Peer Review

Peer Review is an evaluation technique that traditionally plays a central role in scientific research. It is the main form of decision-making in academic publishing, selection and promotion of scientists, and is part of the decision procedures for the allocation of resources and the strategic formulation of research programmes. It is used by all actors in the research field, including government and industry, and has recently gained importance in the evaluation of research institutions themselves and in ex-post evaluation. However, the traditional functions of peer review – ensuring scientific excellence and the autonomy of academia – are increasingly deemed insufficient, as the procedure faces demands to guarantee some degree of accountability to funding bodies and society at large. This supposes to judge research funding, evaluation and to some degree academic appointments not only on scientific excellence, but also on supplementary criteria like economic and social merit. In this context, new forms of peer review including strategic peer review, in-process peer review and merit review are developing. These share the feature of taking into account multiple criteria and/or relying at least partly on non-specialist peers. In the absence of unambiguous definitions, these new forms can be referred to more generally as “extended peer review” processes. This policy brief outlines the development of peer review arrangements over recent decades, and sets out the many challenges faced in implementing such practices.

By Stefan Aykut and Nicolas Baya Laffite

OECD, 2011
The issue in a nutshell

Peer Review is an evaluation technique that traditionally plays a central role in scientific research. It is the main form of decision-making in academic publishing, selection and promotion of scientists, and is part of the decision procedures for the allocation of resources and the strategic formulation of research programmes. It is used by all actors in the research field, including government and industry, and has recently gained importance in the evaluation of research institutions themselves and in ex-post evaluation.

Peer Review should be distinguished from other forms of evaluation used in the research context, including quantitative indicators like benchmarking or scientometrics. It relies on more or less explicitly defined criteria and an often documented formal evaluation process. Its main legitimacy, however, stems from the qualification of selected reviewers. It reflects the specialization of academic fields and the impossibility for outsiders to judge the quality of research in a particular discipline. Peer Review therefore strengthens the autonomy of academia, leaving the definition and enactment of standards of excellence to the practitioners themselves. A widely accepted definition of peer review is the following:

“Peer review is the name given to the judgement of scientific merit by other scientists working in, or close to the field in question. Peer review is premised upon the assumption that a judgement about certain aspects of science, for example its quality, is an expert decision capable of being made only by those who are sufficiently knowledgeable about the cognitive development of the field, its research agenda, and the practitioners within it.” (Gibbons and Georghiou, 1987)

In knowledge-based societies, science becomes increasingly important for innovation processes, as well as for socio-economic choices and political decisions. As a consequence of this, Peer review started to be applied to new domains (project development, in-process evaluation, ex-post evaluation, etc.). This, however, puts on it stark pressures to adapt and take into account wider socio-economic criteria. Indeed, the traditional functions of peer review – ensuring scientific excellence and the autonomy of academia – are deemed insufficient, as the procedure faces demands to guarantee some degree of accountability to funding bodies and society at large. This supposes to judge research funding, evaluation and to some degree academic appointments not only on scientific excellence, but also on supplementary criteria like economic and social merit thus challenging deeply entrenched convictions that funding of ‘basic’ research or ‘pure’ science should be without heterogeneous conditions.

In order to respond such demands, new definitions are being put forward. For example, the Office of Energy Efficiency and Renewable Energy (EERE) in the American Department of Energy (DOE) defines Peer Review as
“A rigorous, formal, and documented evaluation process using objective criteria and qualified and independent reviewers to make a judgement of the technical, scientific, and business merit, the actual or anticipated results, and the productivity and management effectiveness of programs and/or projects.” (EERE Peer Review Guide 2004).

In this context, new forms of peer review including strategic peer review, in-process peer review and merit review are developing. These share the feature of taking into account multiple criteria and/or relying at least partly on non-specialist peers. In the absence of unambiguous definitions, these new forms can be referred to more generally as “extended peer review” processes.

Issues pertaining to socio-economic relevance of scientific research are not the only challenges peer review faces. Concerns also relate to its very functioning, even in its core domains of application. These include: subjectivity of reviewers, conflicts of interest, cost efficiency (direct and indirect costs), availability of reviewers, and discrimination of multi- and transdisciplinary, innovating or heterodox research. These concerns will be addressed throughout the issue brief.

**Rationale and ideas**

Peer review is closely tied to a certain vision of science and scientific work, the so-called Mertonian Norms or CUDOS (Communitarism, Universalism, Disinterestedness, and Organised Scepticism). Following this idea, scientific work is neither individual nor subjective; it is a collective product and has universal validity. In order to achieve this, the functioning of the scientific community is not, like the economy, organized around individual profit. Instead, it rewards the reproduction of experiments and the falsification of well established hypotheses. Although peer review is used in research since the 18th century, the Mertonian claim that the scientific community is capable to manage research autonomously, as well as that it should do so wherever possible, reinforced the status of peer review as the central tool to guarantee the autonomy of the scientific community and the excellence of scientific work.

Furthermore, ideas of “pure science” and “basic research” gained new impetus from the beginning of the 20th century. Vannevar Bush’s 1945 report “Science, the endless frontier” provided justification for unconditional public funding for academic research in order to protect it from any external (governmental or industrial) influence. There is a certain paradox in this claim, as at the same time, unprecedented amounts of government money flooded into Cold War Big Science projects, and Vannevar Bush and his colleagues at the Massachusetts Institute of Technology (MIT) heralded a new model of scientific research organisation in building close academic-industrial links. Nevertheless, the model of science-policy relations defended by the report became common currency for the following decades. Today, pressures to make scientific research more accountable to science policy and to larger society increase for very different reasons. These include:

The growth of scientific funding with the massification of Higher Education and Research in modern knowledge societies puts increasing burdens on State budgets. Closer involvement in and contribution of academic research to innovation, and the intensification of science-industry relationships (“Third Stream” activities) are regarded
as crucial for the competitive position of states and regions in the context of economic globalization.

New social and political challenges with high uncertainties and high urgency like environmental and technological risks or public health issues depend on scientific research and expertise while classical academic research mechanisms prove too heavy and slow to provide the required information in a timely manner.

In the face of “unintended consequences” and “new risks” engendered by research activities (nuclear energy, genetically modified food and organisms, nanotechnology, etc.), public demands also include the democratisation of research decisions in science and technology.

It has been claimed that peer-review, because it insists mainly on internal criteria of relevance and quality, does a poor job in considering the wider context of social relevance and policy implications important in problem-oriented research. This leads to a situation in which scientific excellence is “a necessary but not sufficient selection criterion for establishing research priorities” (Gibbons et al., p. 65). Furthermore, scientific research itself is changing. Some academic observers claim that while traditional research was conducted within the scientific community organized in clearly distinguished disciplines, with little implication of non-researchers (“Mode 1 research”), research is today interdisciplinary, interactive, and closely tied to its economic and political environment, it’s ‘context of application’ (“Mode 2 research”). Some scholarly observers doubt the historical accuracy of this thesis. Notwithstanding these critics, the acknowledgement that academic research should be more closely tied to its context of application has important implications on quality control in science, as well as on strategic decision making about funding and steering of research.

Two challenges to classical peer review arise from these developments: How can interdisciplinary and multidisciplinary research be evaluated properly (1), and how can problem-solving and application-oriented research be taken into account (2)?

(1) Evaluating inter- and multidisciplinary research: In the case of interdisciplinary research, the problem lies in the specialization of academic disciplines. Potential reviewers are usually formed in a specific research context and tend to evaluate submissions (journals, research projects, etc.) according to the criteria of excellence of their respective disciplines. Interdisciplinary research is systematically disadvantaged in such a procedure, as contribution to a specific field might not be the main objective, and reviewers are often not qualified to judge proposals that are rather peripheral to their core research interests. A first solution is to choose reviewers from all concerned disciplines. Although this broadens the set of criteria used to evaluate a proposal, it does not address the basic problem. In order to overcome the tendency to discriminate interdisciplinary research, funding agencies and journal editors increasingly include “interdisciplinarity” as an explicit criterion for evaluation. Other possible measures include “ring-fencing” funds for interdisciplinary proposals and “flagging” of unorthodox research. These forms of “positive discrimination” might be valuable first steps to encourage research collaborations between disciplines.
as crucial for the competitive position of states and regions in the context of economic globalization.

New social and political challenges with high uncertainties and high urgency like environmental and technological risks or public health issues depend on scientific research and expertise while classical academic research mechanisms prove too heavy and slow to provide the required information in a timely manner.

In the face of “unintended consequences” and “new risks” engendered by research activities (nuclear energy, genetically modified food and organisms, nanotechnology, etc.), public demands also include the democratisation of research decisions in science and technology.

It has been claimed that peer-review, because it insists mainly on internal criteria of relevance and quality, does a poor job in considering the wider context of social relevance and policy implications important in problem-oriented research. This leads to a situation in which scientific excellence is “a necessary but not sufficient selection criterion for establishing research priorities” (Gibbons et al., p. 65). Furthermore, scientific research itself is changing. Some academic observers claim that while traditional research was conducted within the scientific community organized in clearly distinguished disciplines, with little implication of non-researchers (“Mode 1 research”), research is today interdisciplinary, interactive, and closely tied to its economic and political environment, it’s ‘context of application’ (“Mode 2 research”). Some scholarly observers doubt the historical accuracy of this thesis. Notwithstanding these critics, the acknowledgement that academic research should be more closely tied to its context of application has important implications on quality control in science, as well as on strategic decision making about funding and steering of research.

Two challenges to classical peer review arise from these developments: How can interdisciplinary and multidisciplinary research be evaluated properly (1), and how can problem-solving and application-oriented research be taken into account (2)?

(1) Evaluating inter- and multidisciplinary research: In the case of interdisciplinary research, the problem lies in the specialization of academic disciplines. Potential reviewers are usually formed in a specific research context and tend to evaluate submissions (journals, research projects, etc.) according to the criteria of excellence of their respective disciplines. Interdisciplinary research is systematically disadvantaged in such a procedure, as contribution to a specific field might not be the main objective, and reviewers are often not qualified to judge proposals that are rather peripheral to their core research interests. A first solution is to choose reviewers from all concerned disciplines. Although this broadens the set of criteria used to evaluate a proposal, it does not address the basic problem. In order to overcome the tendency to discriminate interdisciplinary research, funding agencies and journal editors increasingly include “interdisciplinarity” as an explicit criterion for evaluation. Other possible measures include “ring-fencing” funds for interdisciplinary proposals and “flagging” of unorthodox research. These forms of “positive discrimination” might be valuable first steps to encourage research collaborations between disciplines.
Peer Review

review procedures to judge applications for support of research developed in the early part of the 20th century – but as new research management techniques increasingly rely on competitive procedures for grant attribution; peer review has gained unprecedented significance outside its initial field of application.

3) Government: Ministries do not usually rely on peer review. Exceptions are strategic peer review in the context of Programme Evaluation and regulatory peer review in policy-making. The US Department of Energy for example uses peer review to evaluate research and development activities at the project and programmes level, and the European Union relies on ex-post peer review to evaluate its 6th and 7th Framework Programmes (FP). In the context of regulatory policy, agencies concerned with environmental and sanitary policy frequently rely on peer review procedures (the Environmental Protection Agency (EPA) for example has been using peer review for a long time). The use of peer review in these contexts is however more circumscribed than in academic publishing or the organisation of research careers. While the first mentioned merely affect individual reputations, the peer review of strategic decisions concerning research organisation and priorities concerns have a much greater impact, and decisions over government regulation implicate the economic fortunes of numerous major corporations. Furthermore, much regulatory information is already controversial, due to the interests at stake and the kind of knowledge needed to decide. Thus, the function of peer review tends to be one of consultation and advice, and decision and evaluation are most often strictly separated.

One important distinction concerning the use of peer review consists in its weight in the decision-making process. It has been suggested (Bozeman 1993) to distinguish between pre-emptive peer review (final decision depends entirely on the results of peer review), traditional peer review (peer review is an important element in the decision) and ancillary peer review (decisions are taken by a manager or a board, peer review is a source of information among others). While traditional peer review is the most commonly practiced form of evaluation in publication and research appointments, pre-emptive peer review is usually combined with scoring or ranking models and used by some agencies like the NIH in grant attribution decisions. Major programme evaluations, decisions concerning organisation of scientific research and regulatory policy making most frequently rely on ancillary peer review.

Mechanisms

1) Academic peer review: In academic publishing, where peer review has long been – and still is – by far the most important mechanism for quality assurance, its influence is particularly significant: throughout disciplines, journals with formal peer-review systems are generally rated higher than those without, and peer review is often a pre-condition for journal indexing in scientific databases. Before the review process, the editorial board decides whether or not a paper is important enough to undergo the review process. The following prepublication review (refereeing) normally involves two or three independent reviewers having published in the field in question, who assess a submitted article on a given set of criteria (e.g. competence, significance and originality). Although the final decision for publication is taken by the editorial
board, the review is a central part of the procedure and editors tend to follow the reviewers if their judgements converge. In order to guarantee an orderly review process, reviewers are often anonymous. In some cases, this applies also to the submitting scientist. These procedures are referred to as “single-blind” and “double-blind”. Exceptions to such rules of anonymity are emerging open peer-review procedures, to which we will come back later.

(2) Extended peer review: the NIH ‘Dual Review System’ for grant applications provides a good illustration of the mechanisms involved in extended peer review. In a first level of the procedure, the scientific review group (SRG) focuses on the scientific and technological excellence of the subject under evaluation, and makes recommendations for appropriate level of support and award duration. It does not make funding decisions or set programme priorities. The second level of review by the council focuses on relevance and socioeconomic priorities of research. It makes recommendations on funding, evaluates programme priorities and relevance, and advises on policy. Another possibility to extend peer review consists in the explicit inclusion of socioeconomic criteria in a single-stage peer review process. Early on, this has been suggested in evaluation procedures concerning research funding and organisation. Weinberg (1963) distinguishes “scientific merit” (contribution to overall scientific advancements, including neighbouring fields), “technological merit” (possible applications and industrial uses), and “social merit” (relevance to welfare and other social goals). Other, more precise criteria have been also suggested. These include timeliness, urgency, relevance, exploitability (Scott, 840). As uses of peer review vary widely today, it might be useful to diversify criteria and adapt them to the specific evaluation context. The method of “multi-criteria mapping” responds to that objective.

In addition to the two controversies discussed above concerning evolutions in scientific practice (taking into account problem solving and interdisciplinary research), several pitfalls of peer review closely tied to its mode of functioning have been subject to debate recently. The first set of drawbacks highlights questions of deontology (3); the second is about conservatism of peer review (4).

(4) Deontology issues: As scientists chosen as reviewers are at the same time researchers in the respective field, it is difficult to prevent conflicts of interest. They can be tempted to hinder publication of findings contradicting their own work, or to delay publication of similar research than their own, as there is pressure in the scientific community to “publish first”. In an increasingly competitive academic environment, this is true not only for publication peer review, but also in the assessment of research proposals, programmes, etc.. This raises important questions about accountability and publicity of the review process.

(5) Heterodox research: Another important issue about peer review concern difficulties to acknowledge heterodox or “pathbreaking” research. As peers generally are senior researchers, they might have difficulty to accept new paradigms that threaten the existing structure. Furthermore, epistemic communities largely frame the behaviour of reviewers, who will tend to respect and thus reinforce disciplinary standards. This conservatism can obstruct new disciplines, new methods, and new topics that don’t rank high on the current research agenda. It is supported by two well-known biases of peer-review: the “Matthew effect” described by
sociologist Robert K. Merton 40 years ago postulates that famous and influential researchers tend to have better chances in peer-review processes. Allocation of funds might also be trusted by circles of influent researchers, so-called "Old-Boy networks".

In order to avoid these pitfalls, it has been suggested to:

Pay special attention to the independence of reviewers. Several measures can be taken in order to achieve this goal: the disclosure of financial interests by the reviewed in case of submissions to publication, the presentation of lists of reviewers that submitters consider qualified or not qualified to judge their proposal, etc. In the case of small countries or highly specialized disciplines, the independence of evaluators may not be easy to achieve. In this case, involving foreign reviewers can be helpful. However, this solution poses problems of availability and physical distance, and presents its own drawbacks. Most importantly, foreign reviewers may not be familiar with specific national research contexts, and thus only partly qualified to evaluate specific research contributions or funding requests.

Open up the peer review process through internet-based “open peer-review” procedures. Open Peer Review procedures have been implemented in several scientific journals like the Medical Journal of Australia and Atmospheric Chemistry and Physics. Those journals publish articles having undergone a rapid pre-screening on the internet, in order to allow members of the scientific community to participate in the assessment of scientific manuscripts through interactive comments, in addition to designated reviewers' reports. Recently there have been some experiments with wiki-style, signed, peer reviews, for example in an issue of the Shakespeare Quarterly.

In the context of grant allocation procedures, conflicts of interest are even more important than in publication. Therefore, US government guidelines governing peer review for federal regulatory agencies for example require the disclosure of reviewer's identities under some circumstances.

**Incentives**

*Peer Review pervades science from the beginning to the end* (Scott 2007, p. 828). Participation in peer review therefore is part of researchers' daily lives. Even though it is not usually paid, it provides several advantages for researchers. Firstly, reviewers are provided with up-to-date research on their specialty. Secondly, they collaborate with journal editors or funding organisations and sometimes other reviewers, thus building professional networks.

Nevertheless, editors as well as funding organisations bemoan declining response rates of reviewers, as researcher's time budgets tend to diminish. Several solutions to increase response rates and reliability of reviewers have been proposed:

*Remuneration* of reviewers is a means to acknowledge opportunity costs of the review process. Some organisations remunerate reviewers' home institutions in order to compensate for the time their scientists invest in peer review. This also helps to create incentives for institutions to encourage involvement in evaluation procedures.
Electronic review modes are introduced in numerous institutions in order to increase response rates, opportunity costs for reviewers and to speed up the review process. However, electronic review has the disadvantage to discourage face-to-face deliberation that generally enhances the quality of review.

Peer review colleges are a means to increase response rates, consistency, and continuity in the review process. "College" refers to a group of experts who are contracted by an organisation to provide peer assessment. Membership can be advantageous for the researcher, as it indicates the organization values his expertise. In return, the individual researcher expresses willingness to participate in review processes. Peer review colleges also provide training to reviewers. However, they represent an administrative and financial burden for organisations, and small colleges may have the drawback to lack the necessary expertise in niche disciplines or multidisciplinary areas of science.

On the level of funding organisations, peer review is an important tool to create trust in grant allocation decisions. Albeit the several pitfalls mentioned throughout the brief, it is also seen as the best available procedure to ensure excellence in project selection in the context of lacking in-house expertise, especially in highly specialized research domains. In programme evaluation it presents the additional advantage to create links to the academic community and to increase the acceptance and thus the efficiency of programmes. More generally, an important incentive to use peer review is that it is low cost and relatively easy to implement. However, organization wanting to introduce peer review procedures should be aware of hidden and indirect costs of peer review. These will be addressed in the last section of the brief.

Scope for policy intervention

In most countries, peer review is governed by regulations in two respects. First, federal misconduct regulations apply if a reviewer seriously abuses the review process. Second, peer review for grant attribution sometimes rules out review by individuals with conflicts of interest. Notwithstanding, most parts of peer review are not directly regulated. It is governed instead by guidelines and custom. Guidelines of organisations reviewing research usually relate to confidentiality and conflicts of interest, and sometimes dress specific responsibilities of peer reviewers.

Where peer review is directly implemented by state organisations, policy makers should have in mind that peer review may not be adapted to every situation, and the outcomes of peer evaluations depend heavily on the organisational context in which they are deployed. It is important to know exactly what is to be achieved by introducing peer review, and how its functioning can be optimized.

Several direct and indirect means to improve the process should be considered (see also abovementioned measures in the mechanisms and incentives sections):
Issuance of *guidelines, codes of conduct, and best practices* (for example: Rigby 2002) wherever they are missing, and regular evaluation of those in use can help improve consistency of and trust in peer review.

In some cases, the combination of peer review with *quantitative indicators* may enhance the evaluation process. Bibliometric indicators for example can make peer review more transparent and protect against “old boy networks”. They can also be useful resources in cases of conflicting evaluations by panel members. In addition, they can be used to reveal gaps in the knowledge of reviewers and reveal anomalies in the evaluation process. Quantitative indicators should nevertheless be handled with caution, as they cannot replace the tacit knowledge of specialists.

As said above, *merit review and multi-criteria mapping* can be means to improve responsiveness of peer review to socioeconomic relevance of research. However, the introduction of new criteria in itself can prove to be insufficient to change evaluation habits. The National Science Foundation (NSF) for example updated their merit review criteria for the allocation of funds in 1997. Reaction to the new “broader impacts” criterion however has been mixed and many proposers and reviewers simply ignored it altogether or assigned it little weight. Only after reaction from Congress in 1998 and again in 1999, directing NSF to engage the National Academy of Public Administration (NAPA) to examine and report on the effectiveness of the new criteria, NSF has taken steps to increase attention to the “broader impacts” criterion. This illustrates that *monitoring and adjustment of new criteria*, as well as suggestions of how to interpret them are crucial steps in order to guarantee effective implementation.

Some authors mention issues of *organisational design* that influence the efficiency of the peer review process. Larédo (2008) suggests that peer review’s tendency to favour “mainstream” research plays out more heavily when funding decisions are taken by many different institutions than it is the case when some big funding agencies control the allocation of research resources. This might be worth considering in decisions concerning research organization.

**Assessment**

Assessment of peer review concerns 1) its ability to select excellent research and the consistency of decisions throughout different panels, and 2) its cost-efficiency.

1) The assessment of the quality of peer review procedures can take very different forms. These include monitoring systems, regular evaluations and comparisons, quality assurance mechanisms (ISO and other guidelines) etc. Scholarly observers have proposed different sets of criteria to evaluate the quality of peer review. Chubin (1994) for example proposes seven requirements:

- *Effectiveness* in the allocation of resources and the setting of research proposals.
- *Efficiency* in the use of resources like time and money.
- *Accountability* of science to scientists and the general public.
- *Responsiveness* in supporting new directions for innovation and new research fields.
Rationality through clear procedures and transparency.
Fairness via equal and impartial evaluation of established researchers and newcomers.
Validity. Peer Review should obtain the same results from repeated evaluation (replicability) and remove contingency in the review process.

These value pairs present tensions and there may be trade-offs when implementing peer review. More systematic comparisons between different uses throughout organisations can help to elucidate to what extend basic requirements are respected in different contexts, but they are rather seldom. An exception is the Comparative Assessment of Peer Review project funded by the National Science Foundation (http://csid-capr.unt.edu/). It examines the peer review process at 6 science agencies worldwide (NSF, NIH, NOAA, NSERC, the EU's 7th FP, and the Dutch STW), and aims more specifically at implications of the introduction of broader societal impacts issues into the peer review of grant proposals. Such comparative assessment projects can help to identify best practices of peer review.

2) As peer review is increasingly used in a variety of domains, it is important that the benefits of the procedure outweigh the costs. This is not an easy task because costs of peer review are easily underestimated. In fact, expenses most often incur as indirect costs and implicit opportunity costs, not as explicit payments. They include:

Costs for the editor/funding agency. These tend to be relatively low, except in the case of remuneration of reviewers and/or the use of peer review colleges.

Opportunity costs for reviewers. Time is a precious resource, and as the research environment becomes increasingly competitive, opportunity costs for reviewers – who are generally high-profile researchers with limited time-budgets – should not be underestimated.

Waiting costs for reviewed researchers. This is true for publication submissions and grant attribution, especially when proposals are finally rejected and when the review process implies several corrections.

Costs to the scientific community and the general public. Peer review is a relatively long procedure, and can have the effect to delay the dissemination of new knowledge.

In order to minimize these costs, all stages of the peer review process can be optimized. Possible tools and procedures include the use of electronic templates and screening methods in the application and submission phase; the use of colleges or other networks in the selection of peer reviewers; and electronic meetings and networking for panel discussions. NSF for example reduced the number of final proposals for the evaluation process by one-third by using a screening method in the submission phase. Quantitative indicators can be helpful in this respect, and inform various forms of fast-lane systems. Furthermore, several institutions have created monitoring institutions like NIH’s “Center for Scientific Review” (CSR) so as to maximize efficiency in the evaluation process.
Further resources:


