



Research forum

EFC Research Forum

Evaluation Workshop Series

Saïd Business School, University of Oxford, 9 November 2010

Keeping Researchers in Mind

Discussion Note

Whether aiming to fund the best research or to evaluate its outcomes, there is no substitute for the judgement of experts. However in recent years the 'gold standard' of peer review has come under increasing scrutiny as a result of high-profile cases of bias and fraud, accusations that it stifles innovation and risk, and complaints of 'peer review fatigue' due to the proliferation of funding programmes and journals.

How can Foundations avoid the pitfalls of peer review while aiming to fund high-risk and innovative research? How can the burden on researchers – both as applicants and peer-reviewers – be minimised while ensuring the credibility the system? What is the role of evidence in peer review and how is it best handled? How can Foundations balance the interests of researchers with the need for accountability to their other stakeholders?

These and other questions were addressed at a 1-day workshop hosted by the University of Oxford. Approximately 30 representatives of Foundations, public funding agencies, independent consultancies and think-tanks and the university sector gathered to shed new light on a persistent challenge to the research community. The programme and Power Point presentations are available at <http://www.efc.be/Networking/InterestGroupsAndFora/Research%20Forum/Pages/Evaluation.aspx>. The following note is an attempt to summarise the discussion and cluster it around a number of key issues and concepts.

Diversity in peer review. It is clear from even a cursory survey of peer review systems that despite the simplicity of the core concept there is a good deal of variation in practice. For example, organisations and countries differ on the extent to which it should be left solely to the academic community, and the scope for involvement of users of research and other experts; language requirements vary (organisations that depend on international peer review often require their applicants to apply in English); and organisations and cultures disagree as to whether reviewers should be paid for their services. There may also be goals other than the quality of research, such as disciplinary balance or gender balance, and this will alter the nature of the process and the composition of panels. This diversity of practice is healthy and reflects a variety of different values and goals. As one participant put it, "Your Foundation is the peer review that you do"; or, to put it another way, the peer review process is an answer to the question "what path are you on as a Foundation?" Consequently, there should be a degree of scepticism toward attempts to standardise peer review systems.

Importance of continuous development. The workshop heard about a number of innovations in the peer review system in recent years. As well as those documented in the presentations, mention was made of the MacArthur Fellowships, the so-called 'Genius Awards', based on a system of nominations and no formal application procedure. In general there have been few radical departures from the basic model. "Evolution, not revolution", as one participant expressed it, and comparison was made with requirements necessary to grow a complex eco-system. It was agreed nonetheless that it is important to attempt modifications and improvements. This is less a matter of which elements to review (academic track record, publications list and research proposal being the basic elements), as the appropriate weight to give each element. In particular, there is a good deal of debate as to the relative importance of track record compared to the proposal (see below). There has been a lot of discussion on this point within the Academy of Finland and the Deutsche Forschungsgemeinschaft (DFG), to name only two public funding agencies. Public research organisations can also learn from industry, where there is less focus on publications and track record and more on how the individual contributes to the overall research project.

Professor Caswill noted the 'disconnect' between the evidenced critiques of peer review in the academic literature and the willingness of the science system in general to absorb them. This may indicate an innate conservatism within the system as a whole. However, it would be important to distinguish between peer review as an innately consensual procedure and institutional path dependency (inflexibility of the formal rules of deliberation). It is also important to see the system in its institutional and indeed global context as part of the method by which science delivers important social and economic benefits. Peer review is a messy business built on a negotiated consensus – it is important to periodically review its operation and not to get stuck in 'business as usual'.

Peer Review and Risk. Opinions differed on the effectiveness of these peer review innovations designed to encourage path-breaking or risky research (see presentations by Allen, Hiidenmaa, Morrell and Nedeva). The evidence from the Academy of Finland's 2008 Research Grants scheme – where riskiness and potential for innovation were included as specific criteria – were not conclusive. Peer reviewers may not be best situated to recognise the future applicability of research. In the social sciences and humanities in particular, it was felt that the criteria were somewhat artificial and therefore not possible to apply consistently. Both Hiidenmaa and Hinze pointed to a large degree of conceptual uncertainty within the academic community and amongst funders as to how to define 'risky' research. Hinze presented the results of a recent survey by the DFG of over 4,000 professors in German universities which revealed that a significant number of respondents in all scientific disciplines (up to 1/3) believed that projects designated 'risky' were no different from other research projects. It was suggested that if the question was framed in terms of 'how much of your own research portfolio do you regard as risky' there might be a different response, as researchers are much more willing to apply the distinction in tensioning their own work.

According to a recent survey, a total of 40 European funding agencies claim to have specific programmes supporting novel or 'risky' research projects (Polak, E. et al. (2008): NEST-PROMOSE, Final Report). Taking the long view, however, this concern with 'path-breaking' research is a relatively recent obsession. All research builds on a firm base of scholarship and more mainstream activity. The emphasis on more risky research, or importing unfamiliar conceptions of risk from other domains, may unintentionally distort the research agenda in some fields. For example, in linguistics the application of an established theory to different spoken languages, rather than theoretical development as such, may be considered less 'path-breaking' although it involves substantial risks and may reveal new insights. Finally, it was interesting to note that, despite the increase in such programmes, the lion's share of respondents in a DFG survey believed that there was no need for special funding arrangements for risky research.

As Dr Nedeva showed, dissent between peer reviewers in their deliberations is one indicator of research breaking new ground and that peer review systems that wish to fund path-breaking research, for example the European Research Council, should pay special attention to its presence rather than aim for consensus at all costs. Dr Morrell pointed out that the UK Engineering and Physical Sciences Research Council (EPSRC) currently recognises the role of dissenting opinions in some of its peer review procedures and these are often highlighted in making funding decisions.

It should also be borne in mind that path-breaking research, such as that undertaken by the Howard Hughes Medical Institute (HHMI), builds on the very high average quality of applicants and established success. The average age of HHMI grantees is 42. Looking toward the next generation of innovative research, the HHMI is funding starting grants for early career researchers where the counterfactual criterion or deletion test - what will be the implications for the research field if this person is not funded?

– is substantially relaxed. The intention here is to grow promising talent rather than to reward existing quality.

There is an inherent danger in using awards and prizes as an indicator of path-breaking research, as studies on trends in nominations have shown they often involve past prize-winners picking future winners, creating the risk of ‘closed shops’ and perpetuating a more conservative system overall.

Perceptions of research community on evaluation process: although the evidence for innovations to the peer review system producing better research outcomes is patchy, specific innovations are appreciated by the researchers under evaluation. The EPSRC has experimented with anonymised applications (see presentation by Morrell). The feedback from this process has been positive, with applicants feeling liberated from the need to defend their academic track record and instead focussing on the quality of the ideas. The results of the DFG survey that Hinze presented confirmed that most academics prefer the idea to outweigh the experience of the applicant in reviewers’ deliberations.

The tendency of ‘metrics-creep’ in recent years was observed: metrics that were designed for use at the system-level have been imported into all aspects of research and career evaluation. Studies have shown that this trend is clearly at odds with the dominant professional self-image of the academic community. Professor Walmsley provided a succinct overview of that self-image and the perspective of the academic community on the peer review system. It is a system based on trust, not control. It is a community activity that requires considerable buy-in from its members in order to work properly. Like all such activities, there is a risk of conflict of interest that needs to be managed. It is also a competitive activity and the rules should be designed to ensure a fair, efficient and fast procedure. Individual errors may occur, but it is understood that the system as a whole is on the right track. A distinction between accuracy and precision is helpful here. Although there may be elements of randomness and chance, the system delivers accurate results, while acknowledging that speculative and untested ideas may not always be rewarded. Nevertheless, the academic community is acutely conscious that past performance is not always a good predictor of future success and tries to accommodate this in its judgments and procedures.

The DFG survey confirmed much of the satisfaction with the current peer review process overall, despite dissatisfaction with individual outcomes. So while a large majority of respondents agreed that the composition of the peer review panel strongly affected the outcome of their funding application, a similar majority preferred panel reviewing over individual reviews. The survey also showed that researchers were aware of a random element in the process, with up to 10% of respondents in some disciplines believing that random selection would be an acceptable formal method of proposal selection.

Peer review and bias: Peer review systems have long been accused of containing a ‘hidden bias’. Sometimes this is intentional, where it is alleged that reviewers use the cloak of anonymity to criticise competitors. In other cases the biases are more structural, such as cases where the distribution of gender, race and ethnicity in academia is reflected in peer review outcomes. There is some evidence for the latter bias: for example, the results of a 1997 Swedish study found that there was a gender bias in scoring postdoctoral fellowship applications. This was confirmed in a follow-up study. However the same study also detected a larger bias between ‘insiders’ and ‘outsiders’, i.e. those inside academic networks and others. This would seem to confirm another criticism of peer review, that it is dominated by ‘old boy’s networks’, and is a powerful objection to awards and grants based on a nomination system. However, a similar study of the UK Research Assessment Exercise found no such widespread bias, except to a certain extent in mid-career. The suggestion was made that the problem may be in academic career structures rather than in the nature of the peer review process as such.

Costs of peer review. In response to queries from participants about the cost of peer review, a figure of €3000 per application was mentioned, mainly relating to travel costs of peer reviewers. The UK Research Councils have also produced a report on the efficiency and value for money of their peer review procedures (<http://www.rcuk.ac.uk/reviews/home/Pages/vfmpeerreview.aspx>). Costs and overheads however vary between countries and organisations. Costs should not be seen in isolation, but in relation to the risks that an organisation is willing to take, and its objectives. Direct payments to reviewers increases costs but there is disagreement about the ethics and effectiveness of such arrangements, with many in the academic community believing that the system should be sustained by reciprocal altruism. The view was expressed that the costs of peer review are kept artificially low as the indirect costs are borne by the employing institutions and individuals. In this context it was notable that academic economists have been increasingly reluctant to undertake these reviews. Keeping costs low

may also have implications for the quality of the procedure, i.e. the phenomenon of hard-pressed reviewers carrying out their work on trains or planes on the way to the consensus meeting.

What further research is needed? As noted above, there is a good deal of recognition that the peer review process contains a random element. In principle, it should be possible to test just how random the process is through the use of control trials, although there would be a number of difficulties the implementation of such studies. There has been some pioneering research on the resubmission of articles to refereed journals (Peters, D.P. and Ceci, S.J. (1982) "Peer-Review Practices of Psychological Journals : The Fate of Articles Submitted Again", *The Behavioral and Brain Sciences*, vol. 5 pp. 187-255), but more study is necessary.

Although there have been many studies of peer review outcomes, there have been relatively few studies of the deliberative process itself, characterised by one participant as 'looking inside the black box of peer review'. The DFG is about to embark on such a study.

Finally, the importance of the role and formation of social capital and social networks are clearly important in the peer review system and to any attempts to reform it. More systematic study of these phenomena are needed.

Lessons learned: finally, we try to summarise here some of the key messages coming out of the workshop.

1. Trust, not control: peer review is a community activity built on trust, not control. It is a delicate ecosystem that needs careful tending.
2. Impact on researcher behaviour: following on from the preceding point, funding agencies and institutions should be aware that the specific evaluative tools that they use will have a strong impact on researcher behaviour, often producing perverse incentives that detract from other desirable goals (e.g. teaching or social engagement).
3. Keep the dialogue going: to avoid such unintended consequences, it is important to keep up a dialogue with the research community before and during the implementation of your evaluation scheme. Presenting timely feedback to applicants on their proposals, for example before final funding decisions are made, is one good example of institutionalising such a dialogue. There are a number of funding agencies operating this 'right to reply' procedure as part of their peer review process.
4. Foundations need to find their niche: in keeping with the dictum "Your Foundation is the peer review that you do", foundations need to decide exactly what their role is within the larger research funding landscape and to tailor their peer review systems accordingly.
5. Proportionality: keep whatever evaluation scheme you do decide on proportional to your aims and the uses of the evaluation results. Avoid unnecessarily overburdening researchers with reporting requirements.
6. Learn from other organisations: as the workshop had demonstrated, there is a wealth of experience in other research foundations and public funding organisations, often codified in reports, guidelines and manuals. You should seek guidance from these sources before taking the expensive and potentially unnecessary step of engaging private consultants.