# Issues and Suggestions for Advanced Practices of Expert Review in the Evaluation of Science and Technology[1)]

Dong Hoon Oh[1]

**Abstract**

With the increasing focus on productivity growth and global challenges such as climate change and energy security, there has been a growing focus on the evaluation of public support to R&D and innovation. In that respect expert review is in fact one of the most common methods used to evaluate public funding of science and technology. This article focuses on the role of expert review as a significant methodological tool for *ex post* and *ex ante* evaluations of research policies, programs and public research organizations (PROs). It describes the definitions, the key processes and uses of expert review especially at the program and policy level, summarizes the challenges and suggests potential solutions to these challenges for policy makers. The demand for effective evaluation tools to make decisions on research funding will continue to increase as countries try to enhance competitiveness and improve innovation capacity. In this respect expert review is a useful tool to meet accountability and effectiveness of public support, but it would be needed more research collaboration among researchers, policy makers, and stakeholders. This paper concludes with a list of principles and suggestions for good practices in the use of expert review in programme and policy evaluation that are based on findings from the literature on expert review and the experience of countries.

**Keywords**: expert review, peer review, Science and Technology (S&T), innovation, evaluation

## 1. Introduction

The evaluation of public policy is a constant and characteristic phenomenon of modern knowledge-based economies. Evaluations are useful for designing, monitoring, assessing the effectiveness and legitimising public policy interventions. The area of science and technology is not exempt from this public scrutiny and accountability given the amounts of public spending devoted to R&D and innovation. In many ways, evaluation has been an integral part of a country's national innovation system insofar as it is both a "tool" for policy making but also a process for policy learning and development. The way evaluation is used and institutionalised in S&T policy making, however, varies greatly across countries. In some countries, evaluation is the remit of technical experts and limited to discrete policy interventions. In others, evaluation is used more systemically at different levels of policy making and by different institutions and actors (e.g. funding agencies, PROs, ministries).

The 'expert review' process is perhaps one of

[1] *Office of Science and Technology Policy and Planning, Korea Institute of S&T Evaluation and Planning(KISTEP), 8F Dongwon industry bldg. 68 Mabang-gil, Seocho-gu, Seoul, 137-717, Republic of Korea*
E-mail : smile@kistep.re.kr

1) This paper is based on results of extended literature reviews and the work which was supported by the OECD evaluation project carried in 2007-2008.

the most dominant and common methods used in the evaluation of science and technology. It plays a significant role in many of the key stages associated with research. It is the main mechanism for deciding who gets funded and what type of science is funded; it determines who gets to publish in the scientific literature; and is used in the selection and promotion of individual within research institutions (Scott, 2006). It is also the core tool used in various R&D programmes and innovation policies.

Expert review has many merits. It is as a relatively quick, low-cost, fast-to-apply, well-known, widely accepted, and versatile evaluation method that can be used to answer a variety of questions throughout the project performance cycle, as well as in other applications. It also provides an opportunity for mutual learning. Expert review could very well be the best of all known methods of assessing R&D programmes and policies so long as it is *properly* managed.

There are, however, some concerns that the expert review system is under pressure and losing confidence among users because it depends on the professional but subjective decisions of individuals and it is increasingly time consuming and resource intensive. It is not an exaggeration to say that expert review is currently facing its strongest challenges in several decades. At a higher level, we see external and internal challenges. Externally there is some evidence of discontent among political decision-makers about the ability of expert review to reflect socioeconomic and political priorities. Internally, a hollowing out is occurring as increasing pressure on researchers' time makes it more difficult to find experts willing to undertake reviews. From the perspective of the method of evaluation, it is therefore an appropriate time to assess the status of expert review and to identify the challenges and solutions.[2)]

This paper covers the role and challenges facing "expert review" as a tool for *ex post* and *ex ante* evaluations of research *policies, programmes* and public research *organisations* (PROs).[3)] The purpose of this paper is to provide a comprehensive assessment of expert review at the programme and policy level but also a summary of methodological issues and suggestions which are based on good practices that have emerged from the experience of countries.

## 2. Definitions and Applications

### *2.1 Definitions*

There are several definitions on peer review. Hartmann & Neidhardt (1990) define peer review as various processes to evaluate the quality of research by peer scientists. Chapman & Farina define peer review as "a process of assessment on research proposal by peer scientists" (Chapman & Farina 1983). Kruytbosch (1989) also provides a simple definition of peer review in science as "advice about proposed actions solicited by decision makers from recognized experts in relevant technical areas." Chubin & Hackett(1990) say that peer review is an organized method for evaluating scientific research in order to enhance the exactitude of research process, evaluate the authenticity of results, and allocate scarce resources.[4)] An OECD document provided a comprehensive definition of peer review as follows (Gibbons and Georghiou, 1986):

> *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*

2) For these reasons, the 2005 OECD-BMBF Conference on Evaluation and subsequent meetings have highlighted a number of issues in the domain of peer review of research. For example, see http://www.internationales-buero.de/de/2193.php and http://www.prague forscience.cz/Scientific-Programme.php

3) A research programme is a collection of funded research components. These elements could be subprogrammes, projects, or individual work units. Conceptually, a programme is greater than the sum of its components, just as the living human body is greater than the sum of its component cells. A programme includes the intelligence or inherent logic that links the components to each other and to the programme's overall objectives, just as the living human body includes the intelligence that links the cells to each other and to the homeostatic operation of the body. Thus, the intrinsic quality of a research programme is not merely the sum of the qualities of its component projects, but depends on the quality of the structural relationships among the projects as well (Kostoff, 2004).

4) Some people use the term "peer advice," "peer evaluation," "peer judgment," "quality control," "peer censorship," "merit review," "refereeing" as an equivalent.

*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.*

Peer review in this form is intrinsic to the practice of science, being used in publication, career and resource allocation decisions. It is widely used by industry, government, and academia. Increasingly it is also being used as an instrument for ex-post evaluation. The model of peer review has also been extended to encompass additional criteria, notably socio-economic criteria and the potential to contribute to innovation as well as other considerations of merit beyond scientific quality. According to this trend, EERE's Peer Review Guide (2004) defines in-progress peer review as:

*A rigorous, formal, and documented evaluation process using objective criteria and qualified and independent reviewers to make a judgment of the technical, scientific, and business merit, the actual or anticipated results, and the productivity and management effectiveness of programs and/or projects.*

As we can see above, there is no single definition of peer review used in the evaluation literature. However, all of definitions of peer review adhere to the fundamental concept of a review of scientific or technical merit and socio-economic impacts by individuals with professional competence and no unresolved conflicts of interest (GAO, 1999; Guston, 2001).

Expert review however is a broader concept than peer review.[5] The classical definition of a peer is "A person who has equal standing with another." A peer review, then, could be defined as "A review of a person or persons by others of equal standing." The crucial issue then becomes how "equal standing" is defined. For example, although scientists who participate in an evaluation may be identified as the "peers" of the applicants when evaluating research proposals, in a programme evaluation, experts in other fields in addition to peer scientists should be included. The term "expert review" is therefore more appropriate than peer review for an evaluation of a programme. The term "expert review" could be defined as follows (Ruegg and Jordan, 2007):

*Qualitative review, opinion, and judgment from individuals with professional competence on the subject being evaluated, based on objective criteria.*

The best-known form of expert review is actually peer review, developed from the premise that a scientist's or engineer's peers have the essential knowledge and perspective to judge the quality of research and are the best qualified people to do so. Peer review is commonly used to make many kinds of judgments: about the careers of individual researchers, about the value of their publications, about the standing of research institutions, and about the allocation of funds to individuals and to fields of research (COSEPUP, 1982). Some people therefore often use the term "peer review" instead of expert review.

In conclusion, expert/peer review has distinguishing characteristics such as being a *qualitative* method, judgement by qualified *individuals*, and based on *objective criteria*. Whichever definition one uses, the following three issues might be key for high-quality expert review or peer review. That is, who should be the evaluator? How to enhance the credibility of subjective opinions and judgements of individuals? How to develop and provide materials and criteria to the evaluators for objective evaluation?

### *2.2 Purposes and Applications*

The evaluation of a policy/programme involves assessing one or more of five domains (Rossi, 2004): *i)* the need for the policy/programme, *ii)* the policy/

5) According to COSEUP (1999), 'expert review' could be classified into three types: i) peer review, which is commonly used to make judgments about the careers of individual staff members, the value of publications, the standing of institutions, and the allocation of funds to individuals, organisations and fields of inquiry; ii) relevance review, which is used to judge whether an agency's programmes are relevant to its mission; and iii) benchmarking, which is used to evaluate the standing of an organisation, programme, or facility relative to another.

programme's design, *iii)* its implementation and service delivery, *iv)* its impact or outcomes, and *v)* its efficiency. The general goals of the evaluation relate mainly to programme improvement, enhancement of accountability, or knowledge generation (Chelimsky, 1997). Expert review is one method of evaluation. Therefore, in a basic way, expert review is used to help policy makers reach their goals.

According to the literatures (Kostoff, 2004; Alassaf, 1996; Armstrong, 1997; Cram, 1992; Levine, 1988; Palli, 1993; Rainville, 1991; Ramsay, 1989; Stull, 1989; Wakefield, 1995; Wicks, 1992), expert reviews of projects and programmes serve a broad range of *purposes:*

- It serves as quality filter to conserve scarce resources;
- Papers published in peer-reviewed journals are assumed to be above a threshold of minimal quality, such that the reader can focus limited time resources on the highest quality documents assumed to be contained in journals;
- Projects and programmes selected for initiation or continuation by expert review are assumed to be above a threshold of minimal quality;
- Precious labour and hardware resources can be focused on these high quality tasks selected;
- Expert review has the potential to add value to, and improve the quality of, the manuscript or programme under review;
- Expert review can provide a mark of approval for legitimacy and competency to increase a programme's visibility and support;
- The objectives of expert review range from being an efficient resource allocation mechanism to a credible predictor of research impact; and
- A properly conducted expert review of a research programme can provide research sponsors with a credible indication of the programme's quality, relevance, management, and appropriateness of direction.

Policy makers and programme managers want to know through evaluation whether their research is being done right (*e.g.* has high quality and efficiency); whether the programme's R&D efforts are focused on the right areas; how programme-created knowledge finds varied applications that generate additional benefits to the nation; how collaborations and other activities stimulated by the program have affected the nation's R&D capabilities; and if their past efforts or new planned initiatives are worthwhile, and so forth. A good expert review should be able to provide programme managers and policy makers with answers to these questions. Ruegg and Jordan provide a good summary of *uses* of programme expert review as follows (Ruegg and Jordan, 2007):

- To conduct in-progress reviews of scientific quality and productivity;
- To help answer questions about the relevance, timeliness, riskiness and management of existing programme research activities, and resource sufficiency of new programme initiatives;
- To score and rate projects under review to aid decisions to continue, discontinue, or modify existing or planned project, programmes, or programme initiatives;
- To help assess the appropriateness of programme mechanisms, processes, and activities and how they might be strengthened;
- To integrate across multiple evaluation results and render judgments about the overall success of a programme or programme initiative;
- To provide information to help programme managers make decisions to design or revise their programme, re-direct existing R&D funds, or allocate new funds.

### *2.3 Merits and Limits*

Like other methodologies used in evaluations, expert review has its own strengths and limitations. This section summarises the *merits* and *limitations* of expert review. The merits of expert review can be understood as follows:

- *Expert review is relatively fast and convenient*. Given that the most appropriate experts are selected, expert review, in any form, may be very time-efficient;
- *Expert review may be carried out in diverse*

*situations*. It is also easy to persuade stakeholders for the following reasons. One often finds numerous experts on a given evaluation object; because these experts participate as a third-party, it is easy to persuade the evaluated and the stakeholders;

- *Expert review is relatively cheap*. Because it involves using existing knowledge of the experts, one may reduce the costs of additional analyses;[6)]
- *Expert review provides opportunities for mutual learning to those involved*. There is much discussion and exchange of ideas through expert reviews and one may find intended and/or unintended benefits from such activities.

Despite the merits, there are many limitations to expert review, including:

- *Difficult to ensure accuracy and quality of the resultant evolutions* when expert review is applied to the impact assessments of research and development programmes. Therefore, its usefulness as a method to guarantee reliability and consistency (or repeatability) is limited.
- *The quality of a review is limited by the biases and conflicts of interests* of the reviewers. Although one may reduce biases and conflicts of interests through various measures, in principle, they may never be completely eliminated.
- *Expert reviews tend to perpetuate orthodox and conservative paradigms*, and tend to reject new paradigms that threaten the structure of the status quo.

The second and third drawbacks mentioned above in particular pose challenges to the reliability of or confidence in expert review. These risks mainly concern the review of grant applications or scientific papers (i.e. project level expert review) and have been most frequently examined in the context of "peer review". While reviewers should be as objective as possible, in practice the peer judgment is affected by different factors (e.g. bias, favouritism, conservatism, discrimination, and so on) which have nothing to do with the subject of the evaluation, which raises the risk of a crisis of confidence.

The first *bias* is known as the "Matthew effect". According to the "Matthew effect", the allocation of research funds is more likely to be skewed towards more famous and influential researchers, and researchers who received funds before have a higher propensity of getting funds repeatedly (Merton, 1973). Gustafson (1975) shows that 46% of the entire research funds is awarded to the top 10 research organisations in NIH, and the top one-third of the total funds goes to the top 20 organisations in the NSF. The "Matthew effect" can be a severe problem especially when not enough research funds are available, and such a problem is often pointed out more by those who are unsuccessful in their proposals (Pouris, 1988).

Peer review by definition is not immune to the risk of *cronyism* as the established scientists mutually support each other. *Informal cartels* or personal connections play an important role especially in the evaluation of a major project that may have a large impact on a researcher's reputation. The selection of panel members and their evaluation processes may also be influenced by *favouritism* and *discrimination*. For example, when a member holds a key post too long in the evaluation committee and that person may even appoint his/her successor personally, the evaluation committee may not represent the entire science community but reflect interests of only a certain group. Such a problem can lead to discrimination against certain groups, including women, young researchers, and researchers who work for less renown institutes and universities (Gustafson, 1975). It is therefore very important for an objective and fair evaluation to avoid the effects of the social replication or the so-called "Old-Boys-Network."

The *conservatism* of peer review has also been criticised by many. Peer review can be seen as supporting an orthodox and conservative paradigm whereby it is hard to accept a new and innovative idea that may threaten the stability of the present structure. Given that the goals of evaluation is to

6) However, there are considerable hidden indirect costs in expert review. Also, as regards programme evaluation, the actual cost may increase by a large amount due to additional resources needed to analyse the programme.

promote, support, and indemnify the new innovation and paradigm in science and technology, it is important to choose members of the review panel who value these goals in their evaluation. One of the weak points in peer review is that only specialists who know their own specific fields make up the review panel instead of experts who have a broader view. Established scholars may be satisfied with a present position in the science and technology community and may be against any new paradigm that may threaten the current paradigm. If the review panel is composed of the only mainstream researchers, they will be more concerned with questions such as "is this research successful?" as opposed to more fundamental questions like "is this research really needed?" Such a narrow evaluation will come to support only the views of the mainstream scholars (this is known as "Pied Piper Effect" in the literature (Kostoff, 1996).

Certain authors note that peer review is quite conservative in its analysis and therefore not able to acknowledge the scientific achievements of other fields (Bozeman and Melkers, 1993). The established fields may also have a better chance than new fields to obtain a grant because new fields are placed at a disadvantageous position in accessing the mass media and in lobbying (Pouris, 1988).

Ethical issues also threaten confidence in the peer review process. There are many ethical issues in the scientific community: fraud, plagiarism, fabrication, image manipulation, leakage of commercial confidentiality etc. (Campbell, 2006). For example, plagiarism and wilful delay in the evaluation by reviewers can damage researchers' interests. It is relatively easy for reviewers to appropriate or use the grant applicant's ideas by delaying the evaluation process intentionally especially when reviewers conduct research on the same topics as grant applicants. In addition, a leading scholar in a certain area may not want to see another rival who might challenge his/her authority later so he/she would try to hold him back by criticising a new researcher's work inadequately (Pouris, 1988). Scientific misconduct like this has enormous impacts but is often hard to document and prove. In fact, much scientific misconduct in science and technology originates in the peer review process. The academic world is spending a great deal of effort to prevent such misconduct given that one of the purposes in the peer review is to protect the ethical values in the science and technology community (Goodstein, 1995).[7)]

## 3. Key Processes of Expert Review

Even though the purpose of this paper is to highlight the challenges to expert review and present some emerging solutions, the key processes of expert review of programmes/policies deserve to be touched upon briefly because the evaluation of programmes and policies have a different focus and deal with different uses, different stakeholders and a different level of complications than evaluations of research projects. The section below covers the key processes of the expert review of programmes while highlighting the important aspects of these processes.

Good examples of the process for expert review at programme level are provided in several existing national or institutional guidelines (EERA, 2004; Kostoff, 2003; Kostoff, 2004; Rigby, 2002; The British Academy, 2007; EPA, 2000).[8)] In relation to the review process, Kostoff suggests the following five phases: *i*) initiation of the review; *ii)* establishing the foundations for the review; *iii)* preparing for the review; *iv)* conducting the review; *v)* post-review actions. EERE's guide describes four phases: *i)* preparations; *ii)* pre-review; *iii)* conduct of the review; *iv)* post-review activities. EPA's *Peer Review Handbook* (2000) describes three stages: *i)* planning a peer review, *ii)* conducting a peer review, *iii)* completing a peer review. Rigby (2002) suggests twelve key steps as follows: Setting the Terms of Reference; Overall Time

7) Recent work by the Global Science Forum addresses scientific misconduct in more detail (see: http://www.oecd.org/dataoecd/37/17/40188303.pdf)

8) The EERE guide and Kostoff (2004) provide information and examples useful for planning, conducting, and utilizing expert reviews based on best practices in the US.

Available; Appointment of Panel Chair; Appointment of Panel Members; Appointment of the Panel Secretary or Scribe; Operating Procedure; Schedule of Work of the Panel; Links from Panel to Programme/Client and other Sub-contractors; Identifying the Requirement for External Support; Interim Reporting; Final Reporting; Dissemination. As above, although there are differences in the literature, expert review is generally understood to have the following three phases: Pre-review – Implementing review – Post-review phase. This section describes the main phases and steps of expert review with key steps and actions mainly based on the materials mentioned above.

### *3.1 Pre-Review*

The pre-review phase is as a preparation and planning stage and includes the following three activities: establishing the foundations of the review, selecting reviewers, preparing tools and materials.

#### *3.1.1 Establishing the Foundations of the Review*

*Initiation of the review: Assigning the responsibilities:* A successful R&D programme expert review requires full participation by the unit undergoing a review. With few exceptions, no one likes or wants to be evaluated. How, then, can the stakeholders be motivated sufficiently to participate fully, and insure that the best review product will results? Motivation and participation derive from the actions of organisation's senior management at the initiation of the process. In this process, it is of utmost importance that a senior manager (that is, a senior decision-maker in evaluation agency) sends out an initial letter to all participants including: the purpose of the review and its importance; the goals, objectives, and scope of the review; the identity and responsibilities of the review manager(s), the general responsibilities of the reviewers, and the responsibilities and reporting chain of the reviewers through all phases of the review process etc. (Kostoff, 2004).

*Identifying the purpose and scope of the review:* The first step is to determine the purpose and scope of the review within the context of other review and management activities. Identifying clear the objectives of the review and the boundaries of the programme to be reviewed provides a framework for the remainder of the review. If the purpose is unclear and if the scope is too large, the evaluation gets confusing and the evaluation questions get nebulous. On the other hand, if the scope is too narrow, it is difficult to have a birds-eye view of the programme, and it is hard to draw conclusions on the redistribution of resources and modifications vis-a-vis other programmes. The smallest unit of review also should be determined at this stage. General speaking, at the project level, the review focuses on whether the "projects are being done right" and many of the reviewers have a high level of topical expertise. At the programme level, the focus is on whether the "right things are being done." Evaluation on R&D programme may include in-depth technological reviews of the accomplishments of S&T projects within the programme. It may also fix programmes as the review unit and assess the uniformity of the programme with the policy objectives, the relationship with other programmes, the relevance of the project portfolio, and the relationship with the external environment. Therefore, the review unit needs to be selected in advance according to the objectives and the uses of evaluation.

*Identifying the evaluation criteria and review questions to be used:* Expert review requires pre-established evaluation criteria. Evaluation criteria should be identified and selected primarily by the mission and review objectives as well as the nature of programme and material being reviewed. The criteria and related standards for judging any aspect of the programme reflect the programme's definition of success and characteristics of the programme or projects. The criteria should focus on the right questions and the tough questions, the questions that most need to be discussed by an objective expert group. Criteria and associated questions need to be stated as clearly and succinctly as possible to reduce the likelihood that reviewers will use their own interpretation (EERE, 2004). The fundamental evaluation criteria for a R&D programme are research quality, research relevance,

and overall programme quality. For some evaluations, the fundamental evaluation criteria have been further subdivided into research merit, research approach/ plan/focus/coordination, match between resources and objectives, quality of research performers, probability of achieving research objectives, programme productivity, potential impact on mission needs (research/technology/ operations), probability of achieving potential impact on mission needs, potential for transition or utility, and overall programme evaluation (Kostoff, 1997a; Kostoff, 2004). For example, there are a few criteria that are often recommended and used by the DOE, OMB, NAS, and others. Although programmes may choose to define additional criteria, at a minimum all EERE programmes are expected to use the following three criteria (referred to as "core criteria"). The three core criteria are the following: 1) quality, productivity, and accomplishment, 2) relevance, 3) management. In addition to specific criteria, reviewers often could be asked to provide an overall assessment. The OMB R&D Scorecard of the US provides another example of criteria (US DOE FY 2002 R&D Scorecard): 1) accomplishments, 2) relevance, relevance of future research, 3) approach to performing, technology transfer/collaboration. Asking specific questions has an advantage that it becomes easier for the reviewer to do the job requested. Therefore, evaluation criteria are often presented to evaluators as questions tailored to the particularities of the evaluated project or the programme. Of course, these questions will not be applicable to all programmes.

*Identifying information needed and data collection/ analysis processes:* Once the purpose, scope, criteria, and questions of the review have been determined, attention turns to the review process itself: What type of review should take place? How should one collect and analyze necessary data and transmit them to the evaluators? How should one assemble evaluation results from the evaluators? The focus of information and analysis certainly depends on the particularities of programme/policy and as well as the objectives and uses of the programme/policy. For example, if the main objective of the evaluation is on the performance of a programme, the collected data is focused on the performance of the programme and analysis is focused on output, input and impact of the programme. On the other hand, if the objective was to modify a programme or to decide on the continuation of a programme, analysis on the relevance of the programme as well as its portfolio is important. The data collected must be sufficient for reviewers to judge the set of activities against the standards that have been set by the definition of the criteria and the specific questions. The data includes material that is provided prior to the review and during the review. A balance must be struck between having too much data and not having enough data. To the extent possible, the burden on researchers should be minimized by using materials already developed or planned for other purposes, rather than developing new materials just for the peer review. Depending on the type of programme, data can include the following (EERE, 2004): information on the programme or project mission, goals, and targets and milestones including data on how funding is allocated across key activity areas; summary project reports, plans, and budgets; principal investigator or project manager presentations; lists of publications or patent applications and the results of citation analysis; customer surveys, available impact studies; various reports prepared by other external groups; and/or any additional data and information reviewers may request.

*Identifying the types of review group and the audience:* In programme/policy review, the competence of the review group might be more important than the individual reviewer's technical competence. The selection of the type of review group therefore is an important issue, and should be addressed at the initiation of the review process. Many types of groups are possible in order to achieve the aim of the review: For example, 1) an independent panel which is a group of experts independent of the agency, and typically funded under a contract; 2) external reviewers group which consists of experts individually contracted to the agency. Which type of review group to select depends on the objectives and the particularities of the programme or policy. Generally, in the case of expert review whose purpose is to assess the performance

or the accountability of a programme, an independent panel is frequently used.

A programme review could provide an excellent forum for disseminating programme information and results to a wide audience. A determination therefore needs to be made early in planning about whether or not the public will be invited to be present or participate in the review sessions. Care should be taken to insure that the review audience includes: actual and potential customers, stakeholders and other oversight groups, co-sponsors, users, and other agency representatives (Kostoff, 2003 & 2004).

*Establishing a timeline and determining logistics for the review:* Timing is an important factor because evaluation is not for academic research but for practical use. Consequently, after setting the date of the presentation of evaluation results which will serve as the basis for the timeline, major deadlines of the evaluation process should be clearly determined in advance. The primary intent of programme review is to provide information that assists programme managers and staffs in their efforts to improve programme performance. The timing of when the report becomes available to provide useful input is therefore also important. Of course, resources (time, money, people etc.) need to be considered to identify the logistics for the review. Although in theory, resources (time, money) are determined by programme size, objectives of evaluation etc, in practice these resources are scarce. Therefore, while respecting the definite timeline and the format of the evaluation, one need to take these limitations into account when determining specific logistics and concentrate the limited resources on key issues and fundamental processes.

### *3.1.2 Selecting and Inviting the Reviewers*

*Identifying criteria of selecting reviewers:* When seeking nominations, it is important that the criteria for selection of reviewers be clearly presented. The review manager, working with staff, the external steering group, if any, and others establish qualifying criteria that individuals should meet for selection to the peer review panel. These qualifying criteria include: 1) in-depth knowledge of the subject area for which he/she is being selected; 2) that reviewers have no real or perceived conflicts of interest.

*Developing a list of possible reviewers and nominate:* Once the overall technical description of the programme is generated, and technical descriptions of the technical sub-areas are provided, the identification of the reviewer can be initiated. Sources of candidate reviewers can include: programme manager recommendations, membership lists of prestigious organisations, agency review boards, agency consultant pools, contributors to technical databases (such as journal article authors or technical report authors), and other similar lists. The review manager, working with the external steering group and/or others, develops an initial list of candidate chairpersons and reviewers according to like the following: 1) Arranging for several independent, external, and objective groups familiar with the programme to nominate candidates; 2) Identifying candidate chairpersons and reviewers from experts identified in a bibliometric search of the published literature on the topic, or from their roles in research or management institutions or professional societies; 3) Employing a co-nomination approach for identifying and nominating reviewers, where reviewers are selected from those nominated by more than one external expert in the relevant field.

*Gathering background information and developing an initial selection list:* The review manager develops information on the candidate chairpersons and reviewers using approaches such as the following:

- Reviewing the performance of reviewers in past reviews, noting who did or did not meet selection criteria based on this experience.
- Contacting candidates to determine their general interest and availability; sending them project summary descriptions to further identify interests and possible conflicts; and requesting and reviewing self-assessment forms.
- Obtaining staff and/or public input, as appropriate, to identify candidates that may have known biases or other issues. Considerable care is needed here to prevent gathering of materials or other input

that could unfairly or inappropriately characterise an individual and to make sure that privacy or other concerns are not raised.

*Selecting the chairperson and reviewers from list of nominees:* The review manager should select the chairperson and reviewers from the list of nominees by working with the external steering group, the chairperson (after selection) and/or others, using processes such as the following:

- Arranging for independent, external, unbiased, objective university, professional society, or other groups familiar with the programme, as identified above, to select the chairperson and/or the reviewers from the nominees.
- Selecting from the nominees the review chairperson, who then chooses the rest of the reviewers.
- Identifying the chairperson and the reviewers based on a co-nomination process among the candidates, as described above.
- Using an independent, unbiased, objective contractor to select from the nominees either directly, or in collaboration with the steering group, independent, external, unbiased universities, professional societies, or others.
- The selection process should be carefully and fully documented to ensure transparency, as other aspects of the peer review process are, and included in the final peer review report.

### *3.1.3 Preparing Tools and Materials*

*Developing guidelines and tools for the review:* Both the review panel and the presenters should clearly understand the objectives and guidelines for the review as well as the specific evaluation criteria that will be addressed. The review leader and chairperson should determine how the projects/program would be rated and distributed to both reviewers and those being reviewed a written description (evaluation guidelines) of the evaluation method. These guidelines should describe the purpose and scope of the review, the evaluation criteria and questions, data to be presented, and how the data will be collected from reviewers, analyzed and reported. Rating or scoring systems are often used to improve the effectiveness of the evaluation. In this case, clear standards should be provided. The comparability of ratings across peer reviewers and review groups requires that all reviewers use the rating scale in the same way. Thus, it is imperative that the scale be well defined so that all reviews are calibrated in the same way and an adjective or numerical rating will represent the same cognitive appraisal by different reviewers.

*Developing the presentations:* Although in the case of research project review, presentations may be easily prepared by the project leaders, it is a lot more complicated to present the evaluation results of a programme as one need to take into account various socio-economic factors as well as the numerous components of the programme itself. Therefore, evaluation managers should provide appropriate guidelines on presentation to relevant managers.

*Providing evaluation material:* Before embarking upon evaluation, one needs to provide the evaluators as well as those being evaluated (e.g. presenter, programme manager) clear instructions as to what materials are needed for the evaluation by when. This way, the evaluated can effectively prepare for the evaluation. It is recommended that a variety of background material be supplied to the reviewers (and the invited audience) before the review. When the evaluated submits background materials and analysis results according to the guideline provided by the evaluation manager, these materials and analysis results must be distributed in a timely manner to reviewers with a guideline clarifying evaluation criteria, processes and indicators. It is important to provide sufficient time for the reviewing of these materials in order to ensure the quality of the evaluation. Evaluation managers can provide documents containing programme accomplishments at this time. Although these documents may be provided during an evaluation, it is better to distribute them in advance. The reviewers may request additional materials in advance after having examined the initial materials.

*Creating an expert review record:* The expert review record is established at the beginning and maintained throughout the review process. The record

should contain all the key documents of the review. This record is an important part of transparency of the process and will aid evaluation manager's efforts to continually improve its expert review process.

*3.2 Conducting the Review*

*Providing on-site instructions to the reviewers:* Having provided reviewers with written direction prior to review, it is recommended that the review leader or chairperson reinforce guidelines orally at the opening of the review. This will ensure that the reviewers are clear on what is being asked of them and clarify the purpose of the particular peer review. This provides time to settle any outstanding reviewer concerns or questions before the review begins. And reviewers are instructed to keep all evaluations strictly confidential during and after the review. The specifics of on-site instruction depend on choices made by the review leader, review chairperson, and/or group. However, in general, reviewers could be instructed to: *i)* read and understand the evaluation criteria and peer review procedures; *ii)* evaluate each programme element; *iii)* prepare preliminary comments on the merits of the project/program in accordance with the peer review evaluation criteria; *iv)* be prepared to discuss each project and/or the program at the meeting or assign a rating or ratings that reflect the reviewer's opinion of the merit of the project/program in accordance with the specific evaluation criteria, and; *v)* complete the post-review evaluation form.

*Programme presentation and Q&A:* Given that expert review promotes new ideas through discussions between evaluators and the evaluated and provides mutual-learning opportunities, presentation is a crucial step in this process. Concerned parties from various levels – organisation unit head, programme manager, technical unit head – could give presentations; the content of the presentation depends on the presenter. For example, the broader technical portion of the presentations is initiated by the head of the organisational unit in which the program resides, and it includes the following informational material: the mission and objectives of organisational unit, a list of all programs in organisational unit, a description of objectives of each program, the funds and people associated with each program and with the program to be reviewed, an overview of the accomplishments and transitions of programs not being reviewed, and their relation to the accomplishments and transitions of the organisational unit's mission and potential national impact, etc. And the program manager(s) provides a more detailed overview of the program under review, including: objectives of program under review; requirements to be met and derived target capabilities for the S&T initiative.

*Discussion and judgment:* Reviewer-to-reviewer interaction, for example in a special closed session to discuss their preliminary rating and then finalise each of their individual ratings, can improve the quality of the review findings. This discussion can be useful for clearing up misconceptions or bringing in new information. Such interactions may be particularly important at the higher level program review in order to better understand the full range of issues. The review chairperson needs to ensure that no single reviewer dominates the ratings discussion and to make clear that consensus is not expected. After the discussion among the reviewers, judgment takes place on the level of evaluation panel or evaluation committee. Sometimes individual opinions of the reviewers are merely accumulated and sometimes, a consensual judgment is reached based on the individual review results. On programme evaluation, the latter option is often preferred. What is important is that this choice of final judgment method must be determined in advance in the preparatory stage of the review.

*Synthesising evaluation results from the reviewers:* After discussion and judgment, the evaluation results submitted by individual evaluators or by an evaluation committee are confirmed and synthesized for the final report. In the case of evaluations which determine priorities among different programmes, a rating or scoring system is often used. When this is the case, the type of rating or scoring system which will be in use should be determined in advance.

*Developing review documents and report:* The expert

review report provides managers with an independent assessment of the programme's productivity, relevance, and management. The report should include the following features when applicable: programme/project identification, description, and budget; a narrative summarizing the salient features of the comments of the individual reviewers and their primary reasons for their judgments; support of conclusions with specific observations; summary of reviewers' rating or assessment on each individual criteria as well as the overall assessment; actionable recommendations aimed at improving program performance, including areas where further study is desirable; as appropriate, comments on the status of recommendations made at prior reviews; and appendices with the full text of reviewer input. The review chairperson concurs and signs off on the report, which is often also sent to reviewers for review of the record of their own response. With the conclusion of this report, the "conducting review" phase comes to an end and the report is distributed to stakeholders such as the programme manager.

### *3.3 Post-Review Process*

*Integrating additional comments:* Before the report is distributed publicly, the evaluated like programme manager develop and add their response to reviewers' comments and recommendations. And any additional comments about the review, either from the reviewers, the external audience, or senior management should be considered and integrated into the review report.

*Drafting a final report:* In general, there are two forms of the final report, a long version and a short version. The long version includes all the written material that was generated during the course of the review. It provides an archival record of exactly what was done during the review. The short version would summarize the process details, and would focus on reviewer comments and other significant inputs, conclusions, and recommendations. And the final report should include the viewpoints of all the reviewers, with appropriate weightings given for judgment and expertise of specific contributors.

*Make the report available to the public:* When the final report is presented to policy-level decision-maker or higher-level committee and is recognized as official evaluation results, the report should be available to related parties as well as the general public through publications and the Internet.

*Assigning action items and evaluating responses to action items:* If internal management accepts the conclusions and recommendations of the report, action items should be assigned to the appropriate personnel for responding to problems identified in the report. There are many types of responses possible such as a corrective action or a rebuttal disagreeing with the conclusions and recommendations. The response therefore should be evaluated, and appropriate follow-up action taken. These action items, responses, and follow-up actions should be presented at the introduction of the next review.

*Evaluate the expert review process itself, including the lessons learned:* This step is considered as a type of meta-evaluation. Expert review is used as valuable resource for improving future expert reviews by providing information on problems faced during the process, suggestions and requests by the stakeholders.

## 4. Issues and Suggesed Solutions

### *4.1 The Changing Context*

There are a number of changes in the environment which affect how expert review operates. These changes offer new challenges and opportunities for expert review. The followings are representative examples:

*Emphasis on performance*. There is more emphasis on the evaluation of the results and the performance of public policies and utilization of evaluation results as a result of 'new public administration' promoted since the 1990s in the UK, Australia, New Zealand and the United States. For example, in order to enhance the accountability of government programmes, the US' GPRA (Government Performance and Result Act) requires Performance-based Management, and Performance-based budgeting.

**Table 1** Phases and key actions for the expert review

| Phases | Key actions |
|---|---|
| Pre-Review | *Establishing the Foundations of the Review*<br>• Initiation of the review: Assigning the responsibilities (K)<br>• Identifying the purpose and scope of the review<br>• Identifying information needed and data collection/analysis processes<br>• Identifying the evaluation criteria and review questions to be used<br>• Identifying the types of review group and the audience (K)<br>• Establishing timeline and determining logistics for the review<br><br>*Selecting and Inviting the Reviewers*<br>• Identifying criteria of selecting reviewers<br>• Developing a list of possible reviewers and nominate<br>• Gathering background information and developing initial selection list<br>• Selecting the chairperson and reviewers from list of nominees<br><br>*Preparing Tools and Materials*<br>• Developing guidelines and tools for the review<br>• Developing the presentations<br>• Providing evaluation materials<br>• Creating the expert review record |
| Conducting Review | • Provide final instructions to the reviewers<br>• Programme presentation and Q&A<br>• Discussion and judgement<br>• Synthesizing evaluation results from the reviewers<br>• Developing review documents and report |
| Post-Review | • Integrating addition comments<br>• Writing a final report<br>• Make the report available to the public<br>• Assigning action items and evaluating response to action items<br>• Evaluate the expert review process itself, including lessons learned |

Sources: Adapted with changes form U.S. DOE EERE (2004), EERE Peer Review Guide: Based on a Survey of Best Practices for In-Progress Peer Review, August 2004; Kostoff, Ronald N. (2003), Science and Technology Peer Review: GPRA, Office of Naval Research. Kostoff, Ronald N. (2004), Research Program Peer Review: Purposes, Principles, Practices, Protocols, Office of Naval Research; Rigby, John (2002), Expert Panels and Peer Review," Fahrenkrog, Gustavo, Wolfgang Polt, Jaime Rojo, Alexander Tubke, and Klaus Zinocker eds., RTD Evaluation Toolbox: Assessing the Socio-Economic Impact of RTD-Policies. IPTS Technical Report Series, EUR 20382 EN.

*Progress on international benchmarking and the internationalisation of evaluation*. In many OECD countries, there has been an increase in international benchmarking regarding policies on science and technology (OECD, 2007d). This could be seen as a continued effort to promote the quality and the objectivity of evaluation. As seen by European Science Foundation's member organisations, it is particularly the case among EU member states.

*Development of methodologies*. Recently, there have been more efforts to evaluate programmes/policies using quantitative indicators and these efforts have led to the development of new indicators. Also, various methods have been developed to measure socio-economic impacts of the programme. Therefore, there is growing interest in coming up with ways to effectively complement expert review with other evaluation methods.

*Requirements for greater transparency*. Given the limited resources for research and development, there is greater competition in priority-setting. This requires more transparency in priority-setting processes. Elimination of biases and conflicts of interest in the evaluation process also remains as a challenge.

*Development of information and communication technologies*. There is greater flexibility in expert review with the development of various communication tools such as phone-conference,

video-conference and the Internet. Also, with the expansion of the Internet and the development of online databases, there are no time or spatial limits in accessing and exchanging information. Real time entry and reviewing of evaluation data/information became possible. These developments have contributed to the effectiveness as well as the quality of evaluation and allowed for network-centred expert review. Electronic communications now means that expert review can more easily be an international process, potentially widening the range and number of reviewers.

### *4.2 Methodological Issues and Solution Based on Country Experience*

Over the years peer review has received much attention in the evaluation literature. Studies have suggested a number of challenges, solutions, and issues. And most of these are related to project level evaluation such as grant application, paper publication, and ex-post project evaluation. For example, Wood and Wessley (2007) covers issues in their recent systematic review which are mainly related to grant peer review as follows: Is peer review of grant application fair?; Are peer reviewers really peers?; Is there institutional bias?; Do reviewers help their friends?; Age and getting grants; Gender bias and grant peer review; Misuse of confidential information; Reliability of grant peer review; Does peer review of grant applications serve the best interests of science?; Is peer review of grant application cost effective?; Can peer review of grant applications be improved?; Should peer review of grant application be replaced?

And Kostoff (2004) describes the strength and weakness of major peer review components and issues, including: Objectives and purposes of peer review; Quality of peer review; Impact of peer review manager on quality; Selection of peer reviewers; Selection of evaluation criteria; Secrecy (reviewer and performer anonymity); Objectivity/bias/fairness of peer review; Normalization of peer review panel; Repeatability/reliability of peer review; Effectiveness/predictability of peer review; Global data awareness; Cost of performing a peer review; Ethical issues in peer review; Alternatives to peer review; and Recommendations for further research in peer review.

It is therefore impossible to cover all issues raised. The next section of this paper therefore focuses on a few issues for high-quality expert review in the evaluation of *policy, programmes,* and institutions. Although targeted toward research policy/programme expert review, most of the issues in this report apply to many kinds of expert review including project selection review.

***Issue 1. Consider socioeconomic factors in evaluation: How to effectively reflect socioeconomic and political priorities and link these priorities to decision making in the expert review processes?***

This issue may be one of most important ones in policy making and evaluation. Some decision-makers doubt the ability of expert review to reflect socioeconomic and political priorities. Expert review, in fact, is likely to ignore wider social and economic effects due to its self-oriented and highly scientific approach. Expert review panels are dependent on sound and detailed information on which to base their judgments about a programme's progress or impact, and they are vulnerable to poor and insufficient information. The type of data needed for retrospective impact assessment cannot be created in an expert review panel format. For this reason, expert review tends not to be appropriate for evaluating impacts of programmes (Ruegg and Jordan, 2007).

How can we solve this problem? A couple of solutions could be suggested. To begin with, reviewers could be provided with a pre-analysis of socioeconomic needs and priorities. For example, the Korean government has been informing evaluators of the National Master Plan of Science and Technology about the results of technology foresight, expenditure priorities at a national level, and the status of public R&D expenditure, and analysis of programme's portfolio and performance during the R&D programme evaluation process (Oh and Kim, 2006).

Diversifying the fields of experts could be the most common solution. While it is reasonable to compose a review panel of *peers* from the same field

to assess the excellence of the research proposal and to judge whether or not to award a grant, it may be inappropriate to construct expert panels who only know about their own specialized field or technology, especially when evaluating a programme or a policy aimed at addressing general social and economic problems. To put it simply, it is important to seek *balance* in various aspects when selecting the review panel (U.S. DOE EERE, 2004).

Research & development evaluators should not only have technical expertise but a perspective on the broader issues (for example, the impact of the research, mandate of the programme, economic utility, political and economic effects etc) (Klahr, 1985; Marshall, 1996). Although it would be ideal for an evaluation to have both of these qualities, different evaluators as a team can complement one another to provide the expertise and the broad perspective necessary. Some of experts should have a non-S&T background and have expertise in economics, business, accounting, public relations and policy, industrial policy, and other areas as well. Even in projects related to very specialised technological areas such as biotechnology or nanotechnology, the social and economic impact, needs, relevance, and value that those projects can bring to society as a whole should be treated as important as the scientific merits of technological advances in a programme and policy level evaluation.

Recently, the question of how science can be made more relevant to the needs of society is increasingly central not only in science-policy debate but also in project selection (Scott, 2006; Nightingale and Scott, 2007). In Canada, there is an increased expectation to address political and socio-economic priorities. At NSERC, expert review is mostly used to evaluate applications for research grants. Generally, programme managers guide the work of peer panels regarding the goals, criteria and applicable polices, but they are not involved in the peer review. In some programmes, officers make a recommendation based on the peer review input and analysis of merit relative to the selection criteria. Most panels have members from industry, government and university sectors. There is also diversity in the panels including a mix of national and international geographic representation, stage of career, gender, language and size of the institution. The diversity on panels works well in 'problem/ priority areas' (OECD, 2007d).

Another solution is to establish a dual review committee: One review group focuses on the scientific and technological excellence of the subject of the evaluation, while the other focuses on the relevance and socioeconomic priorities of research. For example, NIH has a 'Dual Review System' for grant applications (Scarpa, 2006):

- The first level of review by a scientific review group (SRG) provides initial scientific merit, review of grant applications, rates applications and makes recommendations for the appropriate level of support and duration of the award.
- The second level of review by council makes a recommendation to institute staff on funding, evaluates programme priorities and relevance, and advises on policy.

Through this dual review system, it is possible to carry out a proper evaluation of the scientific and technical quality of the research as well as its socio-economic value and utility.

"Bicameral review" may also be applied to the evaluation of programmes. According to bicameral review, research grants are assessed with two different, independent criteria. One is the past accomplishments of the researcher and the other is the proposed research project. The former is assessed through peer review; the latter is assessed internally, based on the budget (Forsdyke, 1991; Forsdyke, 1993). These methods could be applied to evaluations with the purpose of setting priorities or the allocation of resources. In other words, it is possible to draw a final conclusion from two independent processes assessing past achievements as well as a country's strategic priorities and budgetary concerns.

It is also possible to use a *Delphi* method, frequently used in technology foresight. In Netherlands, under the assumption that "when evaluating scientific projects, the best standards come from outside the field," when evaluating a grant in the field of physics,

researchers are selected from various fields such as physics, chemistry, mathematics, and astronomy as evaluators. This does not mean that the internal criteria (competence and experience) are not important but that 'relevance' should also be included in the set of criteria. The two-stage Delphi procedure has been adopted in order to preclude bias and misunderstandings. In the first-stage, for each research proposal, the grant applicant is given feedback consisting the questions, criticism and advantages/ disadvantages brought forth by 4-6 evaluators. In the second stage, on the basis of the responses from the grant applicant, the evaluator assesses the applicant's abilities, objectives, methods and the general level of research; he/she then sets the priority accordingly (Pouris, 1988). This allows for improving the quality of the evaluation as the evaluators are aware of others' opinions. It also prevents unnecessary conflict and power struggle among the evaluators as it guarantees the anonymity of the evaluators. Most importantly, it promotes the relevance of projects in national research development policy. With increased emphasis on the issue of relevance, the application of this method in evaluating programme/policy would be very useful.

***Issue 2. Interface of expert review with other means of judgment: How to use objectives indicators or ranking tables effectively in order to enhance the objectivity of evaluation result? How to combine expert reviews with other both quantitative and qualitative methods for evidence-based policy?***

With the rise of indicator-driven judgements and ranking tables, the interface of expert review with these other means to judgement is also of interest. Policy makers and R&D programme managers have attempted to adopt a more quantitative indicator based evaluation system as a complement or substitute to expert review. The problem that occurs in expert review procedures is the identification of relevant performance indicators that are closely linked with the desired outcomes. Programme theory (or a logic model) is used often in order to develop the most suitable performance indicators in programme or policy evaluation. The level of a future target performance of the programme is often presented in advance by objective numbers as well. Developing the more quantitative indicator based on objective numbers rather than qualitative analysis based on subjective opinions by *experts* is becoming an urgent task.

On the other hand, it is also important to raise the accuracy and enhance the subjectivity of the expert evaluation by employing both qualitative and quantitative methodology properly. The qualitative and the quantitative method have their own advantages and disadvantages and taking advantage of the strong points of each can make up for the weak points in the current peer evaluation system.

In fact, there are few examples of combining peer review with other tools. For instance, various methods˜ surveys, case study, sociometric/social network analysis, bibliometrics, historical tracing˜ have been used with expert judgment in ATP programme evaluation (Ruegg and Feller, 2003). Another case shows a similar situation. The technology development programmes in the US Department of Energy (DOE) extensively and successfully utilise expert review to evaluate research and development (R&D) activities at the project and programme levels. In addition to expert review, R&D programme managers in DOE are encouraged to use other evaluation methods in order to obtain information on programme effectiveness and the benefits generated that cannot be provided using the peer review method (Ruegg and Jordan, 2007). Application of quantitative indicators should be based on the notion that expert reviewers need to be provided with condensed, systematic, verified, objective information on the research performance of the groups to be evaluated, and that the grounds for their judgment, or the assumptions underlying it, should become more explicit, thus making the process more transparent (Moed, 2007).

***Issue 3. Cost efficiency of expert review: How to enhance the cost efficiencies of the various parts of the expert review process?***

It is also important to enhance the cost efficiency of the various parts of the expert review process including the administration. Given that the evaluation

process aims at creating new value, the benefits of evaluation should outweigh the costs. Evaluation costs are easily underestimated in a real world because those costs usually incur as an implicit opportunity cost not an explicit payment. For instance, evaluators have to sacrifice their own working time and performance in order to spend their valuable time on evaluating projects assigned to them, but this kind of opportunity cost is often neglected in the benefit-cost analysis in evaluation due to its nature of implicit cost. According to research, indirect costs related to the time value of evaluators, presenters, staffs and visitors are more than ten times as high as the direct costs like travel expenses (Kostoff, 1996). In particular, in the case of the panel evaluation done by renowned experts, the total costs (i.e. sum of implicit and explicit costs) become much higher than direct costs. In sum, given that costs of expert evaluation are not negligible, evaluators should do their best to achieve cost efficiency in the evaluation process.

Efforts should also be made to reduce expenses in each review process. The key process of expert review which we have seen in the previous section (Section 3) suggests the possible ways of minimizing evaluation costs in the process of application, selection of expert reviewers, and panel discussion of discipline committees. Often the size and scope of the programme or project determines the venue for the expert review. Scheduling the event using public facilities, meal planning, and audiovisual requirements, all should be completed well in advance of the actual meeting. Typically, meeting logistics are one of the major costs of an expert review. Expenditures vary depending on the number of projects reviewed, the number of reviewers, whether the meeting is open to the public, and the length of the review. Ways of controlling the cost of the review meeting include the following: Structuring the agenda carefully so that the agenda is focused and people's time is used efficiently; making maximum use of teleconferences, videoconferences, and other electronic media to prepare the review panel. This is particularly helpful when international reviewers are involved (EERE, 2004).

Building an appropriate database of evaluators will have long term consequences in reducing the cost associated with the selection of evaluators. Given today's internationalisation of science, there is much value in promoting international cooperation in building evaluator databases.

There are also many suggestions for reducing the costs in research project evaluation. For example, Klahr (1985) points out that NSF was able to reduce the number of final proposals they should evaluate by one-third using the screening method by comparing the results of "mail review" at the first stage evaluation with the ones of "panel review" at the second stage evaluation. In addition, NIH runs the "Center for Scientific Review (CSR)" to maximize the efficiency in the evaluation process, and CSR also operates "Streamlined Review Procedures (SRP)". SRP was able to save evaluation costs by concentrating on only the quality proposals that rank 50% and above (Lee, Om, and Ko, 2000).

Various types of alternative methods can possibly be employed with the help of various tools supported by the Internet as it delivers real-time news and information and facilitates networking among the persons concerned. For instance, NSF is operating the "NSF Fast Lane System" for more effective, convenient, and faster administration. It has various applications to prepare, submit, and revise research proposals (www.fastlane.nsf.gov). NIH also announces recent policies regarding an assessment of the research proposals officially on the internet through SRP so that researchers can be well aware of the most recent evaluation criteria and policies (www.drg.nih.gov/refrev.htm).

***Issue 4. International frame of reference: How to develop an effective international frame of reference for expert review?***

There are basically two approaches in selecting a panel: intra-national and international. The intra-national panel is composed of local experts in the field, i.e. academics, professionals, policy makers, etc. This type of panel selection is useful in large countries where the possibilities for selection of experts are more numerous due to well-developed S&T systems,

and where the possibility of subjective evaluation is minimal. The international panel is mainly composed of foreign and internationally recognised experts in the respective field. Both approaches can be criticized. The first for its inability to cope with local lobbying of interested groups within the scientific community, and the second for lack of knowledge of external evaluators on certain particularities of the respective country and the possibility that the expertise might be misleading due to the different scientific environments prevailing in different countries (OECD, 1998).

An international frame of reference has been *increasingly* used as the standard for expert review, with the use of foreign experts being seen as the answer both to potential conflicts of interest in small communities and as a means of assuring stakeholders that the work stands up to global scrutiny. In Finland and Portugal, for example, proposals are submitted in English because it increases the number of international reviewers that can be used. As we can see from the examples of countries including Finland, the receptiveness of evaluation results increases when foreign experts are selected as evaluators (Pouris, 1988; OECD, 2007d). The internationalisation of expert panel is needed more in countries that have a small science and technology community. For example, the Korean government is aware that the Korean S&T society and expert pool is very limited. The Government thinks the internationalisation of evaluation would be a solution to enhance objectivity and reliability, and therefore, especially, tries to enlarge the expert pool including foreign experts. But, at the same time, the Korean policy-makers also know that it is very difficult for foreign experts to evaluate Korean R&D programmes because they should have sufficient knowledge of the Korean scientific community, the context of programmes and related policy, and national strategies (OECD, 2007).

The internationalisation of science itself is increasingly important in evaluation at the national level. Research has been internationalising and thus requires international reference points in measuring outcomes. Evaluation needs criteria, standards, and benchmarks to assess the quality and achievements of policy, programme, project, or institutes. In a global innovation system national standards or approaches are limited, and therefore increasingly should be defined internationally. Not only should the performance of an institute be assessed in an international environment but also should the effectiveness of policies and programmes. For these reasons, international indicators, evaluation criteria or benchmarks are needed in expert reviews of research institutions, programmes and projects.

To be sure, the international frame of reference reflects the growing concerns about the role of science in competitiveness and competitiveness in science. However, caution is needed when policy is transferred across different cultures and contexts, particularly when understandings or policies are incomplete. Therefore, there is a need for *taxonomy* of the internationalisation of expert review.

***Issue 5. Managing the conflicts of interest: How to manage the conflicts of interest in the expert review process?***

One of basic hypothesis of expert review is that expert's judgments are trustworthy and reliable, since the persons who do evaluation have judgment, experience, and a professional ethos. However, decisions made by evaluators are easy to be affected by personal relationships with others and this often potentially prevents the entire evaluation process from being impartial and objective. It is therefore to effectively manage the potential or existing conflicts of interest in the expert review process.

The United States Office of Management and Budget's (OMB) Peer Review Standards points out that factors relevant to whether an individual satisfies these criteria include whether the individual: *i)* has a financial interest in the matter at issue; *ii)* has, in recent years, advocated a position on the specific matter at issue; *iii)* is currently receiving or seeking substantial funding from the agency through a contract or research grant (either directly or indirectly through another entity, such as a university); or *iv)* has conducted multiple peer reviews for the same agency in recent years, or has conducted a peer review for

the same agency on the same specific matter in recent years.

One of the direct ways to avoid such a problem is to exclude evaluators who might have interests with proposers in selecting reviewers. It is however nearly impossible to nominate experts to a review panel who have absolutely no interest. Besides the trade-off between choosing reviewers who are indeed peers and the resulting increased chance of a conflict of interest is one of fundamental dilemmas (Wood and Wessely, 2007). Thus, it's better to conclude a mutual agreement out of conflicting views among evaluators in the panel review.

Conflicts of interests could also occur between an evaluation manager and a reviewer. Those conflicts are usually related to the questions such as "who is responsible for the evaluation results?" and "how deep should a manager and a reviewer be involved in decision making?" For example, both an expert reviewer and a manager may want to make the final decision on the proposals, not just supports the other's decision making. The severe disagreement regarding resource allocation decisions in the expert review often comes from the conflict of interests among parties involved in the evaluation process. Proper management of interests and dissolving conflicts among parties would thus enhance the receptivity among them, and it is important to construct peer review mechanism that follows objective evidence, not a personal interest, in overcoming such difficulties.

Concerning the potential conflict of interests, declaration of interests by the evaluators is proposed as a solution (Bozeman, 1993). The UK Research Assessment Exercise requires declaration of interests in order to avoid obvious or potential conflict of interests. It is even argued that the authors of papers should declare their financial interests (RAE, 2001). The scientific journal Nature required authors of papers to declare their financial interests.

Another solution is to internationalise evaluators as foreign experts may have less biases and interests. The Academy of Finland invited a quartet of British, American, West German, and Swedish experts to evaluate the country's progress in Inorganic Chemistry. The assessment believes that 'it succeeded only because the panel came entirely from beyond the frontiers of Finland (Dixon, 1987; Pouris, 1988).

In many review practices, all reviewers must sign a Conflict-of-Interest form prior to the beginning of the review process. In addition, during the review, the reviewer should agree to disclose any actual or perceived conflicts of interest as soon as the reviewer is aware of the conflict.

It is advisable to limit the number of evaluations or the duration of evaluating activities for the participating experts. If an expert participates in too many evaluations, the expert might develop a relationship with the evaluated bodies and might be susceptible to lobbying from them. Moreover, trapped in their judgments from previous evaluations, they might not be able to take a fresh look at similar programmes. However, to limit the duration of evaluating activities excessively may be counter-productive. For example, it might decrease a sense of responsibility. If the expert participates in the evaluation only once, their responsibility might be lower than if they were to participate in future evaluations.

Lastly, it is advisable to prevent an expert who has expertise only in a particular domain from judging the quality or the value of what is being evaluated. However, it is advisable that his opinions are transmitted to other experts in the panel to be used in the joint-decision making. For example, an expert on biotechnology evaluating a biotechnological program might know more about the technical aspects than other experts but cannot judge correctly the socio-economic value of the evaluated program. Also, they might insist on higher allocation of resources to their area of expertise.

***Issues 6. Expert review in the Internet age: What opportunities does the Internet give us for improved and enhanced expert review? Could an Internet based "open evaluation" tool organized by the scientific community be an alternative to the classical approach? Can network-centred expert review replace classical review? Is evaluation possible without expert review panels?***

Technological progress like the Internet provides not only new means and modes of communication but also opportunities for advanced evaluation. Panel review and mail review are the most general type of peer review and both of them are not efficient regards to time and expenses. But using internet in constructing the panel and in evaluating proposals enhances efficiency. Most importantly, peer evaluation systems based on the Internet boost rationality and receptivity dramatically because the Internet conveys all kinds of useful information in real time. That is, all the information regarding text, speech, graphics, music, video, images, 3d-models, and raw data is digitalized and delivered to evaluators, and they can open up necessary information at anytime and anywhere. Besides, the Internet provides evaluators with search engines and alert systems as well as data analysis tools.

Internet could enable a new style of peer review. Whether it's a panel review or a mail review, traditional peer review is a "closed evaluation" by the nominated experts group. An Internet-based "open evaluation" tool organized by the scientific community can be an alternative to the classical approach because the internet can secure additional evaluators around the world without a boundary. For example, a project or an evaluation results can be reviewed by the numerous people once it is posted on the Internet. Also, open evaluation is found to be a very powerful tool to solve data fabrication, which is one of the hot issues in science these days. In fact, the people who first caught the data fabrication in "Hwang's affair" in South Korea were Internet users. Finding errors and data fabrication by researchers are almost impossible to be found during the normal panel review that should be done within a rather short time period, but it is difficult for researchers to deceive all potential reviewers on the Internet at once.

The publishing system of *Journal of Atmospheric Chemistry and Physics* gives a good example, Interactive Open Access Publishing (Mehlhorn, 2006). Papers for *JACP* are handled in two phases:

- In a first phase, the author submits a paper to the editor. The paper is published in *Journal of Atmospheric Chemistry and Physics Discussion* as a paper for discussion. The paper is *openly* reviewed by the scientific community as well as appointed referees with reactions by the author.
- In the second phase, the author is required to submit a revised paper based on comments from referees and the scientific community. The editorial board makes a final decision to publish the final revised version of the paper or not.

The review and publishing system of *JACP* has many advantages. It provides authors, referees and readers with: free speech and rapid publication (authors & readers); direct feedback and public recognition for high quality papers (authors); prevention of hidden obstruction and plagiarism (authors); documentation of critical comments, controversial arguments, scientific flaws, and complementary information (referees and readers); deterrence of careless, useless, and false papers (referees and readers); public discussion and final revision (readers). In short, it could be said that JACP's publishing system provides maximum quality assurance of papers through public, interactive, and collaborative peer review.

Information technology, such as Groupware software, has the potential to significantly improve the efficiency and overall value of the expert review process. Information technology brings real time data entry, screen sharing, data manipulation, and statistical analysis capabilities to the expert review process. Individual reviewers can enter anonymous review and rating data, and the review manager can compute summary rating statistics to share with them in a timely manner. This increased information handling can free up time to permit additional time allocation for important reviewer-to-reviewer or reviewer-to-review manager interactions. Box 1 compares a network-centric expert review with the traditional review process.

***Issue 7. Expert review for policy, programme and/or PROs: What type of expert review is fit for the evaluation of policy, programme, or PROs? Is expert review a relevant tool for evaluating research institutions?***

**Table 2** Comparing a groupware-based peer review with the traditional review process

| Traditional peer review | Network-centric peer review |
|---|---|
| • Data input is via the evaluation form completed during the Q&A session or shortly thereafter. | • All the members of the on-site audience are linked by Group-Ware information technology. All data input is mechanized, and instantly recorded. |
| • Each reviewer completes their evaluation during the session, and the individual and summary result for the panels are computed at the end of each presentation day or after the review has concluded. | • Each reviewer completes their evaluation during the session using the groupware. During the presentations, the reviewers enter final ratings and any additional comments they believe are important based on last-minute observations or insights. Individual and summary results for the panel are made available in real-time and routed back to each individual for further discussion. |
| • Statistical analysis of reviewer comments (summary and integrative statistics, as well as aggregating comments) typically is not available instantly or in time for use in onsite panel discussion. | • Statistical analysis of reviewer comments is completed onsite to provide useful performance data quickly. |
| • Reviewers could meet in closed session to discuss their preliminary reviews. However, during closed session discussion, reviewers often do not have access to the full statistical analysis of ratings for the panel. | • To complement the groupware tool, reviewers could meet in closed session to discuss the preliminary reviews and once the interactive cycle is complete, they may make final changes to their individual review comments and ratings. The groupware technology would enable reviewers to have access to the full statistical analysis of ratings for the panel. |

Sources: www.inform.nu/Articles/Vol2/v2n1p11-18.pdf and Ronald N. Kostoff (2001), Network Centric Peer Review, Office of Naval Research.

This issue is also important for programme level expert review. Because the peer review method is generally used at the project level, it is necessary to consider what type of the peer review (actually, expert review) is appropriate for an upper level decision making in programme, policy, or institution. The outcomes of the review can affect the decision making and those are useful as the reference data as well. For example, an author classified peer review into three categories as follows based on the level of its impact on the final decision making (Bozeman 1993): pre-emptive peer review, traditional peer review, and ancillary peer review. Pre-emptive peer review is the one that the final decision depends entirely on the results of the peer review and a programme manager has no right of judgement. Following the already determined format, scoring model or ranking model is employed in the pre-emptive peer review. The dual review system in NIH is an example.

In the traditional peer review, the decision is also influenced by other factors like the decision of a programme manager while the result of the peer review is still the important factor affecting the final decision. Along with the result of the peer review, an academic standard of the organisation a research proposer is affiliated with and a geographic area are also considered in the traditional peer review. NSF typically uses this method. Ancillary peer review can provide only the partial information out of all the crucial data and thus play a minor role in the decision making process. Assuming that different aspects of evaluation should be considered differently in search of the most suitable evaluation method, economic and political areas and the case of geographical distribution of scare resources are evaluated by different evaluation methodologies from the peer review while the peer review is a suitable method for the science and technology field. These are often used in the major programme evaluation or building up a science complex.[9] Because most programme evaluations have many policy issues to consider, among the three types of peer review mentioned above, the pre-emptive review is rarely used.

On the other hand, selection of the type of review

9) Within the general category of expert review, there are a number of sub-types according to the level of specialisation and professionalisation (Gibbons and Georghiou, 1987; Rigby, 2002): traditional peer review (canonical academic review), direct peer review, modified direct peer review, pre-emptive peer review, indirect peer review, merit review (extended form of peer review), ancillary peer review, expert panels, panel review, professional evaluators.

group is a core issue, and should be addressed at the beginning of the review process. Although there are many types of external expert reviews, two types draw our special attention: the independent panel and the external reviewers group (EPA, 2000; Kostoff, 2003; Kostoff, 2004). The independent panel is a group of experts independent of the agency, and typically funded under a contract. The independent panel has a chairperson, attempts to reach consensus on issues, and generates a written report containing the results of the review and sometimes recommendations. The group of external reviewers consists of experts individually contracted to the agency. The reviewers report to the agency review manager.

In contrast, the external reviewers group does not have a chairperson; the review manager serves in this role. While the group may engage in technical discussions during the course of the review, it does not reach a consensus. While there may be individual written inputs from each group member, there is no group report. The review report is written by the agency review manager based on the individual written inputs plus other considerations. Because of the technical understanding required to write a credible report, as well as select the appropriate mix of reviewers, and conduct all aspects of the review, the review manager should have a solid technical background and some understanding of the subject matter to be reviewed.

Each of the two review group approaches has value for specific applications. The group of external reviewers is less formal, and has fewer reviewer and audience restrictions. It is useful for internal reviews where structural program issues are paramount and need resolution or improvement, and where comparison with other programs is not the major focus. The independent panel is more formal. The independent reviewer panel has more specific reviewer, meeting, and audience selection constraints/requirements. From the agency's perspective, either group has very high utility for addressing the agency's programme improvement needs. From a perspective external to the agency, the independent panel has higher credibility because of its independent nature. For performance evaluation or evaluation for priority setting, the independent panel is more appropriate, because of its perceived independence.

It is also important to determine whether to make the expert review process, for example, the presentation or the contents of the discussions, publicly accessible. Those in favour of having open-to-the-public reviews (EERE, 2004) suggest that having the review meeting open can: help sharpen the questions raised; improve the transparency of the peer review process; help improve or legitimize the technical or management approach; strengthen integration networks for research, deployment delivery, or business management; broaden public learning by providing an opportunity for individuals to hear firsthand what others are accomplishing and how they manage their work; and encourage participants to improve performance due to the pressures of presenting publicly to their peers. It is generally believed that in the case of evaluations of programmes or institutions, making the evaluation process public has a more positive effect than in the case of evaluation for priority-setting.

## 5. Ways to High-Quality Expert Review

### *5.1 Essential Requirements for Good Practice*

What are the key factors for a high-quality expert review? Chubin (1994) suggests seven *requirements* order to enhance the quality and credibility of peer review as follows:

- *Effectiveness*. Peer review should be effective to allocate resources and to set research priorities.
- *Efficiency*. Resources including time, money in peer review should be used most efficiently.
- *Accountability*. Peer review should enhance the accountability of science to not only scientists but also the general public.
- *Responsiveness*. Peer review should be flexible mechanism which can lead the development of new fields and support policy maker's decision of new direction of innovation.
- *Rationality*. Peer review process should be transparent and rational.

- *Fairness*. Peer review should be impartial and observe a social norm that everybody is equal under the law.
- *Validity*. Peer review should obtain the same result from repeated assessment and remove the effect of contingency in review process.

In practice is but all but impossible for peer review to satisfy all requirements mentioned above. Some of these values involve a range of contradictions. Some experts consider that peer review embodies tensions between five 'value pairs' — desirable properties that are in tension with each other (Hackett, 1997; Scott, 2006): *effectiveness and efficiency; autonomy and accountability; responsiveness and inertia; meritocracy and fairness; reliability and validity.*

The most important point here is finding the *optimal* balance between the contradictory requirements. Trade-offs between desirable properties of peer review are inevitable: there is an ongoing challenge for research funding bodies to be able to determine what constitutes a defensible, appropriate, and workable balance (Wood and Wessley, 2007). For instance, although the pursuit of greater effectiveness could enrich the exactitude of evaluation, it requires too much time and resources and reduces cost efficiency. On the other hand, focusing only on cost efficiency may lead to a superficial assessment. Evaluation designers should therefore consider all resources and conditions and choose a more optimal evaluation process and method.

Autonomy is one of key values in the professional community but it often conflicts with accountability. Scientists, as experts, would like to decide what and how to do research by themselves. However, the general public wants to see the output or performance of scientists whose work is supported by taxpayers. Furthermore the dissemination of performance-based budgeting has further increased the emphasis on the accountability of public research. It is very important for a successful peer review to find the optimal balance between the contradictory values in the evaluation process.

While expert review is one of most flexible methods for determining value, capable of application to a wide number of fields, in order to apply it, a number of critical pre-conditions must be met. Much of the literature addresses conditions or requirements of expert review (especially, focusing on peer review of project evaluation) for applications. Rigby (2002) suggested four essential pre-conditions for applying peer/expert review as follows:

- Experts with knowledge of a particular area must be available and be willing to participate. Because it can be difficult for government officials to identify the relevant peers as they not usually part of the social or professional networks of scientific peers, it is important for programme evaluators and responsible bodies to maintain access to such networks.
- The panel of experts cannot be expected to answer questions which are beyond the scope of the available knowledge. Terms of reference need therefore to be set with some sense of what it is possible for the experts themselves to know or to infer and to judge collectively from their specialist knowledge.
- The panel should only be asked to come to a judgment on a single area of knowledge or expertise rather than more than one as peer review is known to be weak where comparative judgments between different fields of expertise have to be made.
- While the costs of peer review are low, sufficient resources should be made available to facilitate the work of the panel. Some panel reviews are often supported by a secretariat.

By definition, a high quality peer review should provide an accurate picture of the intrinsic quality of the research being reviewed, irrespective of whether this intrinsic quality is high or low. The fundamental problem is the lack of absolute standards (analogous to physical standards for primary measurements such as time and length) for measuring research quality. Presently, evaluation of intrinsic research quality is a subjective process, depending on the reviewers' perspectives and past experiences. A high quality review under these imperfect circumstances, then, would occur when two generic conditions are fulfilled:

1) utilization of highly competent reviewers, and 2) no injection of additional distortions in the reviewers' evaluations as a result of biases, conflict, fraud, or insufficient work (Kostoff, 2004).

High quality expert review processes require as a minimum the conditions summarized by Ormala (Ormala 1989): The method, organisation, and criteria for an evaluation should be chosen and adjusted to the particular evaluation situation; Different evaluation levels require different evaluation methods; Program and project goals are an important consideration when an evaluation study is carried out; The basic motive behind an evaluation and the relationships between an evaluation and decision making should be openly communicated to all the parties involved; The aims of an evaluation should be explicitly formulated; The credibility of an evaluation should always be carefully established; The prerequisites for the effective utilization of evaluation results should be taken into consideration in evaluation design.

EERE's *Peer Review Guide* (2004) describes the minimum requirements to be prepared for expert reviews of EERE's R&D programmes as follows:

- *Scope of Review*. All EERE programs in both Technology Development and Business Administration offices and their key projects will be reviewed by qualified and objective peers on a regular basis. This should typically cover 80-90% of RD funding and supporting business analysis and management programs. Earmark projects will be included in the review and treated on the same basis as other activities.
- *Frequency of Review.* All EERE programs and their key projects will be reviewed, on average, every two years, depending on the characteristics of the program and needs for information.
- *Timely Preparation*. Preparation for a peer review will include designation of a review leader, determination of the purpose of the review and the review agenda, and communication of this information to reviewers and those being reviewed in time for them to prepare for the review.
- *Core Evaluation Criteria*. Clear standards for judging the program or projects will be defined prior to the review. This includes the criteria and the kinds of evidence (data) needed to judge those criteria. At a minimum, programs will be assessed on quality, productivity, and accomplishments; relevance of program success to EERE and programmatic goals; and management.
- *Reviewers*. There will be a minimum of three reviewers for each discrete program element or smallest unit that is assessed and reported on. Each reviewer will be independent, competent, and objective, selected by a transparent, credible process that involves external parties. Together the reviewers will cover the subject matter. Reviewers will sign Conflict of Interest forms prior to the review and Nondisclosure Agreements if/when proprietary information is presented or discussed.
- *Plan for Collecting Reviewer Data*. Review leaders will plan ahead for how review inputs will be documented, analyzed, and reported, as well as how individual reviewer comments will be tracked while maintaining their public anonymity. The review agenda will allow sufficient time for a rigorous Question & Answer period for reviewers. Reviewers will be encouraged to support their comments with citations or data wherever possible.
- *Producing the Peer Review Report*. The peer review report will reflect the full range of reviewer comments with high fidelity. The report should also include all individual inputs from the reviewers and will be reviewed by the panel chair and/or the review panel before release.
- *Program Manager Review and Response*. Before the report is finalized and goes to senior management, the program manager/office director will add written responses to peer reviewer findings and recommendations, including actions to be taken to improve the program.
- *Peer Review Report Distribution.* The final peer review report will be promptly communicated to senior management, associated staff and researchers involved with the R&D program or project, and all persons involved in the review, and the report will be made available publicly.

- *Peer Review Record and Ex-post Evaluation.* A peer review record will be established at the beginning of, and maintained throughout, the review process. The record should contain the final form of all the key documents of the review for all phases of the review. An evaluation of the peer review process is necessary to aid continuous process improvement.

Based on the variety of experiences, examining the peer review literatures, and managing hundreds of peer reviews, Kostoff (2004, 2003, 2001, 1997, 1995) concludes the followings as *the factors critical* to high-quality peer review for programme evaluation: 1) Senior management's commitment is the most important factor in the quality of an organisation's S&T evaluations; 2) The second most important factor is the operational manager's motivation to perform a technically credible evaluation; 3) The third most important is transmission of a clear and unambiguous statement of the review's objectives (and conduct) and its potential impact and consequences to all participants; 4) Fourth most important factor is the quality of the technical evaluators themselves, specifically their role, objectivity, and competency. This fourth factor consists of the evaluation *experts' competence and objectivity;* 5) The fifth important factor is selection of evaluation criteria. These criteria will depend on the interests of the audience for the evaluation, the nature of the benefits and impacts, the availability and quality of the underlying data, the accuracy and quality of results desired, the complementary criteria available and suites of diagnostic techniques desired for the complete analysis, the status of algorithms and analysis techniques, and the capabilities of the evaluation team; 6) Every S&T metric, and its associated data, should answer a question that contributes to forming the basis for a decision; 7) The reliability and repeatability of an evaluation is also crucial. To minimize repeatability problems, a diverse and representative segment of the overall competent technical community should be involved in the construction and execution of the evaluation; 8) A sound evaluation processes should in general be seamlessly integrated into the organisation's business operations. Evaluation processes should not be incorporated in the management tools as an afterthought (which is typical practice today), but should be part of the organisation's front-end design; 9) Data awareness is also important. Placing the technology of interest in the larger context of technology development and availability world-wide is absolutely necessary; 10) For evaluations that will be used as a basis for comparison of S&T programs or projects, the next most important factor is normalization and standardization across different S&T areas; 11) Secrecy is as important as normalization: reviewer anonymity and reviewer non-anonymity. "Blind reviewing" has been used for the noble purposes of providing fairer; 12) Cost is also a critical factor for quality of S&T evaluation; 13) The final critical factor, and perhaps the foundational factor in any high quality S&T evaluation, is the maintenance of high ethical standards throughout the process.

### *5.2 Principles and Suggestions for Successful Expert Review*

Some principles or policy recommendations have been suggested for successful expert review (Bozeman, 1993; Rigby, 2002; Ormala, 1989; EERE, 2004; Kostoff, 2004, 2003, 2001, 1997, 1995; Nightingale & Scott, 2007; Moed, 2007; Donovan, 2007; The British Academy, 2007; ESPRC, 2008; Noble, 1974; Gillespie et al., 1985; Bodden, 1982; Porter and Rossi, 1985; GACR, 2007 etc.). Although except for some literature including Bozeman (1993), Kostoff's papers, and EERE (2004), most of the suggestions and principles pertain to the selection of research topic and the publication of scientific papers, these may also be very useful for improving the policy-level or programme-level expert review process. For example, EPSRC of the UK suggests some good peer review principles for reviewing research proposals (See box 3). *OMB of the US provides another example. OMB Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information* (2001) set as general criteria for competent and credible peer review the following: (a) peer reviewers should be selected primarily on the basis of necessary technical expertise,

(b) peer reviewers should be expected to disclose to agencies prior technical/policy positions they may have taken on the issues at hand, (c) peer reviewers should be expected to disclose to agencies their sources of personal and institutional funding (private or public sector), and (d) peer reviews should be conducted in an open and rigorous manner.

What are key principles to high-quality programme/policy expert review? The section below suggests a list of principles and suggestions for good practice of programme/policy expert review based on a number of expert review literatures.

***Principle 1. The philosophy, the focus, the future uses of an evaluation must be understood and agreed upon by the relevant stakeholders in advance.*** This is the foremost principle that applies to not only expert review but to every type of evaluation. Since expert review involves outsourcing to third-party experts by the evaluation manager, there is all the more need for the third-party experts to fully understand and agree upon the reasons for evaluation, the methods and principles guiding the evaluation as well as the utility of the evaluation.

- *High-level policy makers or evaluation managers should clearly define the roles of each actor in the evaluation process and regularly monitor performance.* In all areas of public management, highest-level manager's encouragement and continuous interest in a task is a key factor for success. The evaluation task is more complicated than other tasks, not only because there may be a conflict of interest between the evaluators and the evaluated but also because it involves third-party experts. It is therefore indispensable for the high-level manager of an evaluation to clearly define the roles of each actor and to make sure that agreement is reached among the relevant actors concerning the objective as well as the philosophy of the evaluation well in advance.
- *Provide pre-evaluation training program for the relevant actors.* The training of relevant actors, that is, experts, evaluation staff (the secretariat), the evaluated (e.g. programme managers), processes and criteria could enhance the efficiency, the effectiveness as well as the receptiveness of evaluation.
- *Before evaluation, select objective and useful evaluation criteria.* Evaluation criteria are important as they determine the focus as well as the scope of the evaluation. Consequently, it is imperative to provide clear evaluation criteria before embarking upon the evaluation.

***Principle 2. Qualified experts should be selected as evaluators.*** The quality of the evaluators is dependent on the *professional competence and objectivity of the experts participating in the evaluation.* The panel chair and other experts should all have high professional competence in the areas in which they are required to make judgments in order to instil confidence in the stakeholders of the evaluation.

- *In addition to technological experts, seek experts from diverse domains, including experts in social sciences and the economy.* To make judgements on program's rationality as well as its socio-economic value, in addition to technological experts, it is desirable to have a group of experts from diverse domains, including economy, business management. This is a very important element in policy or programme evaluation
- *Suggestion. Build a sufficiently large database of experts.* For this, there needs to be regular monitoring of research personnel in various research institutions and universities. The data which should be collected on the personnel through monitoring are: past research experience, current research interests, field, affiliation, degree-granting institution, participating academic organisations and other detailed academic activities. These data allows one to infer what area of policy a particular researcher could evaluate based on their qualifications and what contributions they can provide if chosen as an evaluator.

***Principle 3. The risk of bias or conflict of interests should be reduced as much as possible.***

- *Provide a bias statement for reviewers*. That is, make that experts declare their interests to ensure that the panel's reputation for fairness is upheld. In principle, the evaluation manager should not appoint an evaluator who has a vested interest in the evaluated policy/programme or the evaluated institution. To ensure neutrality of the expert panel, make experts declare their interests relevant to the evaluation.
- *Avoid "internal evaluators."* Often, it is useful to include opinions of experts from another field/ region or to have them in the panel. In particular, if there is no language-barrier or no additional cost, it is desirable to include foreign experts in the evaluation. In the case foreign experts are not well aware of the socio-economic conditions of the country in question, it is advised that they focus on the scientific or the technical aspects of the programme.
- *Limit the number of evaluations or the duration of evaluating activities for the participating experts.* If an expert participates in too many evaluations, the expert might develop a relationship with the evaluated bodies and might be susceptible to lobbying from them. Moreover, trapped in their judgments from previous evaluations, they might not be able to take a fresh look at similar programmes. However, to limit the duration of evaluating activities excessively may be counter-productive. For example, it might decrease their sense of responsibility. If the expert participates in the evaluation only once, their responsibility might be lower than if they were to participate in future evaluations. Therefore, one may consider an appropriate duration of evaluating activities for the experts.
- *Prevent an expert who has expertise only in a particular domain from judging the quality or the value of what is being evaluated*. However, it is advisable that his/her opinions are transmitted to other experts in the panel to be used in the joint-decision making. For example, an expert on biotechnology evaluating a biotechnological programme might know more about the technical aspects than other experts but may not be able to judge correctly the socio-economic value of the evaluated programme.

***Principle 4. The review should be conducted in a credible, fair, transparent manner with the highest degree of ethical standards.***

- *Provide transparency in evaluation process and in evaluation results*. Introduce transparency in evaluation principles, criteria, processes and make them accessible to all the actors and stakeholders in the evaluation so that they could prepare for the evaluation properly. After the evaluation, release evaluation results on-line, except those that may be confidential for national security reasons, and let those results be accessible to the evaluated bodies as well as the general public.
- *Maintain high ethical standards*. To ensure that evaluators are free from personal bias, there is the option of requiring bias statements. Such declarations or statements may include clauses on overcoming personal biases as well as on the prohibition of any misuse of information obtained during the evaluation process, such as the use of such information for personal reasons or the release of such information without the permission of the relevant authority.

***Principle 5. The review should be based on objective evidence and information.***

- *Provide, in advance, sufficient information on the evaluated policy/program to the evaluators.* Judgement of the expert panel depends on their comprehension of given information. Therefore, providing sufficient information is as important as selecting qualified experts and one should not ask the experts to provide judgements which go beyond the scope of the provided information. For policy decision making, adjustment of programs, resources allocation, priority setting, it is useful to employ 3P analysis (positioning analysis, portfolio analysis, performance analysis).
- *If indicators or rating are used, test the validity and reliability of those indicators.* Indicators are

important tools to ensure the objectivity of the evaluation. Therefore, before inferring evaluation results, it is necessary to have the evaluators go over the relevance and the reliability of indicators. In case modifications of one or more indicators are necessary, sufficient reason should be provided and such modification must be communicated to the evaluated bodies and be subject to their approval.

- *Encourage a maximum amount of dialogue and discussion.* Ideas arising from discussions and dialogues as well as mutual-learning experience are the biggest advantages of expert review. Therefore, along the evaluation process, encourage as much discussion and dialogues among evaluators and the evaluated (programme managers) as possible. At this time, foregoing the introduction of experts from other domains and expanding the panel as much as possible would be an effective way to encourage productive discussions as well as the creation of new ideas.

***Principle 6. "One size does not fit all."***

- *Complement expert review with quantitative methods to increase objectivity and scientific reliability of the evaluation.* One may increase objectivity and accuracy of expert review evaluation by complementing it with quantitative methods such as bibliometrics or econometrics. Often, expert review is considered a particular way of research evaluation, evaluators often utilise case studies, benchmarking, surveys and other evaluation methods. Therefore, it is difficult to categorise expert-review as a separate evaluation method; it should be understood as a decision-making process that is complemented by various evaluation methods.
- *Seek the type of expert review appropriate for the particular programme/policy.* The review should be tailored to the aim of evaluation and the characteristics of the subject of evaluation. In other words, evaluation methods and results should be differentiated according to evaluation objectives. For example, if the primary objective is to set priorities, a scoring method could be used. If the improvement of a programme is the primary objective, opinions of expert review would be very important. Also, evaluation processes and the form of the final results should be tailored to the particularities of programmes.

***Principle 7. Evaluation efficiency may be increased through various measures.***

- *Increase remote evaluation.* Today, with the development of the Internet, there are many cases in which evaluating institutions distribute IDs and Passwords to the evaluators with which the evaluators may access data and submit reports online. To promote expert review efficiency, devise different technologies to enable evaluators to participate in the evaluation process from a distance.
- *Build and operate evaluation management systems through internet-based technologies.* Evaluation management systems (EMS) should cover fundamental information on the evaluation, including information on the pool of experts, evaluation data, evaluation principles, evaluation protocol, relevant analytic data, and evaluation results. To increase the utilization of EMS, evaluating organisms, evaluators and evaluation object should be able to use it freely. Admittedly, certain restrictions for security reasons may be introduced when necessary
- *Minimize the part of the evaluation cost born by the subject of the evaluation.* Especially, simplify administrative procedures and evaluation formats. In many cases, the evaluation subject is asked by the evaluators to provide administrative information irrelevant to the evaluation as well as unnecessary information. It is advisable to have the evaluation body to extract excessively complicated forms or unimportant information from the basic database and provide these to the evaluators themselves in order to reduce the administrative burden of the evaluates. In addition, having the evaluators provide clear reasons when they ask for additional information from the

evaluation object may enhance cooperation on the part of the evaluation object.

- *Design evaluation processes while paying attention to hidden indirect costs*. In expert review, the indirect costs such as the billable hours spent the evaluators are more important than direct costs such as venue and travel expenses.
- *Improve design of expert panels*. To ensure continuity, it is advisable to appoint someone who has participated in previous panels as the head of a panel. It is advisable that one-third to one-half of a review panel be carried over from one review to the next in order to enhance "new perspectives" as well as continuity.

## 6. Concluding Remarks

In the literature and in various workshops, there appears to be a common view emerging. First, despite problems in expert review including hollowing-out due to time constraints, rising financial costs, and the risk of conflicts of interest among expert reviewers, the expert review process remains as a fundamental mechanism for all stages of research planning and implementation as well as for both ex ante project selection and for ex post evaluation. Second, solutions are available to improve expert review process, including: making the process more transparent, providing clear objectives and guidelines to reviewers, using different tools (e.g. extended expert review processes involving non-scientific stakeholders) and using a variety of metrics and indicators. Thirdly, while indicators can strengthen and inform judgements, they do not form judgements by themselves. Making judgement still requires careful consideration to prevent perverse outcomes. Fourthly, there is a need to facilitate and improve the internationalisation of expert review because of increased international collaboration. However, caution is needed when policy is transferred across different cultures and contexts, particularly when understandings of policies are incomplete. Therefore, there is a need for a taxonomy of the internationalisation of expert review. Finally, one size does not fit all and hence a much better understanding of the design requirements for expert review is needed.

There is another important principle to be added to the emerging views above. A perfect evaluation system cannot exist; one must adapt the evaluation system to the environment. What was ideal in the past may be found to be no longer effective in the future. Also, when improving evaluation systems, it is important to take into account not only the opinions of the evaluation managers but also the opinions of the evaluated. This is because it is easier to find problems as well as solutions when one looks at the question from the perspective of a client. Admittedly, the opinion of the evaluated will differ according to the evaluation results. It is therefore important to strike a proper balance by considering the opinions of policy makers. For this, it is advisable that evaluation institutions carry out regular opinion surveys targeting evaluation participants and evaluates on expert review.

Because expert review involves more individual judgments than any other methods of evaluation, much like the functioning of an orchestra, co-operation among the conductor (review manager), the players (experts) and the audience (stakeholders) is necessary for a successful performance (evaluation). This may be said to be the foremost principle that should be respected for a successful expert review.

## References

Averch, H. (1990), Policy Uses of 'Evaluation of Research' literature, OTA Contractor Report, July 1990.

Barker, K. (1994), R&D Evaluation – Strengthening the impact of R&D evaluation on policy making: methodological and organisation considerations, Research Evaluation, 21(6): 405-413.

Becher, Gerhard and Stefan Kuhlmann eds. (1995), Evaluation of Technology Policy Programmes in Germany, Kluwer Academic Publishers.

Bobe, B. & Viala, H. (1997), Panels of experts of the European Commission in Callon, M., Laredo, P. & Mustar, P. (eds.) The Strategic Management of Research and Technology – Evaluation of Programmes, Economica: Paris.

Bozeman, B. & Melkers, J. (eds.) (1993), Evaluating R&D Impacts: Methods and Practice, Kluwer: Boston.

Bozeman, Barry (1993), Peer Review and Evaluation of R&D Impacts, in Bozeman, B. and J. Melkers (eds.), Evaluating R&D Impacts: Methods and Practice. Massachusetts: Kluwer Academic Publishers.

Branscomb, L., K. Morse, and M. Roberts (2000), Managing Technical Risk: Understanding Private Sector Decision Making on Early Stage Technology-Based Projects, NIST GCR 00-787, Gaithersburg, MD, April 2000.

Butler, Linda and Ian McAllister (2007), Metrics or Peer Review? Evaluating the 2001 UK Research Assessment Exercise in Political Science, http://repp.anu.edu.au/papers/2007_ukresearchassess.pdf

Campbell, P. (2001), Opinion: Introducing a new policy for authors of research papers in Nature and Nature journals, Nature 412, 751 (2001).

Campbell, Philip (2006), Pressures on peer review at Nature and Nature Journals, Presentation in Prague Conference on Peer Review: Its Present and Future State. Oct. 2006.

Chráska, Pavel and Josef Syka (2006), Peer Review Process in the Czech Science Foundation – Experiences and Challenges, Presentation in Prague Conference on Peer Review: Its Present and Future State. Oct. 2006.

Chubin, D. and S. Jasanoff (1985), Peer Review and Public Policy, Science, Technology and Human Values, 10(3): 3-5.

Chubin, D. E. (1994), Grants Peer Review in Theory and Practice, Evaluation Review, 18(1): 20-30.

Chubin, Daryl and Edward Hackett (1990), Peerless Science: Peer Review and US Science Policy. Albany: SUNY Press.

COSEPUP (1999), Evaluating Federal Research Programs: Research and the Government Performance and Results Act. Washington DC: National Academy Press. Available at http://www.nap.edu/catalog/6416.html.

Cozzen, S. E. (1987), Expert Review in Evaluating Programs, Science and Public Policy 14, April.

Dixon, B. (1987), Follow the Finnish Lead in Peer Review, The Scientist, May 4.

Donovan, Claire (2007a), Introduction: Future pathways for science policy and research assessment: metrics vs. peer review, quality vs. impact, Science and Public Policy 34(8), October 2007.

Donovan, Claire (2007b), The qualitative future of research evaluation, Science and Public Policy 34(8), October 2007.

EERE (2004), EERE Peer Review Guide: Based on a Survey of Best Practices for In-Progress Peer Review, August 2004.

Environmental Protection Agency (1997), Peer Review and Peer Involvement at the U.S. Environmental Protection Agency. Available at http://www.epa.gov/ordntrnt/ord/spc/perevmem.htm.

EPA (1998), Science Policy Council Peer Review Handbook. Office of Science Policy. EPA 100-B-98-00. Washington, DC: EPA.

EPA (2000), Peer Review Handbook, Science Policy Council.

EPSRC (2008), Peer Review Principles, http://www.epsrc.ac.uk

ESF MO Forum workshop on Peer Review Website (2008): http://www.esf.org/activities/mo-fora/peer-review.html

Forsdyke, D. R. (1991), Bicameral grant review: an alternative to conventional peer review, FASEB J. 5: 2312-2314.

Forsdyke, D. R. (1993), Bicameral grant review: how a systems analyst with AIDS would reform research funding, Accountability in Research 2: 237-241.

Forsdyke, D. R. (1993), Bicameral grant review: how a systems analyst with AIDS would reform research funding, Accountability in Research, 2: 237-241.

Forsdyke, D. R. (1993), On giraffes and peer review. FASEB.J. 7: 619-621.

GACR (2007), Peer Review: Its Present and Future State, Czech Science Foundation.

Geisler, E. (2000), The Metric of Peer Review, in E. Geisler, ed., The Metrics of Science and Technology. Westport, CT: Quorum Books.

General Accounting Office (1994), Peer Review: Reforms Needed to Ensure Fairness in Federal Agency Grant Selection, GAO/PEMD-94-1.

General Accounting Office (1999), Federal Research: Peer Review Practices at Federal Science Agencies Vary, GAO/RCED-99-99.

Gibbons, M. and L. Georghiou (1987), Evaluation of Research: A Selection of Current Practices. Paris: OECD.

Gieryn, Thomas F. (1995), Boundaries of Science. pp. 393-443 in Sheila Jasanoff, Gerald E. Markle, James C. Petersen, and Trevor Pinch, eds., Handbook of Science and Technology Studies. Thousand Oaks, CA: Sage Publications.

Gillespie, G. W. Jr., D. E. Chubin, and G. M. Kurzon (1985), Experience with NIH Peer Review: Researchers' Cynicism and Desire for Change, Science, Technology and Human Values, 10(3): 44-53.

Godlee, F. & Jefferson, T. (1999), Peer Review in the Health Sciences, BMJ Books: London.

Goodstein, T. (1995), Ethics and Peer Review, Biotechnology,

13(6): 618.

Grigson, D. & Stokes, T. (1993), Use of peer review to evaluate research outcomes, Research Evaluation, 3(3): 173-177.

Grol, R. and M Lawrence, eds. (1995), Quality Improvement by Peer Review. New York: Oxford University Press.

Gustafson, T. (1975), The Controversy over Peer Review: Recent Studies of the Peer Review System Show that its Critics have yet to Make their Case, Science, 190: 1060-1066.

Guston, David H. (2000a), Expanding the Role of Peer Review Processes in the United States, paper presented to the US-European Policy Workshop on Learning from Science and Technology Policy Evaluation, Bad Herrenalb, Germany.

Guston, David H. (2000b), Regulatory peer review. Paper presented at the annual meeting of the American Political Science Association, Washington, DC, 2 September.

Guston, David H. (2000c), What is Peer Review? Report submitted to the American Chemistry Council, May.

Hackett, E. J. (1997), Peer Review in Science and Science Policy, in Evaluating Science and Scientists edited by M. S. Frankel and J. Cave. Budapest/Oxford, Central European University Press/OUP.

Hartmann, I. and F. Neidhardt (1990), Peer Review at the Deutsche Forschungsgemeinschaft, Scientometrics, 19(5-6): 419-425.

Klahr, D., Insiders, Outsiders, and Efficiency in a National Science Foundation Panel, American Psychologist, 40(2): 148-154. Kostoff, Ronald N. (1993), Evaluating Federal R&D in the United States, in Bozeman, B. and J. Melkers (eds.), Evaluating R&D Impacts: Methods and Practice. Massachusetts: Kluwer Academic Publishers.

Kostoff, R. (1998), GPRA Science and Technology Peer Review

Kostoff, Ronald N. (1993), Evaluating Federal R&D in the United States, in Bozeman, B. and J. Melkers (eds.), Evaluating R&D Impacts: Methods and Practice. Massachusetts: Kluwer Academic Publishers.

Kostoff, Ronald N. (1997a), Research Program Peer Review: Principles, Practices, Protocols, www.dtic.mil/dtic/kostoff/index.html.

Kostoff, Ronald N. (1997b), The Principles of Practices of Peer Review, Science and Engineering Ethics, 3(1): 19-34.

Kostoff, Ronald N. (2003), Science and Technology Peer Review: GPRA, Office of Naval Research.

Kostoff, Ronald N. (2004), Research Program Peer Review: Purposes, Principles, Practices, Protocols, Office of Naval Research.

Kruytbosch, Carlos (1989), The Role and Effectiveness of Peer Review, Paper presented at the Ciba Foundation Conference on Research Evaluation.

Lee, M., K. Om and J. Ko (2000), Theory and Practice of Peer Review. Seoul: Pidi.

Luukonen-Gronow, T., (1987), Scientific Research Evaluation: A Review of Methods and Various Contexts of Their Application, R&D Management 17.

Massimo, L. (1997), Lessons from the experience of the European Commission in Callon, M., Laredo, P. & Mustar (eds.) The Strategic Management of Research and Technology – Evaluation of Programmes, Economica: Paris.

McCullough, J. (1989), First Comprehensive Survey of NSF Applicants Focuses on their Concerns about Proposal Review, Science, Technology and Human Values, 14(1989): 78-88.

Mehlhorn, Kurt (2006), Peer Review in the Digital Age: Opportunities and Examples. Presentation in Prague Conference on Peer Review: Its Present and Future State. Oct. 2006.

Merton, R. K. (1973), The Sociology of Science. University of Chicago Press.

Moed, Henk F. (2007), The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review, Science and Public Policy 34(8), October 2007.

National Science and Technology Council (1996), Assessing Fundamental Research. Committee on Fundamental Science, Subcommittee on Research. Accessed online at http://www.nsf.gov/sbe/srs/ostp/assess/start.htm.

National Science Foundation (1999), What Is the State Government Role in the R&D Enterprise? Division of Science Resources Studies: 99-348. Arlington, VA: NSF.

Nature Peer Review Debate: http://www.nature.com/nature/peerreview/debate

Nightingale, Paul and Alister Scott (2007), Peer Review and the Relevanc Gap: 10 Suggestions for Policy-makers, Science and Public Policy 34(8), Oct. 2007.

Noble, J. H. (1974), Peer Review: Quality Control of Applied Social Research, Science, 185: 916-921.

NRC (1999), Peer Review in Environmental Technology Development Programs, Washington DC: National Academy Press. Available at http://www.nap.edu/catalog/6408.html

OECD (1998), Policy Evaluation in Innovation and Technology: Towards Best practices, Paris: Organisation for Economic Co-

operation and Development.

OECD (2006), Evaluation of Publicly Funded Research: Recent Trends and Perspectives, DSTI/STP/TIP(2006)3.

OECD (2007a), Evaluating Government R&D Programmes: Lessons from Korean Experience, DSTI/STP/TIP(2007)7.

OECD (2007b), Revised Proposal for the TIP Work on Evaluation of innovation policies in OECD Countries, DSTI/STP/TIP(2006)17REV1.

OECD (2007c), Peer Review: Its Uses, Demands and Issues, DSTI/STP/TIP(2007)13.

OECD (2007d), Summary of the workshop on Rethinking Evaluation in Science and Technology, DSTI/STP/TIP(2007)14.

OECD (2007e), TIP activity on the Evaluation of research and innovation policies: current Status and next Steps, DSTI/STP/TIP(2007)15.

OECD-BMBF Conference (2005), http://www.internationales-buero.de/de/2193.php

Oh, Donghoon and YoungJun Kim (2006), Overall Coordination of Government-Funded Research and Development Programmes in Korea, Journal of MultiDisciplinary Evaluation, 5 (September 2006): 153-170.

Ormala, E. (1989), Nordic Experiences of the Evaluation of Technical Research and Development, Research Policy, 18: 333-342.

Porter, A. & Rossini, F. (1985), Peer Review of Interdisciplinary Proposals Science, Technology and Human Values, 10(3): 34-42.

Pouris, A. (1988), Peer Review in Scientifically Small Countries, R&D Management, 18(4): 333-340.

Prague Conference (2006), Peer Review: Its Present and Future State, Available at http://www.prague forscience.cz/Scientific-Programme.php

Rigby, John (2002), Expert Panels and Peer Review, Fahrenkrog, Gustavo, Wolfgang Polt, Jaime Rojo, Alexander Tubke, and Klaus Zinocker eds., RTD Evaluation Toolbox: Assessing the Socio-Economic Impact of RTD-Policies. IPTS Technical Report Series, EUR 20382 EN.

Rossi, Peter H., Mark W. Lipsey, and Howard E. Freeman (2004), Evaluation: A Systematic Approach, Thousand Oaks: SAGE Publications.

Roy, R. (1985), Funding Science: The Real Defects Of Peer-Review And An Alternative To It, Science Technology & Human Values 52: 73-81.

Roy, Rustum. (1984), Alternatives To Review By Peers - A Contribution To The Theory Of Scientific Choice, Minerva 22 (3-4): 316-328.

Ruegg, Rosalie and Gretchen Jordan (2007), Overview of Evaluation Methods for R&D Programs, Office of Energy Efficiency and Renewable Energy, DOE.

Ruegg, Rosalie and Irwin Feller (2003), A Toolkit for Evaluating Public RTD Investment, NIST GCR 03-875, National Institute of Standards and Technology.

Scarpa, Toni (2006), Challenges and Opportunities Facing NIH Peer Review: A Vision for Ensuring Its Strategic National Value," Presentation in Prague Conference on Peer Review: Its Present and Future State. Oct. 2006.

Scott, Alister (2006), Peer Review and the Relevance of Science, SPRU Electronic Working Paper Series, No. 145.

Shapira, Philip and Stefan Kuhlmann eds. (2003), Learning from Science and Technology Policy Evaluation: Experiences from the US and Europe. Edward Elgar, Cheltenham.

The British Academy (2007), Peer Review: the challenges for the humanities and social sciences, the national academy for the humanitities and social sciences.

Wood, Fiona and Simon Wessley (2007), Peer Review of Grant Applications: A Systematic Review, in Fiona Godlee, Tom Jefferson eds. (2007), Peer Review in Health Sciences. BMJ Books.

Wood. F. (1997), The peer review process. Canberra: Australian Research Council.

Worthen, Blaine R., James R. Sanders, and Jody L. Fitzpatrick (1997), Program Evaluation: Alternative Approaches and Practical Guidelines, Longman Publishers USA.