Fit for purpose? Assessing research quality for evidence based policy and practice

Annette Boaz
Deborah Ashby

ESRC UK Centre for Evidence Based Policy and Practice
Queen Mary University of London

Email a.l.boaz@qmul.ac.uk

© January 2003: ESRC UK Centre for Evidence Based Policy and Practice

Annette Boaz is Senior Research Fellow at the ESRC UK Centre for Evidence Based Policy and Practice, and Deborah Ashby is Associate Director – Methodology.
Abstract

Efforts to promote quality products, services and practices currently abound in both the public and private sector. The evidence based policy and practice (EBPP) movement cannot exempt itself from this debate. Indeed one of the principal aims of EBPP has been to promote the generation and use of good quality research evidence in policy and practice. Unfortunately, traditional mechanisms for assuring research quality through peer review and quality standards (in the academic sector) and research professionalism (in the government and commercial sector) have failed thus far to deliver consistently high quality research. In particular, peer review has proved unreliable and in some fields there is a lack of consensus as to what ‘counts’ as good quality research.

Although improving the effectiveness of existing quality assurance procedures is an immediate challenge for the research community, this paper goes on to argue that future conceptualisations of research quality need to move beyond a fixation with methodological quality, to address the ‘fitness for purpose’ of research. After all, one of the strengths of research, compared to other sources of knowledge available to decision makers, should be that it is a quality assured product carried out to pre-agreed standards. A broader notion of research quality should help researchers and research users to feel confident about the use of evidence in policy and practice.

Key words: quantitative research; qualitative research; quality; evaluation; methods

Acknowledgements

The authors would like to thank the following people for their extremely useful comments on earlier drafts of this paper: Phil Davies, Lesley Grayson, Andrew Long, Sandra Nutley, Ray Pawson, Jane Ritchie, Bill Solesbury, Rebecca Stanley, Isabel Walter and Ken Young.

The purpose of the Working Paper series of the ESRC UK Centre for Evidence Based Policy and Practice is the early dissemination of outputs from Centre research and other activities. Some titles may subsequently appear in peer reviewed journals or other publications. In all cases, the views expressed are those of the author(s) and do not necessarily represent those of the ESRC.
Fit for purpose? Assessing research quality for evidence based policy and practice

Introduction

Efforts to promote quality are everywhere; from the kite marks on products we buy to awards such as the Booker prize. In the public sector there have been similar initiatives to promote high standards, prompted by the desire to ensure efficient, effective and consistent use of taxpayers’ money, and to demonstrate accountability. These factors lie behind the huge growth in audit and inspection, as well as the promotion of specific tools designed to promote quality including the Charter Mark scheme and Investors in People. Improving quality is a key theme of the current government’s modernising agenda, driving prominent initiatives such as Best Value and Quality Protects. At local level, where most public services are delivered, organisations including the Local Government Quality Group and the Improvement and Development Agency have taken a lead on promoting quality.

The growing interest in using research evidence to inform policy and practice (Cabinet Office, 1999), with its emphasis on identifying, synthesising and applying reputable knowledge to the solution of problems, can be seen as another element in this multi-faceted quality agenda. In turn, this has focused attention on the quality of research itself. In particular, the systematic review of existing research to identify what is reputable (Boaz, Ashby and Young, 2002) has stimulated debate about how and when (and indeed whether) to assess the quality of individual pieces of research for inclusion in a review.

Unlike more traditional reviews, systematic reviews of the literature include a detailed appraisal of the research studies identified (Petticrew, 2001). In healthcare, where the evidence based approach originated, the focus on evaluating the evidence of effectiveness in systematic reviews has led to the emergence of distinct quality standards for appraising research. Developing these standards has been an incremental process, based on considerable methodological endeavour over the past ten years which has focused on a small range of research methodologies, in particular randomised controlled trials. Although considerable efforts have also been made to develop quality standards for other research approaches (Stern, 1979), there is far less consensus on how we might assess the wide range of research approaches employed by public policy researchers.

This paper discusses some of the current debates about research quality. It introduces the mechanisms traditionally used to assure research quality, particularly peer review. The debate about research quality has tended to focus on methodological quality, and some of the key sets of criteria are included in the text. The paper goes on to argue that evidence based policy and practice (EBPP) requires a broader notion of quality that embraces the ‘fitness for purpose’ of the research. After all, unlike other sources of knowledge, research should be able to trade on its commitment to quality assurance. A broader notion of research quality should help researchers and research users to feel confident about the use of evidence in policy and practice.
Quality is quite an abstract concept so in order to ground the arguments we will follow four fictional cases, embarking on research projects:

- A civil servant about to commission qualitative research designed to understand barriers to the take up of means tested benefits by older people (CASE 1)
- A researcher applying for funding to carry out a randomised controlled trial of the effectiveness of a healthy eating intervention (CASE 2)
- A Benefits Agency project worker using action research to evaluate a local project aimed at helping lone parents get back to work (CASE 3)
- A post-graduate student studying the school performance of children living in care (CASE 4)

**Existing structures for quality appraisal**

It would be a mistake to assume that quality appraisal is a new idea for researchers. There are already structures and procedures in place to support and promote good quality research, some of very long standing. These include:

- Published standards and checklists, including sections in research methods textbooks, ‘how to’ guides, funders’ requirements etc.

- Peer review of research proposals. Many research projects pass through a quality filter in the often highly competitive funding process. At this stage proposals are likely to be sent out to reviewers and discussed by a commissioning group of experts. In some fields of research agreed projects can then be subject to the scrutiny of an ethics committee before any fieldwork commences.

- During the research process, researchers often seek advice from their peers and potential research users either informally or through a formal steering group or advisory committee.

- After research is completed, publication peer review then assesses the quality of studies prepared for formal publication in books and journals. Some government departments also send policy or practice research reports out for peer review.

With all this activity why the debate about quality?

**Inadequate application**

One area of concern centres on the failure of quality standards and procedures to be applied effectively in all cases. To take just one mechanism; it is often assumed that publication peer review is a guarantee of quality that does away with the necessity for individual research users to be concerned with applying quality criteria to studies. However, in another paper in this series (Grayson, 2002) Lesley Grayson explores the literature on peer review in biomedicine and reports that it is often:

- Slow
- Expensive
• Prone to bias
• Open to abuse
• Incompetent
• Unable to detect fraud

The last of these deficiencies is of considerable concern, particularly in healthcare research. As the European Science Foundation has commented:

At a time when the need to build trust between science and society is becoming ever more important, it is vital that the conduct of science itself is based on the highest ethical considerations and that misconduct within science itself can be identified and dealt with in an open and transparent manner.

Incidents of misconduct – sometimes revealed by the increasing scrutiny of the research literature attendant on the growth of evidence based medicine – have opened up the debate about the efficacy of peer review, professional self regulation and quality standards. They have also highlighted the potential impact of misconduct on the confidence the public, policy makers and practitioners have in research and researchers (European Science Foundation, 2000).

Other formal and informal quality control procedures have also been criticised on the grounds of abuse or insufficiently rigorous application, although hard evidence is often difficult to come by. For example, there have been accusations of both conscious and unconscious bias in the peer review of research proposals, with reviewers favouring friends, protégés and those of like mind while pronouncing adversely on rivals or those with whom they are intellectually at odds. Similarly, the operation of steering committees and advisory groups during the research process may sometimes be less to do with methodological quality than about ensuring that the project meets the requirements of the sponsor. Sensitivity to the sponsor’s requirements can, of course, contribute to the fitness for purpose of research but can equally well introduce biases that conflict with the aim of producing objective, good quality evidence.

**Differential application**

A second area of concern about quality standards lies in their varying application to different kinds of research setting. For example, in **Case 1**, the commissioning civil servant may have recourse to reviewers’ comments and/or a commissioning group, and can also draw on in-house expertise in deciding what counts as a good quality study. The commissioning group will be in a position to weigh up different dimensions of quality including the methodological rigour of the study design and the fitness of the research for the policy needs of the department. **Case 2**, the academic researcher, is likely to pass through a similar process. However, he or she will also have recourse to well developed methodological standards for how to conduct a randomised controlled trial. For both groups formal publication is a likely outcome and again, at this stage, the research will be subject to peer review.

For **Case 3**, our practitioner-evaluator, recourse to quality assurance procedures is less well defined, although there may be access to advice from colleagues with research skills, or partners in a local university or college, in designing the evaluation. Case 3 is less likely to seek peer reviewed publication of his or her study although may well
report it in a practitioner journal. **Case 4**, the student, is likely to rely on a supervisor as a main source of quality assurance. The student may have benefited from relatively recent training in research methods, but can lack the informal, ongoing peer support available to more established researchers. Our four cases underline the diversity of research practice and access to both formal and informal quality assurance procedures.

**What counts as good quality: hierarchies of evidence**

When it comes to applying notions of quality in practice, the debate about ‘what counts’ as good quality has, to some extent, been fuelled by the widespread influence of the hierarchy of evidence used in health care.

<table>
<thead>
<tr>
<th>Systematic reviews and meta-analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well designed randomised controlled trials</td>
</tr>
<tr>
<td>Well designed trials without randomisation, e.g. single-group pre-post, cohort, time series or matched case-controlled studies</td>
</tr>
<tr>
<td>Well designed non-experimental studies from more than one centre</td>
</tr>
<tr>
<td>Opinion of respected authorities, based on clinical evidence, descriptive studies or reports of expert committees</td>
</tr>
</tbody>
</table>

Source: Canadian Task Force on the Periodic Health Examination (1979)

The hierarchy has some visual appeal as a tool for organising and assessing research, but has been criticised for its oversights (where is qualitative research?) and its judgement about the superiority of certain research methodologies over others (with the systematic review in pride of place at the top of the hierarchy as the ‘gold standard’). However, the model is specifically designed to assess studies providing evidence of the effectiveness of clinical interventions, and not to assess the wide variety of studies conducted to explore different research questions. Researchers in the field of healthcare have had to develop alternative methods for assessing other types of research. Most notably, much effort has been put into assessing qualitative studies.

The broader healthcare literature goes ‘beyond the hierarchy’ and helps to address many of the challenges public policy researchers can relate to. For example, in a
policy context we are likely to be concerned with not only the scientific quality of the research, but also its relevance or fitness for purpose in a policy or practice context.

**Existing criteria for assessing quality**

There is a wide variety of quality criteria for assessing research, ranging from the technical to the philosophical and even the ‘utilisation-focused’. The most established criteria were designed for quantitative research and there is some debate about whether these are appropriate for other sorts of research. Different disciplines also use key terms in different ways to describe dimensions of quality (Ryan et al, 2001). As a consequence there is plenty of lively debate in the research community about research quality, and indeed about whether it is a debate worth having! This section briefly introduces some of the main perspectives and criteria.

### Some key terms

- **Reliability** – are the results repeatable?
- **Validity** – does it measure what it says it does?
- **Internal validity** – do the research results mean what they appear to?
- **External validity** – can the results be generalised to other settings (ecological validity) and to other populations (population validity)?
- **Replicability** – are the results of the study reproducible?

One distinction is between those who seek to extend the use of traditional criteria developed for quantitative studies to all research and those who propose ‘new look’ sociological and philosophical criteria for assessing quality (Popay et al, 1998).

**‘Quantitative quality’**

Most of the criteria developed for evaluating the quality of research are rooted in the quantitative tradition (Bryman, 2001) and focus on reliability, replicability and validity. There has been a broad debate about what should be included as quality criteria. For example, it has been argued that the extent to which the methods are acceptable to participants should also be used as a criterion for evaluating the methodological quality of quantitative studies (Ryan et al, 2001).

Although there is some debate about what the criteria for assessment should be, the use of pre-set criteria is fairly widely accepted by researchers using quantitative methods. They have also been applied to qualitative research, although often their application involves some redefinition of the terms.

The usual canons of good science have value but require redefinition to fit the realities of qualitative research and the complexities of the social phenomena that we seek to understand. (Strauss and Corbin, 1998, p266)
**Good quality experiments**

By far the most energy has been applied to developing quality standards for assessing randomised controlled trials. The case for quality standards has been reinforced by Jüni et al (2002) who show that when quality is compromised in randomised controlled trials, there is a tendency for the results of systematic reviews and meta-analyses to be distorted. The list below is reproduced from Report 4 of the NHS Centre for Reviews and Dissemination (NHSCRD, 2001) where the quality of RCTs is discussed in more depth. The report also references other relevant sources.

---

**Quality criteria for assessing experimental studies**

1. Was the assignment to the treatment groups really random?
2. Was the treatment allocation concealed?
3. Were the groups similar at baseline in terms of prognostic factors?
4. Were the eligibility criteria specified?
5. Were outcome assessors blinded to the treatment allocation?
6. Was the care provider blinded?
7. Was the patient blinded?
8. Were the point estimates and measure of variability presented for the primary outcome measure?
9. Did the analyses include an intention to treat analysis?

---

**‘Qualitative quality’**

As already noted, there is some debate about whether the standards used for quantitative studies should be used for qualitative research at all (Strauss and Corbin, 1998). A number of researchers have developed ‘distinct’ quality criteria for qualitative research and there is a large literature on the assessment of qualitative research and evaluation, particularly in the field of health studies (including Lincoln and Guba, 1985; Mays and Pope, 1995; Boulton and Fitzpatrick, 1994; and the Medical Sociology Group, 1996). For example, the Medical Sociology Group criteria are as follows.

---

**Medical Sociology Group criteria for appraising qualitative studies**

1. Are the research methods appropriate to the question being asked?
2. Is there a clear connection to an existing body of knowledge?
3. Are the criteria for/approach to sample selection, data collection and analysis clear and systematically applied?
4. Is the relationship between the researchers and researched considered, and have the latter been fully informed?
5. Is sufficient consideration given to how findings are derived from the data and how the validity of the findings were tested?
6. Has evidence for and against the researcher’s interpretation been considered?
7. Is the context for the research adequately described and accounted for?
8. Are findings systematically reported and is sufficient original evidence reported to justify a relationship between evidence and conclusions?
9. Are researchers clear about their own position in relation to the research topic?

---

Bryman (2001) argues that most approaches to quality do have parallels with more traditional reliability, replicability and validity criteria. For example, he suggests that
notions of trustworthiness can be broken down into credibility (parallels with internal validity), transferability (parallels with external validity), dependability (parallels with reliability) and confirmability (parallels with objectivity).

There do remain some crucial philosophical differences, in terms of quality assessment, between disciplines. For example, methodological debates in the natural sciences focus on the quest for ‘truth’ and the elimination of bias. In the social sciences the existence of objective truth is often contested, while bias is often an accepted dimension of knowledge, to be acknowledged rather than eliminated. While many criteria are transferable from one research approach to another, Popay et al (1998) feel that there also needs to be a greater understanding of the different ideas about knowledge that underpin different research approaches. They suggest that the following questions need to be addressed in assessing qualitative research.

**Widening the notion of quality in qualitative research – some questions to ask**

| Lay accounts and privileging of subjective meaning: 'Does the research, as reported, illuminate the subjective meaning, actions and contexts of those being researched?' |
| Evidence of responsiveness to social context and flexibility of design: 'Is there evidence of the adaption and responsiveness of the research design to the circumstances and issues of real-life social settings met during the course of the study?' |
| Evidence of theoretical or purposive sampling: 'Does the sample produce the type of knowledge necessary to understand the structures and processes within which the individuals or situations are located?' |
| Evidence of adequate description: 'Is the description provided detailed enough to allow the researcher or reader to interpret the meaning and context of what is being researched?' |
| Evidence of data quality: 'How are the different sources of knowledge about the same issue compared and contrasted?' |
| Evidence of theoretical and conceptual adequacy: 'How does the research move from a description of the data through quotation or examples, to an analysis and interpretation of the meaning and significance of it?' |

**How are quality criteria used?**

As with all guidelines and standards, perhaps the biggest challenge is the issue of utilisation. It is all very well developing new and/or better quality standards, and checklists, but what is it that will drive people to adopt them or use them effectively? Although there are many sets of quality criteria in existence, it is uncertain how far they are used in the research process. Do all journal editors, for example, make existing criteria available to peer reviewers? Are they circulated to commissioning groups or handed out to students?

If researchers are using existing standards and checklists, the impact on research practice seems to be marginal. In a recent systematic review, Harden et al (1999) produced an amalgamated set of seven criteria for assessing qualitative studies.
They examined a set of studies identified for inclusion in a systematic review and found that only two out of 15 met all seven criteria. For example, less than half gave a clear description of the sample and methods used. Often there was a lack of information in the reported studies relevant to these very basic criteria, making it difficult to feel confident about the results. This study highlights the importance of transparency and clear reporting of essential methodological details such as the objectives of the study, a description of methods and a discussion of the main findings of the study. As Jüni et al (2002) conclude:

The assessment of the methodological quality is intertwined with the quality of reporting.

New developments such as online publishing may help to address this problem. For example, publication constraints are sometimes cited as a factor influencing the amount of information provided by authors of journal articles. Some journals (such as the *British Medical Journal*) are now signposting readers to their online archive containing fuller versions of papers printed in the journal.

One of the less frequently asked questions about the use of quality criteria is what sort of tool would be useful for research users when confronted with a piece of research. This is particularly important where the prime focus of the research is producing evidence to inform policy making and practice. There remain a number of important questions that need to be addressed, including:

- How do research users balance notions of utility and methodological quality?
- Are policy makers and practitioners interested in quality?
- Would policy makers and practitioners prefer to make their own quality assessment with a short (or long!) checklist, or would they prefer the quality judgement to be made elsewhere or by someone else?

### Appraising the evidence base

The inability of formal mechanisms such as peer review and quality criteria always to guarantee high methodological standards is not the only research quality concern prompted by the rise of evidence based policy and practice. It has also raised broader questions about the relevance of much research to policy making and practice. By focusing attention on the published research record, the EBPP movement has revealed

---

**Some suggested criteria for assessing the quality of qualitative research**

1. An explicit account of theoretical framework and/or inclusion of a literature review was given
2. Clearly stated aims and objectives
3. A clear description of content
4. A clear description of sample
5. A clear description of methodology, including systematic data collection methods
6. An analysis of the data by more than one researcher
7. The inclusion of sufficient original data to mediate between data and interpretation

Source: adapted from Harden, A; Weston, R and Oakley, A (1999, p26)
not just that research is sometimes poorly executed but that it fails to address issues in the most appropriate way (Hargreaves, 1996).

The uneven quality of the current policy and practice research base has come under attack from a wide variety of sources across different policy sectors. In the field of education, for example, Tooley and Darby (1998) appraised the quality of studies in four key education journals. Using criteria derived from Hargreaves’ critique of educational research (1996) Tooley and Darby took quality to mean research that:

- Makes a real contribution to fundamental theory or knowledge
- Is relevant to practice
- Is co-ordinated with existing research

They found that, although all the papers scrutinised had been through a formal publication peer review process (that is, sent out to at least two experts in the field for comment), the majority did not meet their pre-defined standards of ‘good practice’.

In terms of quality the authors were particularly concerned about the research methods used, the usability of the findings, the quality of non-empirical educational research and the ‘partisan’ nature of much of the research they assessed. They also highlighted a further dimension of quality, that of inadequate reporting, which made it particularly hard for a quality assessment to take place. They conclude that they are not convinced that current educational research is ‘fit’ for the purpose of informing educational policy and practice.

This notion of research being ‘fit’ for the purpose for which it was designed can be particularly useful when thinking about quality. It helps us to build upon existing appraisals of research design, and allows for notions of appropriateness to be added into the assessment process. Research is, after all, carried out for a wide range of purposes ranging from the solution of specific problems to the advancement of knowledge which may inform policy and practice in highly creative, but much more diffuse, ways. In each case there are different implications for the design, conduct and reporting of the findings.

To look at each of our cases again, some of the questions we might begin to ask include:

- In **Case 1** what methods will best access the attitudes of older people?
- In **Case 2** is a trial the appropriate design to address the question of impact of complex initiatives designed to promote healthy eating?
- In **Case 3** what sort of information does the practitioner want to know about the initiative: Did it work? How was it carried out? What did the lone parents think about it?
- In **Case 4** how might the student find out about school performance? Is a comparison group needed to draw conclusions about the impact of care?
Fitness for purpose and utility

As we have seen, methodological quality alone does not provide a sufficient basis for assessing the value and contribution of research to policy and practice. The main dimension of fitness for purpose, as it is currently used, is the ‘fit’ of the methods to the aim of the research. For example, in Case 1, what are the most appropriate research techniques to explore the reasons why older people do not take up benefits?

A further dimension is the ‘fit’ of the research to the ways in which the findings are likely to be used. This concept has been explored by a number of authors, including Lincoln and Guba (1985) who use the term ‘authenticity’ to discuss a wide range of ‘utilization’ dimensions to quality. Edwards et al (1998) suggest that even if the methodological quality of a study is a source of concern, a policy maker or practitioner might, nonetheless, find something of value in the findings. They introduce the concept of comparing signal (the message of the research) with the noise (methodological quality concerns). In the past, they contend, the emphasis has been placed firmly on issues of noise, hence the wide ranging debate and numerous sets of methodological quality criteria. They propose that criteria be developed to assess signal that might include effect size, relevance and applicability. Signal and noise could then be presented in terms of a signal:noise ratio and the research user could use the information to make an assessment about whether or not to trust and use the findings.

Importantly, they suggest that the value of the signal will vary between individuals, depending on their dilemma, priorities etc. Thus the signal could be large for one individual, but small for someone else who finds the findings irrelevant to their work. In public health Rychetnik et al (2002) come to similar conclusions, highlighting the dangers in using the term ‘best evidence synonymously with what is only one aspect of evidential quality, that is study design’. They argue that the extent to which the research addresses the needs of key stakeholders is an important dimension of quality. In social care, researchers have argued that the involvement of service users in the design and conduct of research is also a significant dimension of quality (Fisher, 2002).

The argument that the signal or ‘fitness for purpose’ should form part of quality assessment, has often been met with some resistance from the research community, which argues that research that lacks methodological rigour should be treated with a great deal of caution. However, for research to be of real use to policy and practice, we suggest that these broader dimensions of quality have something useful to add to the appraisal process.

In addition to signal and noise, in assessing research for its contribution to policy and practice we are also likely to be interested in the quality of the transmission. For example, we would be interested in the ways in which the researchers communicate their messages to potential audience(s). We can see in each of our examples that we can add further questions exploring the fit of the findings to their potential use. For example:
• In Case 1 the commissioner might also want to think about how the information about attitudes to ageing is likely to be presented to and used by policy colleagues, and how it links to current policy priorities.

• In Case 2 the researcher might consider collecting (and reporting) information about the nature, delivery and acceptability of the healthy eating intervention to support the use of the approach elsewhere.

• In Case 3 the practitioner will want to think about how the research information on helping lone parents into work can be incorporated in service planning and delivery. He or she might also want to consider feeding back the findings to study participants.

• In Case 4 the student will need to think about how the results will be written up as part of a thesis and subsequently in peer reviewed journals. He or she will also want to think about how they might be shared with schools, care professionals and the research community.

However, in each case we have demonstrated an interest in the quality of the signal (the messages for policy and practice), the noise (the methodological quality), the fit (of the methods to the purpose of the study) and the transmission (the communication of the findings).

The majority of attention has thus far been paid to the upper left hand corner of this diagram. There is plenty of work to be done in exploring each of the three remaining areas and in discussing how we might reconcile these dimensions of quality. For example, if the findings are highly relevant to policy, but the methods are flawed, is there any way that the findings might be used? Should an assessment of the methodological quality always come first?

What next?

The traditional approach to quality assessment has been to focus on methodological rigour. We have discussed a broader definition of quality that pays closer attention to the ways in which the research will be used and the ways in which it is presented. We have identified a number of dimensions of quality that seem to apply to a variety of types of research. Each of these represents an important dimension of quality assessment:
Quality and transparency in reporting
Is the research presented in a way that can be appraised and used by others?

Methodological quality
Was the research technically well executed?

Appropriateness of the methods
Does the research approach match the defined purpose of the study?

Quality of the messages in the research
Does the research address important policy and practice questions in a way that is both useful and useable?

So much has been written about research quality. Hierarchies, typologies, criteria, standards and checklists abound (Oakley, 2000; Bryman, 2001). However, the real challenge appears to be how we pull this accumulated wisdom together and apply it to drive up standards. A current initiative funded by the Strategy Unit in the Cabinet Office is reviewing and synthesising the literature on quality standards for qualitative evaluations in order to develop a set of standards that can be used by policy makers to appraise evaluations. The Social Care Institute for Excellence is funding a project to develop a typology of knowledge in social care that can be used by researchers, commissioners and those interested in using social care research. In addition to mapping the different dimensions of social care knowledge, it will also explore the quality dimensions of different types of knowledge.

There is also potential for us to make better use of existing quality assurance procedures. For example, there is scope to improve the existing system of publication peer review in order to make this potentially powerful mechanism for quality assessment and filtering more effective (Grayson, 2002). The Economic and Social Research Council has also set up a working group to address issues in the peer review of research proposals. Other emerging structures such as university based ethics committees may also have an impact on the quality of research design, development and dissemination.

It is still unclear who could (and indeed should) carry out quality assessments. Systematic reviews usefully include an assessment of the quality of individual studies by researchers, on behalf of research users. However, the additional dimensions of quality discussed above, such as appraising the quality of the messages, are likely to shift the focus of assessment onto the end user. Do research users feel that they have the time, capacity and inclination for appraising research?

Providing encouragement and incentives to use existing quality standards and procedures remains a challenge. Perhaps the biggest incentive should be the potential to demonstrate the distinctive contribution of research outputs as quality assured products. If research evidence is to take a leading role in policy and practice (and individual decisions) then research users need to feel confident about its quality. This need amongst research users, coupled with the requirement for quality assessment emerging from the use of systematic review methods, should act as a driver for the vast majority of researchers (and research users) interested in raising the profile of research.
This paper has outlined some of the debates about research quality in the context of EBPP. It has argued that quality assurance is one of the distinctive strengths of research, as a source of knowledge for policy and practice, but that existing mechanisms need to be used more effectively and consistently. This paper has also argued for a broader notion of quality that goes beyond a narrow concern with methodological quality, to explore the extent to which the research is relevant and fit for purpose. Extending the notion of quality will not be without its problems, particularly in terms of reconciling methodological quality with the quality of the presentation of the research and the ‘fit’ with policy and practice priorities.

These are challenges researchers and research users will need to address in the context of the EBPP debate. However, to end on a note of caution, quality standards and procedures should never be ‘set in stone.’ They should constantly be exposed to discussion and debate so that they can move with the times and continue to offer ongoing support to those doing and using research. Nor should they be seen in isolation from other key debates in the field of EBPP. In particular, attempts to improve the quality of research have clear links to debates about the utilisation of research, and training and capacity building for researchers and research users. These are issues addressed elsewhere in this Working Paper series, and in the work of the ESRC EvidenceNetwork.
References (web addresses accurate at 27.12.02)

Boaz, A; Ashby, D and Young, K (2002) Systematic reviews: what have they got to offer evidence based policy and practice? ESRC UK Centre for Evidence Based Policy and Practice, Department of Politics, Queen Mary, University of London, Mile End Road, London E1 4NS, Jan, 26pp (Working Paper 2) Available via: http://www.evidencenetwork.org


Canadian Task Force on the Periodic Health Examination (1979) The periodic health examination Canadian Medical Association Journal 3 Nov 121(9) pp1193-1254


Mays, N and Pope, C (1995) Qualitative research: rigour and qualitative research British Medical Journal 8 Jul 311(6997) pp109-112. Available at: http://www.bmj.com/cgi/content/full/311/6997/109

Medical Sociology Group (1996) Criteria for the evaluation of qualitative research papers Medical Sociology News 22(1) pp68-71


Popay, J; Rogers, A and Williams, G (1998) Rationale and standards for the systematic review of qualitative literature in health services research Qualitative Health Research May 8(3) pp341-51

Ryan, M; Scott, D A; Reeves, C; Bate, A; Van Teijlingen, E R; Russell, A M; Napper, M and Robb, C M (2001) Eliciting public preferences for healthcare: a systematic review of techniques 186pp (Health Technology Assessment Vol 5 (5)) Available at: http://www.hta.nhsweb.nhs.uk/fullmono/mon505.pdf

