

## **Systematic or Unsystematic, is that the Question?**

**Some reflections on the science, art, and politics of reviewing  
research evidence.**

**Martyn Hammersley**

**Text of a talk given to the Public Health Evidence Steering Group of the Health Development  
Agency, October, 2002.**

Martyn Hammersley

**Systematic or Unsystematic, is that the Question? Some reflections on the science, art, and politics of reviewing research evidence.<sup>1</sup>**

It is important that the practical use of research takes in the whole range of findings on a topic, not just the results from one or two studies. For this reason, reviews play a crucial role as a bridge between research and related areas of policymaking and practice. Indeed, the increasing tendency for the mass media to report evidence from single studies in controversial areas, particularly in the health field, is to be deplored. This can be dangerous in its immediate consequences. Moreover, I suspect that, down the line, it will lead to further erosion in the public authority of scientific research.

This bridging role of reviews has recently been given great emphasis. And it has been argued not just that there have been insufficient reviews to make research findings as widely accessible as they ought to be, but also that many of the reviews that have been done have not been of the right kind. A contrast has come to be drawn between *systematic* reviews and the older form of ‘traditional’ or ‘narrative’ review. However, while the quality of reviews is certainly an important issue, I think that there are some questions to be raised about the notion of systematic review. These start from matters of semantics but move rapidly into the substantive.

‘Systematic’ is a word that, in most contexts, has a positive ring to it. Who would want reviews to be unsystematic, if by ‘systematic’ we mean no more than ‘properly carried out, taking account of all the relevant evidence, and making reliable judgements about its validity and implications’? On this definition, to produce a systematic review is simply to do the job of reviewing well. However, I have heard it suggested that there *are* people who would want a review to be unsystematic, even in this sense: those who want to select and interpret research evidence so as to support their own pre-given views or interests, and thereby to claim scientific backing for these.

Now, there certainly are such people; but it is worth noting that a review constructed in this way would also be systematic in an important sense. It would display what we refer to as *systematic error*. So, the contrast here is not between systematic and unsystematic reviews, but rather between systematic pursuit of two different goals: on the one hand, provision of a sound representation of the current state of knowledge; on the other hand, promotion of a particular point of view.

We can perhaps all agree that research reviews should be systematic in the sense of being done properly, avoiding systematic and other kinds of error. But this hardly exhausts the meaning of the term ‘systematic review’, as currently used, and there is room for reasonable disagreement about some other aspects of it. One of these, often attributed by critics and sometimes rejected by defenders, is that systematic reviews privilege particular types of research, which are treated as the gold standard. Thus, it is undeniable that some advocates of systematic review give primacy to randomised controlled trials (RCTs) in this way. Of course, there are also people involved in the production of systematic reviews who take a different view.

---

<sup>1</sup> Text of a talk given to the Public Health Evidence Steering Group of the Health Development Agency, October, 2002.

For example, there are those who want to incorporate qualitative studies within systematic reviews. And, very often, they want to do this in ways that treat such studies appropriately, rather than simply as the best source of evidence available where there are no quantitative studies, or as providing illuminating detail with which to ‘fill out’ the findings from quantitative studies.

Exploring how qualitative studies can be incorporated in systematic reviews *does* seem to me to be a worthwhile venture, but I think it is easy to underestimate the difficulties involved. And these problems highlight the dangers of thinking about reviews primarily in terms of whether or not they are ‘systematic’.

### *What kind of synthesis?*

The first problem I’ll mention relates to the term ‘review’ as well as to the term ‘systematic’. It seems to be assumed that for a review to be systematic it must involve the *synthesis* of findings or of data from the studies reviewed. This is one of the respects in which systematic reviews are seen as going beyond traditional reviews. But what is meant by ‘synthesis’ here?

There are several meanings of this term, and they are not always clearly distinguished. One key metaphor is aggregation or pooling. Here, it looks as if each study reviewed is to be treated as investigating a different sample of cases drawn from the same population, and what ‘synthesis’ implies is that these various samples are combined to make a much bigger sample of that population. The idea is that this increases the chances that the conclusions drawn are valid for the population of cases as a whole.

This, of course, draws on a way of thinking that is central to survey research in the social sciences. It is less central - though not entirely inappropriate in my view - to both experimental and qualitative research. However, applied to the task of reviewing, I think there are serious questions about whether data from cases studied by different researchers on different occasions often for somewhat varying, and uncoordinated, purposes can be pooled in the way proposed. After all, each of the studies does not simply describe the cases that it is concerned with, it carries out some analysis designed to determine what caused what and why. As a result, there is likely to be variation between studies in what intervening and outcome variables are taken into account, how they are measured, the time period involved, the local context, and so on. So the cases do not necessarily belong to a single population. Nor is it always clear what the intended target population is.

A second interpretation of ‘synthesis’, sometimes mixed up with the first, is that the various studies reviewed can be treated as replications of one another. The argument here is that, to the extent that further studies of the same problem come to the same conclusion as the first one, it can be concluded that the original findings were reliable. This is a logic that is derived from experimental research. And it is important to notice how it differs from the notion of pooling. With pooling, other things being equal, the larger the number of studies the better, and the product is a mean effect size which, it is assumed, gives a statistical prediction of what would be found in other cases. This pooling metaphor encourages the inclusion of unpublished

material or grey literature, not just peer-reviewed studies. By contrast, in principle at least, replication requires only quite a small number of studies, but these must be well conducted and they must all either support or reject the hypothesis if any definite conclusion is to be reached.

The application of this second conception of synthesis to research reviewing may be quite reasonable in principle. However, we should note that the criteria of selection here are *very* restrictive. Each study must be quite similar in its focus and procedures, or must display systematic variation designed to test out specific threats to validity. And this means that the studies must conform fairly closely to the experimental model. The randomised controlled trial is not the gold standard here, but is, in fact, probably the lower limit of what can be included.

A third interpretation of ‘synthesis’ portrays it as the comparative analysis of existing studies with a view to the systematic development and testing of hypotheses. The logic here is obviously that which is central to a great deal of research, both quantitative and qualitative. This is the very core of experimental method, but the notions of analytic induction and grounded theorising used by qualitative researchers also conform broadly to it. Moreover, in principle, there is no reason why this logic should not be applied to extant studies rather than to new data. Indeed, this is more or less the approach that Ray Pawson is attempting in producing theory-driven systematic reviews or realist syntheses (Pawson 2001a and b).

However, while examining existing studies may facilitate the development of theoretical ideas, I doubt whether in most fields this will be enough to provide a rigorous test of those ideas. After all, the logic of theory development and testing is sequential: what’s required to develop and test a theory changes over time as the theory develops, and often can’t be predicted in advance. As a result, very often, the task of testing a theory rigorously will require the collection of primary data, the work already done will not be sufficient.

A fourth interpretation of ‘synthesis’, one that is quite common among qualitative researchers, employs the metaphor of a mosaic or map. Here, putting together different studies means looking at how they can be combined to give us a bigger picture. In this version of synthesis, what each study contributes is distinctive. What is needed is complementarity not similarity. The studies that are reviewed will look at different aspects of the same phenomenon, for example they may examine different types of clinical contact between doctor and patient, at patients’ experience of these kinds of clinical encounter, at the management structures within which such encounters take place, at the history of clinical practice, and so on. This sort of complementarity is often highlighted in traditional narrative reviews, and while it is not unproblematic (see Hammersley et al 1985:48-56) I will not discuss it further here since it is not central to most versions of systematic review.

A fifth possibility has been put forward by some qualitative researchers, who argue that synthesis can involve translation of the findings of some studies into the terms of another, thereby developing key metaphors that will facilitate future understanding of the phenomena concerned (see Noblit and Hare 1988). As far as I understand it, this so-called ‘meta-ethnography’ is similar to what some traditional reviews have done, and it contrasts with what is usually meant by ‘synthesis’ in the

context of systematic review. It amounts to developing a new framework which encompasses existing studies and shows them in a new and illuminating light, or one that is more inclusive or abstract than that in which they were originally framed.

It is worth noting that an important respect in which these different sorts of synthesis may vary is in terms of how far they follow a pre-determined path. Pooling and replication seem actually to require this. The questions to be answered must be clearly specified at the beginning of the reviewing process, and procedures set up to identify all the studies relevant to them. However, some other kinds of synthesis appear to involve a more inductive approach, in which the reviewer may reformulate the focus of the review in the course of doing it. I have to say that my own inclination is to resist any attempt to rule out either approach, since they may both be of value.

So, here we have at least five different meanings that can be given to the term ‘synthesis’, some of which might apply to systematic review and some of which do not. And we need to be clear about which sort of synthesis is implied by the term ‘systematic review’, and why this is to be regarded as more valuable than others.<sup>2</sup>

Over and above noting different types of synthesis, we might ask whether reviews *have* to involve synthesis, in any of these senses. Reviews could simply make judgements about what we do and do not currently know on the basis of careful evaluation of particular studies and their findings; without involving any of the above forms of synthesis. Reviewing can also provide methodological lessons about how we might improve the success of our inquiries in the future.

To be even more provocative, we could ask whether some of these forms of synthesis actually constitute reviewing the literature at all. A few seem to be closer to actually *doing* research, rather than reviewing it. For example, research synthesis involving the pooling of results is sometimes difficult to distinguish from the analysis of secondary data. And there is at least one form of synthesis which, as I indicated, probably cannot be done properly without relying on supplementary collection of primary data, and whose logic is very much that of research designed to develop and test theory.

#### *‘Systematic’ as ‘issue-focused’*

A second component of what it means for a review to be ‘systematic’ seems to be that it is focused on a specific issue – often, the question of whether an identified policy, practice, or technique works well, or whether it works better or worse than alternatives. Now, this is certainly a useful thing for reviews to do. However, it is not the only focus they can have. Often traditional reviews have been concerned with evaluating what is known about some whole field of inquiry. And this too is a perfectly reasonable enterprise.

To say that traditional reviews are unfocused because they don’t concentrate on a specific question is like complaining that a map is of no use because it covers a

---

<sup>2</sup> Incidentally, these five different senses of the word ‘synthesis’ do not exhaust its meaning. Some years ago, Strike and Posner identified 15 kinds of synthesis, along with some forms of what they call ‘quasi-synthesis’ (Strike and Posner 1983).

wider area than the one we are currently interested in. Here, it seems to me, instead of the systematic/unsystematic distinction what is required is a non-evaluative typology dealing with the different kinds of focus that reviews can have.

### *Systematic reviews as 'transparent'*

Perhaps the most important, and distinctive, meaning of the term 'systematic' is the idea that reviewing should be done in a way that makes the procedures employed explicit. In short, the aim is for transparency. However, this too is not unproblematic. The word 'transparency' is even more obviously a hurrah word than is the term 'systematic', at least in recent times. Who could be against transparency, aside from those who have something to hide? Well, at the risk of self-incrimination, I want to raise some questions about the idea of transparent reviews, and transparency generally. There are three points here.

First, explicitness is a matter of degree. If we look at examples of old style narrative reviews, we find that many of them *do* tell us quite a bit about how they were carried out. It may be true that in many cases more such information was required for readers to be able to make a sound judgement about the conclusions of the review. But not all systematic reviews are likely to be perfect in this or other respects either.

The second issue is more fundamental. This is that there is a point beyond which it is impossible to make any activity, including reviewing, explicit. The metaphor involved in the word 'transparency' implies that *anyone* can see, or perhaps can see *through*, what is going on. In other words, it is assumed that there are no audience requirements for understanding and evaluating an activity other than possession of a pair of eyes.

Now, I do not accept that this can ever be true. Explicitness is always relative to a particular type of audience, who have some background knowledge and purpose. Nothing can be *absolutely* explicit or transparent. And when we are dealing with audiences who do not share much of the background knowledge relevant to the activity concerned, we often have to struggle quite hard to help them understand what's involved. Moreover, what we usually do in such circumstances is not so much make things explicit as use parallels and analogies to enable them to relate what they don't understand to what they already know about. In other words, the concept of transparency, and the metaphor of vision on which it relies, can be misleading in thinking about how understanding occurs or knowledge is gained and communicated.

Furthermore, even when we are addressing an audience who *do* know a lot about the activity we are describing, we may not be able to make explicit exactly *why* we did what we did, in other words to specify it in terms of rules. This is a fundamental philosophical argument, one that was developed into what I think is an illuminating account of science by Michael Polanyi many years ago (Polanyi 1959 and 1966). He argues that tacit knowledge necessarily plays a key role in scientific investigation. And if it plays a key role in science, we can be fairly sure that it would play a key role in reviewing scientific evidence as well.

It is perhaps worth emphasising that what I'm drawing on here is not some *avant garde* postmodernist view of science, but a non-positivist view that has been available for 50 years. By contrast, the concept of science that seems to be built into much discussion of systematic reviews is a, by now largely discredited, positivist one, in terms of which scientificity amounts to following an explicit method or set of procedures.<sup>3</sup>

So what I am saying is that there is a point beyond which we cannot go in seeking to make the work of reviewing transparent, and that we shouldn't pretend that any rules we draw up for reviewing simply apply themselves. They do not: people require prior knowledge in order to apply them, and some judgment and discretion is always involved.

The final point about the notion of transparency can also be derived from Polanyi's work on science. The concept of systematic review seems to imply that transparency can be achieved if the task of reviewing is formulated beforehand in terms of a set of *procedures* to be followed. Now what I want to suggest is that, while seeking to specify the approach that will be, or has been, employed in a piece of work may be useful, insisting that people follow a set of procedures rigidly can be a hindrance. Literally, it may rule out any reflection on issues that actually need attention, encouraging a rather pragmatic orientation. And this is especially likely where the time available for producing a review is very limited. As Polanyi points out, science, like many other activities, relies on skilled judgement, the capacity for which is built up through experience; and skilled judgement means taking account of particulars, and will often involve apparently breaking the rules.

In short, following Polanyi, if one tries to reduce science to the following of explicit procedures, one runs the risk of obstructing its progress. And I suggest that, in the same way, producing reviews can't be reduced to the following of explicit procedures: it always involves skilled judgement on the part of the reviewer, about what is and is not likely to be valid, what is and is not significant, and so on. As a result, trying to govern it by transparent rules could lead to a worsening in the quality of reviews, in some key respects, rather than to improvement.

An area where I believe this is a particular danger concerns judgements about the relative value of particular types of study in producing sound findings. One approach to this issue that's built into much systematic reviewing is to specify a hierarchy of research designs, in terms of their relative success in dealing with key threats to validity. Now, it is certainly true that, in principle, different research designs have varying capacities to handle particular types of validity threat. However, a trade-off is usually involved. Any gain in one area is generally at the expense of costs in another. For instance, in the case of the experiment, despite its strengths, there's a huge literature in social psychology which identifies reasons for caution in applying experimental findings outside the experimental situation. In other words, it is recognised that the experimental situation has distinctive characteristics and that the behaviour of subjects will often be influenced by these rather than by, or in addition to, the experimental treatment.<sup>4</sup>

---

<sup>3</sup> Elsewhere, I have spelled out this argument in more detail: see Hammersley 2001.

<sup>4</sup> See the discussion and references in Rosnow 1981.

Some of the same arguments apply to RCTs, along with other problems to do with the way in which selection procedures can render the sample of patients studied unrepresentative of the wider population. Indeed, one of the selection procedures built into some RCTs - that patients should not have other medical conditions - actually means that the results from them are likely to become increasingly unrepresentative of the national population, as the latter ages.<sup>5</sup> My point is that any gain in internal validity, very often, is at the expense of reducing external validity. And the reverse is equally true. In short, given that trade-offs are always involved, a single hierarchy of research designs, by means of which to determine which studies are sufficiently reliable to be included in a systematic review, cannot be justified.

There is a third, more fundamental, point as well, which goes back to the argument about tacit knowledge. This is that it is not possible to judge particular studies wholly in abstract terms, such as according to what research design they employed. While judgements should be informed by guidelines about the strengths and weaknesses of different kinds of methodological approach, *these are only guidelines*. One reason for this is that in assessing the findings of any study we necessarily rely on what we currently take to be sound knowledge. This is not to say that we immediately dismiss evidence which conflicts with what we currently believe, even less that we *should* do this. Indeed, the most interesting studies are, other things being equal, precisely those which challenge what we currently assume we know. My point is that if a study makes such a challenge then it needs support from strong evidence, and our judgement of that evidence will itself necessarily rely, in part, on other things that we take to be sound knowledge. In short, there is no foundation of absolute empirical givens to which we can make ultimate appeal as evidence: we always have to rely on background knowledge.<sup>6</sup>

Summarising what I have said about the idea of transparency in systematic reviews, it's never possible to be fully explicit about how something is done, and trying to proceduralise reviewing could actually make it less rather than more effective in important respects.

#### *What 'systematic/unsystematic' obscures*

The final points I want to make concern aspects of reviewing which the systematic/unsystematic distinction obscures; though I'm not suggesting that those involved in systematic reviewing are unaware of them. Here, first of all, I want briefly to introduce a further non-evaluative typology of research reviews, this time to do with audience.

Much systematic reviewing has been concerned with presenting the fruits of research to policymakers or practitioners. This is clearly an important set of audiences, but it is not the only one. Reviews also have a function within research

---

<sup>5</sup> I owe this point to a discussion with Michael Eraut.

<sup>6</sup> The view I have presented here is more or less that offered by Susan Haack in her book *Evidence and Inquiry*, in which she uses the crossword puzzle as an illuminating metaphor for the logic of scientific investigation (Haack 1993).

communities. I will not say anything about that here, but it needs to be borne in mind in any discussion of the nature of reviews.

Instead, I want to suggest that reviews of research addressed to audiences outside the research community, perhaps more than others, are a matter of art and politics as much as of science. As to art, I don't mean that we should explore presenting reviews in rhyming couplets, as dialogues, or as fictions. What I mean is that whenever we address a particular audience we have to take account of what we can assume they already know, what they would and would not be interested in, and perhaps also what they might misinterpret.

Now in much writing, including the writing of some reviews, we do not know who the specific audience will be. There is not a single target audience. Instead, the review is seen as a resource that could be used by many different audiences. That's a perfectly reasonable way of writing a review. However, there is an important distinction between this sort of general audience review and reviews that are targeted at specific types of audience. It seems to me that the gap that has to be jumped between the world of researchers and that of particular groups of policymakers and practitioners is sometimes so great that considerable bridge-building is required. And I insist that building bridges is as much an art as a science: it depends on skill, judgement and sensitivity to what might be important in the particular situation. What I am saying here is that to think of reviews simply in terms of whether or not they are systematic does not take any account of whether they are appropriate for the audience concerned.

Now I know that many reviewers are all too sensitive to whom their audience is likely to be, and that some reviews are designed for specific audiences. But we need to make this issue more explicit in thinking about the nature of reviews. Moreover, some audiences may not find the standard systematic review very helpful. Or, if they do, they may find it useful for the wrong reason. In trying to explicate what the term 'systematic review' means here I've concentrated on what researchers mean by it. By contrast, for some policymakers it might mean a review which provides a clear recommendation, in the right direction, backed up by impressive-looking scientific paraphernalia that is guaranteed to dissuade journalists from asking too many questions. This cynical point leads me, finally, to politics.

Having just talked about what audiences find most helpful, and suggested that some reviews ought to be designed for particular audiences, we should ask: what limits must be placed on this? At the beginning of this paper, I drew an evaluative distinction between reviews that are designed to present all the relevant evidence, and make reliable judgements about its validity and implications, and those that select and interpret research evidence so as to support a pre-given view or interest, thereby giving this the illusion of scientific backing. Well, while I believe that this distinction is very important, I don't think we can simply condemn what might be called the politically motivated use of research evidence. We need to take account of the politics of how reviews are received and read by audiences. The reception of reviews is often driven by audiences' current preoccupations, and guided by relatively fixed preconceptions. This is almost inevitable because of the very nature and pressures of policymaking, and many kinds of occupational practice.

I should underline that I don't think that it is the proper task of the researcher to try to use research evidence to promote or defend some particular policy or practice. For me that would be to exceed the authority of research. However, I *do* think that it's the responsibility of the researcher who is preparing a review for a specific audience to anticipate the interpretations and evaluations which that audience might leap to, and to point out that what's involved *is* a leap, and is not necessarily justified.

Up to now I have assumed that reviews of research evidence are produced by researchers. But this is not always the case. For one thing, researchers are not always *only* researchers: very often they are also practitioners of some other relevant occupational activity. And they may produce the review with the hat of practitioner rather than that of researcher. Or, more likely, they will wear both hats, without being clear about under which set of auspices they are speaking at any particular time.

For me, this distinction between a researcher's review of research evidence and that carried out by a policymaker or practitioner is perhaps the most basic of all. One reason for this is that some reviews by policymakers or practitioners may well be designed to highlight particular evidence that they believe has been overlooked, and which points in a particular policy direction that, on the basis of their practical experience, they think ought to be given more attention. Such reviewing is not necessarily illegitimate in practical terms; indeed, it is probably an inevitable part of how people in the public sphere routinely read and use research, even though it is not acceptable for researchers to engage in it. We must not pretend that policymakers and practitioners can or ought to operate in the manner that's expected of researchers.

Equally, and this is one respect in which I'm inclined to raise questions about the general idea of evidence-based policymaking or practice, it is important that researchers do not expect that their work, including their reviews of research evidence, will be employed by policymakers or practitioners in exactly the same manner that researchers would (or ought to) use them.<sup>7</sup> Policymaking and professional practice are not the same as research, they do not call for the same cognitive or ethical orientation; because they are not solely or even primarily concerned with producing sound knowledge, but rather with bringing about desirable practical outcomes, and this requires practical wisdom not just valid evidence. It seems to me that *some* accounts of evidence-based practice assume that policymakers and practitioners could act in more or less the same way as researchers are supposed to do: carefully considering all the alternatives, weighing the evidence available and seeking further evidence where it is not, suspending judgement until sufficient evidence is available, and so on. The conditions in which most policymakers and practitioners work do not allow this, most of the time. Of course, I'm not saying that policymakers and practitioners should not draw on research evidence, even less that they shouldn't take care in interpreting and evaluating all sources of evidence. My point is that they can't do this in exactly the same way, or to the same extent, as can researchers. Indeed, were they to try to act in the same manner as researchers they would do their work less well.

---

<sup>7</sup> Elsewhere, I have outlined some other respects in which I have doubts about the notion of evidence-based policymaking and practice: Hammersley 2003.

### *Conclusion*

I have argued that the contrast between systematic and unsystematic reviews is unhelpful in some key respects. The term ‘systematic’ seems to be used to mean several different things, and not all of them are unquestionably good. I noted that systematic reviews synthesise data, but that what’s meant by ‘synthesis’ is open to diverse interpretation. I also observed that the notion of systematic review tends to privilege reviews which have a specific focus on some type of policy or program, rather than reviews which summarise what’s known in a particular field of inquiry. Yet, both types of review are legitimate. Above all, I questioned the notion of transparency, which is perhaps the most central idea behind systematic reviews. I argued that full transparency is impossible and that, beyond a certain point, the attempt to make reviewing transparent, through reducing it to procedures, may actually worsen the quality of the reviews generated. Finally, I pointed to some important aspects of reviewing that the distinction between systematic and unsystematic reviews obscures, concerned with the art and politics of reviewing.

So, what I am saying is that if we are to improve our understanding of research reviews and the functions they can serve, we need to move away from simplistic evaluative contrasts between systematic and narrative reviews. I have focused most of my attention here on the notion of systematic review, since that’s the most influential one in many quarters at the moment. But I could have raised some equally challenging questions about narrative reviewing. Indeed, a colleague and I did so a few years ago, looking at some of the research reviews that had been carried out in the field of education (Hammersley and Foster 1998). So, my argument here is not intended to be part of a battle between advocates of systematic and narrative review. Instead, it is designed to help us move beyond such disputes.

### *References*

Haack, S. (1993) *Evidence and Inquiry*, Oxford, Blackwell.

Hammersley, M. (2001) ‘On “Systematic” reviews of research literatures: a “narrative” reply to Evans and Benefield’, *British Educational Research Journal*, 27, 5, pp543-54.

Hammersley, M. (forthcoming 2003) ‘Some questions about evidence-based practice in education’, in R. Pring and G. Thomas (eds.) *Evidence-Based Practice*, Buckingham, Open University Press.

Hammersley, M. and Foster, P. (1998) ‘A review of reviews: structure and function in reviews of educational research’, *British Educational Research Journal*, 24, 5, pp609-28.

Hammersley, M., Scarth, J., and Webb, S. (1985) ‘Developing and testing theory: the case of research on pupil learning and examinations’, in R. G. Burgess (ed.) *Issues in Educational Research: Qualitative methods*, Lewes, Falmer Press.

- Noblit, G. W. and Hare, R. D. (1988) *Meta-Ethnography: synthesising qualitative studies*, Newbury Park CA, Sage.
- Pawson, R. (2001) *Evidence Based Policy I. In search of a method*, ESRC UK Centre for Evidence Based Policy and Practice: Working Paper 3.
- Pawson, R. (2001) *Evidence Based Policy II. The promise of 'realist synthesis'*, ESRC UK Centre for Evidence Based Policy and Practice: Working Paper 4.
- Polanyi, M. (1959) *Personal Knowledge*, Manchester, Manchester University Press.
- Polanyi, M. (1966) *The Tacit Dimension*, Garden City, NY, Doubleday.
- Rosnow, R. (1981) *Paradigms in Transition: the methodology of social inquiry*, New York, Oxford University Press.
- Strike, K. and Posner, G. (1983) 'Types of synthesis and their criteria', in S. A. Ward and L. J. Reed (eds.) *Knowledge Structure and Use: Implications for synthesis and interpretation*, Pennsylvania MS, Temple University Press.