TECHNOPOLIS

Reviews Reviewed: Lessons from the First Six International Panel Reviews, 1999 - 2004

Erik Arnold
Alessandro Muscio
Rapela Zaman

January 2005

Summary

In 1999, EPSRC– in co-operation with relevant learned societies – began a cycle of international panel-based reviews of the state of UK science in fields corresponding to its major programmes. This short study reviews the reviews at the end of the first cycle, describes the reactions of some participants and the way in which similar exercises are conducted in four other countries before drawing conclusions and making some recommendations about potential improvements to the next cycle.

In general, the reviews have provided potentially useful insights into structural aspects of the health of UK science in ways that the RAE does not. The use of international panels provides a useful critical perspective while their high scientific status puts them in a position to be frank. Through the cycle so far, reviews have learnt from their predecessors, though there appears still to be room for further improvements in the process. A significant concern in the scientific community is that the role of the reviews is not clear and that they have few visible consequences, even if it is said that they do influence the debate and science policy.

International practice in doing these kinds of review varies. Large country practice includes the use not only of peer review, as is the case in the UK, but also involves other techniques including scientometrics of various kinds. Using multiple evaluation tools and more explicitly presenting evidence provides greater coverage of issues and, in our view, makes the reviews more convincing. Achieving this requires that the peer panels involved are supported by specialist competence in research policy and evaluation methods.

We recommend that EPSRC considers

- The role and place of the reviews in the broader national debate about the health of UK science
- Feeding back to the community about the consequences of each report, after a delay of about a year
- Encouraging more debate about the reviews and stressing their role as contributions to policy debate
- Paying particular attention to sub-field coverage in panel composition and securing, where possible, some overlap in competences
- Providing more professional science policy support to the Chair, generating an issues paper ahead of UK visits and helping to clarify questions to those visited for the panel itself
- Using and incorporating evidence from other evaluation techniques than peer review in the reviews
- Feeding back the findings of reviews to the affected communities for discussion, before they are finalised

Contents

1 Introduction and Method

EPSRC asked Technopolis briefly to study its practice of inviting international panels of scientists to review the state of UK science in fields corresponding to its major programme areas. These reviews began in 1999 on a cycle where each field is evaluated every six years. Engineering was the first field to be considered, in 1999, and is going through a second review in 2005.

Our work comprised

- Reading the reviews, and in the case of Chemistry also reviewing the material made available to the visiting panel
- Interviewing some of the people who had been involved, to obtain perspectives from different types of participant
- Looking at the way field reviews are done in some other countries, in order to look for lessons from non-UK practice

We are grateful to the following people for talking (or in three cases e-mailing) with us about their experience with the review process and were impressed at their degree of interest in our questions, their helpfulness and frankness. Of course, none of their individual remarks is quoted or cited in this report.

Exhibit 1 List of Interviewees

	Steering Committee	Panel Chair	Panel Member	Secretariat	Visitees
Engineering (1999)			Robin Paul	Alison Bowen	
Physics (2000)		Alexander Bradshaw		Philip Diamond Tajinder Panesor	Richard Jones Mike Moore
Computer Sciences (2001)		Fred Schneider		Mike Rodd	Nick Fddian
Materials (2002)	Julia Higgins		Christian Colliex	Annie Davison	Neil Alford R Freer
Chemistry (2003)	Steve Ley			Alejandro Palermo	
Mathematics (2004)		Jean-Pierre Bourguignon	Peter Hall	Peter Cooper	Wojtek Zakrzewski

Our report is structured as follows

- How were the reviews done?
- What general lessons emerge from reading the first round of six reviews?
- What do participants have to say about the process and lessons that could be learnt?
- How are similar things done outside the UK?
- What lessons should we draw?

We gratefully acknowledge the support of our interviewees and colleagues at EPSRC and foreign research councils. Naturally, the usual disclaimers apply.

2 How Were the Reviews Done?

The reviews were in each case overseen by a Steering Committee, which set terms of reference and appointed the chair of the panel. Practices differed among reviews as to who then chose the other panellists and the influence of the chair in this choice. (The reports do not describe this.) A secretariat was appointed from one of the organisations supporting the review (**Exhibit 2**), which handled logistics, provided background information to the panel and in some cases helped with drafting.

Exhibit 2 Institutions Supporting Reviews (in Addition to EPSRC)

Review	Institutions Involved in Addition to EPSRC
Engineering (1999)[1]	Royal Academy of Engineering
Physics (2000)[2]	Institute of Physics Royal Asronomical Society PPARC
Computer Sciences (2001)[3]	IEE British Computer Society
Materials (2002)[4]	Institute of Materials
Chemistry (2003)[5]	Royal Society of Chemistry
Mathematics (2004)[6]	Council for the Mathematical Sciences (comprising the London Mathematical Society, the Institute of Mathematics and its Applications and the Royal Statistical Society)

The panel was sent background documentation, to allow it to familiarise itself with data and information about the UK situation. It then convened in London and received presentations about funding structures, academic structures and the field under review in the UK before splitting up into sub-groups to visit universities and other key research-performing institutions. It then reconvened to discuss findings, leaving the chair to draft or orchestrate the drafting of a report and circulate it among the panel members. In at least one case, the chair later returned to the UK to present the findings to members of the scientific community affected.

Exhibit 3 shows the terms of reference of the review panels. While their phrasing varies, the message is rather consistent: assess the comparative international standing of UK science and tell us what needs to be done to maintain or increase its health and its scientific and socio-economic potential. In fact, there is a reference to industry or 'well being' in only four of the six terms of reference, but all the reports underline the importance of their respective disciplines to the economy (based on anecdotes) and all consider aspects of industrial linkage in their discussions of sub-fields, where apparently relevant.

---

1 EPSRC and Royal Academy of Engineering, International Perceptions of UK Engineering Research, 1999

2 EPSRC, PPARC, Institute of Physics, International Perceptions of UK Research in Physics and Astronomy, 2000

3 Fred Schneider, Mike Rodd (eds.), Internatioanl Review of UK Research in Computer Science, EPSRC, IEE, BCS, 2001

4 EPSRC, Institute of Materials, Enabling the Future: A Perspective on UK Materials Research, 2002

5 EPSRC, Royal Society of Chemistry, Chemistry at the Centre: An International Assessment of University Research in Chemistry in the UK, 2003

6 EPSRC and the Council for the Mathematical Sciences, An International Review of UK Research in Mathematics, 2004

Exhibit 3 Terms of Reference

Review	Terms of Reference
Engineering (1999)	The panel should benchmark the quality of UK engineering on an international scale ● consider the available data; assess research quality through visits to specific Academic Institutions ● discuss assessment procedures and current research activities with senior academics and industrialists ● comment on the scale of activity and quality of facilities in academic engineering in the UK ● comment on the degree of collaboration of academic researchers with industrial and other user ● report on the outcome of their assessment
Physics (2000)	To report on the standing, balance and potential of physics research in the UK, and provide a comparison with physics research internationally.
Computer Sciences (2001)	To report on the standing and potential of computer science research in the UK and provide a comparison with computer science research internationally. a) The study will cover all computer science research undertaken in UK universities. b) The study will assess the current standing of computer science research and its likely impact on the UK science base, and on the nation's wealth and well-being. c) The study will draw on existing data, international assessments and site visits. d) The study will result in a report to be published and made available to the sponsors and the UK computer science community.
Materials (2002)	● Report on the current standing of materials research and its likely impact on the UK science base and on the nation's future wealth and well being; ● Cover all materials research supported by the UK Research Councils in UK universities and central laboratories; ● To draw on existing data, international assessments and site visits.
Chemistry (2003)	● To report on the calibre, standing, and research potential in the chemical sciences in UK universities. ● To discuss the impact of chemistry research on the well being of the UK science base, and the wealth and standing of the UK's chemical-based industrial sectors and other areas of the chemical sciences. ● To provide comparisons with international research in the chemical sciences.
Mathematics (2004)	To assess the standing and potential of mathematics research in UK universities and comparable institutions, in comparison with international work in the field. The Review should look at the width and quality of all aspects of research in mathematics and its applications, encompassing all of pure and applied mathematics, and statistics.

The composition of the panels is summarised in **Exhibit 4**. Only in the first exercise were people from UK institutions involved in the panel. Overall, 50% of the panellists had US affiliations while 33% were from the rest of Europe. Only 9% were women. Panels Chairs were evenly divided between the USA and Europe (including the UK). Given the age needed to attain high levels of scientific eminence and the fact that the life sciences are not in scope, the panel composition is very much the white, male club one would expect. In most sciences, demographic change will alter this over time. It was clear from our interviews that efforts had been in negotiations to be inclusive, but the insistence on recruiting to the panels people seen by their peers as being authoritative in their judgements means that they necessarily reflect the composition of existing scientific communities. Age, experience and wide knowledge were seen as prerequisites for getting to useful and credible conclusions from the limited amount of data and 'field' exposure involved in these kinds of exercise.

Exhibit 4 Location of Institutions to which Panelists were Affiliated and Proportion of Female Panelists

Review	USA	UK	Rest of Europe	Japan	Aus, NZ, Canada	Rest of World	Fract-ion Women
Engineering (1999)	1	**3**	5	1	2	-	0/12
Physics (2000)	6	-	**5**	-	-	-	0/11
Computer Sciences (2001)	**10**	-	3	-	-	Israel	3/14
Materials (2002)	**7**	-	4	1	1	-	1/13
Chemistry (2003)	**7**	-	4	-	1	India	2/13
Mathematics (2004)	7	-	**4**	-	2	-	1/13
Total all panellists	38	3	**25**	2	6	2	7/76
Chairs	3	1	**2**	-	-	-	-

Note: Bold figure denotes location of panel Chair's institution

The engineering, computer science and chemistry panels each included one industrial member. All others were academics.

The Materials review pointed out an important difference of perception between US and European panelists. "The Europeans were generally more impressed by the health of the research enterprise than the US members…. The view of the US members is less sanguine and perhaps more strongly expressed in the report. Their perceptions are based on shortcomings perceived relative to a US system that embraces funding diversity and an extensive network for multidisciplinary activity, with well-developed mechanisms for interaction between universities and industry, including an effective role for technology road maps."[7] There may be especial value in US perceptions. Even if these can be overly US-centric, they reflect the position of the world's largest and most successful national scientific community and, often, also an especially good understanding of the connections between science and leading industry.

Sources available to the panels before their visit to the UK typically (but not always) include

- UK policy documents and descriptions or presentations of how the UK research system operates and is funded
- Human resource data for the field in the UK
- Existing field-specific reports, descriptions or 'landscapes'
- Some bibliometric evidence, of varying depth but not much analysed
- RAE scores and reports
- Foresight documentation, where available

The panels tended to request various pieces of additional information during their visits. In three of the six cases, the panel used a survey to triangulate on its own impressions. Outside peer review, the principle of triangulation is fundamental to evaluation. However, as one would expect from the fact that the panellists are good hard scientists, and not specialist evaluators or social scientists, their use of surveys is not especially advanced.

[7] p5

Most of the panels visited very few universities (**Exhibit 5**), mostly focusing on those with high RAE ratings. The scale of UK academia makes this inevitable but, despite the panels' best efforts, it does mean that what they tend to produce is 'perceptions of the **best** of UK research' in various fields.

Exhibit 5 Numbers of Universities and Other Institutions Visited

Review	**Number of Institutions Visited**
Engineering (1999)	6
Physics (2000)	7
Computer Sciences (2001)	8
Materials (2002)	13
Chemistry (2003)	9
Mathematics (2004)	Meetings were held with 42 universities via regional meetings at 8 of them

It is not clear to us that the maths review's practice of meeting groups of universities could be transferred to other disciplines. Mathematics is, according to the review, somewhat fragmented in the UK and does not appear to rely on the kind of infrastructure that has to be physically inspected in order to be appreciated.

Panels made the following critical observations on the process in their reports

- Three of the panels (Physics, CS, MS) said in their reports that they had gaps in their disciplinary coverage.
- Two said there was too little time to accomplish the review (Physics, CS)

3 What Emerges from the Reviews?

This chapter summarises findings. At the end we look across the reviews to see what more general conclusions can be drawn.

3.1 Engineering

The Engineering review (1999) found that

- The quality of engineering research in five and five star rated departments in UK universities rank alongside the quality of the world's best
- In general, engineering research in the UK was in good health; some groups were improving, others declining
- The UK funding system is not yet ideally organised for stimulating the multidisciplinary research cooperation, which international companies are seeking, though the Universities visited are clearly moving in this direction
- EPSRC, the Academy and others offered good funding schemes to encourage cooperation with industry, though the coverage of smaller organisations cold be improved
- Engineering does not work in the same way as basic science, so RAE assessment criteria should reflect this rather than focus strongly on peer-reviewed publications. The panel suggested a number of alternative measures
- RAE criteria obtaining in 1999 were adversely modifying the behaviour of universities, by directing research away from the most relevant engineering problems with applications to industry
- Post-graduate grants and salaries for junior university staff were not competitive with industrial rates, and should be increased in order to prevent a collapse in recruitment

It also stressed the importance of site visits in order to understand quality, which was in some cases changing.

3.2 Physics and Astronomy

The physics and astronomy review (2000) found that

At its best, UK research in physics and astronomy was at the highest level world wide

- Beneath the peaks, quality noticeably drops, largely due to lack of adequate resources. For the same reason, coverage in some sub-fields is inadequate
- Physics research in the UK continues to suffer from low funding levels
- Efforts must be made to attract and retain young people and women, not least since one third of UK academic research physicists would retire within 10 years
- Serious attention must be paid to enhancing the career development of young physicists. Salaries are not internationally competitive
- UK support for international facilities is vital to national scientific strength

- Funders should give physicists the freedom to follow their scientific instinct, and use managed programmes with restraint. Research councils should fund more interdisciplinary work and more young people. UK research policies, especially those for postgraduate education, were too rigid and formulaic

The panel made much of the industrial and economic importance of physics. In the detail of the report, where it comments on sub-fields, the panel sends (on our admittedly lay reading) signals with alarming economic as well as scientific implications.

UK atomic physics has not kept up with new techniques, such as the creation of short pulse light sources, and UK work has moved from being world-leading to being derivative. It warns of the need to keep up photonics and optics. Equipment renewal is needed in plasma physics and in the computer power needed for computational physics. With a small number of honourable exceptions, surface science was weak, as was the area of molecular and nano materials. "The UK physics community must take a proactive role in identifying and supporting such rapidly developing areas … which have potential for wealth generation and new and exciting science."[8] While nuclear physics had recently reinvented itself, nuclear theory was "modest" and academic research in applied areas, such as nuclear waste studies, was essentially absent.

3.3 Computer Science

The Computer Science review found that

- The UK has historically been at the forefront of computer science
- While it retains world leadership in some areas, quality declines are already evident in some areas and more will follow, given current levels of support and the nature of today's university research environment. This decline can have serious industrial consequences, since much of it is in areas with large actual or potential scope for industrial innovation in the UK
- Salaries for postgraduates and faculty are uncompetitive an should be increased, especially for younger people who lack the experience needed to do consulting in addition to academic work
- Computer science does not work in the same way as natural science, even though many of its concerns are very fundamental. "Viewing computer science through the lens of traditional academic disciplines can lead to misconceptions."[9]
- Research infrastructure is inadequate, and the UK style of funding militates against building the 'platforms' needed in important parts of CS
- Response-mode funding is low and projects are often too small
- Assessment processes discriminate against larger projects and more risky types of work, instead favouring safe incrementalism
- The misconception that computer science is 'only' applied science leads to over-emphasis on industrial participation, which also promotes incrementalism. In this industry, radical and economically significant

[8] p14
[9] p1

innovations tend to be made by new companies – not the old ones who have the luxury of partnering with academic projects

- Computer scientists need to be represented in EPSRC panels, in order to remove misconceptions about the field
- IT hardware and Computer Science are very different in the way they do science and in their funding needs. They should be funded via separate programmes

3.4 Materials

The main findings of the Materials review were

- Since the UK cannot lead in everything, it was right to prioritise funding effort into parts of the materials field. However, US-style road mapping would be better than Foresight as a prioritisation mechanism
- The assessment suggests that there are a few world-class individuals and small groups operating effectively in the UK system, but that many sub-fields are in decline compared with the UK's earlier relative position
- The UK's approach to interdisciplinary research has been good. However, the system lacks research nodes that link across university disciplines and to industry (such as Australia's Cooperative Research Centres)
- Nonetheless, more money needs to be put into 'bottom-up' initiatives, large-scale and expensive-to-maintain facilities
- Assessment processes for multidisciplinary initiatives need to be redesigned so to reduce the correlation between novelty and unfavourable panel-style review
- Excessive industrial involvement has reduced the international standing of some researchers
- There is too little funding diversity, with EPSRC and the EU Framework Programme as the main sources
- The UK's approach to IPR is inconsistent and needs to be better organised
- There is too little funding for Materials research in general and for young researchers in particular
- Salaries for young faculty are uncompetitive and should be improved
- Managed programmes can play an important role in building needed research capacity, but are perceived as poor quality and should only form a small minority of EPSRC's spending

3.5 Chemistry

These are the major conclusion[10]s of the committee

- The quality of scholarship in chemistry in the UK *is* comparable with the world's best; levels of innovation and discovery are lower.

---

[10] These are directly quoted from the committee's report. Other review summaries are paraphrases

- The UK has three areas of particular technical strength: i) protein chemistry, ii) synthesis, iii) theoretical and experimental studies of chemical reactivity, properties, and dynamics.
- Facilities in the top universities are world-class.
- The UK lags in two areas widely believed to represent the most important new opportunities for chemistry: chemical biology and materials science.
- The academic community appears not to recruit globally for talented faculty and graduate students as effectively as its competitors. It also lags in building a diverse (in gender and cultural background) workforce.
- Chemistry has relative little contact with chemical engineering in the UK, and chemical engineering (although not explicitly reviewed by the committee) seems more narrowly based here than elsewhere.
- Academic chemistry in the UK is highly dependent financially on direct industrial support. The close relationship between industry and university makes academic chemistry responsive to industrial preferences; this responsiveness is both a strength and a weakness.
- The infrastructure supporting chemistry assumes that it is “small science”. Research is in the form of small, targeted, short-term, single-investigator grants. Mechanisms to support the multigroup, interdisciplinary projects important in materials science and some areas of chemical biology and biomedicine are not well developed.

The committee made three major recommendations to EPSRC:

- Restructure Ph.D. Education and Support for Chemical Research to Encourage Revolutionary and Multidisciplinary Science and Engineering.
  - Redefine the Ph.D. Consider the programme leading to the Ph.D. degree as “education”, not “training”: that is, as a programme intended to generate independent researchers capable of creative and innovative research. Increase the time to Ph.D. to four years, and increase the flexibility of Ph.D. programmes.
  - Develop mechanisms for support that reward discovery and innovation, and that provide the three- to five-year continuity in support required to take risks in research.
  - Develop mechanisms that support multidisciplinary research.
  - Emphasize the expansion of UK chemistry into biology and materials science, conserve historical strengths in scholarship, and build bridges to chemical engineering.
  - Increase mobility of students and faculty among universities.
- Compete for scientific talent globally.
- Work with the chemical community to develop and articulate opportunities, and formulate a strategy for academic chemistry.
  - Develop a rolling strategic plan for academic chemistry.
  - Encourage the academic chemical community to *generate* initiatives, rather than responding to those developed by government.
  - Use the power of the top departments to strengthen departments at the next level and increase the number of internationally first-tier departments; preserve the strength of the current top-tier departments.

In addition to these points (the text above reproduces the executive summary of the review) the body of the report makes interesting additions and amplifications

- UK chemistry has tended to stay with its areas of historical strength, rather than move into the newer sub-fields. This amounts to a commitment (encouraged by the high share – one third – of university chemistry research funding that comes from industry) to incrementalism and to the mature-phase chemical and pharmaceutical industries in the UK. It is especially dangerous for young researchers, who easily become locked into the old sub-fields and short-medium term research questions
- The UK is weak in interdisciplinary research and uses a response-mode funding system that creates fragmentation, rather than the bigger concentrations of capacity needed in the emerging, interdisciplinary sub-fields
- The relationship between UK chemistry and chemical engineering is weak, partly because chemical engineering itself is weak in the UK

3.6 Mathematics

The mathematics review comments at a fairly detailed level on various sub-fields and especially notes UK strengths in many parts of pure and applied mathematics. It points out that some parts of mathematics traditionally seen as 'pure' – notably number theory and combinatorics – now have very important computing applications, and that the links between pure mathematics and computer science in the UK are weaker than elsewhere. Similarly, the mathematics associated with materials science – and especially the important area of nanoscience – was sub-critical and insufficiently connected to materials science. It comments (like many of the other reviews) on weakening mathematics education in schools, the short duration of PhD studies in the UK and lack of support mechanisms for mathematicians in mid career. The panel criticised the RAE, which had useful aspects but involved too much paperwork, reduced researchers' willingness to take risks, tended to prevent intra-UK networking and cooperation and led to too much concentration of the best mathematics in too few departments. It identifies a problem of ageing in the statistics community and argues that EPSRC assessment practice discriminates against statisticians, making it hard for young researchers to become established. The outlook for the UK statistics community was said to be alarming. The UK would need to strengthen training, adopting a more continental model, and increase the number of universities with leading mathematics capabilities in order to secure the health of the discipline.

3.7 Discussion

The reviews have a number of common threads, suggesting there may be generic structural questions to be addressed in UK science

- All the reviews discuss the socio-economic – especially industrial – importance of the field being considered
- They all complain about the low levels of postgraduate stipends and academic wages in the UK, and several argue that the level of debt that students can build up during their first degree is an additional disincentive to working in UK academia
- In several cases, it was said that the EPSRC assessment process for response-mode funding caused scientific conservatism and fragmentation, while in

many areas incentives were needed to greater risk taking and building bigger research groups

- Four of the reviews (CS, MS, Chemistry, Maths) argued that UK efforts to reduce PhD completion times to three years were misguided, producing narrow and insufficiently experienced PhD holders with limited understanding of the 'craft' of research. The Bologna '3+2+3' model was generally preferred
- The CS and MS reviews criticised the confused policies and capabilities of universities in dealing with IPR issues
- Given the modest level of popularity of the RAE[11] among UK academics, it had clearly been much discussed during university visits and elsewhere and it is perhaps not surprising that it came in for a lot of criticism. This related to its potentially distorting effects on science and the university structure, not least through its use of the 'Matthew principle' and the implication that the losers lose the money they would need to build themselves into future winners, the incentives it provides for 'safe' incremental and publishable work, and so on. Most reviews criticised aspects of it, and MS proposed that it should be abolished.
- Three reviews (MS, CS, Chemistry) warned that overdoing industrial collaboration was in certain cases proving detrimental to science and backed this up with specific arguments (rather than merely offering the prejudice that exists in parts of the scientific community about industrial involvement being detrimental to science *per se*)

There is evidence of learning early on in the way the reviews are conducted. The difference between the engineering report and the subsequent reviews is striking, not only because the engineering review is short by comparison with the others but also because it goes into no technical detail. It broadly says that the good departments are good and that the fairly good ones contain some good bits, without making much comment on the state of different sub-fields. It raises a range of important broader policy issues to do with recruitment, pay, proposal assessment and so on. The panel comprised very eminent men who are described as engineering academicians, heads of universities or funding councils. It is not clear to what extent they were still practising researchers, though evidently some were.

The subsequent reports go into the sub-field specifics and therefore produce longer reports with more detail about strengths and weaknesses. The panellists are leaders in their scientific fields but less institutionally eminent: most are affiliated to university departments. For the purpose of digging into the state of the fields, this appears to be a more useful level, and the elimination of UK membership makes the subsequent reviews more credible. By and large, the second review – of Physics – appears to have served as a 'template' for the later ones.

Except in Chemistry (where we have access to the full set of information provided to the panel), we have not been able to examine RAE reports. However, the contrast

[11] Anecdotally, it may be worth noting that in our non-UK R&D evaluation work, we find that the RAE is well known and much discussed among academics in many countries as the 'horror example' of bureaucracy and evaluation gone mad, so visitors to the UK are likely to bring preconceptions with them

between the Chemistry review and what the RAE Chemistry panel had to say[12] at the overall level is striking. The RAE report says that chemistry has substantially improved in UK universities between 1996 and 2001. It points to the conservatism of the peer review techniques used and sounds a warning about under-investment in instrumental techniques. It notes that UK chemistry is starting to enter some of the new areas referred to by the Chemistry review panel, but warns that this should not be done at the expense of the traditional 'core' of chemistry as a discipline.

Clearly, we are in no position to adjudicate between the rather different perceptions of the health of the field offered by the RAE panel and the Chemistry review panel, nor, indeed do we know whether the review simply repeats conventional wisdom about problems in UK chemistry. On the face of it, however, there would seem to be considerable value in having the issues raised by the review exposed for discussion as they have huge scientific and industrial consequences.

A final, somewhat disconcerting impression from a parallel reading of the reviews (Physics, MS, CS, Chemistry and to a small degree Maths) is of a number of traditional strengths in UK science – often in areas clearly linked to industry – either being in decline or not being supplemented by strong entry into new fields.

---

[12] This was part of the background material provided to the Chemistry review panel

4 What Do the Participants Think?

We conducted a small survey among participants in the reviews. The sample we achieved (see Introduction) was not in any sense intended to be statistically representative, given the tight budgetary constraints of this short study. In this section we synthesise the main messages from the interviews.

Very many of the interviewees wanted to see a clearer definition of the purpose of the reviews. While almost everyone said they needed to be clearly differentiated from the RAE, it was well understood that their purpose was more to look at structural issues. Most interviewees felt the reviews' impact had been modest, and this fed into the sense that their purpose was unclear. The role for the reviews that most find valuable is to map performance at the sub-field level and raise structural issues specific to the fields.

The process seems to have been improving, with later reviews learning from the experience of earlier ones. Overall, the following improvements were seen as being needed. (The reader should recall that, because respondents were discussing past experience, some of these may already have been set in train without respondents' knowledge.)

- The reviews need a longer lead-time. Many of the panellists are committed for 1-2 years ahead, so a great deal of notice is needed in order to secure suitable people's participation
- Background materials were hard to absorb and use. Some pre-digestion and synthesis would be helpful
- Better definition of the issues to be tackled was needed before the panels met. This would clarify their brief, focus the acquisition and analysis of background material and allow the universities to be visited to prepare the right kind of presentations. Some had had to abandon their prepared materials and present 'on the hoof' in response to panels' questions prepared at the last minute
- Panels were constrained by the one-week format, and two of the most senior respondents independently suggested that two week-long visits separated by a gap of two to three months would be a practical and improved alternative
- A small number of contextual errors were said to have been made in reports, suggesting that they would benefit from some feedback or review process with the relevant scientific community before being published. The London meeting following the Chemistry review appeared to have been a success
- Views about the six-year interval between reviews varied among disciplines. Materials and computing science are fast changing, so there are scientific reasons for looking at them fairly frequently. It was not clear that physics needed such frequent inspection on scientific grounds. However, it was also pointed out that there can be important changes in the health of individual research groups over 5-6 years

Respondents were emphatic that the capabilities and experience of the Chair are key to the success of the reviews. It was difficult to construct perfect panels and those involved thought this had been achieved reasonably well. Many saw some gaps in capability were as being inevitable. A very small minority of the panellists were seen as non-contributors or as anti-social, within the context of the team. In one case identified to us, shirking by a panel member had meant that coverage of an important sub-field appeared inadequate.

Few interviewees had much to say about the steering committees, whose role is primarily concerned with setting up the reviews and the panels, so their work is invisible to most. The involvement of the learned societies was vital, because they linked the reviews to the scientific communities. Several people argued that, where relevant, the reviews should also involve more than one research council. The reviews should clearly be of ‘maths in the UK’ not ‘EPSRC maths in the UK’.

Panellists were quite strongly motivated to take part. An invitation to join such a panel was seen as an honour, and the reviews were opportunities to learn as well as to contribute. “It’s a scientific experience.” “The panel members benefited enormously.” The two people who suggested that future reviews should take not one but two weeks argued strongly that there were sufficient incentives to ensure that people would participate.

Interviewees were generally happy with the issue coverage in most of the reviews. These dealt with excellence at the sub-field level, providing differentiated judgements that could form the basis for policy or further investigation. They dealt with the structural needs and issues of the fields. They were less good at setting the strengths and weaknesses of the UK in various sub-fields against a view of future developments. We infer that more could therefore be said about needed investment patterns to maintain the health of the relevant communities and enable developments.

Participants were generally content that peer review was the right primary mechanism though a number felt that background information had not adequately been exploited by the panel. Most felt that, although time was very limited, it was possible to do useful work in and around a one-week visit in order to expose the key structural issues, despite the large size of the UK scientific community. The time constraint meant that panels had to focus only on the best-performing research groups. It followed that they learnt little about performance at lower levels, but panellists tended to feel this was not necessary. They saw the reviews as discovering whether the best was good enough, rather than providing a wider survey of capabilities. (It follows that these reviews are alone an inadequate guide to making policy about national capacity.)

Panel composition was generally seen as sound. Almost nobody thought it was a good idea to involve younger researchers, owing to their lack of experience and perspective. The gender imbalance was a function of the composition of the relevant scientific communities. Some people in secretariat roles were concerned about it, but the (predominantly male) panellists were generally not. The mix of nationalities was not seen as problematic, provided panels were not completely dominated by one country (ie the USA). It was important to find leaders in the fields being reviewed, rather than to worry about geographical representation. No one could give instances of a panel being unduly dominated by paricular individuals.

The reviews cost a great deal of time for all involved, so their financial cost significantly understates their opportunity cost. The main issue in discussing whether the reviews provided value for money was whether they had consequences. If they changed funding priorities or otherwise led someone to tackle the structural problems they identified, then they were definitely worth the effort. However, most interviewees felt that the reviews changed little or nothing. Panel chairman were better informed about consequences, but these were often described as ‘the review was used in meetings’.

We interpret this to mean that the way reviews influence policy is indirect and not transparent. For example, the fact that they are considered by EPSRC’s Technical Opportunities Panel, which decides on the division of funds between scientific fields and disciplines, is not obvious to most people. From the scientists’ perspective, these reviews “disappear into a black hole” and are rarely seen again. The scientific community would, of course, like such reviews to lead rather directly to increased research funding. While it appears that those involved are routinely warned that there is no one-to-one relationship between review findings and changes to policy (especially increases in funding), the community seems nonetheless to be disappointed when it discovers that this is true. It might be useful to orchestrate some kind of response or follow-up to the reviews, at least from EPSRC and the learned societies involved in their production.

Some of those we interviewed were aware of equivalent foreign practices, mostly in smaller countries. Except in relation to a rather detailed US review process, none was able to make suggestions for improvement based on foreign practice.

The question EPSRC asked us to put about the utility of involving a professional evaluator in the process raised hackles in a number of cases. Reactions were mostly negative. This would make the reviews ‘mechanistic’; there was no need for such support; non-members of the scientific community involved could not make any useful contribution. Here, as with the absence of use of scientometric evidence in the reviews, we are touching on a matter that is deeply political. These things would intrude on the historical role of peer review panels as sole arbiters on behalf of the scientific community. The ‘scientific alpha males’ (amongst whom we include the women!) who inhabit the review panels do not take kindly to what they perceive as challenges to their position. We return to this question in our conclusions.

5 Lessons from Abroad

In an earlier note to EPSRC, we reported that we had non-exhaustively surveyed the way a number of other countries tackled field or discipline evaluations. We were interested in finding a little diversity so that alternatives and additions can be explored, rather than simply to look for clones of the UK procedure. We agreed with EPSRC that we would look at US, Swedish, Finnish and French practice.

5.1 USA: COSEPUP Benchmarking Experiments

The US Government Performance and Results Act (GPRA) of 1993 has set the tone for much evaluative activity in the USA and influenced practice worldwide to move towards clearer, more quantitative evaluations and performance indicators. As the culmination of various discussions and reports on the goals of US science and the need to benchmark performance, the Committee on Science, Engineering and Public Policy of the National Academies decided in 1997 to do an experiment, which was reported[13] last year.

It commissioned benchmarks of US performance in mathematics, materials and engineering, and immunology, using a range of methods for each field. (Differences in method were driven by differences in the characteristics of the fields.) Via a system of oversight committees, it appointed a panel for each benchmark, comprising eminent scientists from the field, scientists from closely related fields and 'users' of research results. The panels included a significant minority of non-US members. They sought geographical and professional diversity and stressed that industrial membership was essential to provide perspective on the work of leading industrial researchers and on the ultimate utility of research. Each panel was asked to determine

1. What is the position of US research in the field relative to that of other regions or countries?
2. On the basis of current trends in the United States and worldwide, what will be our relative position in the near and longer-term future?
3. What are the key factors influencing relative US performance in the field?

The methods used included

- The **virtual congress**, which was used by all the panels. Each panel divided its remit into sub-fields and asked 5-15 respected leaders in these sub-fields to imagine organising a seminar of 5-20 speakers, and to supply the names of the people they would choose. This information was then analysed to construct tables of US and other countries' performance
- **Citation analysis**
- **Journal publication analysis**. The immunology panel identified the five key journals in the field and analysed the origins of those publishing on immunology
- **Quantitative data analysis** involved simple comparisons of data such as trends in numbers of graduates and PhDs, but struggled with data

[13] COSEPUP, *Experiments in International Benchmarking of US Research Fields*, Washington DC: National Academy Press, 2004

incompatibilities and the absence of non-US data for many of the things in which they were interested

- **Prize analysis**. While recognising that looking at who won scientific prizes in the field would involve issues of small numbers, this information nonetheless proved consistent with that from other sources
- **International congress speakers** - real congresses, this time – were analysed to provide national participation data

Each panel assessed the particular sub-set of these analyses that was relevant to it and went on to make a diagnosis based on the analyses and members' knowledge of causes of problems and factors influencing US performance. They also made judgements about the future likely position of the US science and measures that could be taken to support it. (Note that the recommendation 'increase funding' was forbidden from the outset.)

The reports of the panels are rich in data and analysis. The panels used convergent partial indicators, supplemented by qualitative inputs, to reach conclusions that they regard as robust – despite the fact that many key participants joined in at the outset on the basis of 'suspending their disbelief'. A separate National Science Foundation review of mathematics came up with closely convergent findings to those of COSEPUP's panel. Surprising findings appeared – for example, it had not been understood to what an extent US leadership in mathematics relies on importing mathematicians. The reports were evaluated and process improvements were suggested in the detail of some of the methods. It was decided that the exercise should be repeated, using 50% foreigners and with more standardised methods. A particular source of enthusiasm was that the exercise cost only $50 000 per field – but it should be noted that this was because that almost everyone involved worked on a *pro bono* basis.

5.2 The Academy Of Finland

The Academy is in fact the umbrella research council in Finland, of which the four Finnish research councils are effectively sub-committees. Since 1997, it has prepared an overall review of the state and quality of scientific research in Finland[14] once during each of its 3-year terms. This is prepared by a committee of senior members of the academy and the education ministry, supported by two science policy researchers and the Academy's own analysis department.

The book contains good statistical overviews and interpretation of research funding patterns in Finland, human resources and Finnish science policy. Its analysis of the relative performance of Finnish research is entirely based upon rather aggregated intentional comparisons of bibliometric indicators: publications, impact factors, citations, partly at the national level and partly at the level of six broad disciplinary groupings. Each of the four research councils produces a discursive overview of the state of Finnish science in its own area. The reports use secondary sources complemented by the knowledge of the council members. They are drafted by staff and approved by the councils.

---

14 The latest version is Timo Oksanen, Annamaija Lehvo and Anu Nutungen (eds), *Scientific Research in Finland: A Review of its Quality and Impact in the early 2000s*, Helsinki: Academy of Finland, 2003

The 3-yearly global review of the state of Finnish science is complemented by periodic field and programme reviews, which feed into it. The process is similar to, if less formal than, the EPSRC process. Thus, the astronomy review[15] was organised by three academicians and a Finnish astronomer, who appointed a panel of four foreign (European) astronomers to visit and review the four astronomy research groups active in Finland. An Academy staff member acted as secretary, who sent a self-evaluation questionnaire to the groups ahead of their visits and collated the results for the panel, which visited the four groups across three days. The panel produced a detailed assessment of each group, supported by bibliometric counts and assessments. Despite giving a very good account of astronomy and its importance, and reviewing the groups in detail, there is no overall judgement on the health of the field in Finland and there are few recommendations.

The later review[16] of geosciences, which used a similar method, had considerably more teeth. The panel comprised one US, one Canadian and one European academic and a senior Finnish industrialist. It roundly criticised the organisation of the geosciences and the exclusion of non-university research groups from needed computing facilities, argued that salaries and prospects for young geoscience researchers were inadequate, that PhDs were too short, that responsibility for geosciences should be reallocated within the Academy and a geoscientist appointed to the Research Council for Natural Sciences and Engineering.

The Finnish system has the great advantage that it is possible quite quickly to review almost anything, because the system as a whole is so small. As an approach, the combination of frequent overall health checks with peer reviews of fields and programmes allows rather constant monitoring of the state of Finnish science and lets generic questions such as salaries and career progression be discussed at the national level, rather than being buried in a series of field-specific reviews. Willingness to use bibliometric methods (of varying degrees of sophistication) and to employ Academy staff to do analyses and substantial report writing for the panels means the reviews are strongly evidence-based and convincing, without apparently constraining the panel or its authority. While very broad themes are sometimes attempted, such as the mathematics review in 2000, most of the reviews are less broad in scope than those of the EPSRC.

5.3 The Swedish Research Council

The Swedish Research Council (Vetenskapsrådet – VR), like the Finnish Academy, is an umbrella organisation. It incorporates three research councils and an education committee.

VR's field evaluations serve a rather different purpose to those discussed elsewhere in this report. According to VR's description of its standard approach to these reviews, "The purpose of the evaluations is to inform the Swedish Research Council about the scientific quality of the research projects [it has funded], seen in an

---

15 Academy of Finland, *Evaluation of Finnish Astronomy*, Report 4/00, Helsinki: Academy of Finland, 2000

16 Academy of Finland, Finnish Geosciences Evaluation Report, Report 14/3, Helsinki: Academy of Finland, 2002

international context. The projects are thus the main elements entering into the evaluation. However, the expert panel is encouraged to comment also on structural problems." The panels are international, building on suggestions from those being evaluated and appointed by the relevant sub-council of VR. They are only provided with information about VR-funded projects and are required to review these one by one, providing for each a report on

1 Scientific quality
2 Scientific value and potential improvements
3 Merits of the methods used
4 Capabilities of the project leader and team
5 Adequacy of facilities and human resources in the team
6 Other considerations
7 An overall evaluation of the project leader and a recommendation about whether support should be continued, increased or terminated

These field evaluations therefore function primarily to provide feedback to the Council about individuals' performance. They omit activity funded by others. In practice, they say little about the field as a whole. Thus, the recent review of theoretical chemistry[17] is mostly devoted to 35 individual project assessments on the dimensions indicated above, with a handful of paragraphs at the front of the report that suggest the field should receive more funding, increasing the number of chairs in theoretical chemistry, funding more postgraduate students and encouraging collaborative activity. The chemical engineering review[18] has a similar structure, recommending at the overall level that funding should be increased and that a task force should be set up to investigate (and promote) the role of chemical engineering.

This is the most traditional of the review approaches we discuss. As a way to tackle field evaluation in a larger country, it is clearly infeasible, and it is not clear that the narrowly project-focused approach is the best way to gain insight into the shape and health of an entire field. A wide scope and mission for the panel, combined as needed with other approaches, is a necessity for this.

5.4 France: CNER

The National Committee for the Evaluation of Research (CNER) is responsible to the Ministry of Research for doing large-scale evaluations of fields and government research institutions. It periodically (1997, 2003) produces syntheses considering what it can see of French research as a whole. Thus, in 2003, it commented[19] that French public research was difficult to measure but that despite stifling institutional structures, scientific production was satisfactory – essentially owing to the creativity of the researchers. It pointed to the growing importance of evaluation in industrialised countries and for the need for the nation state to find a new role in research policy that takes account of the European as well as the national level.

---

[17] Vetenskapsrådet, *International Review of Theoretical Chemistry*, Stockholm: Vetenskapsrådet, 2002

[18] Vetenskapsrådet, *International Evaluation of Chemical Engineering in Sweden*, Stockholm: Vetenskapsrådet, 2004

[19] CNER, *La Recherche Publique en France : Une Evaluation*, Rapport au Président de la République, Paris : Documentation Française, 2004

CNER has also produced a long and detailed analysis[20] and set of recommendations for how evaluation should be done, touching on issues such as independence and objectivity, using the appropriate mode of evaluation (with some very specific strictures on the limits to bibliometrics at the level of individuals), separation of the powers and the need for multicriterion approaches.

Field or institutional evaluations are done partly at the request of the Minister and partly at CNER's own instigation. CNER's engineering review[21], for example, focused on the 167 laboratories and 13 000 employees of the CNRS in the area. One group of nine experts appointed by the Committee interviewed 127 researchers and managers at CNRS and 72 companies. A team of peers (mixed French and foreign) reviewed quality and reputation using a mixture of bibliometrics, analysis of conferences and inspection. CNER members themselves considered internal management and evaluation procedures and, with staff support, synthesised the whole into an analytic document covering a wide range of issues from the quality and prominence of work in various sub-fields to management.

The conclusions were that, while there were many research successes, the laboratory structures were locked in, badly needed to be revitalised and better linked to the outside world. The review led to major changes in organisation and planning practices, new strategies, inter-laboratory working and the creation of interface functions to relate better to industry. The political clout of CNER appears to have been important in making it possible to be so critical.

5.5 Discussion

This brief look at a sample of foreign practice suggests that there are some alternative practices worth considering in the UK context.

The US, French and Finnish examples all suggest that careful use of bibliometrics is a useful way to inform reviews in those fields where publication practices make analysis tractable. In other areas, alternative scientometric approaches such as analysis of conferences and other forms of 'publication' can be helpful – at least in the hands of people who know what they are doing with these techniques and who can be clear about their limitations. Both the US and French examples underscore the power of using multiple evaluation techniques in parallel. No single technique is perfect, so looking for convergence among more than one technique is a powerful way to build robust conclusions. The US experiments are especially interesting in that they simultaneously underscore the power of using the knowledge of people in the relevant fields and of being analytic in using multiple techniques.

The Swedish example is of very traditional practice using only peer review, and while this type of project by project evaluation can play a useful role in research councils' detailed work, it does not seem to be very informative at the level of the structural issues tackled by the UK, US, Finnish and French evaluations. Also, the Swedish approach quickly runs into difficulty if applied at significant scale.

---

[20] CNER, *Evaluation de la Recherche Publique dans les Etablissements Publics en France*, Paris : CNER, 2002

[21] CNER, *Evaluation du Département des Sciences pour l'Ingénieur du CNRS*, Paris : CNER, 1997

We draw the following conclusions

- Using additional evidence in parallel with classical peer review helps reviewers generate robust and convincing findings, and is particularly helpful in tackling large-scale evaluations
- Such methods need to be adapted to the subject of the evaluation – one size will not fit all
- Since it involves data and analysis, this kind of evidence should be reported explicitly, in addition to forming inputs to peer judgements so that they can be understood and criticised
- The use of multiple evaluation methods requires professional support and capabilities that go beyond those typically provided to the UK review panels
- Looking periodically at the national level, as well as the field level, can expose generic issues that otherwise can be lost in reviews of **parts** of the national system

6 What Can We Learn?

In this final section we discuss some conclusions and make suggestions for future practice.

The place of the reviews in the overall system. Some of the foreign examples involve, in effect, taking national overviews of the state of science as a whole. It would be helpful if the EPSRC reviews were more clearly positioned in relation to such a process, which would then tackle the more generic issues that are raised in individual field evaluations.

Legitimacy. We were impressed that most of the reviews identified structural issues above and beyond the type of concerns found in RAE-style exercises. The reviews have a distinct and potentially valuable role. Wile those who ought to know told us that the reviews influenced policy, it was hard to track this and most of the people we spoke to were disappointed that the reviews apparently lacked consequences. Those visited do put significant effort into these reviews (even if the RAE means they are better prepared than might otherwise be the case). More of a 'win-win' relationship between reviewers and reviewed would be created if the researchers' 'So what?' question could be answered. This would involve better communication about the purpose of the reviews and a report back from EPSRC (or another appropriate organisation) to the relevant community.

Status of the reviews. It seems to us that the status of these kinds of reviews is special and should be understood. In any evaluation of a complex situation, the evaluator is unlikely to observe, understand or analyse everything. Such evaluations should therefore be seen not as Truth but as systematic and well-informed contributions to a debate – a debate, to which others also contribute. Correspondingly, we should neither demand nor expect a perfect job from the review panels. Encouraging a diversity of qualified input into the debate becomes more important than getting a 'balanced' view in individual contributions. The first two reviews, which entitled themselves 'Perceptions of …' seem to us better named than the subsequent ones. These actually do what it says on the packet. For the reviews to be more than this, they would need to use a wider evidence base.

Panel selection. As the difference between the 'all clear' signal given by the Chemistry RAE report for the field and the rather more critical diagnosis offered by the Chemistry review suggests, the value of the review exercises is not in their thoroughness (with the best will in the world, they cannot be thorough) but in the fact that they present the views and impressions of highly-qualified outsiders. The fact that, after the first review, there have been no UK-affiliated panel members involved is therefore very positive. The panels appear to have been well briefed about the UK system and to have had secretariats on hand able to respond to questions, so there is no further need for UK panellists – who, with the status of peers, would be more likely to temper than to inform the judgements of the others.

Panel selection is an area full of potential pitfalls. The engineering panel was nominated by various academies. As a result, the sub-field coverage was not quality-assured and there were said to be some difficulties in working together.

Overall, three of the panels said in their reports that coverage was inadequate. In most cases, the number of sub-fields is so large that – even if it is possible to cover all the important ones – most if not all will be tackled only by one panel member. So what looks superficially like a review of a field by a dozen or so people turns out to be about a dozen or so reviews of sub-fields by one person each. Both our experience and that of Langfeldt[22] (who studied six peer reviews in Norway) is that members of peer panels defer to other members with more specific knowledge. There needs correspondingly to be overlap, not just an absence of gaps, between panellists' sub-field competences. Operating with a single expert per sub-field also runs the risk that the views of the relevant peer may be eccentric or out of date and that there is no effective quality-assurance mechanism.

In one review, a panel was proposed that was made up almost entirely of US researchers. The steering committee objected and negotiated a more balanced geographical make-up. The advantage of the Chair selecting the panel is that the people chosen then tend to be known at least to the Chair, but often to be part of a common social network. The Chair can use experience to select only people who can work with others and to avoid shirkers, but is likely to choose people with views similar to her own. This reduces the probability that the kind of robustly questioning panel needed for this type of exercise is actually delivered. There is no perfect answer. We suspect a good mechanism could involve

- Administratively obtaining nominations from eminent non-UK members of each sub-field, asking not only about scientific qualifications but also about strategic skills and sociability ('could you bear to have dinner with this person every night for a week?')
- During the course of soliciting new nominations, asking also for views about those already nominated
- Assembling a shortlist of 3/4 candidates per sub-field and asking the Chair to veto any felt to be unsuitable
- While ensuring there are no gross geographical or gender distortions (compared with norms in the field under review), then allowing timetable and logistics to determine which short-listed panellists are actually used
- Either appointing two peers per sub-field or preferentially using people active in more than one sub-field

This implies larger panels or narrower scope for the reviews. Alternatively, a wider mix of evaluation techniques could be used to triangulate the peers' findings, in principle reducing the risks of inadequate sub-field coverage.

Process. The reviews are clearly pressed for time. They already require a very large commitment of time from very busy panellists. We do not know at what point people will balk at being asked for more time. A second, shorter meeting of the panel could provide an opportunity to complete its reflections on the report and to have a feedback session with the relevant community, but would of course add significantly to the cost. What is clear, however, is that a very long lead time is needed. The Chair should be working with a support person ahead of the visits to

---

[22] Liv Langfeldt, 'Expert panels evaluating research: decision-making and sources of bias,' *Research Evaluation* 13(1), 2004

define issues and inform visitees of the questions they will be asked. Using professional research policy or evaluation specialists at this point to generate an issues paper based on evidence such as bibliometrics and surveys would generate a better information basis for the reviews, expose more questions for consideration and relieve the panels of the indigestion the currently experience in tackling a 'data dump' of background materials. Publishing this evidence would increase the credibility of the reviews.

Quality assurance and acceptance of findings. Being evaluated is rarely a comfortable process. It is a bit like going to the dentist – you know it is good for you, but do not look forward to it. Hearing and dealing with criticism is no easier, but the process is less painful if it is consultative. Our own practice is therefore whenever possible, to hold a seminar with those we evaluate where we present our findings prior to delivering a final report. This has several advantages

- Those evaluated can be relied upon to point out any mistakes the evaluator has made, so it increases the quality of the overall review
- Discussion of unpalatable findings at a 'draft' stage makes them easier to accept, because those evaluated have a chance to think about them and air their views and counter-arguments
- Where there is a need for change as a result of the evaluation, this can often be announced at the point where the evaluation becomes public, so that criticism becomes constructive and not mere carping

While it would be hard to arrange a meeting with an entire UK discipline, one with key interested figures is possible, as is use of the Nordic practice of 'remiss': making this kind of report available for comment for a period and recording the tenor of those comments (again, so that evaluation becomes part of the 'debate' we discussed earlier).

Use of evidence. We found the panels' use of the evidence with which they were provided ahead of their visits disappointing. For example. while bibliometrics is – like all other approaches, including peer review – an imperfect way to understand the quality and productivity of scientific research, the technique nonetheless does shed some light on these questions. The brightness of the light varies among disciplines and sub-disciplines, to be sure but we were surprised that none of the panels referred to the bibliometric evidence they had at their disposal. Similarly, other more or less 'hard' data (for example, RAE scores and analyses of grants and of age profiles of research communities, data about conferences) provide useful if partial insights. The reviews are so data-free that they could not possibly be considered for publication in a scientific journal.

It is clear that the panels do not have the time to go to work on these background data. Some panellists indicated that they do not even read them, certainly not in advance of their visits to the UK. The panellists' task is not made easier by the form in which they receive these data. Probably because of the strong tradition of demarcation between decision-making and administration in scientific institutions, the secretariats present the panels essentially with a 'data dump'. They do little synthesis or analysis and offer few (if any) tentative conclusions. So the material is not easy to digest and exploit. There is no tradition in these kinds of peer reviews to be explicit about data or how judgements are made. The eminence of the panellists means they speak as Authorities.

However, just as the chemists reached more interesting conclusions than the engineers because they looked at more of the details, we suspect that more systematic use of existing data would help the panels form more and clearer hypotheses that they could investigate during their UK visits. The non-UK practice we described earlier clearly indicates that it is possible to integrate non-peer evaluation techniques into these kinds of reviews. Having a more explicit factual basis would make panels' reports more persuasive – especially outside their own communities, in the places where the big resource allocation decisions are made. The obvious thing to do is to test our hypothesis by providing professional analytical, and not just secretarial, support to future panel chairs in advance of visits and in report writing. This would also help raise the quality of the surveys used by some of the panels, which are in social scientific terms primitive.

We recommend, therefore, that EPSRC considers

- The role and place of the reviews in the broader national debate about the health of UK science
- Feeding back to the community about the consequences of each report, after a delay of about a year
- Encouraging more debate about the reviews and stressing their role as contributions to policy debate
- Paying particular attention to sub-field coverage in panel composition and securing, where possible, some overlap in competences
- Providing more professional science policy support to the Chair, generating an issues paper ahead of UK visits and helping to clarify questions to those visited for the panel itself
- Using and incorporating evidence from other evaluation techniques than peer review in the reviews
- Feeding back the findings of reviews to the affected communities for discussion, before they are finalised