THE ROYAL SOCIETY

1995

Ref:

Peer Review - An assessment of recent developments

Introduction

Peer review has long been central to decision-making in the administration of research - particularly decisions about the allocation of resources at project level. The processes of peer review are not static: like everything else, they respond to changes in the environment within which they operate. How they respond is a matter of broad concern. It is important to monitor developments in peer review and, so far as possible, to ensure that they increase efficiency and effectiveness without compromising key principles.

The conduct of peer review in the UK has been the subject of extensive debate since the White Paper Realising our potential was published in May 1993. The debate has been triggered by, *inter alia*, the major expansion of the university system, the introduction of new Research Councils with new missions in 1994, the explicit focus on wealth creation and quality of life, the introduction of schemes such as ROPAs (see section 3.3 below) that use non-traditional assessment systems, and the renewed emphasis on efficiency in the conduct of Research Council business. The Council of the Royal Society therefore appointed a group to consider the changes that have actually taken place and to address concerns about how peer review is operating in the rapidly changing environment of public funding for the Science Base.

The Group's remit focused on peer review in the context of judging proposals for funding of projects, programmes or Research Council Units or for fellowship support. It did not extend to the Society's own assessment procedures (which are under separate consideration), nor to decision-making processes in industrial R&D. Peer review (refereeing) of papers submitted for publication was excluded from the remit.

The Group was chaired by Professor Peter Lachmann (MRC Molecular Immunopathology Unit, Cambridge); other members were: Professor David Blow (The Blackett Laboratory, Imperial College), Professor John Coates (Department of Pure Mathematics and Mathematical Statistics, Cambridge), Ms Diana Garnham (Association of Medical Research Charities), Dr Paul Harvey (Department of Zoology, Oxford), Professor Brian Heap (University of Cambridge), Dr Bridget Ogilvie (Wellcome Trust), Dr John Skehel (National Institute for Medical Research) and Professor Brian Thrush (Department of Chemistry, Cambridge). Dr Peter Collins was secretary.

This report is issued as a statement of the Council of the Royal Society. It is intended both to inform and stimulate debate among the UK scientific community and to advise those responsible for developing the practice of peer review in this country.

The Nature of Peer Review

Peer review is to the running of the scientific enterprise what democracy is to the running of the country. It may not be the most efficient system but it is the least susceptible to corruption. The concept of peer review, in spite of all its difficulties, retains the confidence of most working scientists. It also retains the confidence of the great majority of the agencies funding research in the United Kingdom. The Association of Medical Research Charities, whose members in total fund over £360M of research per year, regards the adequate practice of peer review as a condition for membership of the Association and has laid down guidelines about how peer review is to be conducted in order to be regarded as satisfactory for this purpose. The Research Councils have each expressed their continued commitment to peer review; the Director-General of Research Councils in June 1994 confirmed that the means by which we will continue to assure the quality of the research we support is peer review; and in July 1994 the Chancellor of the Duchy of Lancaster said I am delighted that all Councils have come together to confirm their commitment to peer review. The implementation of the policies set out in Realising our potential will not diminish the central importance of peer review in assuring the highest quality of the research supported by Research Councils.

Peer review must be a central element in the scientific community's decision-making process on scientific support. Other elements include assessment of eligibility (i.e. the compatibility of the proposal with the objectives of the scheme in hand) and of affordability. These other elements are largely objective. Peer review involves, *inter alia*, the making of subjective judgements about quality by the proposer's peer group. This group consists of those who are qualified to judge the matter under consideration, both through having the requisite knowledge and training and through having current or recent active research experience (and thus exposed in turn to the same form of scrutiny).

Peer review should be conducted by members of the peer group who do not personally benefit from the outcome of the matter to be decided, and whose judgement can, to that extent, be accepted as impartial and disinterested. This may be difficult to achieve, especially for subjects in which the number of accessible active researchers is small; it may often be helpful to include reviewers from other countries.

The culture introduced by the 1993 White Paper and the missions of the new Research Councils stress both innate scientific excellence and relevance to mission as characteristics demanded of proposals for funding. To use the metaphors of an earlier age, it was Francis Bacon who pointed out that science aims both at increasing light (scientific insight) and at producing fruit (results of practical value) commensurate with the cost of production. Bacon also pointed out that there was no conflict between the search for light and the search for fruit: both were needed for the progress of science (and of society). The 1993 White Paper, too, recognised that first class fruit was not to be had without first class light. The judgement of whether a proposed project is likely to produce fruit is sometimes called merit review, with peer review being used to describe the judgement of potential light.

The Advisory Board for the Research Councils in 1990 published an assessment of peer review (the Boden Report), which concluded that there were no practical alternatives to peer review for the assessment of basic research. A 1991 analysis by the Science and Engineering Policy Studies Unit (SEPSU) of academics views about

techniques for the quantitative assessment of departmental research reinforced the message that peer review was central to any judgement about quality. These and many other reports have described the strengths and weaknesses of peer review in detail, so we shall not cover this familiar ground again here. Rather, our interest lies in what has changed, at a practical level, since publication of the 1993 White Paper Realising our potential.

3 Current Issues in Peer Review

Reducing the cost to the scientific community

(a) The need for efficiency

The traditional process for allocating grants for research involves inviting the scientific community to submit detailed, costed proposals, mostly for projects lasting up to about three years; then asking peers within that same community to review the proposals and forward them to an expert panel that ranks them in order; and, within constraints of eligibility and affordability determined by the funding agency, having representatives of that community decide which of the most highly rated proposals to fund. The establishment and continued support of Units involves similar processes. All this takes a great deal of time, from the assessors and, particularly, from the proposers. This time is generally free of charge to the funding agency, and comes from time that could otherwise be spent on research. It is, in effect, an investment by quality researchers in securing the conditions under which their remaining time can, indeed, be spent doing research. The policy issue is how to optimise the cost/benefit of that investment.

The cost of the paid staff of funding agencies administering the process is also relevant, since administrative costs generally come from the same pot as the money allocated for the support of research. The less costly the administration, the more money there is available to support actual research.

The overall efficiency of resource distribution depends in part on the success rate, i.e. whether the effort put into identifying good proposals is recompensed by those proposals then being funded. As the demand from well rated proposals has increasingly outstripped the supply of funding, the peer review mechanism has become correspondingly more costly, in the sense that an increasing amount of time is consumed for every proposal ultimately funded, and more frustrating and demoralising for all involved.

During the 1980s, the scientific community initially sought to resolve this by focusing on the need to increase the supply of funding to meet quality demand. However, it became apparent that an adequate increase was not going to be achieved: the size of the UK scientific community and the growing cost of modern scientific research meant that the available resources would not meet all reasonable expectations. Attention switched to trying to moderate the demand side.

If peer review is to remain central to the resource allocation process, as we believe it must, then it has to operate at an acceptable efficiency. A funding programme with an excessive rejection rate represents an irresponsible drain on the time of the scientific community. All funding agencies are, therefore, actively seeking ways of streamlining their assessment procedures and of moderating demand.

(b) Moderating demand

Some funding agencies in the UK have looked at ways of moderating the volume of demand so that it is not excessively incommensurate with the supply of funding. One important consideration is to have a clear statement of the remit of each funding scheme, so as to reduce the number of speculative applications. To some extent, demand can be moderated most simply by advertising: the less you advertise, the lower the demand. This approach is open to obvious objections if pushed too far, especially in the context of allocating public money; but it is, arguably, inappropriate to announce the availability of research funds in a way that leads to unrealistic expectations. The strategy used to advertise a given scheme should at least take cognisance of the optimum level of demand for efficient administration and the level of quality expected for an application to be in the running at all.

The Engineering and Physical Sciences Research Council (EPSRC), like the National Science Foundation (NSF) in Washington, has abolished fixed closing dates for grant rounds. EPSRC reports that this has led to a decrease in the number of applications and a noticeable improvement in their quality.

There are more direct approaches. A funding agency could restrict the number of applications that it will accept from a given institution or a given department. This is the practice of the Public Good Science Fund in New Zealand, which uses an institutional limit of 150% of the total number awarded the previous year. With some schemes such as its Prize Studentships, the Wellcome Trust both preselects the eligible institutions and predetermines the number of nominations each of these institutions may make. The Wellcome Vacation Scholarship scheme restricts applications to one per department. Both schemes consequently have high success rates in terms of the proportion of applicants that receive funding. The Natural Environment Research Council (NERC) limits departmental applications for studentships to three times their average allocation over the previous three years. A NERC Working Group has recommended a maximum of one application per principal investigator per grant round for responsive-mode funding.

An analogous approach, for example with studentships or certain kinds of research expense, is to associate eligibility with other kinds of award, so that applicants can apply for the two at the same time. This is practised by a number of funding agencies, e.g. EPSRC (linking studentships to research grants) and Wellcome (linking studentships to programme grants and principal and senior fellowships).

The merit of particular approaches to moderating demand has to be assessed on a case-by-case basis, weighing simplicity and efficiency against potential disadvantages. Attention must be paid to the social consequences within a department: to impose a limit at departmental level may undesirably shift power and initiative from the individual researcher to the departmental head. Care is also needed not to ossify an existing pattern of resource distribution, even though at institutional level year-by-year variations under unconstrained conditions are often not all that great. A major concern in such schemes is that they will tend to prevent talented but inexperienced applicants from embarking on research careers. The cost of monitoring whether departments or institutions stay within any preset limits may not be negligible.

A more brutal approach to moderating demand might be to exclude, for example, all departments other than those graded 4 or 5 in the Research Assessment Exercise.

This would be unacceptable: it would marginalise strong individuals in weaker departments, make it virtually impossible for weaker departments to improve their performance, undermine the purpose of Higher Education Funding Council (HEFC) support of research in grade 2 and 3 departments, and effectively eliminate the checks and balances that come from multiple funding sources.

The main cost of the inefficiency arising from major mismatch between demand and supply falls on the proposer: it takes longer to write a detailed proposal than to review it. To that extent, it is a more serious problem for the universities than for the funding agencies. Insofar as many referees are themselves academic researchers, the cost of their time, too, falls on universities. The search for a solution might therefore focus less on the funding agencies devising externally imposed, top-down solutions, and more on the researchers themselves. There could be a role here for research committees within universities playing a more active advisory part in improving the quality of proposals and providing informed feedback at an early stage, especially for younger staff.

(c) Streamlining assessment procedures

In any grant round, some applications are funded, some are deemed worthwhile but are not funded and some are judged to be intrinsically not worth funding. If an acceptable way can be found to identify this third group without incurring the cost of the complete peer review process, that would constitute a useful efficiency gain.

All funding agencies expect staff to eliminate proposals that fall outwith the terms of reference of the particular scheme. What is potentially controversial is the use of pre-screening to eliminate weak rather than ineligible proposals.

It is essential that pre-screening is conducted by peer review of the specific proposal, albeit more succinctly than the process used in subsequent stages. It would not be acceptable to base pre-screening solely on the track record of the individual or the institution. The Particle Physics and Astronomy Research Council (PPARC) community has long followed the practice of informal pre-screening designed to increase the chances of success when a proposal is finally submitted. A NERC working group has recommended rejection of pre-screening for quality of responsive-mode proposals, though the overall success rate has now fallen to 20%; this recommendation is under consideration.

The Wellcome Trust uses an abbreviated form for pre-screening by peers of potential applicants for a number of its Fellowship schemes; only the stronger applicants are invited subsequently to submit full proposals. The Medical Research Council (MRC) has a similar screen for programme grants. The Biotechnology and Biological Sciences Research Council (BBSRC), however, prefers not to use outline proposals for pre-screening for quality (as opposed to eligibility), because they may contain insufficient information for peer review and because they might encourage larger numbers of speculative proposals, thus defeating the object of diminishing the burden on the peer review system.

(d) Other routes to reducing pressures on peer review

One way of reducing the pressure on the peer review system is to reduce the frequency of demand, by allocating some money through larger blocks of funds (e.g. to support a team or a set of projects rather than an individual project) and/or for longer blocks of time (e.g. five years instead of the typical three, or a series of rolling grants). Larger applications might need to be assessed in different ways, e.g. by a

visiting group, and care would be required to ensure that this did not offset the intended administrative saving. Some form of monitoring would be necessary - progress reports or the like - and these would have to be peer reviewed ahead of any decision about funding for a second term; but this need not be a substantially greater burden than the current monitoring of an equivalent number of project grants.

Simpler ways of marking would ease the work of peer reviewers. The increasingly fine division of the alpha category cannot be the way forward. There must be better ways to deal with the borderline cases; BBSRC, for example, is introducing a new grading structure. A key task for a peer review panel (as opposed to individual peer reviewers), however, is to rank proposals, not simply to grade them.

Consideration could be given to the number of referees consulted about each proposal. A sliding scale, with only one or two referees for the smaller proposals, might be considered. The danger that a single reviewer could be unduly influential may be accommodated by selecting reviewers from a large, defined pool and by limiting individuals to a few years' service at a time (as with EPSRC's colleges of referees).

The emphasis placed in the Research Assessment Exercise on departmental income from Research Councils puts great pressure on individual academic staff to secure funding, which in turn feeds through to large numbers of proposals needing to be peer reviewed. This needs to be addressed in future rounds of the Exercise.

Role of non-peers

Some of the current concern about peer review focuses on the role that administrative staff (as opposed to active researchers) may or may not play in the new procedures being developed by EPSRC. EPSRC has taken pains to be open and consultative about its approach, though conflicting interpretations of its intentions persist.

In terms of allocation of funds at project level in the responsive mode, the key unit of organization within EPSRC is the programme area. There are 16 programme areas, each headed by a programme manager who is a full-time employee of EPSRC and has a scientific background (but is not considered peer of the researchers seeking support). Five of the programme areas (clean technology, facilities, high performance computing, marine technology and the innovative manufacturing initiative) do not operate in the responsive mode, though they still make use of peer review. The remaining eleven programme areas have set up a total of 16 colleges of referees (following extensive consultation with their respective research communities). The colleges range in size from 35 to 180 members, about one third from non-academic backgrounds; all members are based in the UK. EPSRC has published full details of the colleges.

An investigator submitting a proposal for funding identifies three referees of his/her own choice. The programme manager selects one of these plus two additional referees from the appropriate college, and may also consult further experts if the college does not adequately cover the specialism under consideration. The proposal is then sent to all three (or more) referees for assessment. Proposals judged sufficiently favourably by the referees then go forward to a panel, selected by the programme manager from the college when a critical mass of proposals in a given field has accrued. The panel, and its chairman, are appointed on an ad hoc basis and

are charged with rank-ordering the proposals according to their quality (guided by the three initial referees) and their pertinence to EPSRC strategic objectives.

The panel is purely advisory. The programme manager is responsible to the Chief Executive for meeting the objectives of his/her programme. On receiving the panel's advice, the programme manager constructs a list of successful proposals on the basis of: (i) the money made available by the EPSRC Council, (ii) the rank ordering produced by the panel, and (iii) the guidance of the panel chairman (particularly in borderline cases). While this appears to put a good deal of responsibility onto programme managers, it has been emphasised repeatedly that they will not vary the quality rankings made by the peer reviewers.

We have described the EPSRC procedure in detail because it is the one that has attracted most comment. The controversial point is the role of the programme managers who, despite having no direct involvement in the making of value judgements about the scientific quality of proposals, are seen as potentially exercising undue influence over the process as a whole. In BBSRC, the existence of subject-based committees removes the need for programme managers to appoint ad hoc panels; however, it is proposed that programme managers should be responsible, in consultation with committee members, for the selection of referees from a database. In the Wellcome Trust, senior staff may select referees and may serve as deputy chairmen to grants committees.

Non-peers, then, currently play a variety of roles in the conduct of peer review. The EPSRC approach has generated some concern within the community it serves. Whether this concern is justified will be seen in due course, when the new system has had time to settle down. The new system is most likely to command support if: (i) the programme manager seeks expert advice when appointing a review panel; (ii) the membership of the panel is in the public domain; (iii) there is some continuity between panels so that consistent criteria are applied to the rank-ordering of proposals (though a fairly rapid rotation of membership guards against persistent prejudice impeding the chances of any one individual, group or scientific view); and (iv) investigators have confidence in the programme manager's choice of additional referees. A further confidence-building measure would be an explicit commitment to monitoring the performance of programme managers: the experience of the NSF in using external visiting committees for this purpose should be examined.

ROPAs

The ROPA (Realising Our Potential Award) initiative has attracted a good deal of attention, in part because of its declared intent not to use conventional peer review. Eligibility for funding under the initiative depends on industrial support of a particular kind. Those eligible may put forward proposals for blue skies research. Proposals are assessed by small panels of scientists and others, generally guided by referees' comments. However, this assessment is on the basis of originality and technical feasibility, not quality in the sense of scientific excellence. Moreover, the expert panels are asked only to accept or reject proposals, not to place them in rank order. So the assessment process for ROPAs differs significantly from that for normal research grants.

The ROPA initiative has some innovative features and some apparent drawbacks. In the context of peer review, we note that ability to attract £25K of industrial support (the sum judged by the Research Councils as the minimum needed for a piece of strategic research) is neither a necessary nor a sufficient indicator of ability to conduct

first class basic research. On the other hand, the ROPA approach may avoid the pitfalls of conservatism to which conventional peer review can be prone (see section 4.2 below). There is pressure to complete ROPA grant rounds extremely quickly, which inhibits panels from seeking advice on proposals outside their expertise. This same pressure means that, in general, no feedback is provided to unsuccessful applicants. The dual criteria of originality and feasibility may in some cases be mutually incompatible; indeed, the object of the proposal may be to explore feasibility. The risk-taking inherent in blue skies research implies that funded proposals must be allowed to fail. Given the uncertainties at this stage in the development of the ROPA scheme, Research Councils should tell applicants as much as possible about how the decision-making process works in practice.

Both the processes and the outputs of the ROPA scheme should be reviewed by an independent body before the scheme is expanded further.

Merit review

As mentioned in section 2 above, the term merit review is sometimes used to describe the process of judging whether a proposed project is likely to lead to practical and affordable outcomes, with peer review being used to describe judgement of scientific quality. The question may then arise whether these two processes are really distinct and, if they are, how they can be combined into a single decision about whether a proposal needing to meet both sets of criteria should be funded. It is, however, unhelpful to draw too strong a distinction between peer review and merit review. The point of principle is simply that, in both basic and applied work, funding decisions must be based both on peer review (to establish quality) and, where appropriate (e.g. within directed programmes and other initiatives aimed specifically at a practical objective), on merit review.

Further Issues

Leakage

One shortcoming commonly ascribed to peer review is leakage of ideas: peer reviewers, deliberately or subconsciously, may take on board ideas set out in the proposals they are reviewing and incorporate them into their own work. This possibility is inherent in the system.

How much leakage of this sort really occurs, and whether it causes problems, is unknown. There is no reason to suppose that it occurs on anything like the scale that would bring peer review into disrepute, or that it has increased significantly in recent years. But complete respect for confidentiality and avoidance of all (deliberate or inadvertent) plagiarism are central to the successful operation of peer review.

Conservatism

Probably a more important shortcoming of peer review is that the hurdle of gaining peer reviewers' approval can inhibit researchers from putting forward their most radical ideas for funding through conventional channels. It may be difficult to gain support for a proposal that challenges received wisdom or current fashion. A virtue of a scheme like the ROPA initiative, in principle, is that by emphasising originality and bypassing peer review it may reduce this hurdle.

A variant on the conservatism theme is that peer review may discourage established researchers from moving into new fields where they have no previous track record. This is a major drawback.

Neutrality

It is not self-evident that peer reviewers can be wholly neutral, especially in highly specialised fields. There is a strong case for using peer reviewers from other countries who are not competing for funds from the same source. The Wellcome Trust and many other charities already make frequent use of overseas reviewers, particularly where the charity's interest is in a single disease. There is also a case for using some reviewers knowledgeable in the general area of the proposal rather than, or as well as, expert in the particular specialism.

Casting the net more widely in order to increase the perceived neutrality of a peer review panel would also make it easier to select individuals who are competent at the specific work of peer review. Not all researchers are naturally able at this particular task.

However, perfect neutrality, like perfect consistency between review panels or between successive meetings of a single review panel, is neither demonstrable nor strictly necessary to the reasonable operation of the peer review system.

Since it is often not possible to assemble a peer review panel that is both knowledgeable and neutral, it is essential for all panel members to declare any vested interests in some explicit manner at the outset.

Transparency

One of the pre-requisites for confidence in the integrity of peer review (and, indeed, of other decision-making processes) is that its workings should be transparent. There are several levels to this. At the minimum, the membership of peer review and other decision-making panels should always be in the public domain. However, we recognise that the practice of normally preserving referees' anonymity at the level of individual proposals is most likely to maximise the value of their reports. Funding agencies should normally be willing to offer some degree of feedback to applicants: expert criticism is an important and valued component of the scientific process. Transparency is increased if applicants are able to see and comment on referees' reports before decisions are made on their proposals.

Fraud

Deliberate fraud is almost certainly rare, but because so much scientific research now takes place a number of well publicised examples can be found in the recent medical and scientific literature. It is not evident that peer review, or any other method of judging research proposals, can prevent fraud. While peer review panels should always have in mind the possibility that not all the information presented to them is true, their chances of detecting the rare cases of deliberate fraud (as distinct from plagiarism) are low. Moreover, an attitude of excessive suspicion is likely to do more harm than good, by exacerbating the conservatism mentioned earlier.

Conclusion

The confidence of the scientific community, and of the taxpayer, that the public investment in the national science base is well managed, can be sustained only if an effective form of peer review holds. This is well recognised by the Office of Science

and Technology. The system is, nevertheless, under pressure, and it is in everyone's interest to improve its efficiency and adapt it to meet scientific needs within the context of current policy objectives. This report suggests some general principles to further that end.

The Royal Society will continue to monitor developments in peer review, in order to ensure that the efficient and effective application of peer review can continue to serve the needs of the scientific community and its customers.