The use of peer review as an evaluative tool in science

Ilse Eigelaar

Thesis presented in partial fulfilment of the requirements for the degree of Master of Philosophy at the University of Stellenbosch, South Africa

Supervisor: Prof. Johann Mouton
December 2001

Declaration

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Abstract

Peer review as an institutional mechanism for certifying knowledge and allocating resources dates back as far as 1665. Today it can with confidence be stated that it is one of the most prominent evaluative tools used in science to determine the quality of research across all scientific fields.

Given the transformation within the processes of knowledge production, peer review as an institutionalised method of the evaluation of scientific research has not been unaffected. Peer reviewers have to act within a system of relevant science and find themselves responsible to the scientific community as well as to public decision-makers, who in turn are responsible to the public. This dual responsibility of reviewers led to the development of criteria to be used in the evaluation process to enable them to measure scientific excellence as well as the societal relevance of science.

In this thesis peer review in science is examined within the context of these transformations. In looking at the conceptual and methodological issues raised by peer review, definitions of peer review, its history and contexts of application are examined followed by a critique on peer review. Peer review in practice is also explored and the evaluation processes of four respective funding agencies are analysed with regards to three aspects intrinsic to the peer review process: the method by which reviewers are selected, the review criteria by which proposals are rated, and the number of review stages within each review process. The thesis concludes with recommendations for possible improvements to the peer review process and recommended alternatives to peer review as an evaluative tool.

Opsomming

Portuurgroep-evaluering as 'n geïnsitutsionaliseerde meganisme in die sertifisering van kennis en die toewys van hulpbronne dateer terug so ver as 1665. Huidiglik kan dit as een van die mees prominente metingsinstrumente van die kwaliteit van navorsing in alle wetenskaplike velde beskou word.

Die transformasies wat plaasgevind het binne die prosesse waar kennis gegenereer word, het ook nie portuurgroep-evaluaring as 'n geïnstitusionaliseerde metode van evaluering ongeraak gelaat nie. Portuurgroep-evalueerders bevind hulself binne 'n sisteem van relevante wetenskap. Binne hierdie sisteem het hulle 'n verantwoordelikheid teenoor die wetenskaplike gemeenskap sowel as die publieke-besluitnemers wat op hul beurt weer verantwoordelik is teenoor die publiek. Hierdie dubbele verantwoordelikheid het tot gevolg die saamstel van kriteria waarvolgens evalueerders wetenskaplike uitmuntendheid sowel as relevansie tot die breër samelewing kan meet.

Hierdie tesis ondersoek portuurgroep-evaluering teen die agtergrond van hierdie transformasies. Die konseptuele en metodologiese aspekte van portuurgroep-evaluering word ondersoek deur eerstens te kyk na definisies van portuurgroep-evaluering, die geskiedenis daarvan en kontekste waarbinne dit gebruik word. Tweedens word gekyk na kritiek gelewer op portuurgroep-evaluering. Portuurgroep evaluering binne die praktyk word ook ondersoek waar vier onderskeie befondsingsagentskappe se evaluerings prosesse geanaliseer word. Hierdie analise word gedoen in terme van drie essensiële aspekte binne portuurgroep- evaluering. Hierdie drie aspekte is as volg: 1) die wyse waarop evalueerders geselekteer word, 2) die evalueringskriteria waarvolgens navorsingsvoorstelle gemeet word en 3) die hoeveelheid evalueringsfases binne die protuurgroep-evaluerings proses. Laastens word aanbevelings ter verbetering van die portuurgroep-evaluerings proses as ook voorstelle tot moontlike alternatiewe tot portuurgroep-evaluering as 'n evaluerings instrument gebied.

Acknowledgements

First of all I have to thank God for the opportunity as well as the ability to undertake this thesis. To my parents for supporting me all the way and who made this possible as well as the NRF for their financial support. A special thank you to Maryke Hunter and Zoë Dowling for the inspiration when I had none of my own and help when my fund of ideas ran dry. I would also like to thank Enver Ravat for his help in obtaining data for one of my chapters. And then of course to my study leader, Prof. Johann Mouton, who directed my thoughts and from whom I learned more than could ever be reflected in this thesis.

Table of Contents

List of Tables and Figures

CHAPTER 1

Peer review: Changes within a dynamic system

Various authors such as Gibbons (1994) and Weingart (1997) have recently claimed that a transformation within the processes of knowledge production has taken place. Although these authors agree on the transformation of the knowledge production processes, they differ in their descriptions of the scope of the changes that have taken place. These differences are encapsulated in the different terms the authors apply to this transformation. Gibbons speaks about the different modes of knowledge production and argues that there has been a total transition from Mode 1 to Mode 2 knowledge production. "Post-normal science" authors such as Funtowicz and Ravetz argue for a qualitative transformation in science. This qualitative transformation will impact on the definition of subject matter, methods, and social functions of science. Weingart, on the other hand, argues that these new forms of knowledge production only pertain to a specific section of the research system and cannot be generalised to science as a whole (Weingart, 1994). The aim of this chapter is to show that these transformations within the processes of knowledge production have a significant impact on the character of peer review as an evaluative practice.

The first part of this chapter will be devoted to a discussion of the changing character of science by exploring the different regimes of science. The debate on the different forms of knowledge production, that is the changing from Mode 1 to Mode 2 forms of knowledge production, will be discussed as an aspect of this change. In the second section of this chapter I will argue that the changing nature of peer review is closely related, if not directly linked, to the changes in modes of knowledge production.

1.1 Science as a dynamic process

Since the Second World War science has been viewed as a crucial factor in the development of a country's social welfare. President Franklin Roosevelt, in closing his letter to Dr Vannevar Bush, then Director of the Office of Scientific Research and Development, wrote: "New frontiers of the mind are before us, and if they are pioneered with the same vision, boldness, and drive with which we have waged the war we can create a fuller and more fruitful employment and a fuller and more fruitful life" (Bush, 1980:4).

In turn, Bush wrote the following to the president in his final report: "Advances in science when put to practical use mean more jobs, higher wages, shorter hours, more abundant crops, more leisure for recreation, for study, for learning how to live without the deadening drudgery which has been the burden of the common man for ages past. Advances in science will also bring higher standards of living, will lead to the prevention or cure of diseases, will promote conservation of our limited national resources, and will assure means of defense against aggression. But to achieve these objectives – to secure a high level of employment, to maintain a position of world leadership – the flow of new scientific knowledge must be both continuous and substantial" (Bush, 1980:1).

The above correspondence clearly shows the almost devout belief in science as the ultimate power and in its being the only source of development. Scientists were believed to have the power in their hands, owing to the knowledge at their disposal, to ensure a brighter and better future for all. Scientists had the responsibility to build a powerful country – more powerful than any other country - in order to ensure global political and economic power.

Today the focus is on relevant science. Science that has societal value and is proactive in the sense that it addresses economic and environmental problems. Science as a process of knowledge production has developed over time, driven by different social forces within society. In the following section we will look at the different "regimes" of science, each with its own characteristic foci and roleplayers. Regimes here refer to different "time frames" or rather phases within the development process of science with regard to its responsibility or use as we conceptualise it today.

1.2 From "Science the endless frontier" to "Strategic science"

During the regime of "Science the endless frontier" government was seen to have the dual responsibility of promoting the flow of new scientific knowledge and developing scientific talent among young people. Control of the types of inquiry, policy, personnel, method and scope of research was left to internal controls by the research institutions themselves. "It [the state] should recognize that freedom of inquiry must be preserved and should leave internal control of policy, personnel, method and scope of research to the institutions in which it is carried on. It should be fully responsible to the President and through him to Congress for its program" (Bush, 1980:9). Science was believed to be a closed entity that, in order to ensure its development and survival, had to manage itself free from governmental and societal interference. The author of *The Republic of Science*, Michael Polanyi, writes that "any attempt to organize the group of helpers [scientists] under a single authority would eliminate their independent initiatives and thus reduce their joint effectiveness to that of the single person directing them from the centre" (Shils, 1968:3).

The focus during the regimes of "Science the endless frontier" and "The republic of science" was on the development of basic science with as little interference as possible and as much support as possible (mainly financial) by the government. "Any attempt at guiding scientific research towards a purpose other than its own is an attempt to deflect it from the advancement of science...You can kill or mutilate the advance of science, you cannot shape it. For it can advance only by essentially unpredictable steps, pursuing problems of its own, and the practical benefits of these advances will be incidental and hence doubly unpredictable" (Shils, 1968:9).

According to this perspective funding agencies, although formally part of the government, were in truth "governed" by the scientific community. Under the regime of "Science the endless frontier", the state as final sponsor delegated most of its tasks to the funding agency as an intermediary sponsor, and the funding agency delegated refereeing and ranking to the scientists. Scientific peers were equated with competent referees. In a sense the scientific community captured the funding agencies while they were formally part of the government. The need for accountability to the final sponsor has always been important for politicians and government officials (as was clear in their original reluctance to delegate). This, however, was short-circuited by the argument that peer review was selecting the best research and this was what the final sponsor wanted, or at least should want (Rip, 2000).

The above view of science came under heavy attack during the 1970s with an ever louder call for relevant science from governments and a more knowledgeable society. Science now became depicted as a problem solver, functioning within a context of industrial competitiveness with the aim of economic growth. These calls for relevant science as well as an increasingly proactive science policy were combined with hopes for the potential of the scientists to overcome the economic and environmental problems evident in the 1980s. This crystallised into a new regime of science in the late 1980s and early 1990s - the regime of "Strategic science" (Rip, 2000).

The call for relevant science and a proactive science policy, together with the realisation of the responsibility of science towards economic and environmental problems, were only some of the forces driving science towards the new regime of "Strategic science". In his book *Perspectives on Science Policy in South Africa* Marais (2000) refers to another five forces that had a part in this change in thinking about science as well as in its relationship with government. The first force was the democratisation of institutional management, which resulted in a devolution of decision-making as well as a diversification of missions at lower hierarchical levels. The second force identified by Marais is the information technology (IT) revolution. This revolution provided scientists with the means for effective inter-institutional and international communication. The increase in competition for funding was another force encouraging researchers to seek alliances with prominent researchers and institutions. The fourth force was the emerging of a new world order following the end of socialism in Eastern Europe, which resulted in the permeability of national borders and the softening of the obsession with secrecy. The fifth force identified by Marais was the ever greater pressure on government for accountability to the public with regard to the spending of taxpayers' money. This induced a climate requiring researchers to justify their research in terms of societal relevance.

"Strategic science" is manifested at three levels: knowledge production, institutional management and national research systems. The whole vision of "Strategic science" can best be summarised by the following attributes. In the first place it acknowledges the importance of societal relevance as a valid force determining priorities, together with an associated role differentiation. Secondly, there is greater focus on strategic research programmes. There is thus a greater focus on basic research as having the responsibility for generating knowledge that would form a base for the solution of practical problems. This links with the third attribute, where the criterion for good research has become enriched by the criterion of societal relevance. There is also an outward orientation on the part of researchers, resulting in increasing collaboration across institutional, interdisciplinary and national borders. This outward orientation is enhanced by the fifth and sixth attributes, which are greater mobility of researchers between and among alliances and centres and the increasing notion of joint funding of collaborative research programmes (Marais, 2000).

1.3 Different regimes of science and Mode 1 – Mode 2 knowledge production

The discussion on the different regimes of science can be seen as central to the debate on the transformation to new forms of knowledge production.

In their book *The new production of knowledge*, Gibbons *et al.* (1994) explore changes in the mode of knowledge production in contemporary society. They contrast two modes of knowledge production, where Mode 1 precedes Mode 2 knowledge production.

When speaking about Mode 1, Gibbons *et al.* refer to the traditional meaning of the word "science". "Its cognitive and social norms determine what shall count as significant problems, who shall be allowed to practice science and what constitutes good science. Forms of practice which adhere to these rules are by definition scientific while those that violate them are not" (Gibbons *et al.*, 1994:3). The problems in Mode 1 are set and solved in a context governed by the academic interests of a specific scientific community.

Knowledge production within Mode 2 is carried out in the context of application. Mode 2 is characterised by heterogeneity, it is trans-disciplinary and less hierarchical and transient than Mode 1. In contrast, Mode 1 is discipline based, homogeneous, more hierarchical and tends to preserve its form. Each mode of knowledge production employs a different type of quality control. Mode 2, when compared to Mode 1, is more socially accountable and reflexive. "It includes a wider, more temporary and heterogeneous set of practitioners, collaborating on a problem defined in a specific and localised context" (Gibbons *et al.*, 1994:3).

Gibbons *et al.* argue that Mode 2 is a totally new mode of knowledge production, where Mode 1 will gradually be assimilated into Mode 2. "We believe that Mode 1 will become incorporated within the larger system which we have called Mode 2 and other forms of knowledge production will remain dynamic" (Gibbons *et al.*, 1994:154).

Weingart, on the other hand, in his article "From 'Finalization' to 'Mode 2': old wine in new bottles?", critically discusses the notion of Mode 1 and Mode 2 knowledge production as presented by Gibbons *et al.* He acknowledges the changes within the forms of knowledge production, but views the Mode 1 - Mode 2 argument as being too simple and over-romanticised. He also comments that changes in the processes of knowledge production were already noted by scientists two decades ago, when a couple of scientists publicised "the finalization thesis". However, at that stage other scientists did not meet this specific thesis with much enthusiasm. In fact the finalisation thesis was bitterly opposed by scientists and policy-makers (Weingart, 1997).

In his commentary Weingart points to some parallels between the descriptions of science by Gibbons *et al.* and the "Finalisation thesis". These parallels are as follows:

- The increasing interrelation in networks and the transitory character of research teams and networks;
- Knowledge production within the context of application, with concrete utility and clients in mind;
- Research is characterised by transdisciplinarity;
- The criteria and quality of research are no longer determined by the disciplines nor carried out by "peer review" alone; additional social, political and economic criteria emerge in contexts of application;
- Knowledge production becomes socially accountable and reflexive. Research is increasingly orientated towards social values and political objectives as well as to the media.

(Weingart, 1997)

Be that as it may, whether we speak of Mode 1 – Mode 2 knowledge production, different regimes of science, or refer back to the :"Finalization thesis", there is perhaps sufficient consensus emerging on significant changes taking place within the realm of science. And though they might speak from different theoretical points of view, I dare say that most scientists recognise this fact. Science can no longer claim that it is an entity functioning on its own and ruled by its own criteria. It has become accountable to society on various levels as well as to government, while still needing to remain true to itself with regards to innovation and quality.

1.4 The notion of change and its impact on the evaluation of research

The main issue that is of specific interest here, is how these changes that have taken place in science up to this point have impacted on the evaluation of research, specifically peer review.

If we take Gibbons *et al.* seriously on their prognosis that the form of knowledge production as found in Mode 1 will gradually disappear into Mode 2, then we are left with the notion that "peer review" as we know it - i.e. scientific experts evaluating work for their scientific excellence - will also gradually disappear and social, political and economic factors will be the only true criteria of what good science entails. Thus no more use for basic science and peer review?

During the regimes of "Science the endless frontier" and the "Republic of science" (Mode 1), scientists were the only evaluators of what is regarded as good science and demanded that others steer clear of their domain, insisting that research institutions govern themselves. In the *Republic of Science* Michael Polanyi argues the case for research institutes to govern themselves, stating: "Any attempt to organise the group of helpers [scientists] under a single authority would eliminate their independent initiatives and thus reduce their joint effectiveness to that of the single person directing them from the centre" (Shils, 1968:3). The regime of "Strategic science" (Mode 2) introduces reflexive science relevant to the context of application and responsive to a wider audience, with the new buzz words being participatory science and the incorporation of lay knowledge into research projects and initiatives.

Scientists acting as evaluators are now faced with a dual responsibility. They have to be responsive to the scientific community as well as to public decision-makers, who in their turn are responsible to the public for their legitimacy. The scientific community is often dependent on funds provided by the public decision-makers and thus frequently find themselves with the obligation to be more sensitive to the needs of society in order to gain public support and thus political support. The media have done a great deal in publicising research projects and scientific efforts, leading to a more knowledgeable citizenry, which means that scientists have to account not only to their peers and politicians but also to the public. However, it is important to remember that this scenario only pertains to certain areas of knowledge and knowledge production, i.e. those areas that are relevant to policy-making and have emerged as political arenas as a result of "scientification". These areas include environment, health, communication, privacy and reproduction (Weingart, 1997).

According to Weingart, "In arenas like environment and reproduction, there is the requisite technology assessment, in health there are ethics committees, in communication it is data protection. These, indeed, are new institutional arrangements which require new types of knowledge bearing the characteristics of 'Mode 2'. For areas of knowledge having no immediate connection to social values and subjective risk perceptions these conditions do not apply: high-energy physics, astronomy, and paleontology lie outside the concerns of citizens' groups and, at best, end up as issues in priority debates" (Weingart, 1997:603).

A trend that cannot be denied (and which may already have come to mind in reading the above) is the coupling of science and politics. We have witnessed how governments, NGOs, supranational organisations and supranational scientific associations form international consortia and launch international research programmes. Examples of these are the World Climate Research Programme (WCRP), the World Meteorological Organisations the light of criteria such as societal relevance and its reflexive nature which is the domain of policy-makers, the evaluation of scientific excellence will remain the domain of the scientist. There seems to be an interactive relationship between the two domains, societal and political on the one hand and scientific on the other, where both contribute to the development of science but both respect each others' domains and fields of expertise.

The weaving of scientific excellence and the societal relevance of science into one scientific process, resulting in both societal and scientific development, is an endeavour also evident in South Africa. The National Research Foundation's (NRF) new mission statement focuses on equity and excellence. It emphasises partnerships and interactions and embraces wide-ranging research activities that investigate fundamentals, while also focusing on science within the context of application and encouraging the academic community to link up with the world of commerce and industry. "The business of the NRF is to support and promote research in order to facilitate the generation of knowledge, innovation and development" (NRF newsletter, 2000).

The mission statement of the Research Support Agency of the NRF appears in the same newsletter:

A dynamic, quality-driven organisation that provides leadership in the promotion and support of research and research capacity development in the natural, social and human sciences, engineering and technology to meet national and global challenges through:

- Investing in knowledge, people and infrastructure
- Promoting basic and applied research and innovation
- Developing research capacity and advancing equity and redress to unlock the full creative potential of the research community
- Facilitating strategic partnerships and knowledge networks

(WMO), the Intergovernmental Oceanographic Commission (IOC) and the Human Dimensions of Global Environmental Change Programme (HDP). These programmes have as their main function the standardisation of data formats, standards of measurement and methodologies that are all essential for the transnational communication and transfer of consensual knowledge into politics. They represent the coupling of science and politics in the sense that they imply political goals and interests that may be legitimating or economic. These research projects are further characterised by negotiations between scientists and policy-makers where programme objectives and research priorities are at stake which are of interest to both sides: research funds in exchange for relevant knowledge and experience (Weingart, 1997).

Transdisciplinarity is another reality for scientists today. Scientists are increasingly expected to collaborate with others from different disciplines in so-called "hybrid communities". "Scientists move about in the diffuse border area between science and politics, between research and advising policy-makers. The interrelationships and overlaps between these areas can even be demonstrated in the respective discourses" (Weingart, 1997:600).

In the light of the changing environment in which science has to function the issue that this thesis investigates is the status of basic research and peer review within science. Do these changes mean that basic research and peer review will be abolished as the transformation within science develops? According to Weingart the answer to this question is that they will not. "[T]he point of democratisation, which is that virtually all political and interests have acquired access to scientific knowledge. But these groups resort to and use scientific experts as a political resource rather than claiming authority for their own (lay) knowledge" (Weingart, 1997:605). Thus, although science is to be evaluated in

- Upholding research excellence.

(NRF newsletter, 2000)

The South African government's vision for science and technology also reflects the above and is seen as central to the success of the Growth and Development Strategy (GDS) which is to address the needs of all South Africans. "The development and application of science and technology within a national system of innovation (NSI) in South Africa will be central to the success of the Growth and Development Strategy of the government as it seeks to address the needs of all South Africans" (*South African White Paper on Science & Technology*, 1996:5).

1.5 Peer review: A system under strain and attack

Science as an institution does not stand on its own, oblivious of changes within the societal and political arena. These realms directly influence science by determining the paradigm in which research and development takes place and thus also determines the role of the peer reviewer. The changing regimes of science or modes of knowledge production have placed peer review under mounting pressure, with criticism from scientists on the one hand, and administrators and politicians on the other. Politicians are calling for a science that is relevant for economic and societal development. Scientists, on the other hand, increasingly find themselves having to protect science as an institution. Administrators are standing right in the middle, dependent on politicians and society for funds, on the one hand, and responsible to the scientific community for the development and protection of good science, on the other.

The factor essentially responsible for exposing the shortcomings of the peer review system is the reduction of research funding. Much of the current pressure on peer review results from over-subscription to research programmes, with the result that the success rates of applications have fallen drastically. This results in a strain on reviewers as well as on the system with regards to its assumed fair operation.

1.6 Main points of this chapter

The main focus of this chapter was the changes that have taken place in science thinking since the Second World War. We have seen a change in science, or rather the evaluation of science, from being an enterprise responsible only to its scientific peers to being responsible to society at large. It has the responsibility of sustaining scientific excellence together with societal relevance. As this chapter has shown, this dual responsibility impacts significantly on the way in which science is evaluated, with the focus on peer review as an evaluative tool. The role of scientific evaluators has changed from only having a responsibility towards their scientific peers to also include a responsibility towards society at large. We find scientists and technologists occupying an advisory role in contemporary society, where they are expected to offer a unique and powerful kind of expertise. "They must not only provide specialised and detailed research findings, but must also have the knowledge, skills and understanding to relate these findings to social and political concerns about health, safety and other problems that affect all citizens" (Rip, 1985: 95). In contemporary society scientists are expected not only to convey their scientific knowledge, but they also find themselves increasingly fulfilling the role of expert advisers. Expert advice goes beyond the "truths" of scientific knowledge and becomes involved in the complexities and uncertainties of the world outside the laboratory (Rip, 1985). Peer review as an evaluative tool has not lost it importance; it has only changed in character.

CHAPTER 2

PEER REVIEW: CONCEPTUAL AND METHODOLOGICAL ISSUES

In this chapter peer review as a concept will be analysed. In Section A the discussion will focus on definitions of peer review, its origin, the contexts in which it functions as well as the notion of 'peers'. The aims and objectives of peer review as well as the different types of peer review will also be discussed in this section. In the second section, Section B, of this chapter the focus will be on critiques of peer review as an evaluative mechanism.

SECTION A

DEFINITIONS OF PEER REVIEW, ITS HISTORY AND CONTEXTS OF APPLICATION

Peer review can be described as a "generic term" which includes merit review, refereeing, quality control, peer judgement, peer advice and peer evaluation (Australian Research Council, 1997). Chubin and Hackett (1990) define peer review as an "organised method for evaluating scientific work which is used by scientists to certify the correctness of procedures, establish the plausibility of results, and allocate scarce resources (such as journal space, research funds, recognition and special honour" (Chubin and Hackett, 1990:2).

The origins of peer review as an institutional mechanism for certifying knowledge and allocating resources can be traced back as far as 1665. The establishment of the first professional scientific societies resulted in the publications of *The Philosophical Transactions* of the Royal Society of London and the *Journal de Scavans* in 1702. The formalisation of peer review as an advice mechanism in determining project support by federal research grant-giving agencies is essentially a post-World War II phenomenon. This is the case in most Western countries and coincides with the systematising of government support for scientific research (Australian Research Council, 1997).

Although peer review is most commonly associated with the editorial review of manuscripts for publication and the review of research proposals by grant-giving agencies, it now also operates within broader socio-economic concerns. Peer review is not only used by funding agencies and editors of scientific journals but also by government, policy-makers, university boards and the industrial community. Peer review as an evaluation mechanism has skilfully infiltrated different contexts of "quality" evaluations. To name but a few of these contexts: journal manuscripts, preperformance evaluation of proposals, evaluation of persons to be appointed at universities, evaluations of research programmes at universities, selection of industrial research, disciplinary evaluation committees and regulatory practices in the USA (Van der Meulen, 1992).

The "infiltration" (perhaps even "invitation"?) of peer review into these different contexts is not solely due to the competence of scientists to evaluate but also to the authority that these peers bring into the evaluation process. The functioning of peer review within a more extended context has as one result the reconceptualisation of "peers". Kostoff (1999) defines a peer as "A person who has equal standing with another" and thus defines peer review as ". . . a review of a person or persons by others of equal standing" (Kostoff, 1999:2).

The changing and extending contexts in which science and thus peer review operates, has as result that peers are not just people "who are equal in ability, standing, rank or value" (*The Concise Oxford Dictionary*, 1990). Peers are increasingly considered to be the experts (Van der Meulen, 1992). This "new' conceptualisation of peers is already visible in the review procedures of some scientific journals. To quote but two: "A referee must be a specialist in his field, recognised and accepted as an expert". "A referee must be someone who is regarded as an authority in the specialised subject of the paper" (Van der Meulen, 1992:71).

Whilst peer review was an autonomous mechanism within science, evident in the regime of 'Science the endless frontier', it would be naïve now to consider it as such. "Considering that peer review emerges at the instance where science depends on others for its resources, it would be naïve to consider it solely as one of the mechanisms to autonomize science" (Van der Meulen, 1992:71). Increasingly science is dependent on others, such as government and funding agencies, for resources. This implies that for scientists to get access to the resources, they have to abide by and adapt their research goals and initiatives to "fit" the goals and mission statements of the sponsors. Autonomy is thus only one possible outcome and mostly only partially accomplished, as the patrons are eager to keep some control over the decision process (Van der Meulen, 1992). "Dependency has to be acknowledged as the patrons take part in the construction of the peers, and hence of the autonomy" (Van der Meulen, 1992:71).

"The term peer review is somewhat of a misnomer – it is not intended to mean evaluation by one's equals but evaluation by those who are at the forefront of their respective research fields and acknowledged as possessing the necessary expertise and judgement to distinguish between the relative quality of research accomplishments and proposals and make discernments about where the potential for advancement is greatest" (Australian Research Council, 1997:9). The new context within which evaluation takes place is a multidisciplinary context where neither of the positions, other than that of the scientific object, are restricted any longer to scientists. This makes it imperative that, instead of equals evaluating each other, peers should be those considered to have the expertise and authority to make the evaluation legitimate and credible.

A.1 PEER REVIEW OBJECTIVES/PURPOSES

From the discussion at the beginning of this chapter it is evident that peer review supports diverse purposes. In the remainder of this section, I aim to discuss the different purposes of peer review. These are:

- To serve as a quality filter
- To add value and improve the quality of the manuscript, proposal or programme under review
- To provide a stamp of legitimacy and competency
- To indicate programme quality, programme relevance, management quality and appropriateness
- To ensure accountability
- To function as a mechanism for policy-makers to direct scientific effort
- To function as a valid and reliable measure of scientific performance.

A1.1 Peer review as a quality filter

Peer review functions as a quality filter to conserve resources. Readers of papers published in peer-reviewed journals assume that these papers are above a minimal quality threshold. The readers can thus focus their limited time resources on the highest-quality documents assumed to be contained in these journals. The same holds for projects and programmes selected for initiation or continuation by peer review. These projects and programmes are also assumed to be above a minimal quality threshold, which allows precious labour and hardware resources to be focused on these high-quality tasks selected (Kostoff, 1999).

A1.2 The function of peer review in adding value and improving the quality of the manuscript or programme under review

Recent studies evaluating the effects of peer review and editing on manuscript quality have shown that peer review and editing improve the quality of reporting as well as the readability of original articles and their abstracts (Kostoff, 1999).

Kostoff (1999) identifies three stages where value is added during the peer review process of a research programme. The first stage is the period between the reviews. The researchers do their work knowing that it will be subjected to a high-quality review. This results in researchers maintaining a higher level of performance quality owing to their knowledge of the forthcoming expert review. The second stage is the period of review preparation. The preparation period provides a focal point for discussions in order to arrive at a quality presentation. The value added here is the substantive increase in the intrinsic programme quality. The actual review makes up the third stage. During the review independent viewpoints are injected in a public forum, high-quality research is re-affirmed and strong recommendations are provided for poor research.

Another factor which influences the amount of value added during the peer review process is the level of reviewer and reviewee anonymity. "There is a hierarchy of levels of reviewer anonymity which produce different degrees of frankness and honesty in the reviewer's response" (Kostoff, 1999:8). This degree of frankness and honesty directly influences the amount of value added with every peer evaluation as it determines the quality of the review.

A1.3 Peer review: Providing a stamp of legitimacy and competence

Peer review is readily referred to when the quality of science becomes the topic of discussion where it is displayed as the "showpiece of science's capability to judge what is good science" (Van der Meulen, 1992: 71). Although peer review is best known as an evaluative mechanism in contexts such as journal manuscripts, preperformance evaluation of proposals and the evaluation of persons to be appointed at universities, it has also entered into many other contexts. Examples of these are evaluation of research programmes at universities, selection of industrial research, disciplinary evaluation committees introduced in science policy in the 1970s and regulatory practices (Van der Meulen, 1992).

Van der Meulen (1992) attributes this spreading of peer review as an evaluative mechanism to the many contexts in which it is applied not solely to the competence of scientists to evaluate, but also to the authority that peers bring into the evaluation process. Peers are seen as the experts and it is their expertise that gives them authority. This authority carried into the evaluation process by the experts is the factor that gives authority to the peer review conducted by them. Thus, to enhance the authority of peer review it is imperative that peers be not merely equals but experts. "Instead of being equals to those evaluated, peers should foremost have the expertise and authority to make the evaluation legitimate and credible" (Van der Meulen, 1992:74).

A1.4 Peer review: Ensuring accountability

Science is increasingly measured against the yardstick of relevance and accountability. Scientists are constantly reminded of their responsibility towards their sponsors as well as towards society, where they are held accountable for their actions as well as their financial expenditure. In 1993 the United States passed the Government Performance and Results Act (GPRA) as part of its effort to improve the accountability of government-funded programmes. The three major components of the GPRA are strategic plans, annual performance plans and metrics to show how well the annual plans are being met (Kostoff, 1997).

The mechanism by which accountability is to be promoted is peer review. "Peer review should also promote accountability in science, ensuring that funds have been expended appropriately, that laws and regulations governing human subjects, animal and lab safety, accessibility of raw data, and hiring practices have been observed, and that the proposal work has indeed been carried out" (Chubin & Hackett, 1990:44).

A1.5 Peer review as an indicator of programme quality, programme relevance, management quality and appropriateness

Giving an account of programme quality, programme relevance, management quality and appropriateness is all part of science being accountable to its sponsors and broader society. Kostoff (1997) depicts a more appropriate accountability approach for basic research to comprise of three components; 1) articulation of a rational investment strategy; 2) long- and short-term retrospective studies that show the diverse benefits from past research and potential future benefits; and 3) quality control of expert peer review.

In a system such as the GPRA expert peer reviews' function is to validate the soundness of investment strategy and the importance of the research accomplishments and subsequent impacts on technology. "Peer review properly designed to support GPRA would provide credible indication to the research sponsors of intrinsic program quality, program relevance, management quality, and appropriateness of direction, and has the potential to improve the quality of the research program as well" (Kostoff, 1997:652).

Also important here is the evaluation criteria by which peer review is to take place. Research evaluation criteria function as a reflection of an organisation's strategic and policy research objectives. Thus, when review criteria are documented, this also serves as tangible indicators to external groups that strategic objectives are being implemented, contributing to the accountability of science (Kostoff, 1999).

A1.6 Peer review as a mechanism for policy makers to direct scientific effort

"Peer review is expected to be responsive, providing a mechanism for policymakers to direct scientific effort and for new science to shape future research" (Chubin & Hackett, 1990:45). This statement by Chubin and Hackett reflects the dual responsibility of peer review where, on the one hand, it has to ensure the economic and social relevance of science and on the other, it has to preserve science from political, social and scientific currents while encouraging continuity in research.

In the current regime of "Strategic science" basic research is expected to provide a broad base of knowledge which is to form the basis for the solution of recognised current or future practical problems (Rip, 1991). Governments rely on science to act as a compass to guide them towards social and economic development. "The White Paper on Science and Technology" in South Africa illustrates this point in its vision: "The vision is one where, on the one hand, South Africa uses science and technology to become economically competitive on a global scale and, on the other hand, to provide essential services, infrastructure and effective health care for all South Africans" (Mtshali in *White Paper on Science and Technology*, 1996:i).

For science to comply with this vision, government is responsible for creating an environment in which innovation can flourish (Blankley, 1997). In creating this environment government in its turn is dependent on scientists to direct policies and funding towards the governments' needs. The mechanism used by scientists and government to direct science policies and funding to aid governments' needs is peer review. In the past peer review was solely the mechanism by which scientific excellence was measured, but now it has also become a mechanism used by politicians to further their interests.

A1.7 Peer review as a valid and reliable measure of scientific performance

Although peer review also functions within the political sphere, it is above all the mechanism by which scientific performance is measured. It is thus imperative that peer review adheres to the technical standards of good measurement, that is validity and reliability. "The term validity refers to the extent to which an empirical measure adequately reflects the real meaning of the concept under consideration" (Babbie, 1998:133). So validity in peer review refers to valid measurements of the quality it is claimed to measure. "Thus if peer review is to measure the scientific merit of a proposal, it should not instead measure the visibility of its author or the prestige of its institutional origins" (Chubin & Hackett, 1990:46).

"Reliability is a matter of whether a particular technique, applied repeatedly to the same object, would yield the same result each time" (Babbie, 1998:129). With regards to peer review as an evaluative mechanism, this would imply that if a proposal or proposed journal article be rejected by one review panel, it would ultimately be rejected again when presented to another panel. Reliability with respect to peer review would thus entail the limiting of chance in selections.

A.2 THE QUALITY OF THE PEER REVIEW PROCESS

The extent to which peer review can fulfil the above objectives/purposes entirely depends on the quality of the review process. A high-quality peer review has as overall objective the provision of an accurate picture of the intrinsic quality of the research/proposal or manuscript reviewed. The lack of absolute standards for the measurement of research quality, however, presents a problem and often results in the evaluation of intrinsic research quality in a subjective manner, depending on the perspectives and past experiences of the reviewers (Kostoff, 1999). However, despite this dependence on past perspectives and past experiences of reviewers, high-quality reviews are still possible. Kostoff (1999) supports this view when he argues that high-quality reviews are possible on the condition that two generic conditions are fulfilled. The first condition, according to Kostoff, is the utilisation of highly competent reviewers and the second is the elimination of additional distortions in the reviewers' evaluations as a result of biases, conflict, fraud or insufficient work.

In another document Kostoff goes further to identify three factors imperative in ensuring high-quality peer reviews: " . . . three of the most important intangible factors for a successful peer review are motivation, competence, and independence" (Kostoff, 1996:250). With regards to motivation Kostoff specifically stresses the role of the review leader. "The review leader's motivation to conduct a technically credible review is the cornerstone of a successful review" (Kostoff, 1996:250). The review leader is the most important link in the review team in the sense that it is he/she that selects the reviewers in the first place. Second it is he/she that summarises their comments; thirdly, the reviewer leader guides the questions and discussions in a panel review and, finally, makes recommendations on whether or not a proposal should be funded or an article published. In the selection of reviewers the review leader needs to consider the individual reviewer's competence in the subject area, as well as the competence of the review group as a whole to cover the different facets of research issues. The review leader also needs to know the biases and conflicts of each reviewer selected as the biases and conflicts of the reviewers limit the quality of a review (Kostoff, 1996).

Another factor also identified by Kostoff is a review panel compiled from a broad range of reviewer expertise in order to enhance review results. "The presence of reviewers with different research target perspectives and levels of understanding on one panel provided a depth and breadth of comprehension of the different facets of the research impact that could not be achieved by segregating the science and utility components into separate panels and discussions" (Kostoff, 1996:251).

The recipe for a successful, high-quality peer review, according to Kostoff, is thus determined by an interplay of factors: a team leader motivated to conduct a technically credible review by selecting competent reviewers, who will function as competent individual panel review members, but who will also contribute to the broad range of reviewer expertise needed in a review panel to enhance review results.

A.3 DIFFERENT TYPES OF PEER REVIEW

In the above section several factors ensuring high-quality peer review were discussed. However, there is one factor not mentioned in the above section which also has an influence on review quality and this is the role of the level of anonymity in the quality of the peer review process.

It has already been stated in the above section that what is desired from the peer reviewer is an honest viewpoint on the quality of the research, proposal or manuscript under review. There are two sides to the issue of anonymity within the peer review process, as illustrated by Kostoff (1999): "Having the reviewer and reviewee present during the review (and this applies to manuscript, proposal, and program review; 'present' just must be interpreted differently in each case) will sharpen the quality of the technical discussion details, . . .However, having the reviewer and the reviewee present during the review will, in many cases, obviate the expression of the reviewer's deepest convictions about the quality of the research" (Kostoff, 1999:7).

Kostoff identifies a hierarchy of levels of reviewer anonymity, which according to him produces different degrees of frankness and honesty in the reviewer's response (Kostoff, 1999). This hierarchy of levels of reviewer anonymity as discussed by Kostoff also links with the different types of peer review and will be discussed in the following part of this section.

A3.1 Telephone reviews

Telephonic reviews can be placed at the top of the hierarchy with regards to reviewer anonymity. In this type of peer review the reviewer is provided with information on the research, typically written, after which feedback is provided orally over the phone with total anonymity with regards to the identity of the reviewer. It is exactly this high level of anonymity which, according to Kostoff (1999), results in telephonic reviews offering the most honest and straightforward reviewer response.

A3.2 Written reviews

The next level in the hierarchy of reviewer anonymity is written reviews. Within this type of peer review the reviewer is asked to provide his or her comments in writing. Although the reviewer is completely anonymous to the reviewee, reviewers tend to moderate the frankness of their comments when asked to provide them in writing. This moderation of frankness could, however, be extenuated or even prevented if the reviewers fully trust the review manager to protect their anonymity (Kostoff, 1999).

A3.3 Face-to-face review with total confidentiality

The next level of anonymity is face-to-face review with total confidentiality. During this review both reviewers and reviewees are present during the research presentation. At the close of the presentation reviewers meet behind closed doors in a closed session, after which written and oral comments on the evaluations are presented, but without any attributions, thus ensuring total confidentiality (Kostoff, 1999).

A3.4 Face-to-face review in the absence of anonymity

The final level in the hierarchy of reviewer anonymity is face-to-face review in the absence of anonymity. During this review process both the reviewers and the reviewees are present throughout the duration of the whole process and all the verbal and written comments are provided with full attribution. Although it might be argued that this approach to peer review is better than no review at all, it must be noted that this approach does not begin to utilise the full potential of what expert peer review can offer (Kostoff, 1999).

A3.5 Blind and double-blind reviewing

In the different peer review types above, anonymity was linked throughout to the anonymity of the reviewer. However, there is another side to the coin of review anonymity.

Blind reviewing refers to partial anonymity where the reviewees' name and affiliation is withheld from the reviewer. The argument for this kind of review is the assurance of fairer reviews, especially with regards to work done by unknown researchers or researchers from less prestigious institutions as well as the elimination of bias on the basis of personal characteristics such as gender. Double-blind review in its turn takes place in a context of total anonymity where both the reviewer and reviewee are anonymous to each other (Kostoff, 1999).

Section B

Critique of peer review as an evaluative instrument

Changes within the scientific environment are not the only constraints on the peer review processes. Peer review is also often subject to critique from the very community it is to serve and protect. Criticisms levelled against peer review are more often than not the result of discrepancies perceived between the rhetoric regarding peer review and its actual practices (ARC Report, 1997).

Some scientists view peer review as an unreliable system, saturated with bias, a system that is too slow and too leaky, allowing peer reviewers to gain research advantages unfairly. Peer review also often finds itself under scrutiny from scientists, on the one hand, and politicians and administrators, on the other. Often-cited criticisms of peer review from politicians and administrators include the following:

1. Peer review takes the decision-making power out of the hands of selected officials and their appointees and puts it into the hands of people who are not accountable to the public.
2. It enables the scientific community to use public funds for its own purposes, that is, "pure" research, while ignoring the pressing needs of society that might benefit from "applied" research.
3. It discriminates against scientists working in small science departments at low-prestige universities and colleges.

4. It does not weight adequately the opinions of non-academic scientists on the merits of proposals.

5. It fails to screen out proposals of questionable merit.

(Chubin & Hackett, 1990:29)

It is clear from the above criticisms that there is a very visible tension between politicians and administrators on the one hand, and scientists on the other with regard to the functions and their expectations of peer review. Scientists expect peer review to evaluate scientific merit, while politicians and administrators attach social responsibilities and social accountability to the role of peer review. Administrators and politicians have established specific criteria to combat some of the above criticisms, including drawing up criteria such as the naming of the kind of projects that are more viable for funding than others, with the focus on applied and relevant science.

The remaining part of this section will be devoted to criticisms of the peer review system as expressed by the scientific community. The greater part of this critique is based on a research report on the peer review process published by the Australian Research Council in 1997 (ARC Report) as well as two special publications in the *Journal of Behavioral and Brain Sciences* on the peer review process published in 1982 and 1991 respectively. Other authors' works are also used to substantiate arguments.

B.1 CRITICISMS OF PEER REVIEW MECHANISMS FROM THE PERSPECTIVE OF SCIENTISTS

Peer review has been criticised on several accounts by the scientific community. Problems perceived are:

1. The partiality of peers
2. Unreliability of reviewers
3. Problems in peer selection and performance
4. Perceptions of bias
5. Lack of transparency
6. Inherent conflict of interest
7. Potential for misuse of confidential information
8. Costs.

B1.1 The partiality of peers

How objective are reviewers really? The concentration of research facilities in fewer, larger centres makes it increasingly difficult to find peer reviewers with no vested interest in the review (King, 1989). The social structure of science is such that scientists are bound to meet each other at scientific conferences, serve on scientific societies together and work together on projects from time to time. It is obvious that subjectivity and partiality in these circumstances are difficult to say the least (Labuschagne, 1985).

Subjectivity and partiality are also manifested in the so-called "old boy" network. "The 'old boy' network often results in older entrenched fields receiving greater recognition than new, emerging areas of research, while declining fields may be protected out of a sense of loyalty to colleagues" (King, 1989:5). This protection of older entrenched and declining fields can be very dangerous to the development of science. The reason for this is simple; peer review that functions within an "old boy" network is ineffective as a mechanism for restructuring scientific activity, and if there is no restructuring this implies there is no change, thus no development.

B1.2 Reliability of reviewers

The issue of reliability is really an issue of consensus as to what constitutes good science. Low reliability implies low agreement concerning research priorities, research methods and the rating of the quality of research methods. If peer review is considered as a proxy for assaying the standards of the scientific community, unreliability within this system could have detrimental consequences for the value attributed to scientific merit in the decision-making processes of peer review (Blissett, 1982).

Peer review functions well in classifying documents like grant applications into categories such as excellent, fair and poor. The difference in opinion, however, emerges in the selection of documents all classified as excellent. Reviewers are then forced to select proposals and articles from a big pool of excellent documents due to limited funds and journal space. As Fred Delcomyn writes: "Expecting it [peer review] to allow one to distinguish between applications that are all in the excellent category, however, is like expecting to be able to measure the diameter of a nerve cell with a meter stick" (Delcomyn, 1991:144).

Although the above situation has led some scientists to believe that the peer review process is the best way of placing bets as chance is considered to be inherent in the system, unreliability does not seem to influence science as such. The system might be unfair to individual applicants, but the best article or proposal is still selected, as peer review is quite consistent on flawed papers or proposals (Chicchetti, 1991). Chicchetti goes as far as to argue that disagreement among reviewers can sometimes serve a useful purpose, considering that one referee may detect a flaw in reasoning that a second referee has failed to uncover. It is important to consider the possibility that one of the main reasons for referee disagreement is that the negative review is usually considered to be either misinformed or wrong by those denied grants or journal space (ARC report, 1997).

Musacchio, writing about the National Institute of Health, identifies three causes for variability regarding the assignment of priority scores by reviewers:

1) Reviewers have different backgrounds and perspectives, so they have different opinions as to how a question should be best approached and answered. Moreover, reviewers may be highly qualified to evaluate only some, but not necessarily all, the multiple aspects of a given application. It is paradoxical that the inclusion of members with different backgrounds and perspectives, which is a desirable characteristic of the system, is also what introduces a wide range of variability into the scores. The spread of the votes is what ultimately makes it impossible to find statistical differences between the best applications.

2) Inexperienced and ad hoc reviewers do not always use the same criteria as the experienced members of the initial review group, and some tend to be hypercritical. This is not necessarily a negative factor, because it may represent a different scientific perspective.

3) Unsympathetic or unfunded reviewers tend to pick on applications, sometimes focusing on non-essential points. This may lower the priority of an application enough to put it in the unfunded category. By contrasts, sympathetic reviewers tend to do exactly the opposite (ARC, Report, 1997:20).

B1.3 Problems in peer selection and reliability of review process

The knowledgeability of peers selected as reviewers has often evoked critique from those reviewed. Selecting knowledgeable peers has indeed become one of the main difficulties faced by peer review, because of the increasing multidisciplinary character of proposed research or journal manuscripts. This multidisciplinary character often results in sometimes very small scientific communities within a specific multidisciplinary field from which reviewers have to be selected (ARC Report, 1997).

With regards to the reliability of the review processes one of the debating issues is the question on the number of reviews sufficient for each application. Some commentators, such as Chicchetti (1991), argue that a larger sample of reviewers will enhance the reliability of their combined judgements. However, when discussing the question of a sufficient number of reviews, two issues come to mind. The first is, as mentioned above, the size of the scientific community and the second is the implication with respect to the costs of the review process. More reviews might enhance the reliability, but it will also increase the review costs, that is reviewer time, administrative costs as well as potential assessor overload (ARC Report, 1997). Before the sufficient number of reviews can be determined, it is important to consider the implications and to decide what carries the most weight.

Horrobin (1982), in his commentary on a study done by Peters and Ceci (1982) on the peer review practices of psychological journals, argues that the concept of peer review is based on two myths, which are both relevant here. The first myth is that all scientists are peers, thus people who are roughly equal in ability. The second myth is that the ordinary scientist always instantly recognises genius and smooths its path. "Most scientists are not the peers of the very best, and most scientists follow the crowd when it comes to the recognition of brilliance" (Horrobin, 1982:218). It is important to realise that there are scientists who, because of their brilliance and eminence, have no peers. But it must also be mentioned that these "scientists without any peers" are few and far between.

Another serious concern in the ever more competitive nature of science is that it is often the case that a scientist's peers will also be the scientist's competitors or direct rivals. It is this situation that make some scientists wary of presenting their most original research ideas in grant proposals or in articles for journal review. The way in which editors and research funding agencies determine and deal with potential conflict of interest will very definitely affect their credibility.

B1.4 Perceptions of bias

Peer review has often been accused of bias with regard to gender, status of institutions and eminence of scientists. Bias within the review process can have a very negative effect on the quality of a review as, "The quality of a review is limited by the biases and conflicts of the reviewers" (Kostoff, 2000:4).

Horrobin, when commenting on bias in peer review, recalls an experimental study conducted in the United States of America. A number of women complained to the Modern Language Association in the United States that there were surprisingly few articles by women in the association's journal compared to what would be expected from the number of members that are women. The association vigorously denied any accusation of biased reviews, but under growing pressure was forced to institute a blind reviewing procedure. Under this procedure the names of the authors and their institutional affiliations were omitted from the material sent to the reviewers. The results showed a dramatic rise in the acceptance of papers by female authors (Horrobin, 1982).

Horrobin introduces the concept of "built-in bias against highly innovative work" (Horrobin, 1982:217). Although uncontested, possible reasons for this "built-in bias" could be the schooling of the reviewer together with the challenging of accepted paradigms and theories by the reviewee. Scientists are trained and schooled within a specific paradigm and in most cases scientists approach scientific questions or problems on the basis of this paradigm. It would frequently be the brilliant mind who will challenge these established paradigms. This challenging of accepted ways leads to scepticism from other scientists as well as the "peers" responsible for the upholding of scientific standards and excellence. The challenging of prevailing paradigms by certain types of innovative or cross-disciplinary research often result in the experience of difficulties with regards to the competition for research funds and/or publication space (Horrobin, 1982).

Institutional bias is probably best explained by the "Matthew effect", first defined as such by Merton. The Matthew effect is based on the writings in the Gospel of Matthew; *"for unto every one that hath shall be given, and he shall have abundance, but from him that hath not shall be taken away even that which he hath."*

The "Matthew effect" refers to the accumulation of ever-greater augmentation of recognition to scientists of considerable repute for particular scientific contribution and the withholding of such recognition from scientists who have not yet made their mark (Merton, 1968). Merton also points to the paradoxical character of the "Matthew effect", where it is functional for certain aspects within science as a social system and dysfunctional for certain individuals within this social system. "A scientific contribution will have greater visibility in the community of scientists when it is introduced by a scientist of high rank than when it is introduced by one who has not yet made his mark. In other words, considered in its implications for the reward system, the Matthew effect is dysfunctional for the careers of individual scientists who are penalised in the early stages of their development. However, considered in its implications for the communication system, the Matthew effect, in cases of collaboration and multiple discoveries, may operate to heighten the visibility of new scientific communications" (Merton, 1968:59).

Another aspect that is related to the "Matthew effect" is the so-called "Halo effect". The "halo effect" refers to the raising of the reputation of all the members of a group simply through their association with one person of high reputation who is part of that specific group (Johnes, Taylor & Francis, 1993). The "halo effect" is seen to be functional when applying for grants or journal space as it is assumed that peer review is biased towards eminent names.

B1.5 Lack of transparency

Inadequate provision of feedback

In a context where an increasing number of highly rated proposals go unfunded and an increasing number of highly rated articles go unpublished, the inadequate provision of feedback has become an aspect for increased critique. Applicants have the right to query how the choice among the excellent was made. "Proposers should have the right to know the rationale behind the decision taken on their applications and should be able to get it in writing" (ARC Report, 1997:32).

Adequate provision of reviewer comments has the potential to add value to the peer review process, to enhance transparency regarding the process and hence confidence in its integrity. An important factor to keep in mind regarding the peer review process is the variability in assessors' reports. It is not sufficient to merely provide the reviewee with copies of assessors' reports verbatim as this could potentially leave the funding agency open to further criticisms. Making it clear to the applicant how the assessors' comments were used can circumvent further criticisms (ARC Report, 1991).

Lack of an appeals mechanism

A formal means for addressing applicants' concerns, be it for grant submissions or journal publication, has not traditionally been a feature of the peer review system (ARC Report, 1997). However, in the light of decreasing journal space as a function of increasing numbers of applicants and decreasing funding made available for worthy research endeavours, it has become an important issue.

Cicchetti (1991) lists eight variables that have been used by editors to justify rejection of a given journal submission:

1. Documentable factual errors made by the reviewee
2. Faulty data analyses (which are readily reparable)
3. Prior publication in non-refereed conference proceedings
4. Replication of previously published work
5. Null or negative findings (despite very favourable reviews)
6. Subject inappropriate
7. Ideas insufficiently novel or original
8. Insufficient space to accommodate the submissions.

(Cicchetti, 1991:132)

In respect of the first five variables, a carefully documented, dispassionate letter could result in a journal editor's honouring a request that the rejected manuscript be resubmitted to a new set of referees. These referees would review it independently and without prior knowledge of its rejected status. Should this appeal fail, the author is free to submit the manuscript to other, equally prestigious journals (Cicchetti, 1991).

An increasing number of funding agencies have incorporated a formal appeals mechanism as part of their peer review processes. Recently the National Institute for Health (NIH) and the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA) established formal peer review appeal systems. The National Science Foundation (NSF) has a formal appeal system whereby a principal investigator (PI) whose grant has been disapproved can contact the designated programme officer. The programme officer then provides the PI with a detailed account of the specific reasons for disapproval. If the PI is still not satisfied, other options are available such as requesting that the Assistant Director of the NSF reconsider the grand disapproval based on arguments set forth by the PI (Cicchetti, 1991).

Classes of problems that might justify an appeal as stipulated by Cicchetti (1991) include:

- The granting agency's refusal to accept an application
- An applicant's disagreement over the assignment of his grant (whether to a specific study section or institution)
- Author's doubts with regards to the level of knowledge of specific study section members
- Evidence of bias associated with the peer review.

In the light of increased critique of the fairness, reliability and transparency of the peer review process, developments towards an effective appeals system is very important. Cicchetti also states that "These developments portend a healthier climate for handling legitimate claims about perceived unfairness or incompetence; they become especially important as less and less funding is made available to support worthy research endeavours" (Cicchetti, 1991:131).

Panel membership and operation

As mentioned in the previous section, one of the criteria which has to be met by science in the current regime of science is accountability. The functioning of peer review within a closed system of review, however, often creates the impression that it is sidestepping this requirement. In order to keep funding agencies more accountable for their operations, greater attention is being directed to the functioning and performance of peer review panels. Questions asked include:

- What is the optimal size of the panels?
- What are the appointments and review procedures for members?
- What considerations apart from the reputation and authority of the scientists should determine panel composition (e.g. minority group status, gender, age, institutional affiliation and geographical location)?
- How are the panel deliberations documented for accountability purposes?
- What regulations exist regarding conflict of interest of panel members?
- What sorts of procedures do the panels, especially regarding voting, follow?

(ARC Report, 1997:33)

Another area of concern is the length of the term and the number of times a scientist can serve as a reviewer. The size of the scientific community is the determining factor here. If a scientific community is large, the turnover can be frequent, but if a scientific community is relatively small, as is the case with some scientific fields, a large turnover is just not possible. Rapid turnover, however, also has its pitfalls as it may result in a lack of policy continuity within panels (ARC Report, 1997).

B1.6 Inherent conflict of interest

Assumption of agreement

Peer review is grounded on the assumption that there is consensus among scientists as to what constitutes good-quality of work, who is doing it, and where promising lines of inquiry lie. It is exactly this assumption that is being questioned by scientists where disagreement among assessors has been identified as a major source of dissatisfaction with the peer review system (ARC Report, 1997).

In my view the assumption about and desire for agreement is simply naïve and even detrimental to science. It is important to remember what the function of peer review is. John Bailar's commentary on a study conducted by Cicchetti on the reliability of the peer review process as published in the journal *Behavioral and Brain Sciences* is very relevant to this discussion. "Cicchetti consistently misses the mark. The purpose of peer review is not reliability, but to improve decisions concerning publication and funding . . ." (Bailar, 1991:137). Kiesler, in his commentary on Cicchetti's paper, reminds us that "Reviewing manuscripts for publication and grant proposals for funding is merely a means to an end, not the end in itself" (Kiesler, 1991:151). As an editor himself, Kiesler comments that "a high correlation between reviewer ratings of submitted manuscripts should neither be expected nor desired. The expectation that these ratings should be highly correlated is naïve; it almost assumes the reviewers are randomly drawn by the editor" (Kiesler, 1991:151).

From the above it is clear that undue agreement rather than disagreement within the peer review process should function as a warning device. "Do scientists agree? It is not only unrealistic to suppose they do, but probably just as unrealistic to think they ought to" (Harnard, 1982:185).

Perceived difficulties for some groups (such as young researchers)

Young scientists often claim difficulties in competing for research grants and journal space with older, more eminent scientists. Substantial budgetary pressure, an ever-growing scientific community and declining journal space are all factors contributing to this dilemma faced by young scientists (ARC Report, 1997).

Those who value equity over scientific merit perceive under-representation of young scientists as discriminatory. For those, however, who value scientific merit as the only criterion, under-representation of young scientists does not seem to be a problem (ARC Report, 1997). It seems that the real issue at stake is not merely the under-representation of young scientists as a discriminatory practice, but the under-representation of young scientists as a function of the various aims of science.

Shouldn't one of the aims of science be to train young scientists in such a way that they can learn through experience so as to mature into good scientists, able to conduct scientific endeavours characterised by excellence? How is this to be accomplished if young scientists do not get the opportunity to do research and publish their articles? Part of developing as a researcher is reacting to and learning from the critiques and comments made by peers in the relevant scientific field. This cannot be accomplished if young scientists do not get their work published, as they have to compete with other prominent names already known for their excellent scientific endeavours. Scientific merit is important, but so is investment.

B1.7 Potential for misuse of confidential information

The issue of confidentiality

Peer review as an evaluative mechanism assumes a high level of objectivity, disinterestedness and honesty on the part of the reviewers. This view of peer review, however, is not shared by all: "... some critics have suggested that such presumption is naïve and that the system allows for the misuse of privileged information or more bluntly the stealing of ideas from grant applications" (ARC report, 1997:2 4).

Within the problem of confidentiality lies an issue of the legal rights of the reviewees as well as the ethical "constraints" of the reviewers. What instructions do reviewers receive regarding confidentiality? Recently a lawsuit was launched by academic researchers in the United States against a biotechnology company, claiming reviewer misuse of manuscript data. The researchers accused the biotechnology company of using data from a manuscript that one of its staff members reviewed for the prestigious journal *Nature*. The lawyer of the biotechnology company brought forward the question of the legal validity of the academic notion of peer review. He argued that "there is no rule – legal or otherwise- that prevents a reviewer from using data in an academic paper he or she reviews" (Marshall 1996:1162).

Marshall examined this issue and compared the instructions of a number of leading journals in the terms of their explicitness. He summarises his findings as follows:

- *Nature:* One line in the letter sent to reviewers states "Colleagues may be consulted (and should be identified for us), but please bear in mind that this is a confidential matter."
- *Science:* Reviewers are explicitly forbidden to disseminate or exploit information in the paper: "We expect reviewers to protect the confidentiality of the material presented. Please ensure that the enclosed manuscript is not disseminated or exploited. If you find it necessary to discuss this paper with a colleague, please specify the particulars in a letter to the editor."
- *The New England Journal of Medicine:* Reviewers are told that "The manuscript should be considered a privileged communication. You should not show it to another person without calling us, and you should not photocopy it" (ARC Report, 1997:25).

The above instructions given to reviewers by the various journals clearly indicate that confidentiality is seen as an academic notion rather than a legal issue. "If one could gain a material advantage from knowing what is in a manuscript, one has a potential conflict of interest and should not review the paper" (Brogard in Marshall, 1996:1163).

"When to declare a conflict of interest and step aside versus downgrading a competitor's proposal to improve one's own future funding prospects are questions of ethics on what constitutes proper conduct. This is relative both to research community and individual standards" (Chubin in ARC Report, 1997:25). Within the peer review process the scientists is allowed to oscillate between the roles of reviewer and grant proposer. This oscillation between reviewer and grand proposer fortifies the potential for conflict of interest and subsequently complicates the ethics at work in the peer review process.

B1.8 Costs

The cost of peer review is not merely monetary. It amounts to time, money and opportunities lost as researchers spend more time on proposal writing and review and less on the actual conduct of research.

Time as an expense is probably more significant for a grant application than a manuscript. 20-80% of manuscripts rejected by the journals to which they are submitted are ultimately published elsewhere. An article that is rejected by one journal possibly only means a delay in publication of two to four months. However, a delay of two to four months for a grant resubmission can have detrimental results as researchers are dependent on the grant for the continuation of their research project (Kiesler, 1991). "A delay in four to six months necessitated by a grant resubmission (which would not be unusual and may be minimal) may force an investigator to shut down a research team that had been carefully built up over a period of years" (Kiesler, 1991:151).

Kostoff claims that "for . . . serious panel-type peer reviews, where sufficient expertise is represented on the panels, total real costs will dominate direct costs. The major contributor to total costs is the time of all the players involved in executing the review. With high-quality performers and reviewers, time costs are high, and the total review costs can be a non-negligible fraction of total program costs, especially for programs that are people intensive rather than hardware intensive" (ARC Report, 1997:30).

Determining the real costs of scientists' involvement in review positions, however, is not that straightforward as appointments to committees/panels usually carry with them a high degree of prestige and influence (ARC Report, 1997).

B1.9 Main points of this chapter

The focus of this chapter was the conceptualisation of peer review as well as the examining of peer review as an evaluative tool. Peer review, which dates back as far as 1665, is an organised method by which scientific work is examined with the aim of certifying the correctness of procedure, establishing the plausibility of the results, and allocating scarce resources (Chubin and Hackett, 1990).

We have looked at the purposes and objectives of peer review, arguing that the manner in which peer review can perform high-quality evaluation is dependent on the manner in which it can fulfil its purposes and objectives. The different types of peer review in relation to the different levels of anonymity were also discussed and finally a detailed discussion on critiques levelled against peer review concluded this chapter.

In the remainder of this thesis we will look at peer review in practice; the peer review processes of four funding agencies will be examined (Chapter Three) and the closing chapter (Chapter Four) will offer recommendations for improvements in peer review as well as discuss possible alternatives to peer review.

CHAPTER 3

PEER REVIEW IN PRACTICE: FOUR CASE STUDIES

In this chapter our discussion will centre on peer review in practice. The peer review processes of four funding agencies (i.e. the National Science Foundation, National Institute of Health, Medical Research Council in Canada, and the National Science and Engineering Research Council of Canada) will be discussed in depth; an analysis of three aspects of the peer review process will follow that discussion.

The three aspects which are intrinsic to the quality of the peer review process and that are often unique to the respective funding agencies are the method by which reviewers are selected, the review criteria by which proposals are rated and the number of review stages within the review process.

3.1 The peer review process at the National Insitute for Health (NIH)

The National Institute for Health was founded in the United States in 1887. It is known today as one of the world's foremost medical research centres and also forms the federal focal point for medical research in the U.S. The NIH is comprised of 27 separate institutes and centres and is one of the eight health agencies of the Public Health Service which in turn form part of the U.S. Department of Health and Human Services.

The source of all the information in this section can be found at the following website: www.drg.nih.gov

The goal of the NIH is to acquire new knowledge to help prevent, detect, diagnose and treat all diseases and disabilities. "The NIH mission is to uncover new knowledge that will lead to better health for everyone" (*www.drg.nih.gov*, 2001). The NIH works towards this mission ". . . by conducting research in its own laboratories; supporting the research of non-Federal scientists in universities, medical schools, hospitals, and research institutions through out the country and abroad; helping in the training of research investigators; and fostering communication of medical and health sciences information" (*www.drg.nih.gov,* 2001).

The NIH is responsible for the cyclic evaluations of proposals received by the Centre for Scientific Review (formerly the Division of Research Grants). The contents of approximately 10,000 applications are reviewed by more than a dozen Referral Officers. During evaluation of the applications the Referral Officers have to decide on an Initial Review Group (IRG) best suited to assess the application on scientific merit as well as the institute(s)/ centre(s) of the NIH best suited to fund the application, should it be considered sufficiently meritorious.

A chartered Centre for Scientific Review (CSR) section is composed of 18 – 20 individuals. These individuals are nominated by the Scientific Review Administrators (SRA) from the active and productive researchers in the biomedical community, to serve for multi-year terms. For review applications requiring special expertise, Special Emphasis Panels (SEP) are established on an ad hoc basis.

The assignment process is a collegial one with interaction, when necessary, on a case-by-case basis between Referral Officers, study section Scientific Review Administrators (SRA), Institute programme representatives, and applicants. Although there are official guidelines defining the content and boundaries of science reviewed in each study section, there is often an overlap.

The SRA reads through the proposals, analyses the content and checks for completion so as to decide which study section members are either best suited to review each application or to act as discussants. Usually two or three study section members are assigned to provide written reviews of each application, with one of two additional members to serve as discussants[1.] In special cases, Special Emphasis Panels (SEP) are established on an adhoc basis to review applications requiring special expertise, or due to special circumstances such as conflict of interest.

A week before the convening of a study section, the SRA solicits from all members a list of research project grant applications (R01) believed not to rank in the top half for scientific merit. After the individual lists have been combined, a final list is produced at the outset of the study section meeting.

These so-called "Streamlined" applications are not discussed at the meeting, but summary statements[2] are provided and the applicant may subsequently revise and resubmit the application.

---

1 A chartered CSR study section is composed generally of 18 - 20 individuals, nominated by the SRA from among the active and productive researchers in the biomedical community, to serve for multi-year terms. The purpose of this process is to have the group's combined knowledge span the diversity of the subject matter assigned to the study section for review. In many cases, however, this is hard to accomplish as the study section membership is often supplemented by temporary members and written outside opinions.

2 Summary Statements: a combination of the reviewer's written comments and the SRA's summary of the members' discussion during the study section meeting. They include the recommendations of the study section, a recommended budget, and administrative notes on special considerations.

"Streamlining" is not regarded as disapproval, but represents a decision by the study section that the application would not rank in the top half of applications generally reviewed by that study section.

All regular CSR study section meetings follow the same format to a large degree. Meetings often last two days with members convening around a conference table to maximise interaction. The chairperson, who is a member of the study section, and the SRA sit together and are jointly responsible for conducting the meeting. Observers (programme representatives or other staff from the different NIH institutes) may attend, but do not participate in the discussions. Only after the assigned reviewers and discussants have provided their evaluations are any outside opinions read. Following the general discussion, members mark their priority scores[3] privately for each application on scoring sheets provided by the SRA. At the end of the meeting the SRA or an administrative assistant collects these sheets.

The next step is for the proposals to proceed to the Nation Advisory Council (NAC) of the relevant institute. The NAC of each institute is composed of approximately eight independent scientists and four people from the lay community. The function of this council is to determine whether the proposal received a fair hearing from the study section and provides its own evaluation based on the proposal's relevance to the institute's overall goals. The NAC can only make recommendations on proposal funding and cannot alter priority scores.

---

3 Priority Scores: a numerical rating that reflects the scientific merit of the proposed research relative to the "state of the science".

Feedback to applicants is regarded as very important. The generating of summary statements however, is a lengthy process – it takes up to eight weeks to generate an average of 80 summary statements.

After the summary statements are generated, they are transmitted to the appropriate NIH Institute for funding considerations. At this point the SRA's control over the review of those applications ends. The individual's application now moves to the next grant application cycle, where the Institute programme officials become the applicant's link to the NIH in terms of interpretation of the reviews and the disposition of the applicant.

All the priority score information is entered into the application database, after which the computer-generated priority scores and percentiles are automatically mailed to applicants.

Study sections meet between mid-February and mid-March and summary statements are prepared by late April/May. Institute Advisory Councils, the second step in NIH peer review, meet in May/June to consider the study sections' recommendations and successful applicants usually receive funding several months later.

Although this section describes Research Project Grant applications (RO1/R21), other kinds of grant applications reviewed in the CSR are dealt with in a similar manner, but with some differences. Different kinds of applications (e.g. Small Business Innovation Research (SBIR) and fellowships) receive expedited review and have receipt deadlines one to two months later than RO1s. Further more, Special Emphasis Panels always review SBIRs and fellowships are "streamlined".

3.1.2 REVIEW CRITERIA FOR EVALUATING GRANT APPLICATIONS

Review of grant applications at the NIH is focused on the quality of the science as well as their possible impact on the field. In the rating of grant applications reviewers are required to (a) address the five review criteria listed below and (b) assign a single, global score for each application. The score should indicate the overall impact the proposed research could have on the field based on consideration of the five criteria, with the emphasis on each criterion differing from one application to another, depending on the nature of the application and its relative strengths.

NIH-supported research strives towards the further understanding of biological systems, improving the control of disease, and health. Reviewers are instructed in the written comments to judge whether the proposed research would have a significant impact on the pursuit of these goals. In allocating the overall score, each of the criteria are considered and addressed, weighting them as appropriate for each application. Note that the application does not need to be strong in all categories to be judged likely to have a significant scientific impact and thus deserve a high priority score.

3.1.2.1 The five review criteria

1. *Significance:* Does the study address an important problem? If the aims of the application are achieved, how will scientific knowledge be advanced? What if the aims of the application are achieved, how will scientific knowledge be advanced? What will the effect of these studies be on the concepts or methods that drive this field?

2. *Approach:* Are the conceptual framework, design, methods and analysis sufficiently developed, well integrated, and suitable to the aims of the aims of the project? Does the applicant acknowledge possible problem areas and consider contingency plans?
3. *Innovation:* Does the project employ novel concepts, approaches or methods? Are the aims original and innovative? Does the project challenge existing paradigms or develop new methodologies or technologies?
4. *Investigator:* Is the investigator suitably trained and does he/she have the necessary qualifications to conduct this research? Is the proposed project appropriate to the experience level of the principal investigator and other researchers (if any)?
5. *Environment:* Does the scientific environment in which the work will be conducted contribute to the probability of success? Do the experiments exploit any of the unique characteristics of the scientific environment or employ useful collaborative arrangements? Is there evidence of institutional support?

While the review criteria are supposedly aimed for use principally with unsolicited research project applications (e.g. R01, R29, P01) they also form the basis of the review of solicited applications and non-research activities. However, for specific activities (e.g. construction grants) use of these criteria as stated may not be appropriate.

Apart from the above-mentioned criteria, all applications are also reviewed in terms of the following:

- The suitability of plans to both include genders, minorities, as well as their subgroups as suitable for the scientific goals of the research. Plans for the recruitment and retention of subjects will also be evaluated.
- The reasonableness of the proposal budget and duration in accord with the proposed research.

- The adequacy of the proposed protection for humans, animals or the environment, to the extent they might be negatively affected by the project proposed in the application.

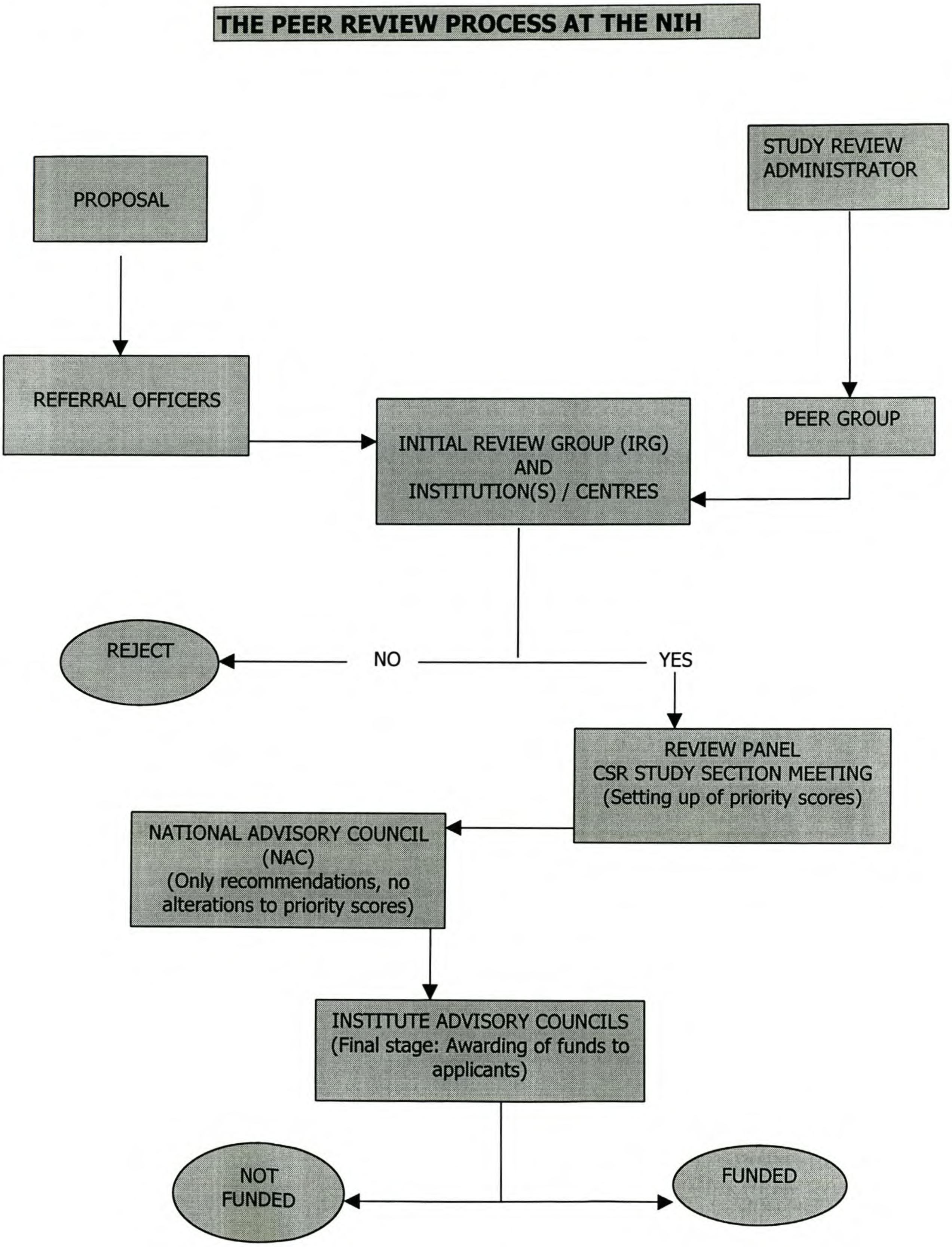
3.1.3 A summary of the peer review process at the NIH for the evaluation of proposals

- The Centre for Scientific Review evaluates the proposals at the NIH.
- Referral Officers assign applications to the appropriate Initial Review Group an also decide which institute(s)/ centre(s) of the NIH would be best suited to fund the application, should it be considered sufficiently meritorious.
- A unique application number is assigned to each application.
- The Scientific Review Administrator (SRA) determines the study section members and discussants best suited to review each application. In some cases Special Emphasis Panels (SEP) are established on an ad hoc basis to review specific applications.
- A week before the convening of a study section, the SRA solicits from all members a list of "streamlined applications", which are not discussed at the study section meeting.
- At the CSR study section meetings, members mark their priority scores privately for each application on special scoring sheets.
- Proposals then proceed to the National Advisory Council (NAC) of the relevant institute.
- NAC determines whether the proposal received a fair hearing from the study section and provides its own evaluation based on the proposal's relevance to the institute's overall goals. NAC may not alter priority scores, but can make recommendations.
- Feedback to applicants is considered as important and is a lengthy process. (On average it takes up to eight weeks to generate 80 summary statements).

- After the summary statements are generated and transmitted to the appropriate NIH Institute for funding considerations, the SRA's control over the review of those applications ends.
- The individual's application now moves to the next grant application cycle. At this point it is the Institute programme officials who become the applicant's link to the NIH in terms of interpretation of the reviews and the disposition of the applicant.
- All priority score information is then entered into the application database. Computer-generated priority scores and percentiles are then automatically mailed to applicants.
- Institute Advisory Councils meet to consider the study sections' recommendations and successful applicants usually receive funding several months later.

A visual representation of the evaluation process at the NIH appears on the next page.

Figure 1


THE PEER REVIEW PROCESS AT THE NIH
PROPOSAL
STUDY REVIEW ADMINISTRATOR
REFERRAL OFFICERS
INITIAL REVIEW GROUP (IRG) AND INSTITUTION(S) / CENTRES
PEER GROUP
REJECT
NO
YES
REVIEW PANEL CSR STUDY SECTION MEETING (Setting up of priority scores)
NATIONAL ADVISORY COUNCIL (NAC) (Only recommendations, no alterations to priority scores)
INSTITUTE ADVISORY COUNCILS (Final stage: Awarding of funds to applicants)
NOT FUNDED
FUNDED

3.2 The peer review process at the National Science Foundation (NSF)

The NSF was established by the National Science Foundation Act of 1950 as an independent agency of the United States government. The Foundation consists of 24 members of the National Science Board and a Director. The President, with the advice and consent of the U.S. Senate, appoints each member of the NSF. Other senior officials include a Deputy Director and eight Assistant Directors. The mission statement of the NSF is to promote the progress of science; to advance the national health, prosperity and welfare; and to secure the national defense.

The cornerstone of the National Science Foundation (NSF) proposal review system is the participation of "knowledgeable peers" from outside the Foundation in the review of proposals. The NSF identifies potential reviewers from several sources such as applicant suggestions; references attached to proposals and published papers, input from mail reviewers, panelists and visiting scientist. With regards to the applicants' suggestions, they are invited to provide a list of names of people they believe to be suitably qualified to review their proposal or individuals they would prefer not review the proposal. Another source used by the NSF is a large and expanding foundation-wide reviewer database from which reviewers are selected.

The judgement of the peer reviewers with respect to the extent to which proposals address established criteria are essential for informing NSF staff and influencing funding recommendations. For this reason the NSF's system of proposal review can best be described as "merit review with peer revaluation".

The source of all the information in this section can be found at the following website: www.nsf.gov

The NSF Proposal Processing Unit receives proposals which are issued to the correct NSF programme for acknowledgement should they meet the NSF requirements for review. Proposals are thoroughly reviewed by a scientist, engineer or educator serving as a NSF program officer, and usually by 3 to 10 other individuals outside the NSF who are experts in particular areas represented by the proposal. Applicants are invited to provide a list of names of people they believe to be suitably qualified to review the proposal or individuals they would prefer not review the proposal. Program officers may, if they chose, receive comments from assembled review panels or from site visits before recommending final action on proposals. Senior NSF staff further review recommendations for awards.

3.2.1 REVIEW CRITERIA AT THE NSF

On 28 March 1997 the National Science Board approved revised criteria for evaluating proposals. The approved criteria are structured to be useful and relevant across the NSF's many diverse programmes. Special criteria will be applied as required to highlight the particular objectives of certain programmes and activities.

The Merit Review Criteria are listed below. Each criterion is followed by potential considerations that the reviewer may apply in the evaluation. These are suggestions and not all will apply to any specific proposal. Each reviewer is requested to address only those relevant to the proposal and which he/she is qualified to evaluate.

Criterion 1:

What is the intellectual merit of the proposed activity?

How important is the proposed activity to advancing knowledge and understanding within its own field or across different fields? How well qualified is the proposer (individual or team) to conduct the project? (If appropriate, the reviewer will comment on the quality of prior work) To what extent does the proposed activity suggest and explore creative and original concepts? How well conceived and organised is the proposed activity? Is there sufficient access to resources?

Criterion 2:

What are the broader impacts of the proposed activity?

How well does the activity advance discovery and understanding while promoting teaching, training and learning? How well does the proposed activity broaden the participation of under-represented groups (e.g. in terms of gender, ethnicity, disability, geographical region, etc.)? To what extent will it enhance the infrastructure for research and education, such as facilities, instrumentation, networks and partnerships? Will the results be disseminated broadly to enhance scientific and technological understanding? What may be the benefits of the proposed activity to society?

3.2.2 The review process at the NSF

External peer review is obtained primarily by two methods: mail and panel reviews. Site visits by NSF staff and external peers are frequently used in addition to mail and panel reviews to review proposals for large facilities, centres and systemic reform initiatives. NSF programme officers can use their discretion in the specific use of review methods, subject to supervisory approval.

In "mail-only" reviews peers are sent proposals and required to submit written comments to NSF by postal mail, facsimile, electronic mail or through Fast Lane, NSF's Web-based system for electronic proposal submission and review. The NSF programme officer then uses these mail reviews directly to support a recommendation for award or decline.

"Panel-only" reviews refers to that process of soliciting reviews only from those peers who meet in a panel review setting to discuss their reviews and provide advice directly to the programme officer. Most programmes that use this process mail proposals out to panelist and receive their reviews prior to the panel meeting. Other programmes provide panelist with access to the proposals at the start of the panel meeting, allowing a period of time during which they prepare their reviews at the meeting.

Most proposals submitted to NSF are reviewed using some combination of these two processes (mail and panel review). The programmes that employ the mail-plus-panel review process have developed several different configurations, such as:

- A peer is required to submit a written mail review and also serve as a panelist in basically contributing two reviews for each proposal; and
- A peer is asked to participate only as a panelist, with responsibility only for reviewing and discussing mail reviews written by others and providing verbal and/or written advice to the programme officer.

Note: One of the conditions stated by NSF policy is that each recommendation for final action on a proposal must be accompanied by at least three external reviews, unless the requirement has been explicitly waived.

3.2.3 Reviewers

A fundamental component of the NSF merit review system is the diversity of the reviewer pool. This diversity assists in ensuring that a wide range of perspectives is taken into account in the review process. The selection of a review panel is the responsibility of the Programme Director. In selecting the panel the Director has to consider the following guidelines.

1. Scientific expertise: Does it match the science represented in this round of proposals?
2. Diversity: Keep a balance in professional development areas.
3. Distribution: In terms of age, geographical distribution, institutional type, gender and ethnicity.
4. Industrial representation is encouraged, if possible.

Several processes are in place to assist in reviewer diversity such as explicit policy guidance, the use of a large and expanding Foundation-wide reviewer database[4], mandatory training for all program officers, and directorate initiatives.

Potential reviewers are identified from several sources including applicant suggestions, references attached to proposals and published papers, and input from mail reviewers, panellists and visiting scientists. Participation in the peer review process is voluntary. Panellists are reimbursed for expenses; mail reviewers receive no remuneration.

---

4 NSF maintains a central electronic database of 250,000 reviewers.

The written narratives provided by reviewers, the deliberations by panel members, and the expert opinions provided by program officers all are important components of the merit review system where no particular component is allowed to take priority over the others.

3.2.4 Reviewer proposal ratings

Within the NSF merit review system greater importance is placed on reviewer narratives than on summary ratings. Summary ratings represent only one indicator of reviewer judgement of the proposal quality.

3.2.5 NSF programme officers

The narrative comments and summary ratings provided by external reviewers are fundamental inputs in the NSF's merit review system. Upon receipt, these inputs guide the evaluation of the programme officers who make awards and decline recommendations to NSF's senior management.

Program officers are scientists, engineers and educators from whom NSF seeks to obtain expert judgement and programme management. In recommending to award or decline proposals, these expert individuals produce and manage a portfolio of awards addressing the NSF's strategic goals and associated factors such as:

- Contributions to human resource and institutional infrastructure development;
- Support for "risky" proposals with potential for significant advances in a field;
- Encouragement of interdisciplinary activities; and
- Achievement of programme-level objectives and initiatives.

Despite the increase in proposal pressure and general workload the number of programme officers employed by NSF has remained steady at just over 400 for the past five years. Programme officers are classified as assistant programme director, associate programme director, or programme director, depending on their professional experience. They can be either permanent or temporary NSF employees. Some temporary officers are "on loan" as visiting scientists, engineers and educators for up to three years from their host institutions. Other are employed through grants to the home institutions under the terms of the Intergovernmental Personnel Act.

Each programme officer's recommendation to award or reject a proposal is dependent on a programmatic review by a higher-level reviewing official (usually the division director) and an administrative review by a grants officer in the Office of Budget, Finance, and Award Management (BFA). All award recommendations that exceeds $1,5 million in any one project year of $6 million over five years must be reviewed by the Director's Review Board (DRB). Awards exceeding a $3 million commitment during a project year, or $15 million over five years, require approval by the National Science Board.

3.2.6 Summary of the Merit Review Process at the NSF

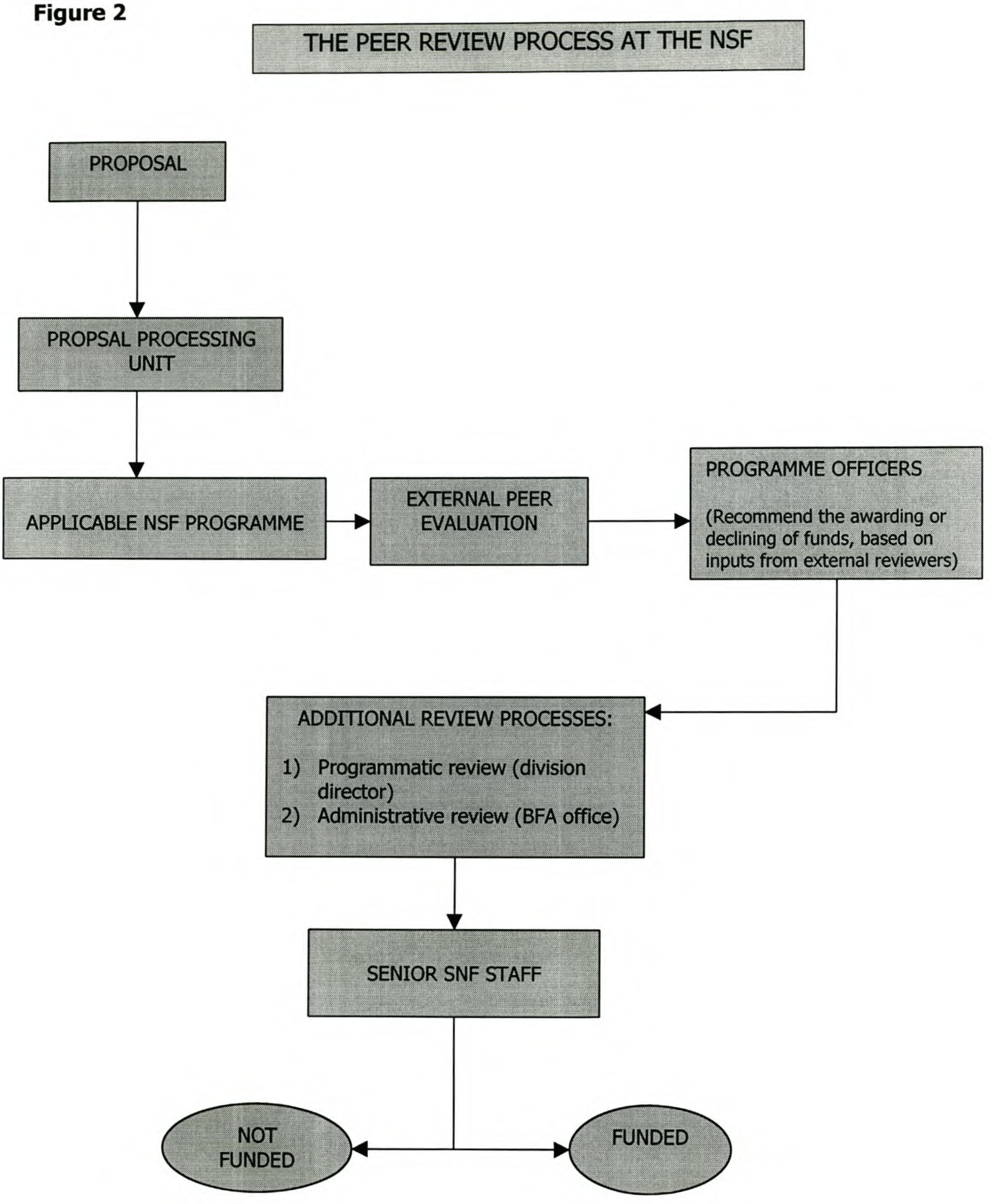
- Proposals received by the NSF Proposal Processing Unit are issued to the correct NSF programme for acknowledgement and, if they meet the NSF requirements, for review.
- Directorates use one of the following methods to review proposals:
  - Mail-only reviews
  - Panel-only reviews
  - Mail-plus-panel review
- A least three external reviewers evaluate the proposal according to the following criteria:

- What is the intellectual merit of the proposal activity?
- What are the broader impacts of the proposed activity?

- The reviewer produces reviewer narratives and summary ratings in the evaluation of the proposal quality.
- The narrative comments and summary ratings provided by external reviewers are fundamental inputs in NSF's system of merit review. Upon receipt, these inputs guide the evaluation of the programme officers who make awards and decline recommendations to NSF's senior management.
- Programme officers may, if they choose, receive comments from assembled review panels or form site visits before recommending final action on proposals.
- In making recommendations to award or decline proposals, programme officers produce and manage a portfolio of awards addressing NSF's strategic goals and associated factors.
- Each programme officer's recommendation to award or reject a proposal is dependent on a programmatic review by a higher-level reviewing official (usually the division director) and an administrative review by grant officers in the Office of Budget, Finance, and Award Management (BFA)

A visual representation of the review process at the NSF appears on the next page.

Figure 2


THE PEER REVIEW PROCESS AT THE NSF
PROPOSAL
PROPSAL PROCESSING UNIT
APPLICABLE NSF PROGRAMME
EXTERNAL PEER EVALUATION
PROGRAMME OFFICERS
(Recommend the awarding or declining of funds, based on inputs from external reviewers)
ADDITIONAL REVIEW PROCESSES:
1) Programmatic review (division director)
2) Administrative review (BFA office)
SENIOR SNF STAFF
NOT FUNDED
FUNDED

3.3 The peer review process at the Medical Research Council of Canada (MRC)

The MRC is a federal agency responsible for funding bio-medical research in Canada. This agency aims at promoting, assisting and undertaking of basic, applied and clinical research in Canada in the health sciences. It plays a major role in supporting research training of health scientists and functions as an advisory body on health research to the federal Minister of Health as well as promoting co-operation between university and industry. The council supports research conducted by scientists in universities, hospitals and research institutes across Canada and does not employ its own scientists nor does it run laboratories of its own.

3.3.1 Committee and Council policies

The governor in Council appoints members of the Medical Research Council. The MRC Council controls the expenditure of the Council's funds, it establishes programmes and policies, and determines the committee structures, membership, policies and procedures. Although committee members and staff work within policies determined by the Council, committees can and will provide opinions on Council policies. Committees may not, however, take actions that go against the Council's established policies.

The source of all the information in this section can be found at the following website: www.mrc.ac.uk

The Chairperson and Scientific Officers have the responsibility of ensuring that the most appropriate committee for the evaluation of applications is compiled. The duties of the Chairperson and Scientific Officers are to:

- Identify external referees for each application in their committee;
- Identify the committee members who shall review each application;
- Propose additional committee members for a meeting if they feel that the committee does not have adequate expertise for one or more applications.

MRC committees are requested by the Council to review applications and to make recommendations on their relative merits, after which the Council decides on the applications that are to be supported.

3.3.2 The evaluation process

The evaluation process at the MRC takes place on four levels. Two weeks after the deadline for proposal submission the Chairperson and Scientific Officers of the Grand Committees send each application to a specific Grand Committee, to assign external and internal referees and to assign the applications to Committee members. The first and second levels of evaluation entail sending proposals to external and internal reviewers for evaluation. The third level is evaluation by committee members, who receive copies of all the applications as well as copies of all the external referee reports. The fourth and final level is where the MRC council meets to consider the committee recommendations to reach their final decisions.

3.3.2.1 Chairperson

The committee chairperson is directly responsible to Council to ensure that the committee operates well, effectively and objectively according to Council's policies for the grants programme and its committees. He/she creates a positive, constructive, fair-minded environment in which the research proposals are to be evaluated.

The chairperson:

- Must be familiar with MRC influencing the procedures of MRC Grants Committees;
- Ensures that Grants Committees function in line with MRC policies and procedures;
- Ensures that each application receives equal attention by the Committee;
- Ensures that opinions expressed by external referees are fully integrated into the discussion of each application;
- Ensures the involvement of the entire committee with recommendations for each grant;
- Summarises, together with the Scientific Officer the discussion of each application before the rating;
- Ensures that particular functions of ethics, and other Council obligations are addressed;
- Works together with the Scientific Officer to assign applications to committees and referees.

3.3.2.2 Chairpersons and Scientific Officers

Chairpersons together with Scientific Officers are tasked with the responsibility of ensuring that the most appropriate committee is selected to evaluate the applications received by the Council. Upon accepting an application for review by their committee, it is they who accept responsibility for ensuring that the committee performs the review.

The responsibility of chairpersons and scientific officers is to:

- Identify external referees for each application in their committee;
- Identify the committee members who shall review each application;
- Propose additional committee members for a meeting if they feel that the committee does not have adequate expertise for one or more applications.

3.3.2.3 Scientific Officer: Additional duties

In addition to the duties shared with the Chairperson the Scientific Officer:

- Supports the Chairperson in his/her role during the committee meeting;
- Assumes the chair when the Chairperson cannot be in the room for reasons of conflict of interest;
- Presents points raised by the external referees that have not been raised by the internal reviewers and identifies the external referees;
- Takes notes of the discussion as it is proceeding which will be sent to applicants with their letter of decision;
- Works with the Chairperson to present a brief verbal summary of the discussion on each application before the application is rated;
- Ensures that issues of ethics, eligibility and other concerns that need to be flagged for the attention of Council are raised at the committee.

3.3.2.4 The Grants Committee

The MRC Grants Committees has two main functions: the evaluation and the rating of applications submitted for a specific competition. Proposals are rated so that they can be ranked in order of priority and also to recommend funds needed to support the research if the application is approved.

The committees flag applications with special concerns for ethical issues, laboratory safety, eligibility, compatibility with Council's policies, etc. These concerns, however, are not intended to prevent a Committee from evaluating an application it has received and, unless they directly influence the scientific merit, neither are they intended to influence the rating. When making the final decision on whether or not to approve the grant, however, the Council will take these concerns into account.

Committees also have to identify proposals that represent new, emerging areas of research with the potential to attract private sector funds and/or are appropriate for public interest stories. Again it should be noted that this information should not affect the rating of the application, but is intended solely for planning and communications purposes.

Council considers recommendations from Committees in reaching its final decisions. Usually the Council does not depart from the priority recommendations of the Committee, and changes are only made to budget recommendations or budgetary restrictions.

3.3.2.5 External and internal referees

The guidelines by which referees are to review proposals are based on the MRC's scientific banding system. The definitions for this system are as follows:

Bands	Description
ALPHA-A	Work which is at the forefront internationally, or nationally where there are no international comparators, and which is judged to possibly have an important and substantial impact on understanding, practice or policy.
ALPHA-B	Work which is at the forefront of the UK effort in the field, is internationally competitive in a significant proportion of the research proposed – where such comparisons can be made, and will make a significant contribution
ALPHA-C	Work which is nationally competitive and will make valuable contributions to addressing important scientific and/ or policy questions
ALPHA-D	Work which is nationally competitive but which is at a lower priority in the competition for funds.
DECLINE	Reject: (1) work which would add to understanding but is not competitive; (2) work which is judged to be seriously flawed in design; (3) work which is not suited to the form of support requested.
Not Applicable	For use for schemes were reviewers are not asked to band the application e.g. Programme Grants. Also for use by overseas referees who are not asked to submit a banding.

External referees as well as members of the MRC Advisory Board also follow the Scheme-specific Assessment Factors to assess applications. These Scheme-specific factors/ research funding are:

- Programme Grants
- Co-operative Group Component Grants
- Co-operative Group Development Grants

- Strategic Grants
- Centre Grants
- Centre Component Grants (and Centre Development Component Grants)
- Trial Outlines
- Trial Full Proposals
- Co-operative Group Core Grants.

For all scheme-specific factors/ research funding schemes there are assessment factors according to which the referees order their evaluation. Although the assessment factors are the same for all the scheme-specific factors (there are minor variations), their definitions differ according to the objectives of each. The assessment factors are:

- Summary of assessment of the proposed research
- Significance of the topic
- Details of proposal
- People (i.e. researchers involved in proposed research)
- Environment
- Value for money
- Ethical and other implications
- Renewal where relevant (Have the aims and objectives of research been met?)
- Public understanding of science
- Commercial exploitation
- Dissemination of research results.

More questions to be taken into account by the referees when reviewing a proposal are listed below. It must be understood, however, that referees and Committees will consider questions such as these differently from one application to another.

The questions are:

- How important and/or original is the hypothesis or the questions to be addressed, and how clearly are they formulated?
- How important and original are the contributions expected from the research proposed? What is the possibility for significantly new observations or knowledge?
- How well will the proposed experiments address the hypotheses of questions?
- How suitable are the methods to be applied and the proposed analysis of data?
- How well will the applicant implement new methods (to science or to the applicant) which are to be introduced and/or explored? How well have the applicants anticipated difficulties in their research and developed contingencies?
- How critically is the relevant literature covered and evaluated?
- With regards to the applicant's productivity, experience and training the question of how suitable is the training and/or track record of the applicant(s) to the proposed research? How significant and original is the recent productivity of the applicant(s)? How much confidence is there that the applicant can do the work proposed?

The internal referee report has to briefly describe the aim(s) of the project and give a description of the project in no more than 60 – 100 words. The report must provide an evaluation of the application in a clear and concise manner, employing objective and non-inflammatory language, while at the same time stating opinions together with reasons for the opinions held. The internal referees must also comment on views of the external referees indicating areas of agreement and disagreement again with clearly stated reasons for disagreements. It is very important that the statements of the external referee are not merely repeated by the internal referee as the applicant receives all referee reports.

The internal referee report must also summarise views of the strengths and weaknesses of the application as well as comment on the past productivity of the applicant. If necessary internal reviewers have to comment on factors such as ethical acceptability of the research, the presence and suitability of any ethics forms, possible overlaps of the application with other sources of funding, possible issues of eligibility or any other issues. These concerns, however, should not affect the rating of budget recommendations, unless they relate to the scientific merit of the application.

3.3.2.6 The evaluation process by the Committee

All Committee members must be familiar with the aims of the projects, the experience of applicants of all applications to be evaluated by their Committee as well as with the external referee reports briefly summarised by the internal referees.

The evaluation of each application starts with internal reviewers announcing their initial ratings to one decimal place. All applications rated 2.5 or below by the internal reviewers and about which the external reviews are usually in agreement, will not be discussed further by the committee unless a committee member requests it (triage). In these cases, the applicant will receive a copy of all internal and external reviews, but there will be no Scientific Officer notes. Committee members do not vote on the rating; it is calculated as the mean of the initial ratings of the two internal reviewers.

Unless an application has been triaged, discussions proceed with the first internal referee presenting his/her evaluation of the application. During this presentation the strengths and weaknesses, extent of the originality, applicant's training, experience and recent productivity are discussed. The second referee addresses the same issues, focusing on points of agreement and disagreement with the first and elaborates on points that may not have been addressed by the first referee. One of the internal referees or the Scientific Officer then addresses the comments of the external referees, detailing areas of agreement and disagreement with the internal reviewers. The reader may want to comment further on issues that have already been raised or may want to raise additional issues as appropriate.

The Chairperson leads the discussion of the committee members to ensure that both the quality and originality of the research proposal as well as the recent productivity of the applicant(s) are considered; this includes views expressed in the external referee reports. The Scientific Officer has the responsibility of introducing points that were not satisfactorily addressed by the two internal reviewers.

The Scientific Officer or Chairperson summarises the discussion, briefly identifying the strong and weak point(s) in the proposed research and commenting on the recent productivity of the researchers whose proposals are under review. Following the discussion summary, the Chairperson seeks a "consensus rating" from the two internal reviewers. If a consensus cannot be reached, the mean value of the two ratings is used (rounded off, if necessary, to obtain a single decimal point). All committee members, including the two internal reviewers, then cast individual confidential votes within +/- 0.5 of the consensus rating. To ensure consistency, Council expects committees to comply with a common scale and apply the same convention in assigning ratings. To facilitate this, the following scale and descriptions should be applied:

Range	Descriptors
4.5 – 4.9	Outstanding
4.0 – 4.4	Excellent
3.5 – 3.9	Very good
3.0 – 3.4	Solid/significant research
2.6 – 2.9	Needs revision
2.0 – 2.5	Needs major revision
1.0 – 1.9	Seriously flawed
0	Not acceptable

A budget is recommended and the term of support for the grant is established. Particular concerns of eligibility, ethics, overlapping sources of funds or other points are discussed and, if necessary, flagged for MRC staff to address.

The final review of the entire meeting and a discussion of policy present a key element of any committee meeting. Committee members are thus obliged to stay for the entire meeting and staff and the Committee Chairperson should provide advance guidance on the probable length of the meeting.

3.3.2.7 The role of the MRC staff in the evaluation process

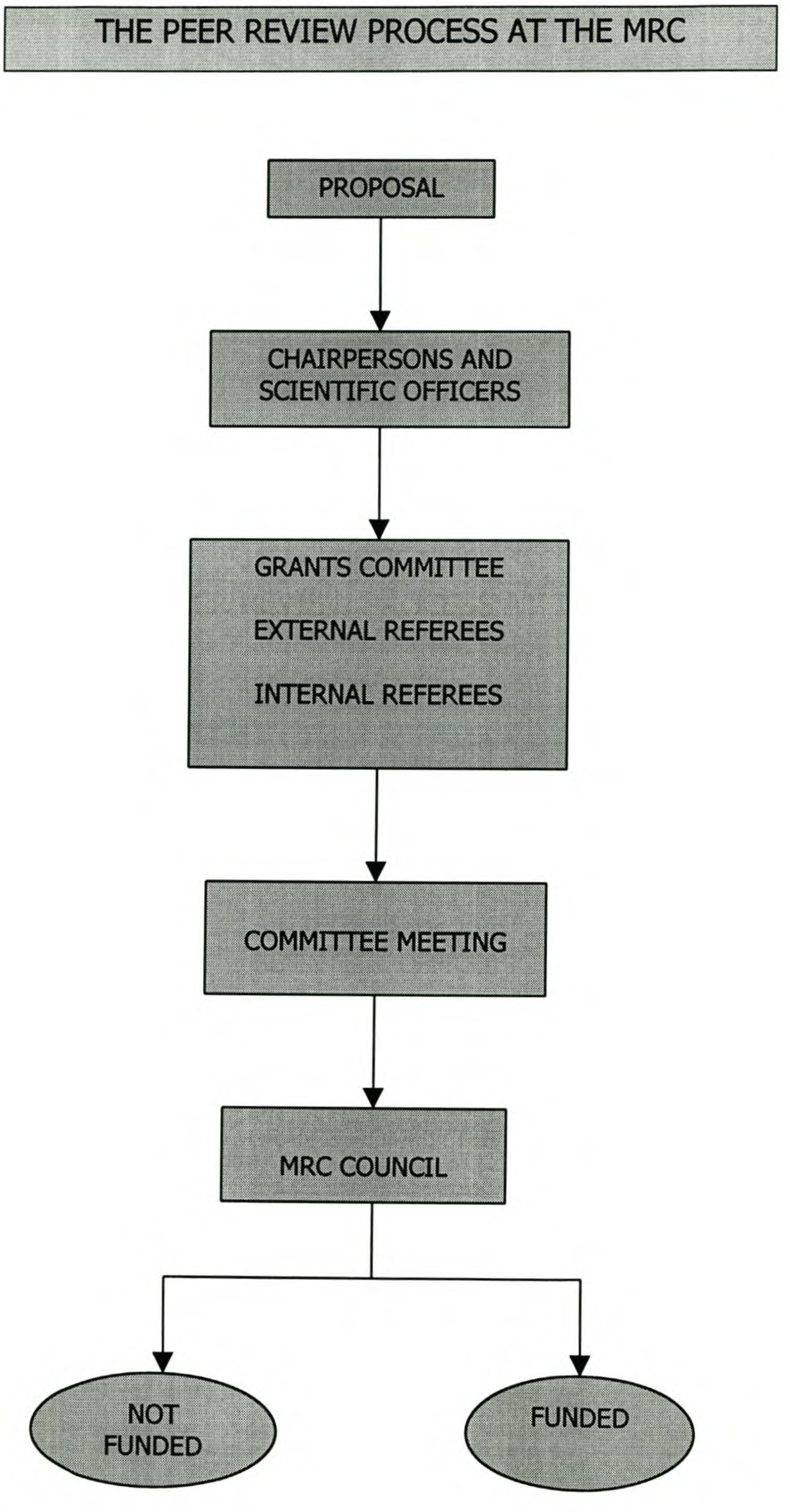
The role of the MRC staff entails dealing with the four aspects that they are responsible for. In the first place the MRC staff has to give advice to the Committee specifically with regards to Council policies. Secondly, they keep notes on the procedural aspects of the Committee's functions. Their third responsibility is to record the rating category and budget recommendations made by the Committee for each application. Their fourth and final responsibility is to note problems raised by the Committee on issues requiring later attention by staff (i.e. overlap of funding, ethics and eligibility).

3.3.3 A summary of the process of peer review evaluation for grant applications

- All applications received by the specific deadline date are entered into the competition. (Material arriving after the deadline will only be considered if it relates to the outcome of manuscripts submitted with the application, or of other applications submitted by the applicant.
- Two weeks after the deadline date the Chairperson and Scientific Officers of the Grant Committees meet to:
  - Assign each application to a specific Grants Committee;
  - Assign external referees (3 or more) to whom the application will be sent for written reviews;
  - Assign the application to three Committee members: two internal referees who will lead the review of the application at the Committee meeting, and a reader who will read the application in the "red book", but need not provide a written review. The reader will act as a discussant as necessary.
- Applications are then sent to external referees and committee reviewers by MRC staff. Steps are taken to ensure that at least one external review is received for each application. Internal and external referees receive copies of the applications assigned to them. Each committee member receives all applications (excluding appendices) to be considered by the Committee and copies of all the external referee reports. Committee members receive applications for which they are to act as internal referees at least six weeks before the committee meeting and books containing all the applications before their committee at least two weeks before the meeting. External referee reports are sent as far ahead of the meeting as possible; however, some arrive late and may be available only on the day of the meeting.
- Council considers the Committee recommendations as soon as possible after the committee meetings to reach their final decisions.
- Applicants are usually informed within three weeks after the Council meeting.

A visual presentation of peer review process at the MRC appears on the next page.

Figure 3


THE PEER REVIEW PROCESS AT THE MRC
PROPOSAL
CHAIRPERSONS AND SCIENTIFIC OFFICERS
GRANTS COMMITTEE
EXTERNAL REFEREES
INTERNAL REFEREES
COMMITTEE MEETING
MRC COUNCIL
NOT FUNDED
FUNDED

3.4 The peer review process at the National Sciences and Research Council of Canada (NSERC)

The NSERC is a national institution that invests in Canada's science and technology capacity. It supports both basic university research through research grants and project research through partnerships of universities with industry, including advanced training of highly qualified people in both areas. The NSERC offers three types of research grants, each covering different aspects within the research scene. *Research Grants* support ongoing programmes of research. In these grants the creativity and innovation of research advances, whether made individually or in groups, are recognised. *Equipment Grants* are specifically aimed at enhancing research and research training capabilities of university researchers by supporting the purchase and installation of research equipment. *Major Facilities Access (MFA) Grants* support researchers' access to major regional or national research facilities.

The NSERC is a separate employer of the Government of Canada and reports to Parliament through the Minister of Industry. The council is governed by the university and consists of 22 members selected from the private and public sectors. Within the council there are three overall committees: 1) *Selection committees* are comprised of hundreds of volunteers. These selection committees are responsible for the awarding of grants and scholarships through a competitive peer review process. 2*) Standing Committees* serve as advisory bodies to the NSERC on particular policies and also oversee the work of the selection committees. 3) *The Executive Committee* of the council acts as a planning, allocation, internal audit and personal committee between council meetings.

The source of all the information in this section can be found at the following website: (www.nserc.ca)

3.4.1 Procedures for the evaluation of the eligibility of researchers for funding

For applicants to be eligible for funding the applicant must hold, or have a firm offer of, an academic appointment at a Canadian university for a tenured, tenure-track or life-time Professor Emeritus position, or a term position of at least three years.

The appointment must be ratified by the person(s) body responsible for approving academic appointments or their delegate(s) and must be in accordance with university statues. The position held by the applicant must enable him/her to conduct research that is not under the direction of another individual as well as authorise the applicant to supervise or co-supervise students registered in an undergraduate degree programme or post-doctoral fellows. The applicant must conduct research in the natural sciences or engineering and must not hold a full-time position (academic or other) outside Canada. It is also required that the applicant must not be under an NSERC sanction resulting from a finding of financial or scientific misconduct.

At the time of accepting the reward the applicant must not be enrolled in a graduate programme in the natural sciences or engineering. Neither must the applicant be paid out of NSERC (or other federal granting council) grant funds or hold federal granting council fellowships or scholarships. An exception to these rules occurs where the person applying for a grant is a co-applicant with no Canadian academic appointment and is involved in collaborative research or group grants (it is expected that the applicant will be qualified to conduct research independently). The applicant, however, is not eligible to receive NSERC funding and is expected to provide his/her own resources in the collaboration.

3.4.2 Application procedures

The first step to be taken when applying for a research grant is to submit a "Notification of Intent to Apply", including a list of the applicant's research contributions for the last six years. For Subatomic Physics Project Research Grants requesting $500,000 or more, the following must then be submitted:

- An application for the grant and a Personal Data Form for the applicant and each co-applicant;
- Samples of contributions such as reprints, preprints and/or manuscripts, excerpts from thesis, technical reports, etc. These documents should be chosen to represent the applicant's most significant recent contributions, or those most relevant to the proposed work in the last six years as they will be used as an indicator for the reviewers as to who is to assess the quality of the applicant's work;
- The list of samples of contributions the applicant is submitting with his/her application.

3.4.3 Review process and selection criteria

Grant selection committees (GSC) review applications with input from external referees when necessary.

Applications are assigned to grant selection committees on the basis of the research field, the aims of the proposed research programme, and input from applicants and committees. Although the NSERC makes the final decision, the applicant is allowed to suggest to committee whom she/he believes to be most appropriate to assess his/her application.

In cases where the nature of the research is interdisciplinary, a process is put in place to ensure a suitable review of all components of the application and if necessary, additional interdisciplinary or multidisciplinary committees are established.

Applicants cannot submit the same proposal concurrently to both the NSERC and any other Canadian science council such as the Social Science and Humanities Research Council (SSHRC) or the Medical Research Council (MRC).

The selection criteria used by reviewers/referees to evaluate proposals are:

1. Excellence of the researcher;
2. Contributions to research;
3. Assessment of the quality of the researcher;
4. Contributions to collaborative research;
5. Contributions to training of highly qualified personnel;
6. Merit of the proposal.

Excellence of the researcher refers to the applicant's and co-applicant's ability and track record over the past six years by focusing on their contributions to and impact on the speciality profession. These include significant research contributions and other evidence of contributions to the profession or impact on the field.

To assess significant research contributions selection committees and panels evaluate the quality and overall impact of the contributions made by the applicant(s) and co-applicant(s). In addition to refereed journal publications, conference proceedings, monographs, books and other contributions to engineering and applied sciences can be considered and evaluated according to their quality and significance (i.e. patents protecting technological products, processes or services are fundamental indicators of creative applied research). Excellence and impact on the field of speciality can also be demonstrated by such activities as membership on code, selection from other committees, advisory boards, boards of directors and research interactions.

Contributions to research are measured by using dissemination of results as an indicator. The research process is not complete until the results are validated and transmitted to a suitable audience. For many specialities the most common and effective means of disseminating results is through publication of articles in refereed journals. Other means of dissemination do exists, however, and it is up to the researcher to select the most suitable means to ensure maximum impact on his/her field.

The NSERC recognises that there are occasions where it is impossible or undesirable for researchers to publish significant results of their research prior to reapplying for NSERC support. It is also recognised that research productivity may differ during periods of pregnancy or early childcare, whether or not a formal leave of absence is taken, or as a result of other personal circumstances. Administrative leave, disability and other situations may also result in publication delay. Applicants should clearly detail any circumstances that influence dissemination of research results on the Personal Data Form. NSERC advises its committees to be sensitive to the impact of these circumstances on the level of productivity, while maintaining that the quality of the research remains competitive. Each case is reviewed on its own merits.

The best means to assess the quality of any research contribution is to test its significance and use by other researchers and end users by determining the extent to which it affects their direction of thought and activity in the target community. Although arbitrary, this assessment is a key feature of peer review.

To *allow evaluation of excellence of the researcher* on the quality and impact of his/her recent contributions to research, applicants are required to identify their five most significant contributions in the last six years. They also have to describe how these contributions have influenced their field and/or the activities of users.

A few researchers conduct most of their work in a *collaborative mode*. Publications are usually prepared jointly with students, postdoctoral fellows and other researchers. Applicants are required to describe their intellectual contribution to collaborative work or joint publications in their Personal Data Form. The assessment of researcher excellence must fully consider the overall quality and impact of these collaborative activities.

Advanced training in science and engineering is a key component of university research. The NSERC's mission is to advance the discovery and application of knowledge, thus the emphasis is on the researcher's *contribution to the training of highly qualified personnel.*

Individuals trained in science and engineering are favourably situated to profit from the new ideas and technologies developed in Canada as well as internationally, irrespective of the sector in which they are employed. Professionals in science and engineering contribute to Canada's understanding of the natural and physical environment, leading to improvements in the standard of living and quality of life for Canadians.

Training supported by NSERC ranges from undergraduate theses and summer projects to the postdoctoral level and include technical and other research personnel. The level and content of the training should be suitable to the research speciality, with opportunities for interaction and collaboration with other researchers inside and outside the university where necessary. Undergraduate student participation in final-year projects and summer projects forms a key part of research training and plays a primary role in encouraging excellent students to pursue research careers. For technicians and others who have been in long-term positions, gaining new techniques and knowledge is a significant contribution to training. NSERC is aware that not all research is suitable for training and that there will be circumstances when training will not be possible. In these cases, the onus is on the applicant to provide an explanation of the absence of a training aspect.

The fact that an applicant has trained, is training or plans to train students, technicians or post-doctoral fellows is not in itself a necessary criterion for awarding a grant. A researcher's contribution to training will be evaluated in terms of its quality and impact, and not merely in terms of the number of people supervised. Moreover, the applicant and the research proposal must meet the standards of excellence of the programme.

The following questions will serve as a guide for evaluating the applicant's involvement in research training:

- Have the resulting contributions been of high quality?
- Have the students and other personnel gone on to further research training positions (e.g. Ph.D. programme, postdoctoral position)?
- Have the people trained by the applicant proceeded to become respected professionals in disciplines related to science and engineering in all sectors?
- Examples of professional contributions:

a) Transferring new knowledge and expertise from the universities to the Canadian private sector;
b) Starting businesses, creating jobs and new economic opportunities;
c) Maintaining Canada's international competitiveness in research in science and engineering, renewing Canada's intellectual resources;
d) Developing and implementing policies, standards and regulations on issues of national interest; or
e) Maintaining and enhancing the national framework for competitive R&D through teaching, administration and research dissemination.

- In the context of the research field and the applicant's capabilities, is the past level of training activity appropriate? If not, has appropriate justification been provided?

From information provided in the application, the following questions must be answered:

- Are the projects feasible and appropriate for the training proposed?
- What opportunity will there be for trainees to contact, collaborate or train with other sectors, if appropriate?
- What opportunity will there be for training in a collaborative or interdisciplinary environment, if appropriate?
- Will trainees be able to make an original contribution to knowledge?
- If little or no training is planned, has an appropriate justification been given?

To *assess the merit of the proposal* the following factors are taken into account:

- ❖ Originality
- ❖ Anticipated significance
- ❖ Clarity of long- and short-term objectives
- ❖ Suitability of proposed methodology, and

- Feasibility.

The major focus is on the stage of innovation and the potential to make an important contribution to the area. In engineering and the applied sciences consideration is given to the potential for technological impact and the extent to which the proposal deals with the potential applications of the research.

Applications dealing with a particular product or process should indicate collaboration or interaction with industry or a government department, and the type and extent of the expected benefits. These should be documented in the application form.

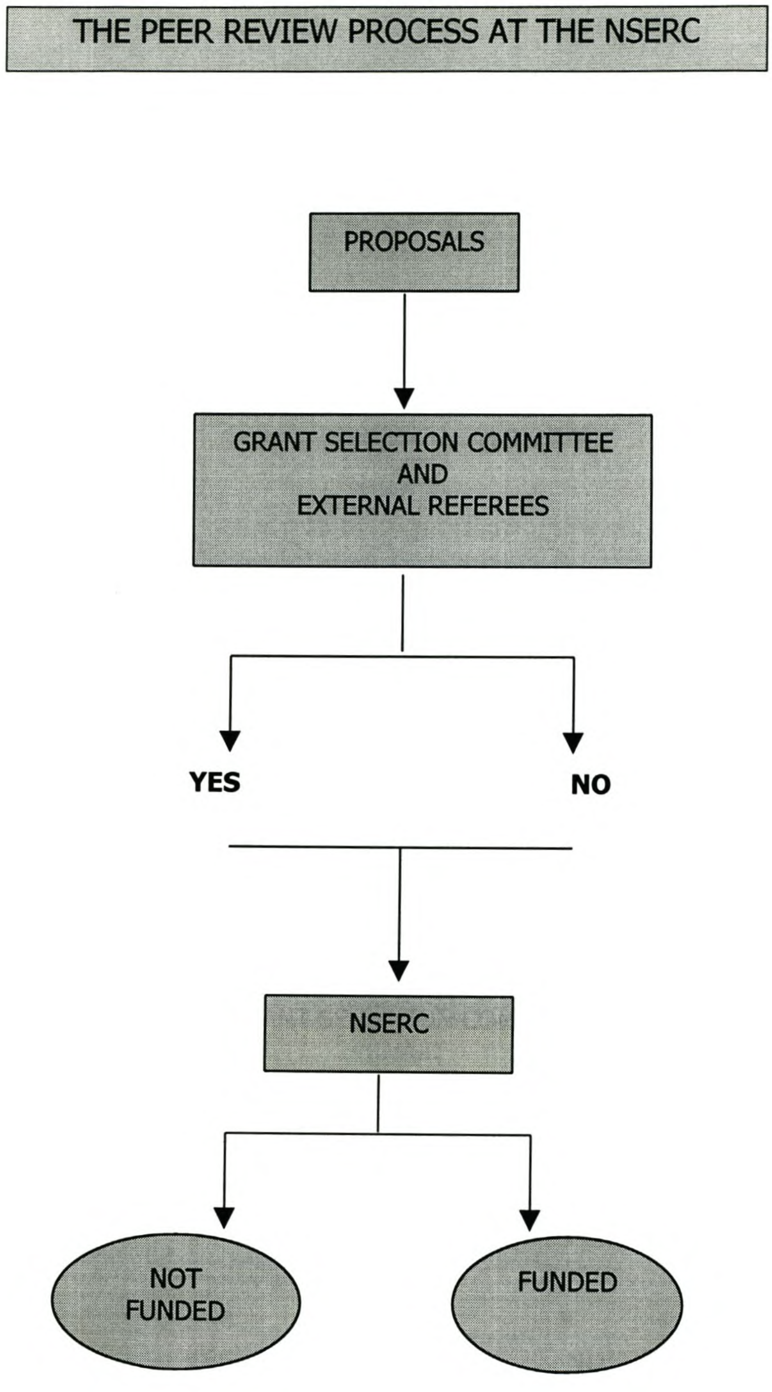
3.4.4 A summary of the peer review process at the National Sciences and Engineering Research Council of Canada

- Grant selection committees review applications. There may also be input from external referees.
- NSERC assigns applications to grant selection committees on the basis of the research field, the aims of the proposed research programme, and input from applicants and committees.
- Selection committees evaluate applicants according to the following criteria:
  - Contributions to research
  - Assessment of quality
  - Contributions to collaborative research
  - Delays in the research and dissemination of results
  - Contributions to training of highly qualified personnel
  - Merit of the application.

- When assessing the merit of the proposal the following factors are taken into account:
  - Originality
  - Anticipated significance
  - Clarity of long- and short-term objectives
  - Suitability of proposed methodology, and
  - Feasibility.

A visual representation of the evaluation process at the NSERC appears on next page.

Figure 4


THE PEER REVIEW PROCESS AT THE NSERC
PROPOSALS
GRANT SELECTION COMMITTEE
AND
EXTERNAL REFEREES
YES
NO
NSERC
NOT
FUNDED
FUNDED

3.5 Comments on the case studies

At the beginning of this chapter I referred to three aspects of peer reviewing which are in my view, intrinsic to the quality of the peer review process. Keeping the above four case studies in mind, a few brief comments on these three aspects follow.

3.5.1 The selection of the reviewers

During an ARC workshop on peer review the workshop members had, amongst other things, the following question: who should be an assessor? The workshop members came to the conclusion that the most appropriate assessor/ reviewer is the scientist who is both sufficiently experienced and is still actively involved in research. It was also considered important to select a review team with complementary skills and backgrounds (ARC Report, 1997).

In the case studies it is clear that all four funding agencies do their utmost to ensure that the best reviewer is selected for the review of the proposal at hand. Funding agencies such as the NSF allow reviewees to suggest possible reviewers to review their proposals. The NSF as well as the MRC makes use of internal as well as external reviewers. The NIH's alternative to external reviewers is that of discussants. The importance of external reviewers or discussants is to get as wide a range as possible of input with regards to the quality and relevance of a proposal. This also enhances, the ARC workshop members suggested, the possibility of convening a review team with complementary skills and backgrounds. All this is done to ensure a review process of the utmost quality.

3.5.2 Review Criteria

Review criteria are probably the most important aspect of, as well as an absolute requisite for, the peer review process. Review criteria provide the reviewer with guidelines along which he/she is to review the proposal and make a judgement on its quality and relevance.

Although there are aspects common to the review criteria of different funding agencies, there are also aspects of the criteria unique to every funding agency. These "unique aspects" in most cases correspond with the respective funding agencies' policies and mission statements. The MRC, for example, has as part of its mission statement the promotion of public engagement with medical research. This correlates with one of the assessment factors reviewers have to keep in mind when reviewing proposals for programme grants, i.e. the public understanding of science. The "public understanding of science" basically entails that the proposal must show how the proposed research will promote public understanding of science sufficiently in order it to be reviewed as having merit to be funded by the MRC.

3.5.3 Number of review stages within the review process

Within a review process there is often more than one review stage. The advantages of making use of multiple review stages are twofold. On the one hand, it ensures that a proposal receives reviewing of the highest quality in that it is studied in depth by a number of reviewers with complementary backgrounds and skills. On the other, it ensure that all aspects of the proposal receive the same amount of attention as the different stages of the review focus on different aspects of the proposal.

The review process at the NSF is a very good example of a review process that is comprised of multiple review stages. The review process of the NSF consists of four stages. The first stage is external peer evaluation, after which the comments and proposals are sent to programme officers. Here recommendations are made on whether a proposal should be awarded or declined funding based on the comments of the external reviewers. A higher level of reviewing officials then again evaluates the recommendations by the programme officers. It is only after this stage of evaluation that a final decision on whether to award or decline funding is made by the NSF staff based on all the comments from the prior evaluation stages. In a review process such as this not much can be added to ensure a higher quality of review.

3.6 Main points of this chapter

In this chapter we have examined the review processes of four funding agencies. The focus was on the review process as such, the manner in which and criteria by which reviewers are selected, as well as the number of stages within the review process. It is evident from this study that, although there are certain aspects unique to every funding agency, there are a number of aspects that are common. Table 1.1 shows the respective criteria taken into consideration when scientists are selected to act as peer reviewers in the review process and Table 1.2 summarises the variety of review criteria used by these peers when evaluating research.

Table 1.1

Peer selection and number of review stages at the NIH, NSF, MRC and NSERC

Funding Agencies	NIH	NSF	MRC	NSERC
Focus on knowledgeable peers	X	X	X	X
Internal and external reviewers/ discussants	X	X	X	X
Suggesting reviewers	X			
Single stage of review within funding agency				X
Multiple stages of review within funding agency	X	X	X	

Table 1.2
Summary of evaluation criteria for proposals at the NIH, NSF, MRC and NSERC

Evaluative criteria	NIH	NSF	MRC	NSERC
Importance of research question	X	X	X	X
Originality/ innovative character of aims	X	X	X	
Implementation of new methods		X	X	
Research design	X	X	X	
Protection of humans, animals and environment	X			
Collaborative arrangements	X	X	X	X
Qualification of scientist	X	X	X	X
Scientific environment (contributing to the probability of success)	X	X		
Proposal budget & duration of research	X			
Comprehensiveness of literature review			X	
Advancement/ contributions to scientific and technological knowledge	X	X	X	X
Promoting training and teaching		X		X
Inclusion of minority groups	X	X		

CHAPTER 4

RECOMMENDATIONS ON THE PEER REVIEW PROCESS AND CONCLUDING COMMENTS

The previous chapters have dealt with various aspects within the peer review process. We have looked at peer review as a concept, the critiques of peer review as well as the peer review practices of four respective funding agencies. There are two more aspects, however, that still need to be discussed: recommendations for improvement to the peer review process and recommended alternatives to peer review as an evaluative tool. In this closing chapter these two aspects will be discussed, followed by some concluding comments.

4.1 Evaluation of the peer review process

In order to make recommendations for improvements to the peer review process, it is necessary to establish criteria according to which peer review can be evaluated. Peer review will never be error free and this is why it is so important that funding agencies and scientific journals periodically examine their review processes. They need to assess how well the process is working to support high-quality research and to experiment with different ways of improving current processes (ARC Report, 1997).

In order to establish criteria or a yardstick against which to measure peer review, it is necessary first to establish the desired characteristics of peer review. Chubin and Hackett (1990) summarises the desired characteristics of peer review as follows:

- An effective mechanism for allocating resources and communicating priorities to scientists;

- A mechanism that promotes accountability in science;
- A mechanism that provide policy-makers with a guide to direct scientific effort as well as for new science to shape future research;
- A fair system, adhering to societal norms of equitable treatment as well as scientific norms of universalism and disinterestedness;
- A system that measures scientific performance with respect to standards of good measurement: validity and reliability.

With the above desired characteristics of peer review in mind as well as the concerns and critiques raised by scientists and other parties as mentioned in Chapter 2, it is possible to construct recommendations with the aim of improving the peer review process. Not all of the recommendations proposed are practical and certainly some recommendations bring about their own difficulties and possible critiques. Nevertheless, different evaluation processes require different procedures and ultimately it is up to the funding agency or journal editor to decide what recommendations are necessary, possible and practical in each situation.

In what follows recommendations for peer review practices will be presented. In the first section the recommendations that apply to both manuscript and grant reviews will be addressed. Thereafter recommendations that apply only to manuscript review will be considered and thirdly recommendations only applicable to grant reviews.

4.2 Recommendations for peer review practices valid for manuscript and grant reviews

4.2.1 Construction of a more reliable and accountable system of peer review

The use of referee reviews could prove to be one method to combat unaccountability in peer review effectively. In referee review the author of a submitted manuscript or applicant for a grant proposal will be given the opportunity to evaluate their referees in terms of bias or fairness, carefulness, constructiveness or whatever other factors are deemed relevant. The information would be recorded, filed and periodically reviewed by the editor or grant officer. Referees who receive repeated low ratings might then be eliminated as future peer reviewers (Cicchetti, 1991).

In an effort to improve reliability, the Advisory Board of the Australian Research Councils recommended the continual evaluation of their referees. This evaluation will be done by considering response rates, usefulness of replies, accuracy of "field" or subject descriptions. These evaluations will then form the basis for the extension, pruning or editing of their reviewer database (ARC Report, 1997).

4.2.2 Blind reviewing

A number of arguments have been offered concerning the case for and against blind reviewing. Some believe it to be a valuable method in regaining the confidence of authors and grant applicants in the review process. Others believe it to be a mechanism for authors and grant applicants that can be used for their own benefit. Be that as it may, blind reviewing is viewed by most as a possible solution to bias in peer review. "If bias proves to have validity, a possible solution will be to establish blind reviews as standard journal policy" (Peters & Ceci, 1982:194).

It has been argued, however, specifically with regards to journal review that blind reviewing is not an effective mechanism to ensure the anonymity of the author. An example of the failure of blinding occurs when names and affiliations are removed on the face sheet, but a footnote identifying the senior author and the institution at which the research was conducted is not. This assumption with regards to the failure of blinding could possibly be overstated, as research done by Peter and Ceci on the feasibility of blind review showed. Peter and Ceci randomly selected 180 reviewers for 6 psychology journals and found that, although the reviewers predicted that they could correctly identify authors of manuscripts in 72% of the cases, their actual "hit-rate" was only 36% (Peters & Ceci, 1982).

Blind reviewing is definitely not a foolproof mechanism, but it is one that has the potential to aid the regaining of confidence in the peer review system. By removing authors' and grant applicants' names and affiliation to institutions, effects such as the "Halo effect" and "Matthew effect" can potentially be reduced to a measurable degree.

4.2.3 Computer-assisted referee selection

Selecting reviewers from a large computer-based database could prove to be a very effective mechanism in reducing lengthy delays in the review process as well as counteracting potential bias. Using a computer-based mechanism to select reviewers assists editors and grant officers to overcome possible biases they may have toward an article or research proposal (Bernard, 1982).

Bernard comments that long delays in the review process can often be attributed to the fact that editors simply can't find enough reviewers who will respond quickly. The computer-based mechanism can briefly be described as a database where names and specialisations are entered in free format into a logical text editor. The text editor is used as a database management device to find potential reviewers with any number of simultaneous specialties. "Use of this technique cut the review time down to less than two months, on average. If a reviewer fails to respond within three to four weeks, I just send out copies of a manuscript to alternative referees. The computer is utterly unintimidated by my requests for names of more potential referees" (Bernard, 1982:202).

4.2.4 Pre-acceptance interactions

Tax and Rubinstein (1982), argue for pre-acceptance interactions between authors and reviewers (or grant applicants and grant officers) to be the best mechanism to ensure fairness and responsibility in reviewing. These pre-acceptance interactions would function by journal editors (grant officers) sending out papers (proposals) to reviewers, after which signed comments will be sent to the authors (grant applicants). Authors (grant applicants) and reviewers would then be free to correspond directly, with copies sent to the editor (grant officer) to make sure that the originator of new conceptual material was properly credited. Authors (grant applicants) have the choice of arguing the adequacy of their submitted presentation, revising their submitted presentation after correspondence with the reviewers or to withdraw their contribution. In the case of manuscript review, should authors persist, editors would need to judge at what point, if any, papers meet standards of acceptability for their journal. They need to exercise their editorial position in deciding on the basis of the developing correspondence what the disposition of the paper ought to be. In the case of proposal review it would be the responsibility of the grant officer together with the review panel to make the final decision in respect to the funding, or not, of the grant applicant (Tax & Rubinstein, 1982; Chubin 1982, Cicchetti, 1991).

4.2.5 Allowing the submission of names of potential reviewers by authors or grant applicants

Another mechanism which could be effective in reducing possible bias in reviewer selection for a particular manuscript or grant proposal is allowing authors or grant applicants to present the editor with names of potential reviewers. Names of reviewers not to be asked to review the submitted manuscript or proposal should also be submitted. In stating the names of those reviewers not to be asked, it can be expected of the author or grant applicant to state reasons why the specific reviewers are not to be asked. It is, however, up to the editor or grant officer to decide whether such information is required or not (Cicchetti, 1991).

4.2.6 Signed reviews

The signing of reviews is an aspect that has been heavily debated by various scientists, who remain divided as to whether this is a desirable practice or not. Chubin and Hackett (1990) argue in favour of reviews being signed by reviewers and referees as, according to them, this would hold reviewers publicly accountable for their decisions and would take a step toward acknowledging the value of reviewers' work. "No longer would it be convenient for a reviewer to trash another's work. Nor would it be advisable to endorse unexamined work" (Chubin and Hackett, 1991: 204).

Partiality of peers has been cited as one of the many criticisms of peer review. It was stated earlier that the social structure of science often makes impartiality and objectivity among peers very difficult to achieve. It is exactly because of this social structure that some scientists do not agree with the practice of signed reviews. Young reviewers asked to evaluate a prominent scientist's manuscript or grant application are placed under unnecessary pressure and may suffer recriminations when having to sign their reviews. It is also possible that younger scientists will mute criticisms if forced to acknowledge their source. On the other hand, it is argued that insisting that reviewers acknowledge their identities is nothing but fair and just: ". . . I would simply point out that hiding behind the cloak of anonymity opens the door to the worst sort of blackballing. Stating opinions that one must stand by – and defend, if necessary – is part and parcel of what the social dimension of science is all about" (Rourke, 1991:161).

4.2.7 Greater use of electronic media

Research funding agencies and scientific journals are increasingly making use of the electronic media to communicate their policies and peer review procedures and to make templates for grant proposals or reviewer comments available (ARC Report, 1997). In South Africa two examples are the National Research Foundation and the *South African Journal of Science* which both make use of the internet to communicate important information. The home pages of The National Research Foundation provides a complete guide to its policies with regard to research support and stipulate the evaluation procedures and criteria for participation in their focus areas. On its home page the *South African Journal of Science* has important information for authors, information on current issues, subscriptions, advertising and either e-mail or physical addresses for possible inquiries or comments. "The use of new technologies for communication by research funding agencies not only increases the availability of a very wide range of information and documentation by also helps to substantially redress past concerns regarding transparency of operations" (ARC Report, 1997:47).

4.2.8 Training referees to increase the quality and reliability of their reviews

With regards to the selection of reviewers, Kostoff (1999) cites two dimensions of competence that should be considered in the selection process. Although these dimensions have been referred to earlier in this thesis, I will briefly state them again. The first dimension Kostoff identifies is the importance of the individual reviewer's technical competence of the subject area, and the second is the competence of the review group as a body to cover the different facets of research issues. The training of reviewers in the technology, mission considerations and infrastructure of the respective funding agency as well as the political and social issues relevant for the environment in which science is to be practised is very important. The reason, as pointed out by Kostoff, is very simple: "The quality of a review will never go beyond the competence of the reviewers" (Kostoff, 1999:4).

Training of referees in manuscript reviews can be undertaken by sending potential reviewers copies of remarks made by existing reviewer(s) after reviews have been submitted. "What is important is seeing a colleague's opinion of the papers, to see if you missed an important point, or for younger reviewers, to see how someone else handles the entire review process" (Delcomyn, 1991:142). Training referees on how to phrase criticisms so that they do not seem like a personal attack on the author could also aid the peer review process. Thus it could be a learning rather than a humiliating experience for the author and in this way contribute to the reliability of the process.

In addition to the training of referees, one could develop procedures for monitoring the continued accuracy of trained reviewers, as proposed by Cone (1982). Other proposed mechanisms to enhance reviewer accuracy is the screening of reviewers to ensure they are competent before sending them reviews, the screening of reviews as to eliminate gratuitous and irrelevant comments as well as the implementation of a reward system to encourage consistent high-level reviewing. This reward system could either entail paying reviewers for their work or the establishment of a performance-dependent selection and retention on editorial boards (Cone, 1982; Perloff & Perloff, 1982, Cicchetti, 1991).

4.3 Recommendations for peer review practices in manuscript reviews

4.3.1 Standard rating form in which an explicit set of evaluative criteria is listed

Although the assumptions concerning agreement among peers in their evaluation of science have already been discussed in Chapter 2 they are discussed again here as they are also relevant in this context. When evaluating research, scientists view it from their own unique "science frame". "Science frame" refers to their own ideas, mostly shaped by their own training, as to what constitutes good science. As Cone writes: "Researchers agree on the 'normative' criteria to apply in judging a paper's scholarly worthiness; they disagree on the application of these criteria to given manuscripts and on the publishability of given papers" (Cone, 1991:142). Goodstein goes further by commenting: "Given the fact that we are asking our peers to make important decisions, using complex and ambiguous criteria, it should come as no surprise that the interjudge agreement of these decisions is low, more or less bordering on chance" (Goodstein, 1982:212).

One way to eliminate, or at least lessen, low inter-judge agreement is by supplying reviewers with explicit criteria and/or clear guidelines to guide them in the review process. The *Journal of Behavioral & Brain Sciences* instruct their reviewers to assess manuscripts by rating them on the following criteria: significance, originality, generality, presentation, scholarship, data/methods, reasoning, theory and length (*Instructinos to Behavioral and Brain Sciences referees*, 2000).

43.2 "Open peer commentary" / "open review" system

An open peer commentary/open review gives authors of accepted refereed articles the opportunity to openly respond to criticism and for readers to form their own opinions in respect of the merit of an individuals work (Peters & Ceci, 1982; Harley, 1982; Tax & Rubinstein 1982). "With the article, commentaries (from first-round referees and others), and the author's formal response published together in their entirety, readers of the journal can have a chance to examine and appraise this process of 'creative disagreement' and form their own opinions as to the merit of an individual's work" (Peters & Ceci, 1982:194).

4.3.3 The selection and activities of editors

Editors have been accused of being merely puppets in the hands of reviewers, relying fully on the evaluation of the reviewer and not using their mandate as editors to evaluate reviews and finally deciding which manuscripts to publish or not. This may be a very harsh criticism and even unfair, but it is still worthy of note. It is very important that editors truly be editors. "It would help if editors would truly be editors rather than clerks who tabulate referees' recommendations and let the 'vote' decided the outcome or submissions. Editors should evaluate the performance of referees, making proper adjustments for the varying strictness of the different ones, and should make an honest effort to achieve consistency in the standards used to judge different papers" (Goodstein, 1982:212).\

The way editors are elected must also be viewed critically as it is important that editors be chosen for their "editorial wisdom", their sagacity and willingness to work hard at the editorial tasks. Research competence and political connections are not valid criteria in electing an editor. Being a good scientist does not necessarily make a person a good editor, and thus it might be worthwhile to invest in editor training to produce the necessary levels of judgmental reliability (Goodstein, 1982).

4.3.4 Multiple reviewers

It is argued that the use of multiple reviewers in the review process improves the reliability of peer reviews and facilitates responsible consideration of the data presented in accepted papers (Tax & Rubinstein, 1982; Cicchetti, 1991; Greene, 1991; Zentall, 1991). By using more than one reviewer the editor can be reasonably sure that major theoretical or methodological errors are not missed, as one reviewer will most probably detect errors missed by another. Another advantage of multiple reviews is that they provide an editor with a wider variety of opinions regarding the scientific merit or relevance of a manuscript. This aids him or her in making an informed decision, a decision not merely based on one reviewer's opinion. It also facilitates responsible consideration of the data and ideas presented in accepted papers.

4.4 Recommendations for peer review practices for proposal review

4.4.1 More meaningful scoring procedures

The meaningfulness and potential for manipulation in scoring procedures for referees and panel members have generated much debate among scientists. In a recent report compiled by the NIH on grant ratings, attention was drawn to the problem of scores being treated as if they represent a higher degree of reliability and precision than they actually have. Funding agencies have gone trying to address this problem in different ways. Some National Science Foundation divisions have asked their panel members to ignore the scoring system and to place the proposals into three categories: "fund the proposal under any circumstances"; "fund if sufficient money can be found" and "don't fund". In respect to variation in assessment between reviewers of the same proposal, the National Commission on Research in the United States suggests the use of categorical as well as overall ratings to allow better comparisons among reviewers of the same proposal and among reviews by the same reviewer. The assessors for the Australian Research Council Large Grants applications are asked to provide ratings across a range of criteria as well as summary assessments for both the quality of the proposes research and the quality of the applicant(s). The National Institute for Health (NIH), in its report on the rating of grant applications, recommends that assessor reports explicitly address a set of criteria in their written comments and provide a separate numerical rating on each criterion. Global ratings of scientific merit, however, are not recommended (ARC Report, 1997).

4.4.2 Greater attention to internal data collection

Research funding in current regimes of science has to meet multiple objectives such as societal relevance, accountability and the development and sustaining of scientific excellence. To ensure that all the objectives are met, the grant process must be both monitored and examined through the systematic collection of information regarding both the awards process and its outcome. The Welcome Trust, in its effort to assess whether its research funds have been invested sensibly, established a Research Output Database. This database provides quantitative data on the published output of researchers and links this output to the sources of research funds (ARC Report, 1997).

4.4.3 Streamlining of the peer review process

Streamlining of the peer review process in respect to grant applications includes shorter proposals, increase in the discipline base of experts who assist in the review and greater use of overseas assessors. It also incorporates lengthening the average duration of basic research grants, non-renewable extensions of a limited number without full review, provision of referee reports before interviews, and the abolition of fixed closing dates for grant applications.

In February 1995, after a series of pilot studies and as part of the streamlining of its peer review process, the Centre for Scientific Review (CSR) in the United States implemented the "triage procedure". Briefly, the "triage procedure" entails the categorisation of applications by reviewers in either an upper half or lower half in respect to quality. Applications categorised in the upper half are given to a full advisory council for second-level review. However, those applications categorised in the lower half are not discussed nor scored at the study section meetings. This "triage procedure" is claimed to enhance the review process in several substantial ways:

- Only those applications judged highly meritorious are discussed at the study section meeting. Since it is intended that only approximately half of the applications will be placed in this category, there is ample time for in-depth discussion of each of these applications.
- In many instances, the process results in significantly shorter study section meetings and some savings in costs.
- Assigned reviewers are encouraged to focus written critiques primarily on major strengths/weaknesses, issues and concerns. This provides applicants with a better understanding of the major concerns they must address if preparing a revised application.
- As part of the streamlined review process, the summary statements consist primarily of reviewers' prepared critiques, essentially unaltered. Production of summary statements in this fashion allows for prompt feedback to applicants. (CSR, 2000)

4.4.4 Plurality of funding sources and alternative grant forms

Although not proven to be efficient, a plurality of funding sources has been suggested by a number of commentators. Smith (in ARC Report, 1997), the editor of the *British Medical Journal*, indeed argued that multiple funding sources will do little to offset problems such as randomness in the review process or systematic bias against individuals, institutions or fields of research, associated with the peer review process.

A proposed alternative to the traditional grant system is the instigation of a post-project research grant. Competition and awarding of these grants will only follow completion of the research project. Attention has also been drawn to the importance of seed funding, in which grant awards will support research thrusts in new areas. "In general, grant awards have tended not to cover serendipitous factors in research, such as the opening up of potentially important new avenues of inquiry or the need for new equipment or the development of a new technique to deal with unforeseen problems" (ARC Report, 1997:49). In recognition of this problem the Australian Research Council has introduced "out of cycle" grants.

4.4.5 Controlling proposal demand

One method of controlling proposal demand, as proposed by the Chief Executive of Britain's Engineering and Physical Sciences Research Council, is the promulgation of rejection rates for research grant applications by universities. This could potentially reduce the huge number of projects turned down as feedback is introduced back into the system, "so that if applications fail it is not just neutral but can work against the institution" (ARC Report, 1997:49.) Another direct method that could be implemented to control proposal demand is the limiting of the number of resubmissions.

4.5 Alternatives to peer review

4.5.1 Formula funding

An alternative model for distributing research funding is offered by Rustum Roy. He claims that this model is not dependent on peer review and is biased to performance and not promise. The formula in Roy's model funding would look like this:

Allocation to university department (or equivalent)

= Ax (number of MS degrees + 3x number PhD degrees)

+ Bx (number of papers published in refereed journals)

+ Cx (amount of research support from US and state mission agencies)

+ Dx (amount of research support from industry)

(ARC Report, 1997)

This model by Roy measures performance in terms of output (that is the number of MS and PhD degrees, number of papers published in refereed journals) and amount of research support (that is support from the state and state mission agencies as well as research support from industry). It is based on a rolling average for the preceding three years and data are to be collected on an annual basis.

4.5.2 Cash prizes for solutions to problems

This alternative proposed by Horrobin involves the government offering substantial cash incentives for solving particular problems. This will, according to Horrobin, ultimately save a substantial amount of money and has the potential of involving a large inflow of private venture capital into research (ARC Report, 1997).

4.5.3 Lottery

A lottery has been suggested by some as an alternative to peer review. "Peer review should be replaced by a system in which proposals would be assessed only to ensure they were technically competent. Every proposal which passed the test would then be put into a lottery, and funded if it came out of the hat before all the cash available was spent" (American Association for the Advancement of Science in ARC Report, 1997:53).

According to Chubin and Hackett (1990), however, a lottery system will result in the abandonment of meritocratic decision-making and, perhaps more important, the quality-assurance function of peer review; also the societal notions of fairness would be offended as luck is not a source of success. Success is a result of hard work, talent, persistence and dedication and not of luck. "Thus, for reasons that range from threats to the professional identity of scientists through the practical benefits of (admittedly modest levels of) quality assurance to consideration for the norms of science and society, it would be unwise to replace the peer review system with a lottery" (Chubin and Hackett, 1990:198).

4.5.4 Block grants to universities

Allocating government funds through block grants to universities is another alternative suggested to competitive grant schemes. A scheme as proposed by McCutchen, for example, suggests that: "The schools would decide who to support however they wished, using any system they wished, from despotism to democracy . . . Each year, universities would go to the federal government and argue for support. Let them bring citation scores, rumors of Nobel Prizes almost awarded, whatever they want. Out of this free-for-all, a formula would emerge, no doubt with loopholes and exceptions, and the negotiators would return home exhausted and tell the troops how they made out. The mutual dependence of scientists and brass would develop the loyalty upward and downward that makes institutions bearable to their members" (McCutchen in ARC Report, 1997: 53).

4.5.5 Bypassing peer review through the "pork barrel"

"Pork barrel" funds refers to the earmarking of congressional funds in the United States for capital grants and research facilities for specific institutions outside the authorisation of the traditional peer review process. Although the earmarking of funds in this way has become an increasing feature of appropriations for academic science in the United States, there has been substantial criticism of the "pork barrel" phenomenon both from within and outside of the scientific community. One of the concerns is that pork-barrel funding will allow for the production of second-rate research (ARC Report, 1997). As Nicholson comments: "Excellence, which most scientists would say should be the principal criterion, is not even mentioned. Economic competitiveness, while important, is not a good basis for selecting individual projects in basic research because, as any scientist knows, basic research is too unpredictable" (Nicholson & Richard, 1992:1497).

4.5.6 Bibliometrics

The importance of the output of research increasingly results in the augmenting of peer review practices with quantitative indicators such as bibliometrics. Bibliometrics can be defined as a process that systematically measures scientific output by using "counts of publications, patents, citations and other potentially informative items to develop science and technology performance indicators" (Kostoff, 1999:11). Bibliometrics, however, is not without its limitations. In particular there is the problem of cross-discipline comparisons of outputs. The idea is not necessarily to replace the peer review of project proposals with bibliometrics, but rather to use it as an aid in the decision-making process for the allocation of resources (ARC Report, 1997).

4.5.7 The use of multiple performance indicators

Peer review is an evaluative tool within an evaluation process. An evaluation process is the prepatory phase of decision-making. Evaluation can either be evaluation of the present (*ex ante evaluation*) or evaluation of the past (*ex post evaluation*) (De Greef & Frijdal, 1989). Peer review functions within both evaluation spheres: *ex ante evaluation,* where it deals with assessment of future research activities (research proposals), and *ex post evaluation* in the assessment of manuscript reviews.

When reflecting on evaluation processes and the whole principle of research performance assessment, the question that emerges is: to what extent can the outputs of research groups and centres be assessed in such a way as to make the whole evaluation process reliable and trustworthy?

Research is by nature multi-dimensional in character and it is exactly this characteristic of research that "unsettles" the evaluation process by exposing it to sometimes very fierce criticisms by those subject to the evaluation result.

A method that has been proposed by a variety of authors such as Martin (1996), Greef & Frijdal (1989), Yorke (1991) and Tognolini, Adams & Hattie (1994) is the implementation of multiple indicators in the research performance assessment process.

Their motivation is that few would dispute the claim of research being multi-dimensional. Greef & Frijdal (1986) identify four possible dimensions of research. These dimensions are 1) Scientific, which refers to the contributions made to the stock of knowledge;

2) Educational, which refers to the contributions in terms of skills and trained personnel; 3) Technological, referring to the development of new/ improved technologies; and 4) Cultura,l which refers to contributions made to the wider society (Greef & Frijdal, 1986).

Another reason for the introduction of multiple indicators in the research performance assessment process is that, given the multi-dimensional character of research, no single indicator of research output or performance can be considered conclusive, as it will never reveal more than a small part of the multi-dimensional picture. Thus a performance indicator can never be conclusive, but can only function as a partial indicator. Although selective and careful use of these indicators is probably better than no evaluation at all, the most fruitful approach would involve the combined use of multiple indicators.

A counter-argument to the converging of indicators is the argument that such convergence is in actual fact misleading as at least some of the indicators are related. Should this be the case, the dimensions measured are not orthogonal but overlap to some extent. The response to this argument would be that the various indicators that are used in assessing research performance are not measuring the same thing and neither are they measuring research along a single dimension (Bird and Oppenheim in Martin, 1996). This leads to the conclusion that a result based on the convergence of several indicators, preferably including extensive peer evaluation, is likely to be more reliable than one based on a single bibliometric indicator or on peer review alone.

The use of multiple indicators also minimises the risk that scientists will in some way or other "play the game" and manipulate the indicators to their advantage. When applying a number of indicators in the assessment process it becomes much more difficult, if not unlikely, to manipulate all the indicators without at the same time improving one's research (Martin, 1996).

Although much of the research evaluation community has come to believe that the simultaneous use of various evaluation techniques is the preferred approach, it is interesting to note that this belief is not necessarily mirrored in actual research performance practices. Martin (1996) did a study where a selection of articles published in *Scientometrics* was examined to establish how many of the articles were evaluated by using multiple indicators. The sample consisted of 12 recent issues of *Scientometrics* (Volumes 31-34 published in 1994-95) and 12 issues published a few years earlier (Volumes 14 and 15 published in 1988-89). The results of the study showed that 70% of the articles published in 1988-89 were evaluated by using only one or two indicators, whereas only 4% were evaluated using five or more. For the articles published in 1994-95 the results were much the same, where 64% were evaluated using only one or two indicators. A summary of the findings is shown in Table 1 below.

Table 2

(Author of table: Martin, 1996:353)

Numbers of indicators reported in a sample of *Scientometrics* papers

Number of distinct Indicators	Empirical papers published in *Scientometrics* 1988-89	1994-94	TOTAL
1	24	25	49
2	14	18	32
3	8	15	23
4	6	3	9
5 or more	2	6	8
TOTAL	54	67	121

One possible reason for this is the easy accessibility and fairly inexpensive nature of publication counts, citation counts and journal impact factors. "[T]he first point to note is that publication counts, citation counts and journal impact factors tend to be used a lot because they are there" (Martin, 1996:359). When proposing the use of multiple indicators in the assessment of research performance, one is also implying a process that is much more time consuming and expensive. However, in the quest for an effective research performance evaluative system which is characterised by fairness and reliability, time consumption and expense should not be our fist concern.

CONCLUDING COMMENTS

In this thesis we started by looking at the notion of change within science and its impact on the evaluation of research. In the development of the different regimes of science we saw how in the regime of "Relevant science" the role of the scientist has changed from entailing responsibility only to the scientific community to also having a responsibility to the broader society. Evidently this dual responsibility of scientists also had an influence on the criteria used in the evaluation of science. Scientific excellence is not regarded as the sole criterion. Science, in order to be regarded as important, needs to contribute to the benefits of society as a whole. This entails contributing to society's well-being, society's development as well as society's scientific understanding.

In Chapter 2 we examined conceptual and methodological issues related to peer review, followed by cases studies on the peer review practices of four funding agencies, two in the United States and two in Canada. In the fourth and final chapter we looked at recommendations and alternatives to peer review. This brings us to a final question: where do we place peer review as an evaluative tool in research performance assessment?

Despite all the critique directed at peer review, it is still believed by most scientists to be essential to the evaluation process. However, the tendency today in the light of the multi-faceted character of research is to supplement peer review with other performance indicators such as citation analysis and publication counts. In a study conducted by Martin in 1990-1992, the views of scientists on the use of multiple indicators to assess their research were investigated. One of the questions asked of university staff by Martin was whether departmental research performance was best assessed using peer review alone, performance indicators alone, or some combination of the two. Those choosing the last option (some combination of the two) were also asked whether equal weight should be given to the peer review and performance indicator components, or whether more emphasis should be attached to one of them. The results were as follows.

Table 3

(Source: Martin 1996:358)

Relative importance of peer review (PR) versus performance indicators (PI)

Number of interviewees expressing a particular view

	Physics	**Biochemistry**	**Mathematics**	**Chemical engineering**	**TOTAL**
Peer review (PR) only	0	0	1	0	1
PR and PI – more weight to PR	11	7	13	2	33
PR and PI – equal weight to each	9	11	5	6	31
PR and PI – more weight to PI	6	13	7	2	28
Performance Indicators (PI) only	0	0	1	2	3

From the above table it is clear that scientists prefer the use of peer review together with performance indicators. Only one person suggested that peer review should be used without any other performance indicators. The same applies for using only performance indicators, where only three people argued that performance indicators completely supplanted the need for any element of peer review. With regards to the option where peer review is combined with performance indicators, 96% where in favour. This 96% were split approximately equally between those favouring similar weight given to the two elements (32%), those advocating more weight to peer review (34%), and those pressing for more emphasis on performance indicators (29%).

Thus a response to the question as to where to place peer review as an evaluative tool in research performance assessment today would be the following: peer review as an evaluative tool in research performance assessment is one of the cornerstones, if not the cornerstone, of evaluative practices. Peer review has been and will always be an essential element in the evaluation of research and has shown through the changing regimes of science that it has the capacity to change and to adapt to the changing needs of the scientific as well as the broader social community.

BIBLIOGRAPHY

Australian Research Council, 1997, 'The Peer Review Process', *Commissioned Report No. 5.*

Babbie, E., 1998, *The Practice of Social Research*, Wadsworth Publishing Company: United States of America.

Bailar, J.C., 1991, 'Reliability, fairness, objectivity and other inappropriate goals in peer review', *Behavioral and Brain Sciences'*, Vol. 14(1).

Instructions to Behavioral and Brain Sciences Referees, 2000, www.cogsci.ac.uk/bbs/instructions.for.referees.html

Bernard, H.R., 1982, 'Computer-assisted referee selection as a means of reducing potential editorial bias', *Behavioral and Brain Sciences*, Vol. 2 (5).

Blankley, W., 1997, 'Research funding and priorities: The new paradigm', Unpublished paper represented at a workshop on *"Restructuring Higher Education"*, Bifsa, Midrand, South Africa.

Blissette, M., 1982, 'Peer review and the structure of knowledge', *Behavioral and Brain Sciences*, Vol.5 (2).

Busch, V., 1980, *'Science – The Endless Frontier'*, New York: Arno Press.

Chubin, D.E., 1982, 'Reforming peer review: From recycling to flexivity', *Behavioral and Brain Sciences*, Vol. 5 (2).

Chubin, D.E, Hackett, E.J, 1990, *'Peerless Science: Peer Review and U.S. Science Policy'*, State University of New York Press: Albany.

Cicchetti, D.V. 1991, 'The reliability of peer review for manuscript and grants submissions: a cross-disciplinary investigation', *Behavioral and Brain Sciences*, Vol. 14, pp. 119-186.

Cone, J.D., 1982, 'Criterion problems in journal review practices', *Behavioral and Brain Sciences*, Vol.5 (2).

Center for Scientific Review, 2000, *Streamlined Review Procedures Used in the CSR*, www.csr.nih.gov/REVIEW/streamln.htm

De Greef, J.P., Frijdal, A., 1989, 'Evaluation of scientific research profile analysis, a mixed method', *Higer Education*, Vol. 1 (1), pp. 83-90.

Delcomyn, F., 1991, 'Peer review: Explicit criteria and training can help', *Behavioral and Brain Sciences'*, Vol. 14 (1).

Gibbons, M., *et al.,* 1994, *'The new production of knowledge'*, Sage publications: London.

Goodstein, L.D., 1982, 'When will the editors start to edit?', *Behavioral and Brain Scientists*, Vol. 5 (2), pp. 212-213.

Green, R., 1991, 'Is there an alternative to peer review?', *Behavioral and Brain Sciences*, Vol. 14 (1), pp. 149-150.

Harnard, S., 1982, 'Peer commentary on peer review', *Behavioral and Brain Sciences*, Vol. 5 (2), pp. 185-186.

Horrobin, D. F., 1982, 'Peer review: A philosophically faulty concept which is proving disastrous for science', *Behavioral and Brain Sciences,* Vol. 5 (2), pp. 217-218.

Johnes, J., Taylor, J., Francis, B., 1989, 'The Research Performance of UK Universities: a Statistical Analysis of the Results of the 1989 Research Selectivity Exercise', *ROYAL STATISTICAL SOCIETY*, Vol. 156, Part 2, pp. 271-286.

Kiesler, C.A., 1991, 'Confusion between reviewer reliability and wise editorial and funding decisions', *Behavioral and Brain Sciences'*, Vol. 14 (1), pp. 151-152.

King, J., 'A review of bibliometric and other science indicators and their role in research evaluatior', *ISI*, (1989), Vol. 21, (17), p. 4-10.

Kostoff, R.N. 1999, *The Handbook of Research Assessment*, http://www.dtic.mil/dtic/kostoff/Handweb71V.html

Kostoff, R.N., 1997, 'Peer review: The Appropriate GPRA Metric for Research', *Science*, Vol. 277, pp. 651-652.

Kostoff, R.N., 1996, 'The Principles and Practices of Federal Research Impact Assessment', *Encyclopedia of Library and Information Science'*, Vol. 58, pp. 243-274.

Kostoff, R.N., 2000, 'The Handbook of Research Impact Assessment', *www.dtic.mil/dtic/kostoff/Handweb71.html*

Labuschagne, H.A., 1985, *'Evaluering van die produktiwiteit van akademici as navorsers met verwysing na doelwitte vir evaluering en indikatore van produktiwiteit: 'n verkennende studie'*, Potchefstroom, Buro vir Navorsing: PU vir CHO

Marais, H.C., (2000), *Perspectives on Science Policy in South Africa*, Pretoria: Business Print.

Marshall, E., 1996, 'Trial Set to Focus on Peer Review', *SCIENCE*, Vol. 273, (5279), pp. 1162-1164.

Martin, B.R., 1996, 'The use of multiple indicators in the assessment of basic research', *Scientometrics*, Vol. 36 (3), pp. 343-362.

Medical Research Council, 2001, *WebPages*, www.mrc.ac.uk

Merton, R.K., 1986, 'The Matthew effect in science', *SCIENCE*, Vol. 159, pp. 56-63.

National Research Foundation Newsletter,2000, *news@nrf*, July.

Nicholson, Richard, S., 1992, 'Congressional Pork Versus Peer Review', *SCIENCE*, Vol. 256, Iss. 5063, p. 1497.

National Science and Engineering Council, 2001, *WebPages*, www.nserc.ca

National Science Foundation, 2001, *WebPages*, www.nsf.gov

National Institute for Health, 2001, *WebPages*, www.drg.nih.gov

Peters, D.P., Ceci, S.J., 1982, 'Peer-review practices of psychological journals: The fate of published articles, submitted again', *Behavioral and Brains Sciences'*, Vol. 5 (2), pp.187-255.

Perloff, R.M., Perloff, R, 1982, 'Improving research on and policies for peer-reviewed practices', *Behavioral and Brain Sciences*, Vol. 5 (2), pp. 232-233.

Polanyi, M. 1986, The Republic of Science: Its political and economic theory. In E. Shils (Ed.) *Criteria for scientific development: Public policy and political authority.* Cambridge, Mass.:MIT. Pp 1-20 (Originally published in *Minerva*, 1962, 1(1): 54-73).

Report to the Advisory Board for the Research Councils from the Working Group on Peer Review,1990, *Centre for Studies on Science, Technology and Society, 'De Boerderij'*, University of Twente.

Rip, A., 2000, 'Higher Forms of Nonsense', *European Review*, forthcoming, December 2001.

Rip, A., 1991, *'Strategic Science – Evolution of a Science Policy Category'* (Unpublished paper).

Rip, A., 1985, Experts in public arenas, in: Peltu, M., Otway, H., eds., 'Regulating Industrial Risks', London: Butterworth publications, pp.94-110.

Rourke, P., 1991, 'Toward openness and fairness in the review process', *Behavioral and Brain Sciences*, Vol. 14 (1), pp. 161.

Shils, E., 1986, Introduction. In E. Shils (Ed.), *'Criteria for scientific development: Public policy and political authority'*, Cambridge, Mass.: MIT. pp. v-xiv

South African White Paper on Science & Technology, Department of Arts, Culture, Science and Technology (1996).

Tax, S., Rubinstein, R.A., 1982, 'Responsibility in reviewing and research', *Behavioral and Brain Sciences*, Vol.5 (2), pp. 238-240.

Tognolini, J., Adams, K., Hattie, J., 1994, 'A methodology to Choose Performance Indicators of Research Attainment in Universities', *Australian Journal of Eductaion*, Vol.38 (2), pp. 105-117.

Towards a Science and Technology Policy for a Democratic South Africa, 1992, A report by the ANC, COSATU and SANCO.

Van der Meulen, B.J.R., 1992, *Evaluation Processes in Science: The construction of quality by science, government and industry*, University of Twente: FEBO-druk.

Weingart, P., 1997, 'From "Finalisation" to "Mode 2" old wine in new bottles?', *Social Science Information*, Vol. 36 (4), pp. 591-614.

Weingart, P., 1994 *'Scientific expertise and political accountability – paradoxes of science in politics'*, (Unpublished paper).

White Paper on Science and Technology: Preparing for the 21st century, *Department of Arts, Culture, Science and Technology*, South Africa.

Yorke, M., 1991, 'Performance indicators: towards a synoptic framework', *Higher Education*, Vol. 21, pp. 253-248.

Zentall, T.R., 1991, 'What to do about peer review: Is the cure worse than the disease?', *Behavioral and Brain Sciences*, Vol. 14 (1), pp. 166-167.