, Gitte Sommer (2011). More than method? A discussion of paradigm differences within mixed methods research. *Journal of Mixed Methods Research*, *5*(2),150-166.

Hatch, Amos (2007). Trustworthiness. In George Ritzer (Ed.), *The Blackwell encyclopedia of sociology* (pp.5083-5085). New York, NY: Wiley.

Heyvaert, Mieke; Hannes, Karin; Maes, Bea & Onghena, Patrick (2013). Critical appraisal of mixed methods studies. *Journal of Mixed Methods Research*, *7*, 302-327.

Kuntz, Aaron (2020). Standing at one's post: Post-qualitative inquiry as ethical enactment. *Qualitative Inquiry*, *27*(2), 215-8.

Lather, Patti (2013). Methodology-21: What do we do in the afterward?. *International Journal of Qualitative Studies in Education*, *26*(6), 634-645.

Lather, Patti & St. Pierre, Elizabeth (2013). Post-qualitative research. *International Journal of Qualitative Studies in Education*, *26*(6), 629-633.

Leavy, Patricia (Ed.) (2019). *Handbook of arts-based research*. New York, NY: The Guilford Press.

Lee, Justin (2014). Genre-appropriate judgments of qualitative research. *Philosophy of the Social Sciences*, *44*(3), 316-348.

Lincoln, Yvonna S. (1995). Emerging criteria for quality in qualitative and interpretive research. *Qualitative Inquiry*, *1*(3), 275-289.

Luft, Sebastian (2002). Husserl's notion of the natural attitude and the shift to transcendental phenomenology. In Anna-Teresa Tymieniecka (Ed.), *Analecta Husserliana The yearbook of phenomenological research: Phenomenology worldwide*, *80* (pp.114-119). Dordrecht: Springer.

Mays, Nick & Pope, Catherine (1995). Qualitative research: Rigour and qualitative research. *British Medical Journal*, *311*(6997), 109-112.

Merton, Robert & Lazarsfeld, Paul (Eds.) (1950). *Continuities in social research: Studies in the scope and method of "The American soldier"*. Glencoe, IL: Free Press.

Murphy, Elizabeth; Dingwall, Robert; Greatbatch, David; Parker, Susan & Watson, Pamela (1998). Qualitative research methods in health technology assessment: A review of the literature. *Health Technology Assessment*, *2*(16), 1-260, https://doi.org/10.3310/hta2160 [Accessed: March 11, 2022].

O'Cathain, Alicia (2010). Assessing the quality of mixed methods research: Toward a comprehensive framework. In Abbas Tashakkori & Charles Teddlie (Eds.), *The Sage handbook of mixed methods in social and behavioral research* (pp.531-556). Thousand Oaks, CA: Sage.

Potter, Jonathan (1996). *Representing reality discourse, rhetoric and social construction*. London: Sage.

Potter, Jonathan & Hepburn, Alexa (2005). Qualitative interviews in psychology: Problems and possibilities. *Qualitative Research in Psychology*, *2*(4), 281-307.

Power, Michael (1997). *The audit society*. Oxford: Oxford University Press.

Reicher, Stephen (2000). Against methodolatry: Some comments on Elliott, Fischer, and Rennie. *British Journal of Clinical Psychology*, *39*(1), 1-6.

Reichertz, Jo (2019). Method police or quality assurance? Two patterns of interpretation in the struggle for supremacy in qualitative social research. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 20*(1), Art. 3, https://doi.org/10.17169/fqs-20.1.3205 [Accessed: March 11, 2022].

Sandelowski, Margarete (1978). The problem of rigor in qualitative research. *Advances in Nursing Science*, *8*(3), 27-37.

Sandelowski, Margarete (1993). Rigor or rigor mortis: The problem of rigor in qualitative research revisited. *Advances in Nursing Science*, *16*(2), 1-8.

Schoonenboom, Judith & Johnson, R. Burke (2017). How to construct a mixed methods research design. *Kölner Zeitschrift für Soziologie*, *69*, 107-131.

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

Shadish, William; Cook, Thomas & Leviton, Laura (1991). *Foundations of program evaluation*. Newbury Park, CA: Sage.

Smith, John K. (1990). Alternative research paradigms and the problem of criteria. In Egon Guba (Ed.), *The paradigm dialog* (pp.167-87). Newbury Park, CA: Sage.

Smith, John K. (2004). Learning to live with relativism. In Heather Piper & Ian Stronach (Eds.), *Educational research: Difference and diversity* (pp.45-58). Aldershot: Ashgate.

Smith, John K. & Hodkinson, Phil (2005). Relativism, criteria, and politics. In Norman K. Denzin & Yvonna S. Lincoln (Eds.), *Handbook of qualitative research* (3rd ed., pp.915-32). Thousand Oaks, CA: Sage.

Smith, John K. & Hodkinson, Phil (2009). Challenging neorealism. *Qualitative Inquiry*, *15*, 30-39.

Spencer, Liz; Ritchie, Jane; Lewis, Jane & Dillon, Lucy (2003). *Quality in qualitative evaluation*. London: Cabinet Office, https://www.gov.uk/government/publications/government-social-research-framework-for-assessing-research-evidence [Accessed: March 11, 2022].

Symonds, Jennifer & Gorard, Stephen (2010). Death of mixed methods? Or the rebirth of research as a craft. *Evaluation and Research in Education*, *23*(2),121-136.

Tashakkori, Abbas & Teddlie, Charles (Eds.) (2010). The *Sage handbook of mixed methods in social & behavioral research* (2nd ed.). Thousand Oaks, CA: Sage.

Author

Martyn HAMMERSLEY is emeritus professor of educational and social research at The Open University, UK. He has carried out research in the sociology of education and the sociology of the media, but much of his work has been dedicated to the methodological issues surrounding social enquiry. His books include (with Paul ATKINSON) "Ethnography: Principles in Practice" (4th ed., Routledge, 2019), "The Politics of Social Research" Sage, 1995), "Taking Sides in Social Research" (Routledge, 2000), "Questioning Qualitative Inquiry" Sage, 2008), Methodology Who Needs It?" (Sage, 2011), "The Myth of Research-Based Policy and Practice" (Sage, 2013), "The Radicalism of Ethnomethodology" (Manchester University Press, 2018), "The Concept of Culture" (Palgrave Macmillan, 2019), and "Troubling Sociological Concepts" (Palgrave Macmillan, 2020). In *FQS*, he published "A Historical and Comparative Note on the Relationship Between Analytic Induction and Grounded Theorising" (https://doi.org/10.17169/fqs-11.2.1400, 2010) and "Understanding a Dispute About Ethnomethodology: Watson and Sharrock's Response to Atkinson's 'Critical Review'" (https://doi.org/10.17169/fqs-20.1.3048 2019).

Contact:

Martyn Hammersley

The Open University
WELS
Milton Keynes, MK7 6AA, UK

E-mail: m.hammersley@open.ac.uk

Citation