RESEARCH PROGRAM PEER REVIEW: PURPOSES, PRINCIPLES, PRACTICES, PROTOCOLS

DR. RONALD N. KOSTOFF

OFFICE OF NAVAL RESEARCH

800 N. QUINCY STREET

ARLINGTON, VA  22217

INTERNET: kostofr@onr.navy.mil            PHONE: 703-696-4198           FAX: 703-696-4274

(THE VIEWS IN THIS REPORT ARE SOLELY THOSE OF THE AUTHOR AND DO NOT REPRESENT THE VIEWS OF THE DEPARTMENT OF THE NAVY)

TABLE OF CONTENTS

I.  ABSTRACT
II. EXECUTIVE SUMMARY - PEER REVIEW PRINCIPLES
III. INTRODUCTION, DEFINITIONS, AND BACKGROUND
IV. PEER REVIEW PURPOSES AND PRINCIPLES
V. PEER REVIEW PRACTICES
VI. PEER REVIEW PROTOCOLS
   VI-A. APPENDIX I - REVIEW PANEL SELECTION APPROACHES
   VI-B. APPENDIX II - PROGRAM PEER REVIEW PROTOCOL
   VI-C. APPENDIX III - USE OF PUBLISHED PAPERS IN RESEARCH PROGRAM EVALUATION
   VI-D APPENDIX IV – NETWORK-CENTRIC PEER REVIEW
   VI-E APPENDIX V – SIGNIFICANT PEER REVIEW RESOURCES
      V-1. PROPOSAL PEER REVIEW
      V-2. PEER REVIEW GUIDE
      V-3. BIOMEDICAL PEER REVIEW CONGRESSES
   VI-F LARGE AGENCY PEER REVIEW
   VI-G DETAILED PEER REVIEW PROTOCOL
VII. BIBLIOGRAPHY AND RELATED REFERENCES

KEYWORDS: peer review; GPRA; evaluation criteria; metrics; research assessment; retrospective studies; roadmap; data mining; text mining; research merit; research approach; blind review; research evaluation; research impact; bibliometrics; decision aids.
1. REPORT DATE  
01 JUN 2004  

2. REPORT TYPE  
N/A  

3. DATES COVERED  
-  

4. TITLE AND SUBTITLE  
Research Program Peer Review: Purposes, Principles, Practices, Protocols  

5a. CONTRACT NUMBER  
-  

5b. GRANT NUMBER  
-  

5c. PROGRAM ELEMENT NUMBER  
-  

5d. PROJECT NUMBER  
-  

5e. TASK NUMBER  
-  

5f. WORK UNIT NUMBER  
-  

6. AUTHOR(S)  
DR. RONALD N. KOSTOFF  

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  
OFFICE OF NAVAL RESEARCH 800 N. QUINCY STREET  
ARLINGTON, VA 22217  

8. PERFORMING ORGANIZATION REPORT NUMBER  
-  

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)  
-  

10. SPONSOR/MONITOR’S ACRONYM(S)  
-  

11. SPONSOR/MONITOR’S REPORT NUMBER(S)  
-  

12. DISTRIBUTION/AVAILABILITY STATEMENT  
Approved for public release, distribution unlimited  

13. SUPPLEMENTARY NOTES  
-  

14. ABSTRACT  
See report  

15. SUBJECT TERMS  
-  

16. SECURITY CLASSIFICATION OF:  

<table>
<thead>
<tr>
<th>a. REPORT</th>
<th>b. ABSTRACT</th>
<th>c. THIS PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>unclassified</td>
<td>unclassified</td>
<td>unclassified</td>
</tr>
</tbody>
</table>

17. LIMITATION OF ABSTRACT  
UU  

18. NUMBER OF PAGES  
299  

19a. NAME OF RESPONSIBLE PERSON  
-
I. ABSTRACT

The purposes, principles, practices, and protocols of research program peer review are described. While the principles are fundamentally generic, and apply to peer review across the full spectrum of performing institutions, as well as manuscript/proposal/program peer review, the focus of this report is peer review of proposed and ongoing research programs in federal agencies.

Following the self-contained Executive Summary of factors for high-quality peer reviews, the report addresses potential implications of the implementation of the Government Performance and Results Act of 1993 on federal agency research program peer review practices. Then, the report describes strengths and weaknesses of major peer review components and issues, including:

Objectives and Purposes of Peer Review;
Quality of Peer Review;
Impact of Peer Review Manager on Quality;
Selection of Peer Reviewers;
Selection of Evaluation Criteria;
Secrecy (Reviewer and Performer Anonymity);
Objectivity/ Bias/ Fairness of Peer Review;
Normalization of Peer Review Panels;
Repeatability/ Reliability of Peer Review;
Effectiveness/ Predictability of Peer Review;
Global Data Awareness;
Costs of Performing a Peer Review;
Ethical Issues in Peer Review; and
Alternatives to Peer Review.

The report then presents different federal agency peer review practices, and sample protocols and processes for conducting a successful research program peer review. Some peer review variants, such as the Science Court and Network-Centric Peer Review, are described, and research requirements to improve peer review are discussed. The final section is an extensive bibliography of over 3000 references that includes not only text references but related references for further reading as well.
II. EXECUTIVE SUMMARY - PEER REVIEW PRINCIPLES

The Government Performance and Results Act of 1993 (GPRA, 1993) requires federal agencies to develop strategic plans, annual performance plans, and performance measures to gauge progress in achieving their planned targets. A precursor paper in *Science* (Kostoff, 1997b) recommends that peer review be the dominant metric GPRA applies to basic research. However, for research program peer review to be used effectively and efficiently for GPRA, it must be understood, developed, and standardized well beyond its present status. Program peer review should also be integrated seamlessly into an organization's business operations evaluation processes in general, and in particular into its peer review processes. It should not be incorporated into management tools as an afterthought, which is today's common practice, but should rather be part of the organization's front-end design. This allows optimal matching among requirements for generating, gathering, and reviewing data. It helps avoid the present practice of force-fitting evaluation criteria and processes to whatever data are produced from non-evaluation requirements. This report focuses on the underlying principles necessary for high-quality peer review. Although targeted toward research program peer review, most of the principles this report enunciates apply to many kinds of peer review. The author's experience, based on examining the peer review literature, conducting many peer review experiments (e.g., Kostoff, 1988), and managing hundreds of peer reviews, leads to the following conclusions about the factors critical to high-quality peer review (Kostoff, 1995, 1997a, 2001b):

1) Senior Management Commitment
Senior management's commitment is the most important factor in the quality of an organization's S&T evaluations. The relevant senior positions are those with evaluation decision authority, and their most significant contributions lie in the rewards and incentives they institute to encourage high-quality evaluation. Senior managers' commitment should include not only assurance that a credible need for the evaluation exists, but also a strong desire that the evaluation be structured to address that need as directly and completely as possible.

2) Evaluation Manager Motivation
The second most important factor is the operational evaluation manager's motivation to perform a technically credible evaluation. The manager:

a) sets the boundary conditions and constraints on the evaluation's scope;
b) selects the final specific evaluation techniques used;
c) selects the methodologies for how these techniques will be combined, integrated, and interpreted, and
d) selects the experts who will perform the interpretation of the data output from these techniques.

In particular, if the evaluation manager does not follow, either consciously or unconsciously, the highest standards in selecting these experts, the evaluation's final conclusions could be substantially determined even before the evaluation process even begins. All the evaluation
processes considered (peer review, retrospective studies, metrics, economic studies, roadmaps, data mining, and text mining) need experts, and this conclusion about expert selection holds for every one of them.

3) Statement of Objectives
Third most important is transmission of a clear and unambiguous statement of the review’s objectives (and conduct) and its potential impact and consequences to all participants. This statement should occur at the very beginning of the review process.

4) Competency of Technical Evaluators
Fourth most important factor is the quality of the technical evaluators themselves, specifically their role, objectivity, and competency. While the requirements for experts in peer review, retrospective studies, roadmaps, and text mining are obvious, there are equally compelling reasons for using experts in metrics-based evaluations. Metrics should not be used as a stand-alone diagnostic instrument (Kostoff, 1997b). Like lab tests in a medical exam, even quantitative metrics results from suites of instruments require expert interpretation to be placed into proper context and gain credibility. Evaluation resembles diagnosis more than it resembles accounting. The metrics results should make a subordinate contribution to an effective peer review of the technical area being examined.

Thus, this fourth critical factor consists of the evaluation experts' competence and objectivity. All the experts should be technically competent in their subject area, and the competence of the total evaluation team should cover the multiple S&T areas critically related to the present interest. The evaluation team's focus should not be limited to disciplines related only to the present technology area (that tends to reinforce the status quo and provide conclusions along very narrow lines). It should be broadened to disciplines and technologies that have the potential to impact the overall evaluation's highest-level objectives (that would be more likely to provide equitable consideration to revolutionary new paradigms).

5) Selection of Evaluation Criteria
The fifth most important factor is selection of evaluation criteria (Delcomyn, 1991; Sutherland, 1993; Weinberg, 1989). These criteria will depend on the:

- interests of the audience for the evaluation,
- nature of the benefits and impacts,
- availability and quality of the underlying data,
- accuracy and quality of results desired,
- complementary criteria available and suites of diagnostic techniques desired for the complete analysis,
- status of algorithms and analysis techniques, and
- capabilities of the evaluation team.

For evaluating basic research proposals, the three main criteria are research merit, research approach, and team quality (DOE, 1982; Kostoff, 1992, 1997a). For research sponsored by a
mission-oriented organization, a fourth criterion related to mission relevance is useful. To ensure that this mission relevance criterion does not filter out the more basic research oriented proposals, a very liberal interpretation of mission relevance is necessary. For basic research, a nearer-term relevance criterion, such as transition or utility, correlates better with overall proposal quality score than does a longer-term criterion (Kostoff, 1992). Use of a fifth criterion for overall research quality is essential, and makes it possible to incorporate the effects of unlisted criteria that the reviewer feels is important for considering a specific proposal. For example, reviewers might feel that an agency proposal is more appropriate for sponsorship by industry than by government. In this case, the proposal could receive a low overall rating, even though the listed component technical criteria were rated very high.

6) Relevance of Evaluation Criteria to Future Action
Almost every metrics briefing the author has attended—in government agencies, industrial organizations, and academic institutions—has violated a principle of evaluation selection criteria. Although stated in terms of metrics-based evaluation, it applies to all evaluation techniques:

**Every S&T metric, and its associated data, should answer a question that contributes to forming the basis for a decision.**

Metrics and associated data that do not perform this function become an end in themselves. They offer no insight to the central questions of a well-structured study or briefing, and they contribute nothing to decision-making. They dilute any study, and over time they devalue the worth of metrics in credible S&T evaluations. Because of:

1) the political popularity and subsequent proliferation of S&T metrics;
2) the widespread availability of data; and
3) the ease with which these data can be electronically gathered, aggregated, and displayed,

most S&T metrics briefings and studies are immersed in data geared to impress rather than inform. While metrics studies provide the most obvious examples, this conclusion can be easily generalized to any of the evaluation methods.

7) Reliability of Evaluation
The reliability or repeatability of an evaluation is also crucial. To what degree would an S&T evaluation be replicated if a completely different team were involved in selection, analysis, and interpretation of the basic data? If each evaluation team were to generate different evaluation criteria, and in particular generate far different interpretations of these criteria for the same topic, then what meaning or credibility or value can be assigned to any S&T evaluation (Cole, 1981)? To minimize repeatability problems, a diverse and representative segment of the overall competent technical community should be involved in the construction and execution of the evaluation.

8) Evaluation Integration
A sound evaluation processes should in general be seamlessly integrated into the organization's
business operations. Evaluation processes should not be incorporated in the management tools as an afterthought (which is typical practice today), but should be part of the organization's front-end design. This allows optimal matching between data generation, gathering and evaluation requirements, as opposed to the present practice of force-fitting evaluation criteria and processes to whatever data are produced from non-evaluation requirements.

9) Global Data Awareness
Also important is data awareness (Kostoff, 2003). Placing the technology of interest in the larger context of technology development and availability world-wide is absolutely necessary. Failure to do so tends to be a central deficiency of most management decision aids. Lack of S&T documentation, inaccessibility of S&T that is documented, inability to retrieve S&T documents due to poor retrieval methods, inability to extract information from large retrievals, and general lack of interest and will in global data awareness, mitigate against attaining comprehensive global data awareness.

10) Normalization across Technical Disciplines
For evaluations that will be used as a basis for comparison of S&T programs or projects, the next most important factor is normalization and standardization across different S&T areas. For S&T areas that have some similarity, use of common experts (on the evaluation teams) with broad backgrounds that overlap the disciplines can provide some degree of standardization (Kostoff, 1988, 1997a). For very disparate S&T areas, some allowances need to be made for the relative strategic value of each discipline to the organization, and arbitrary corrections applied for benefit estimation differences and biases. Even in this case of disparate disciplines, some normalization is possible by having some common team members with broad backgrounds contributing to the evaluations for diverse programs and projects (Van den Beemt, 1997). However, normalization of the criteria interpretation for each science or technology area's unique characteristics is a fundamental requirement. Because credible normalization requires substantial time and judgment, it tends to be an operational area where quality is sacrificed for expediency.

11) Secrecy
Secrecy is as important as normalization: reviewer anonymity and reviewee non-anonymity (Altura, 1990; Clayson, 1995; Gresty, 1995; Neetens, 1995). If honest and frank viewpoints on the intrinsic quality of the research under review are desired, the reviewer must remain anonymous to all but the review manager. Rewards are few for a reviewer making strong negative statements about a proposal (or research paper or program), and resulting retribution and resentment against the reviewer may far outweigh the intrinsic benefits to science of honest and forthright statements of judgment.

"Blind reviewing," the withholding of the reviewee's name and affiliation from the reviewer, has been used for the noble purposes of providing fairer reviews of work by unknown researchers or by researchers from less prestigious institutions, and to eliminate bias based on such personal characteristics as gender (Ceci, 1984; Laband, 1994; Cox, 1993; Nylen, 1994). However, studies of proposed and existing research evaluations have shown that team quality was the most important variable in determining overall project quality (DOE, 1982). Removing the identity of
the reviewee from the research under review is akin to solving an equation after eliminating the dominant term. Rather than eliminate the key variable of researcher identity, it may be more important to select additional reviewers who will broaden the review group's perspective and address the "right job" aspects of the research project. This will help insure that outmoded, albeit frequently cited, research is not promulgated in perpetuity, and that fresh perspectives of new paradigms will receive the attention they deserve.

12) Cost of S&T Evaluations
The next critical factor for quality S&T evaluations is cost (ASTEC, 1991; Buechner, 1974; Hensley, 1980; Kostoff, 1995, 1997a). The true total costs of peer review can be considerable, but tend to be ignored or understated in most reported cases. For high quality peer reviews, where sufficient expertise is represented on the review group, total real costs will dominate direct costs (Kostoff, 1995, 1997a). The major contributor to total costs is the time of all the individuals involved in executing the review, including staff, reviewer, and presenter time. If a substantial audience is in attendance, then audience time should be included in review costs. With high quality performers and reviewers, time costs are high, and the total review costs can be non-negligible. For sponsor environments where a large number of proposals are rejected, and where multiple proposals to different sponsors are the norm, peer review costs per funded proposal increase dramatically in proportion to the ratio of proposals reviewed to proposals funded. Accurate cost analyses should not be neglected in designing a high quality proposal, manuscript, or program peer-review process.

13) Maintenance of High Ethical Standards
The final critical factor, and perhaps the foundational factor in any high quality S&T evaluation, is the maintenance of high ethical standards throughout the process. A plethora of ethical issues surround evaluation: technical fraud, technical misconduct, betraying confidential information, unduly profiting from access to privileged information, and other pitfalls (Fielder, 1995; Goodstein, 1995; Gupta, 1996; Keown, 1996; Moran, 1992). This stems from an inherent bias or conflict of interest in the process when real experts are desired to participate in every aspect of an S&T evaluation. The evaluation managers need to be vigilant for undue signs of distortion aimed at personal gain.

REFERENCES FOR EXECUTIVE SUMMARY:

Clayson-DB, "Anonymity in Peer-Review - Time for a Change - Comment",


III. INTRODUCTION, DEFINITIONS, AND BACKGROUND

INTRODUCTION

In 1993, Congress enacted the Government Performance and Results Act (GPRA) into law [GPRA, 1993]. GPRA applies to all federal outlay programs, and has three components: strategic plans, annual performance plans, and metrics to show how well the annual plans are being met. Since the plan became law, many federal interagency meetings have sought to ascertain how the third requirement of the plan—performance metrics—could be implemented to portray the progress and accomplishments of research properly, especially basic research. The emerging consensus from the basic research sponsor and performer communities holds that the stated requirements of GPRA and what is required to determine the health of a research program are badly mismatched.

However, GPRA states that if "it is not feasible to express the performance goals for a particular program activity in an objective, quantifiable, and measurable form, the Director of the Office of Management and Budget may authorize an alternative form" [GPRA, 1993]. A precursor article in Science [Kostoff, 1997b] proposed that peer review be used as the dominant basic research program health diagnostic for GPRA, supplemented by bibliometric and other measures. There is a growing consensus in the larger research community that use of peer review is a more appropriate tool to measure basic research program performance in order to satisfy GPRA requirements. If the GPRA oversight agencies agree, then the volume of research program peer reviews across the federal agencies will increase dramatically.

However, not only the volume of program peer reviews will change, but the conduct of the reviews will also change. If GPRA is fundamentally a budgetary instrument [Brown, 1996], then the performance evaluation results that input to the performance budgeting process must be of the highest quality. The methods chosen to obtain these performance evaluation results, program peer review and the supplementary quantitative performance measures, would require more rigorous and standardized operational characteristics (Process selection, reviewer selection, etc.).

The purpose of the present document is to bring to the attention of the relevant research sponsoring, oversight, managing, and performing communities the underlying issues and concerns surrounding research program peer review. If these issues can be addressed comprehensively prior to full scale GPRA implementation, then procedures could be developed to conduct peer review in a manner that will not only support the performance budgeting process but could add value to the research program being reviewed as well. To insure that the present document reflects the experiences and findings of the larger research evaluation community, principles and findings from the manuscript and proposal peer review literature will be utilized, where applicable, to illuminate the research program review issues and help bridge the gaps in the research program review literature.

There are four major components of the present report. The main body of the text (Sections II, III, IV) addresses the underlying issues surrounding research program peer review. Section V summarizes research program peer review practices for selected federal agencies. Section VI describes in detail a peer review process protocol that embodies the best practices of federal
agencies and many of the principles espoused in the main body of the present text. Finally, Section VII, the bibliography contains an extensive list of primary and related references to the peer review literature. First, some definitions and background will be presented, to set the stage for detailed examination of issues surrounding peer review.

DEFINITIONS AND BACKGROUND

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

Review of a research program can then be viewed as consisting of two elements: 1) "review of a program of research", which examines the nature of the component projects, and is commonly referenced as an in-depth technical review, and 2) "review of a research program", which examines the nature of the structural relationships among the projects and between the projects and their external environment, and is commonly referenced as a management review. These two elements could be merged operationally into a single review, or could be performed separately.

A program could be single research discipline intra- or inter-agency; multiple discipline intra- or inter-agency; multiple discipline vertically integrated intra- or inter-agency; multiple discipline multi-agency multi-national; or other variants of the above. The nominal program discussed in this report is assumed to be intra-agency; the nominal review is assumed to be intra-agency. Some organizations review by disciplines, some organizations review by multi-discipline management unit, and in some organizations disciplines coincide with management units.

Peer Review Definition
The classical definition of a peer is "A person who has equal standing with another." A peer review, then, is a review of a person or persons by others of equal standing. The crucial issue then becomes how “equal standing” is defined.

Most research peer reviews with which the author is familiar--whether of journal research manuscripts, research proposals for funding, or research project performance reviews--tend to employ peer reviewers who are experts in the specific research area of the person or group under review. Depending on the relative levels of expertise between the reviewers and reviewees, the reviewers may or may not be de facto peers. Applied to research program review, such experts are most competent for the in-depth technical subset defined above as "review of a program of research." The focus of this subset is on the intrinsic nature of the collection of research projects
within the program, especially on their quality, accomplishments, ongoing problems, unexpected findings and discoveries.

The focus of the management review subset defined above as "review of a research program" is on the structural relationships among the research projects within the program. This subset addresses issues such as mission relevance, budget adequacy, program staff, objectives, and procedures. To address the issues of this subset, additional types of peers to those of the first subset are required.

For the purposes of the present document, a more liberal interpretation of a peer than normally employed will be used to encompass the requirements for addressing both subsets of research program peer review. This expanded definition of a peer describes the types of reviewers that the author has tended to choose in conducting research program peer reviews that combine both subsets of program review into a single process. In this more inclusive definition, a peer may be a person expert in the specific technical area of the research being reviewed, in allied technical areas to the research being reviewed, in technology areas that may be impacted eventually by the research being reviewed, and in systems and operational areas that may be impacted in the future by the research being reviewed. These different types of peers are required to examine the different facets of a research program that could have impacts far beyond the specific research area being reviewed.

**Research Program Peer Review Background**

Research evaluation methodologies can be divided generically into three groupings [Kostoff, 1995b, 1996a]: Qualitative (e.g., peer review); Semi-Quantitative (e.g., retrospective); and Quantitative (e.g., bibliometric). Peer review of research is overwhelmingly the method of choice in practice in the U.S., as well as the rest of the world [Salasin, 1980; Logsdon, 1985; Chubin, 1990; Chubin, 1994; Kostoff, 1995b; Stamps, 1997a; Wood, 1997]. Presently, the major applications of research peer review are, in order of decreasing usage: journal manuscript submission review; proposal review; project and program review; faculty performance review; and dissertation review.

Most of the peer review literature has focused on manuscript and proposal review. For example, a 1993 literature survey [Speck, 1993] compiled 780 abstracts of papers on peer review, of which 643 papers were on journal peer review. According to Armstrong [Armstrong, 1997], 101 of these provided empirical evidence. Relatively few studies have been done on the issues and principles underlying project or program review and reported in the open literature. This conclusion, complemented by Speck's and Armstrong's findings, was confirmed most graphically by a recent peer review literature survey conducted by the author. Over half the documents retrieved were either letters to the editors of journals, or editorials (or their equivalent). The papers on program review tended to be reports of technical and statistical results of the review, with little or no focus on the principles and issues underlying the peer review components. Whatever papers existed on peer review component principles related to manuscript reviews (mainly) or proposal reviews.

Peer reviews of research programs, when done at all, are not nearly as consistent across the research sponsoring organizations as are the manuscript and proposal reviews. Program reviews tend to range from very informal personal discussions to tens of formal panel reviews. Most of the people
who conduct program reviews do not document them in the literature, and most of the principle and concept papers in the peer review literature are written by people who have never conducted a research program peer review. Consequently, there are two major gaps in the literature on research program peer review. First, there are quantitatively few papers published, and second, most of the concept and principle papers that do exist bear little relation to the reality of conducting a program review.

To identify and address some of these gaps, a number of peer review issues will be examined now. These issues were selected from a taxonomy of categories generated by the author's recent peer review literature survey, as well as from previous assessments of problems with peer review and other research evaluation approaches [Kostoff, 1996a]. The headings of the topical issues addressed in the main body of this text immediately following the present section include:

Objectives and Purposes of Peer Review;
Quality of Peer Review;
Impact of Peer Review Manager on Quality;
Selection of Peer Reviewers;
Selection of Evaluation Criteria;
Secrecy (Reviewer and Performer Anonymity);
Objectivity, Bias, and Fairness of Peer Review;
Normalization of Peer Review Panels;
Repeatability/ Reliability of Peer Review;
Effectiveness/ Predictability of Peer Review;
Global Data Awareness;
Costs of Performing a Peer Review;
Ethical Issues in Peer Review;
Alternatives to Peer Review;
Recommendations for Further Research in Peer Review.

IV. PEER REVIEW PRINCIPLES

OBJECTIVES AND PURPOSES OF PEER REVIEW

Global funding for science and technology (S&T) is approaching one trillion dollars per year. The S&T products resulting from this funding are the engines that drive today’s global economies and militaries. It is important that this trillion dollar investment be used efficiently to maximally accelerate S&T progress. One way is to insure that efficiencies are implemented at all stages of the investment cycle.

The S&T investment cycle progresses from a planning phase to a proposal phase, to a selection phase, to an execution phase, to a review phase, and then returns to a planning phase. There are continual feedbacks among the phases. Underlying each of the phases is an ongoing S&T evaluation process, to aid in both the tactical and strategic decisions required for efficient operation of the phase. This ongoing evaluation process has three components, the balance among which
depends on the specific phase. First, retrospective evaluation of S&T assesses program performance, identifies S&T products that can be taken to the next stage of development, and identifies the management and performance environment most conducive to producing high-quality S&T. Second, real-time evaluation of ongoing S&T is used to modify management, performers, and resources as required in order to maximize progress and efficiency. Finally, evaluation of potential S&T identifies how resources should be reallocated in the future to select S&T portfolios with the highest estimated returns.

A spectrum of methods is available to perform these evaluations. Ideally, evaluation methods would be selected on the basis of how well they contribute to the objective of accelerating the progress of S&T efficiently. Specifically, evaluation methods would be chosen on their capability to identify and eliminate or overcome the barriers to efficient S&T progress. These barriers or deficiencies include:

- Risk-averse S&T;
- Short-term horizon S&T;
- Over-emphasis on evolutionary rather than revolutionary S&T;
- Poorly coordinated S&T;
- Lack of interdisciplinary S&T;
- Unawareness of parallel or previously performed S&T;
- Insufficient documentation and dissemination of S&T products;
- Emphasis on tactical S&T management at the expense of strategic S&T management;
- S&T resource allocations made for reasons other than technical merit;
- S&T manpower imbalances and deficiencies;
- Maintenance in perpetuity of costly S&T infrastructures and facilities;
- Reluctance to share new ideas openly;
- Reluctance to terminate low-output S&T.

This spectrum of evaluation methods ranges from quantitative (metrics) to semi-quantitative (anecdotes) to qualitative (peer review). While all three classes are represented in the published literature, peer review is in practice the overwhelming method of choice. A properly conducted peer review can surface many of the barriers or deficiencies for programs or organizations undergoing review, especially barriers or deficiencies in the first half of the list. However, peer review is not an end in itself; it is a means to the end of accelerating S&T progress efficiently. How well does peer review serve as a mechanism to achieve the objectives stated above? This report attempts to answer that question. This report also describes potential improvements in the peer review process that could help eliminate or reduce the barriers to efficient accelerated S&T progress.

In practice, peer review supports many diverse purposes:

- It serves as a quality filter to conserve resources.
- Papers published in peer-reviewed journals are assumed to be above a threshold of minimal quality, such that the reader can focus limited time resources on the highest quality documents assumed to be contained in these journals.
• Projects and programs selected for initiation or continuation by peer review are assumed to be above a threshold of minimal quality.
• Precious labor and hardware resources can be focused on these high quality tasks selected.
• Peer review has the potential to add value to, and improve the quality of, the manuscript or program under review.
• Peer review can provide an imprimatur of legitimacy and competency to increase a program's visibility and support.
• The objectives of peer review range from being an efficient resource allocation mechanism to a credible predictor of research impact.
• A properly conducted research program peer review can provide credible indication to the research sponsors of program quality, program relevance, management quality, and appropriateness of direction [Alassaf, 1996; Armstrong, 1997; Cram, 1992; Gabel, 1992; GERMANY, 1988; Kessler, 1992; Levine, 1988; Palli, 1993; Rainville, 1991; Ramsay, 1989; Stull, 1989; Wakefield, 1995; Wicks, 1992].

The literature contains some quantitative studies that indicate some value added by peer review. For example, mid-1990s studies evaluated the effects of peer review and editing on manuscript quality [Goodman, 1994], and the effects of peer review and editorial processes on the readability of original articles [Roberts, 1994]. They concluded that peer review and editing improve the quality of medical research reporting, as well as the readability of original articles and their abstracts. They did not address whether the quality of the research was improved, nor do other literature articles.

From the author's experience, there are three times during the research program peer review process when value is added. First is the period between reviews, when the researchers do their work knowing that it will be subject to high quality review. The value added during this performance phase is that the researchers will maintain a higher level of performance quality because of the knowledge of the forthcoming expert review. For example, performers will be less inclined to essentially work on extending their theses for decades if they know that they will be evaluated periodically. Program managers will be more likely to continually update the balance and relationships among their component projects, rather than allow poor performers to languish, if they know that a review is forthcoming.

The analogy is to a well-known speed trap on a highway. The knowledge that a stretch of road is well-policied is sufficient to keep the average speed within the posted limit. The fact that the officers write relatively few tickets in this area is not a measure of effectiveness of the speed trap. It would be useful if studies were done comparing the quality of research of periodically reviewed programs to infrequently ad hoc reviewed programs to see if this value added component is experimentally verifiable.

Second is the period of review preparation, particularly the “dry runs” for reviews that include presentations. This is an extremely valuable experience, both for the managers and the researchers, and would by itself justify the cost and effort of the total review. Especially for research program peer review, the preparation period provides a focal point for discussion of unresolved issues and
priorities, and fuels substantive discussions in order to arrive at a quality presentation. The value added is not in the superficial presentation form improvement, but in the substantive increase in the intrinsic program quality.

Third is the actual review. Here, independent viewpoints are injected in a public forum, high quality research is re-affirmed, and strong recommendations are provided for the fate of poor research.

A fourth time of value added could be postulated as well, depending on the review results. If the review outcome was very favorable, and eventually resulted in additional program funding, then value was added, at least to the funding recipients and hopefully to the larger society as well.

Finally, it should be remembered that any of the review processes involve real-time judgments of the quality of research, not expressions of the intrinsic quality of the research. The passage of time is required to follow the evolution of research to ascertain whether it achieves its promise. How well these peer review judgments relate to the actual impact of the research on science and technology and society is an important measure of long-term peer review value, and is addressed to some extent in the later section on Predictability.

Another taxonomy of the potential values added by peer review can be summarized as follows [Chubin, 1994]:

1. an effective resource allocation mechanism;
2. an efficient resource allocator;
3. a promoter of science accountability;
4. a mechanism for policymakers to direct scientific effort;
5. a rational process;
6. a fair process;
7. a valid and reliable measure of scientific performance.

Much of the remainder of the main body of this report examines the intrinsic and arbitrary roadblocks to achieving these desirable goals in a research program peer review. Many of the negative aspects of program peer review will be addressed, such as potential bias, cost, and protection of the status quo. The present section concludes by examining briefly another potentially negative aspect of peer review not addressed by the literature; namely, whether the knowledge of periodically scheduled reviews would stifle the pursuit and presentation of very innovative but far-out ideas. Would performers be reluctant to present these ideas in a public forum, where either the credibility of the performers could be challenged for these ideas or the ideas themselves could be usurped by the reviewers? In other words, does the practice of peer review, and especially panel-based program peer review, effectively result in self-censorship of radical ideas? This is an area where research is needed to ascertain whether ideas have been suppressed in periodically reviewed programs, and then to determine how this problem could be surmounted if it exists.

**QUALITY OF PEER REVIEW**
The studies related to peer review that have been reported in the literature range from the mechanics of conducting a peer review, to examples of peer reviews, to detailed critiques of peer reviews and the process itself. In addition to descriptions of peer reviews and processes contained in the reviews and surveys referenced above, other examples of processes and critiques can be found in [Armstrong, 1997; Chubin, 1990; Chubin, 1994; Barker, 1992; Cicchetti, 1991; Cole, 1981; DOE, 1993; Frazier, 1987; Kostoff, 1996a; Wood, 1997].

While the reported studies of peer reviews present the process mechanics, the procedures followed, and the review results, the reader cannot ascertain the quality of the findings and recommendations of the review. In practice, procedure and process quality are mildly necessary, but nowhere sufficient, conditions for generating a high quality peer review. Many useful peer reviews have been conducted using a broad variety of processes, and while well documented modern processes (e.g., [DOE, 1993]) may contribute to the efficiency of conducting a review, more than process is needed for high quality. Many intangible factors enter into a high quality review [Evans, 1990; Friedman, 1995; Goodman, 1994; Lundberg, 1991; Luukkonen-Grunow, 1990; McNutt, 1990; Vandenbroucke, 1994], and some of the more important factors will be discussed.

The underlying hypothetical postulate of this section is that there exists an intrinsic quality inherent in every basic research task. By definition, a high quality peer review should provide an accurate picture of the intrinsic quality of the research being reviewed, irrespective of whether this intrinsic quality is high or low. The fundamental problem is the lack of absolute standards (analogous to physical standards for primary measurements such as time and length) for measuring research quality. Presently, evaluation of intrinsic research quality is a subjective process, depending on the reviewers’ perspectives and past experiences. A high quality review under these imperfect circumstances, then, would occur when two generic conditions are fulfilled: 1) utilization of highly competent reviewers, and 2) no injection of additional distortions in the reviewers’ evaluations as a result of biases, conflict, fraud, or insufficient work.

High quality peer review processes require as a minimum the conditions summarized from Ormala [Ormala, 1989]:

1. The method, organization and criteria for an evaluation should be chosen and adjusted to the particular evaluation situation;
2. Different evaluation levels require different evaluation methods;
3. Program and project goals are an important consideration when an evaluation study is carried out;
4. The basic motive behind an evaluation and the relationships between an evaluation and decision making should be openly communicated to all the parties involved;
5. The aims of an evaluation should be explicitly formulated;
6. The credibility of an evaluation should always be carefully established;
7. The prerequisites for the effective utilization of evaluation results should be taken into consideration in evaluation design.
The impact of a peer review on decision-making is considered a measure of its effectiveness, not its quality. Poorly conducted peer reviews could have major influences on decisions, and well conducted peer reviews could have minimal influence on decision-making. It is important to distinguish quality from effectiveness.

A corollary aspect of peer review quality, although in the author's judgment not a primary contributor to nominal research program peer review quality, is the commission of errors by the reviewers. The author is not aware of published studies that have examined the commission of errors by research program peer reviewers. In a 1997 paper [Armstrong, 1997], different studies of errors and superficial work by peer reviewers of journal manuscripts are described. The conclusion one draws from these results is that the problem of manuscript reviewer error production is not insignificant. Armstrong does make the point that journal manuscript peer reviewers typically receive no extrinsic awards, are typically anonymous, and therefore in some cases may not feel motivated to exert the effort required for a high quality review. Additionally, there is something of an imbalance in this author-reviewer symbiosis, since the journal article author spends hundreds of hours performing the work and is required to place his reputation on the line when submitting the article for publication, while the reviewer spends relatively few hours at his task with essentially little chance of damage to his reputation for mediocre performance. The legal system recognizes the existence of these human frailties, and has a multi-level hierarchical appeals system established to handle possible errors by judges and juries. Both the medical and legal professions have effectively established an appeals procedure through their malpractice system. Perhaps the scientific profession needs a more formal appeals system to level the playing field for manuscript authors and others subject to peer review, and to insure that in the end justice will be served and quality maintained. A 1997 paper [Stamps, 1997b] reviews the literature on conflict resolution, and describes a process (dialectical scientific brief) for resolving disputes from manuscript peer review in scientific journals. This or some alternative procedure could be modified to apply to other types of scientific peer review as well.

In most research program peer reviews, commission of technical errors by reviewers due to the relaxed standards resulting from anonymity and lack of financial incentives is probably not nearly as serious as in manuscript reviews. While a small fraction of program reviews may be carried out by anonymous mail reviews from experts (if this is done at all, it would apply when the program is evaluated by reviewing each of the projects separately), the vast majority of program reviews are carried out with the use of expert panels. In some cases, the panel members may receive modest compensation, but in any case, they are no longer anonymous. Their reputations are on the line as they participate in these panels. In the author's experience, panel members tend to suppress overt expressions of biases, and they typically make statements they are able to defend. Whether this translates into more conservatism relative to the anonymous journal manuscript reviews depends on how the review process is structured, and is discussed in more detail later in the section on Secrecy. In any case, studies of the extent of errors committed by research program peer reviewers remain to be done, and if these panels eventually have substantial input to the budgetary process, then some sort of appeals system for program reviews may have to be established.

**IMPACT OF PEER REVIEW MANAGER ON QUALITY**
From the author's perspective, the single most important factor in producing a high quality research program peer review is the dedication of an organization's senior management to the highest quality objective review, and the associated deployment of rewards and incentives to encourage such reviews. The second most important factor in producing a high quality review (and in fact the cornerstone of a successful review) is the motivation of the person managing the review to conduct a technically credible review. This review leader selects and manages the review process, selects the review criteria, selects the reviewers, guides the questions and discussions in a panel review, summarizes the reviewers' comments in a mail or panel review, and makes recommendations about whether a program should be initiated, continued, or modified.

The direction of the assessment may be heavily influenced if the review leader consciously or unconsciously exercises biases, especially while selecting reviewers. In an extreme case of bias, the review's results could be determined completely by reviewer selection before the reviewers ever meet. This conclusion is valid for the manager of a program or project review, the manager of a proposal review, or the editor in charge of a journal manuscript review. The author is not aware of any of these types of reviews where the reviewers are selected by a random process, which would eliminate much of the selection bias. Because of this potential intrinsic bias due to the conscious reviewer selection by the review manager, unless random reviewer selection is operable in conducting a review, any mathematical correlations [e.g., Cicchetti, 1991] among reviewers' scores and review outcomes (illuminating and insightful though these correlations may be) must be opened to question.

**SELECTION OF PEER REVIEWERS**

Even with the strongest support from an organization's top management, and the direction of an unbiased and competent review leader, the quality of a review will never go beyond the competence of the reviewers. Two dimensions of competence that should be considered for a research review are the individual reviewer's technical competence for the subject area, and the competence of the review group as a body to cover the different facets of research issues (other research impacts, technology and mission considerations and impacts, infrastructure, political and social impacts) [Kostoff, 1995b, 1996a; Garson, 1980; Klahr, 1985; Marshall, 1996]. The quality of a review is limited by the biases and conflicts of the reviewers. The biases and conflicts of the reviewers selected should be known to the leader and to each other. One common error in panel selection is limiting the choice of research experts to those who have specific expertise in the subdisciplines of the existing program. This provides an answer to the question of whether the job is being done right, but not to whether the right job is being done. The former question relates to detailed technical quality, while the latter question relates more to investment strategy in the broadest sense (investment strategy is the rationale for the prioritization and allocation of resources among the program components). To answer the latter question, people with broad expertise in the area covered by the overall program's highest level objectives should also be selected. They would be able to address the investment strategy more objectively, and determine whether the mix of subdisciplines, and the allocation of resources among the subdisciplines, is appropriate. The review group, then, would be able to address the central question of whether the right job is being done
One of the major criticisms of peer review, whether manuscript, proposal, or program, is that it tends to perpetuate orthodox and conservative paradigms, and tends to reject new paradigms that threaten the structure of the status quo. If one of the objectives of a research program peer review is in fact to ensure that innovation is recognized, that truly revolutionary research with attendant new paradigms will be promoted and rewarded, then this selection of reviewers to address the right job issue in parallel with reviewers to address the job right issue becomes of paramount importance.

Many present research program peer reviews remain severely deficient in the concentration of panel experts on the issue of doing the job right and the effective absence of experts on doing the right job. This can lead to the situation that the author has termed "The Pied Piper Effect" [Kostoff, 1996a]. This phenomenon was defined initially for the specific case of interpretation of journal paper citations, but it is applicable to any conclusion resulting from any type of peer review as well: journal, proposal, or program. Its initial bibliometric definition, and then extrapolation to program peer review, follows.

Using citations as a stand-alone measure of quality and impact has raised concerns about the potential bimodal interpretation of the numerical results. The traditional bimodal interpretation is that a paper could receive high citations because of its high quality, or because the citers disagree with it. However, there is a third interpretation: the "Pied Piper" effect. It may be the most insidious, and further precludes citations being utilized in stand-alone mode.

Assume there is a present-day mainstream approach in a specific field of research; for example, the chemical, radiiological, and surgical approach to treating cancer (See [Kostoff, 1996a] for a more detailed example of the "Pied Piper Effect"). Assume the following hypothetical scenario:

- There are alternative approaches to treatment not supported by the mainstream community;
- In fifty years a cure for cancer will be discovered;
- The curative approach has nothing to do with today's mainstream research, but is perhaps a downstream derivative of today's alternative methods;
- It turns out that today's mainstream approach sanctioned by the mainstream medical community was completely orthogonal or even antithetical to the curative approach.

Then what meaning can be ascribed to research papers in cancer today that are highly cited for supposedly positive reasons?

In this case, a paper's high citations are a measure of the extent to which the paper's author has persuaded the research community that the research direction contained in his paper is the correct one, and not a measure of the intrinsic correctness of the research direction. It is analogous to firing a missile precisely at the wrong target. It is the essence of the difference between precision and accuracy. In fact, the high citations may reflect the deliberate desire of a closed research community (the author and the citers) to persuade a larger community (that could include politicians and other resource allocators) that the research direction is the correct one.
This is the "Pied Piper" effect. **The large number of citations in the above hypothetical medical example becomes a measure of the extent of the problem, the extent of the diversion from the correct path, not the extent of progress toward the solution.** The "Pied Piper" effect is a key reason why, especially in the case of revolutionary research, citations and other quantitative measures must be part of and subordinate to a broadly constituted peer review in any credible evaluation and assessment of research impact and quality.

The extrapolation of the "Pied Piper Effect" to research program peer review becomes obvious. Many technical communities are comfortable with the status quo, have large personal and infrastructure investments in the mainline orthodox approaches, and feel threatened by new paradigms that could render their investments obsolete. If the peer reviewers represent only the community of the specific research approach being reviewed, then the debate will typically center around the correctness of the miniscule details of the approach (job right) rather than whether the approach should be used at all (right job). The net effect of such a limited review is to provide a stamp of approval (analogous to the high citation rates described above) to continuance of the mainline approach, and to close the door to revolutionary thinking. Appendix I describes a method for selecting peer reviewers that approximates the best practices in use today. While it is not a pure random selection process, it does remove much of the bias of present selection practices, and would be appropriate for the large scale program peer reviews discussed here.

**SELECTION OF EVALUATION CRITERIA**

Research evaluation criteria are one instrument through which an organization promulgates strategic and policy research objectives. Detailed responses to the criteria by reviewers are valuable as inputs for downstream decision-making. When documented, review criteria also serve as tangible indicators to external groups that strategic objectives are being implemented [Delcomyn, 1991; Eibeck, 1996; Kellie, 1991; Martin, 1981; Sutherland, 1993; Weinberg, 1964, 1989].

Individual criteria can be viewed mathematically as the components of a vector. The complete vector, or figure of merit of the review, can then be constructed as the weighted sum of the scores of its components. For example, assume two criteria, Research Merit (RM) and Mission Relevance (MR), are generated by the evaluating organization to be used by reviewers for research program evaluation. Assume each criterion is weighted equally by the evaluating organization. Then, in the absence of further constraints, the final figure of merit, overall program quality (OPQ), is computed as \( \text{OPQ} = 0.5 \times \text{RM} + 0.5 \times \text{MR} \).

Problems arise, however, because the stated criteria are seldom the only criteria the reviewers consider important. In the case above, the evaluating organization selected only two criteria that it felt were important and that it wanted the reviewers to address. It also selected the weighting to be assigned to each criterion, and the figure of merit algorithm. Conflict arises because each reviewer has his or her own view of:

- what criteria are important for evaluating research,
- how these criteria should be weighted for a particular program, and
• how they should be integrated for a final figure of merit.

In the author's experience covering hundreds of different types of peer reviews, evaluators actually conceive a Gestalt, or view of the integrated nature, of the total research package when performing the evaluation. The component criteria provided serve to stimulate reviewers' thinking in specific areas, and insure that the reviewers include issues deemed critical to the review managers.

In the example case, there is the potential for serious mismatch between the final figure of merit vector obtained by the organization's algorithm and by the reviewers' mental algorithm. The two vectors could be sufficiently different that one could completely misrepresent the other. For example, assume the organization provided the algorithm above to the reviewers, and also assume that the definition of Research Merit (importance of the problem to science) did not include Research Approach (approach taken to solve the problem). Assume the reviewers felt that the RM and MR were high quality for a program being reviewed. However, assume that the reviewers felt the Research Approach taken was extremely poor in the program under review, and that Research Approach was the most important criterion in deciding the overall value of this particular research program. In this case, use of the organization's criteria and algorithm will provide a conclusion orthogonal to that desired by the reviewers. Even if the organization provides the additional flexibility of allowing the reviewers to provide their own weighting to the criteria, in the example shown the reviewers' desired conclusion will still be orthogonal to that obtained using the organization's algorithm with criteria of arbitrary weighting.

The author has found that expert reviewers are usually individuals of integrity, and the way they resolve the above dilemma is through the principle of compromise rather than the compromise of principle. Operationally, the reviewers develop an intuitive judgment of the worth of the total research package under review, then "reverse-engineer" the weighting and scoring of the criteria sub-consciously (if not consciously) until the evaluation algorithm comes closest to their desired intuitive overall result.

Based on these observations, the author recommends (and uses) inclusion of an overall project/program quality criterion as well. This “bottom-line” score makes clear the reviewers' judgments about the total research package presented, and incorporates the effects of any unstated criteria (e.g., organizational appropriateness) that a reviewer feels are important determinants of overall research quality. This approach reduces the necessity for “reverse engineering” to arrive at displaying the reviewers' deepest convictions. If the evaluating organization still wants to use only its own criteria to arrive at the final figure of merit, then, by comparing the reviewers' vector and the organizational algorithmic vector, the organization can identify the trade-off in reviewer-perceived quality that resulted from ignoring reviewer-relevant criteria.

The later section in this report on agency peer review practices discusses the more detailed studies performed by the author and others on selection and importance of research program evaluation criteria. In general, these studies show that the most influential criteria relative to a reviewer's final evaluation rating are research merit, research approach, and performer quality. In addition, a relevance criterion is important in mission agencies. Nearer-term relevance, such as transition to
technology (or utility), tends to be more influential on a reviewer's final overall rating than longer-term relevance to the sponsor's downstream mission. Finally, as stated above, inclusion of a single "bottom-line" criterion is crucial.

SECRECY: REVIEWER AND PERFORMER ANONYMITY

The issue of reviewer anonymity was discussed briefly in the section on Quality, with the conclusion that anonymity did not help the detailed technical quality of the reviewer's product. From the author's viewpoint, this negative aspect pales compared to the benefits resulting from reviewer anonymity, although there is not a unanimity of opinion on this conclusion in the literature [Altura, 1990; Berezin, 1994; Clayson, 1995; Debackey, 1990; Frei, 1993; Gresty, 1995; Knox, 1981; Neetens, 1995].

What is really desired from a peer reviewer is an honest viewpoint on the intrinsic quality of research under review, supported where possible by rigorous technical analysis. 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 and eliminate many of the types of errors the studies report [Armstrong, 1997] discussed earlier in the Quality section.

However, having the reviewer and reviewee present during the review will, in many cases, tend to inhibit the expression of the reviewer's deepest convictions about the quality of the research. Rewards are few for making strong negative statements about a research paper, proposal, or program, and resulting retributions and resentments may far outweigh the intrinsic benefits of stating judgments honestly and forthrightly. In research program peer review in particular, the situation is more complex than a manuscript peer review. In program review, the program manager is in a real sense being reviewed, as well as the research. If the reviewers are "bench-level" experts in the field of the manager's research program--as one assumes they typically are--and at some point in the future would have an interest in participating in the manager's specific research program, then forthright but negative reviews could damage their prospects of obtaining future funding from the program manager. Finding true peers to serve as research program reviewers in this case may be extremely difficult, and requires judicious care in the selection process.

The author has conducted program and proposal reviews that ran the gamut from complete reviewer anonymity to complete reviewer presence with reviewee and audience. In the author's experience, there is a hierarchy of levels of reviewer anonymity that produce different degrees of frankness and honesty in the reviewer's response.

The most honest and straightforward reviewer's opinions result from phone reviews where the reviewer is completely anonymous to the reviewee. In this case, the reviewer has been provided information about the research (typically written) and provides feedback orally over the phone. The frankness of response is most evident in evaluating the right job function, where the integrity of the total research approach is at stake. Reviewers are less reluctant to be more open when critiquing the job right function, since major direction and infrastructure changes will not be at risk, and the
reviewee's defenses will not be as vociferous.

Next in the hierarchy are written reviews where the reviewer is completely anonymous to the reviewee. Some reviewers will tend to moderate the frankness of their comments when asked to provide them in writing. However, if the reviewers trust the review manager to protect their anonymity, they will still be quite frank in their write-ups.

The next level of anonymity occurs when the reviewers and reviewees are both present during the research presentations, but the reviewers meet in closed session to provide oral and written evaluations of the research, with these evaluations not for attribution. Even the presence of the anonymity during the closed session will provide much frank discussion and exchange of heartfelt opinion.

The final level is the absence of anonymity, where both reviewers and reviewees are present throughout the total process, and all verbal and written comments are provided with full attribution. While it may be argued that this type of review is better than having no review, from the author's experience this approach does not begin to utilize the full potential of what expert peer review can offer.

The other side of the secrecy coin is withholding the reviewee's name and affiliation from the reviewer. This process has been called "blind reviewing" [Blank, 1991; Ceci, 1984; Cox, 1993; Evans, 1990; Fisher, 1994; Johnson, 1995; Laband, 1994; McNutt, 1990; Nylenna, 1994; Rosenblatt, 1980; Shaughnessy, 1988; Sly, 1990]. Its objectives are to provide fairer reviews of work by unknown researchers or by researchers from less prestigious institutions [Armstrong, 1997], or conceivably to eliminate bias based on personal characteristics like gender. Blind reviewing (and its corollary "double-blind" reviewing, when both the reviewer and reviewee are anonymous to each other) is probably most applicable to manuscript review. Some studies of blind reviewing for journal manuscripts have been reported [Fletcher and Fletcher, 1997; Fisher, 1994; Laband, 1994]. Reviews by blinded reviewers were judged by the editors to have higher quality; the blinded reviewers gave better scores to authors with more previous articles, and articles published in journals using blinded peer review were cited significantly more than articles published in journals using non-blinded peer review.

Unfortunately, removing the identity of the reviewee from the research under review is like solving an equation after eliminating the dominant term. The DOE peer review study of the quality of its Office of Basic Energy Sciences' research program [DOE, 1982], which is probably the classic study of research program quality using a statistical sampling of component project quality, concluded that team quality was the most important variable in determining overall project quality. Based on these, and other similar results, evaluating proposals without reviewee identity could provide misleading results. There are many good proposed research topics. The high quality researcher will develop a track record of not only addressing good research topics, but will make substantial progress toward solutions through perseverance and critical thought. Today, many consulting firms help researchers prepare funding proposals. These consultants are very aware of the appropriate “buzzwords” and politically correct terminology, and what type of formatting and
proposal organizational structure will appeal most to decision makers. Judging such proposals independent of the researcher will eventually allow form to predominate over substance.

In any case, blind reviews probably have minimal applicability to research program reviews. In most cases, panel reviews are used, and extraordinary precautions would have to be taken to protect the identity of the reviewees. Coupled with the inability to use the team quality criterion, there appears to be little motivation to employ this process in program peer review. There appears to be nothing on this topic related to program review in the literature.

**OBJECTIVITY, BIAS, AND FAIRNESS OF PEER REVIEW**

Probably the most criticized aspect of all types of peer review is the role of bias, and its subsequent impact on fairness, in the reviewers’ final recommendations. Peer reviews have received written and verbal accusations of having gender bias, race bias, institutional bias, geographic bias, age bias, and especially a conservative bias toward protecting the “old boys” network of the status quo. Much research effort has been focused on this issue of bias and fairness [Armstrong, 1982, 1997; Bailar, 1991; Daniel, 1993; Ehlen, 1996; Ernst, 1994; Ramasarma, 1995; Spitzer, 1994]; Armstrong [Armstrong, 1997] makes the point that almost half of the empirical papers on journal reviewing in a massive 1993 study [Speck, 1993] address these issues.

The findings are mixed. A 1994 study [Gilbert, 1994] assessed whether manuscripts received by the JAMA possessed differing peer review and manuscript processing characteristics, or had a variable chance of acceptance, associated with the gender of the participants in the peer review process. The study concluded that gender differences exist in editor and reviewer characteristics at JAMA with no apparent effect on the final outcome of the peer review process or acceptance for publication.

Another study [Peters, 1982] found that reviewers were biased against authors from unknown or less-prestigious institutions. A study in which NSF proposal reviews were re-evaluated by a different panel [Cole, 1981] included institutional reputation, professional age, academic rank, geographic location, and other variables. It concluded that the peer review system employed by NSF was essentially free of systematic bias. A study of the DOE Office of Basic Energy Sciences [DOE, 1982] stated that the conclusions concerning the laboratory and non-laboratory projects were not distorted by reviewer biases.

A 1992 report elaborates on the concerns of bias and conflict in a section describing guidelines on a common framework for organizing Federal investments [NAS, 1992]. Its Principle 6 (Program Evaluation) contains the statement: "Current efforts to review government R&D programs have suffered, in some instances, from the fact that annual reports to Congress or the executive branch have been conducted by mission agency employees with a direct interest in having projects they evaluate continue. Technical evaluations of the R&D work and of the contributions to national economic welfare of pre-commercial R&D programs should be conducted by nongovernmental groups that do not have a direct role in program management or funding decisions".
The underlying paradigm of the bias/fairness issue is that all reviewees should be treated the same; there should be a level playing field for all players. The rationale for fairness is that decisions made on the basis of other than technical merit can impede the main objective of accelerating S&T progress efficiently. Unfortunately, in the implementation of this noble philosophy, the rules of scientific evidence take second priority to the rules of political correctness. This motivation toward perceived increased fairness is probably the main driver for peer review concepts such as 'blind reviewing', which was addressed in the previous section of this report on Secrecy. It was concluded that the downside to "blind reviewing" was the elimination of the key reviewer criterion of track record (team quality) and the subsequent degradation of the review process quality.

However, assigning overwhelming importance to track record, as proposed by some researchers in the later Alternatives section of this report, shifts the functional balance toward emphasizing the job right aspect of the research as opposed to the right job aspect, and is in many respects a double-edged sword. It presents serious obstacles for young researchers with little track record who may have very good ideas for solving difficult research problems and may be very capable of addressing these problems, and has the potential for maintaining the “old boys” network and the status quo. This can have very serious consequences, as the discussion of the "Pied Piper Effect" showed. The solution is not to eliminate the key variable of researcher identity, but rather to select reviewers such that the perspective of the panel is broadened. Use panelists who are able to address the right job aspects of the research target, to insure that outmoded but prolific and well-cited research is not promulgated in perpetuity, and that the pool of expertise is being continually refilled.

NORMALIZATION OF PEER REVIEW PANELS

Peer review is a diagnostic process that can be applied in isolation on a body of research, or can be used for comparing many different types of research. When applied for comparative purposes, a key issue centers on how the results of different panels evaluating different technical disciplines can be normalized such that comparisons across disciplines and panels become meaningful. How, for example, can the differences in intrinsic quality of the different types of research being reviewed be separated from different panel biases, different panel interpretations of criteria, different severities of panelists in applying the criteria, when only scores and comments that include all these factors are presented. This normalization issue is perhaps the most difficult aspect of peer review, and normalization difficulty also applies to other aspects of research evaluation such as bibliometrics [Braun, 1982; Kostoff, 1997c; Schubert, 1996].

Most studies that examine peer reviews across disciplines present the results for the major discipline categories separately [e.g., DOE, 1982; Cicchetti, 1991; Cole, 1981]. They essentially finesse the problem. While this separation of categories is valid when research is viewed from a strategic viewpoint, where disciplines are selected and maintained for their importance to an organization's mission, this discipline separation reduces the value of peer review as a quality comparative yardstick considerably. Quantitative evaluation approaches, such as bibliometrics, develop reference standards for different disciplines and then construct appropriate scaling procedures for ranking the research [Schubert, 1996]. This does allow for comparison of relative rankings across disciplines in a broad generic sense, but questions arise [Kostoff, 1996a] as to the
applicability of reference standards defined for a discipline (e.g., acoustics) to programs being compared within the discipline (e.g., underwater acoustics vs aeroacoustics).

The author has not seen any fully satisfactory peer review normalization approaches due to the presence of the many variables listed previously. However, one interesting normalization approach is used by the Dutch STW for evaluating research proposals [Van den Beemt, 1991, 1997]. Technical comments, but not quality ratings, are provided by technical peers. The comments, and proposer responses, for twenty different proposals are then provided to twelve people from a variety of disciplines. This 'jury' of twelve provides the scores through an independent mail review. Essentially, the normalization is provided by having the twelve jurors common to all proposals.

The author has used two approaches to improve normalization across panels somewhat. First is the utilization of some individuals common to all panels. In a series of competitions for new accelerated research programs that was held in the late 1980s [Kostoff, 1988], the author served as chairman of all the different discipline panels. This resulted in some small measure of normalization among the different panels. Use of more individuals common to all panels would have provided an extra measure of normalization, and in this sense the presence of senior management during the reviews provided additional measures of normalization. Obviously, the more closely the panels are related topically, the more valuable is the technical contribution of individuals common to the different panels.

Second, it was assumed that the difference in aggregated average scores for major disciplines (e.g., physical sciences and life sciences) was due to two factors: differences in intrinsic quality of the programs proposed and differences in the scoring severity of the reviewers. To normalize, a fraction of the differences in aggregated average scores for the major disciplines was removed. This was assumed to eliminate the scoring severity difference. Trial and error showed a fifty percent correction factor provided results that appeared intuitively reasonable to the relevant audience members who had attended all the reviews. This normalization procedure had the added benefit of preserving and insuring representation from disciplines that had strategic value to the organization.

This approach to normalization could have a second interpretation. If the research is viewed as having a strategic component and a quality component, with the reviewers' scores viewed as addressing the quality component only, then the correction could be perceived as adjusting for the presence of the strategic component. For example, assume a Life Sciences panel produced an average program score of five, and an Engineering Sciences panel produced an average score of ten. Assume further that each discipline had equal strategic value to the organization, and that the strategic value was of equal importance to the reviewers' scores (assumed to be a total program quality score that includes mission relevance). Then the normalized total score can be computed as $FOM = 0.5*STRAT + 0.5*SCORE$, and the difference between the two panels' scores would be reduced from five to 2.5. This correction factor can then be applied to the raw score of each program within the discipline to arrive at a final 'normalized' score.

If peer review is eventually used to support GPRA, then some sort of normalization procedure will
be required for credibility. Given the very limited validity of existing schemes for normalization, especially across disparate disciplines, this will be difficult. If GPRA is used to affect research budgets, valid procedures to normalize scores will be essential, and they do not exist now. This is a very fertile area for peer review research.

REPEATABILITY AND RELIABILITY OF PEER REVIEW

In a physical system experiment, one of the main questions asked to gauge credibility of the results concerns the repeatability of the results. Can the same experiment be run at different laboratories under the same controlled conditions and yield the same results, or some reasonable facsimile thereof? The analogous issue in peer review has been termed alternatively reliability, repeatability, consistency, uniformity, etc., and has received much focus in the literature [Bailar, 1991; Ceci, 1982; Cicchetti, 1976, 1979, 1991; Cole, 1991; Colman, 1991; Crothers, 1993; Daniel, 1993; Gorman, 1991; Halpin, 1986; Kiesler, 1991; Kraemer, 1991; Laming, 1991; Luce, 1993; Marsh, 1989; Roediger, 1991; Rosenthal, 1990, 1991, Rubin, 1992]. The meaning is the same.

There are two corollary concepts in physical systems that unfortunately are not always carried over to peer reviews. These are the concepts of precision and accuracy. Precision represents the degree to which a measurement value can be replicated, while accuracy represents the relation of the measurement value to some absolute value or standard.

In a very comprehensive study of the reliability of peer review for manuscripts and grant proposals [Cicchetti, 1991], which included hundreds of references, reliability was defined generically by different measures: internal consistency, inter-referee agreement (degree of agreement among referees), and stability across time. Reliability by these definitions appears to be the analogue of precision as defined above, and the issue of accuracy does not appear to enter the definition. The study stated that the most common measure is inter-referee agreement at a given point in time. The study essentially concluded that, across the various science disciplines examined: 1) agreement is better on manuscript and grant submissions of perceived poor quality than on submissions of good quality; 2) better defined (specific and specialized) areas of scientific inquiry have higher acceptance rates and use fewer reviewers than less well-defined (general and less focused) areas of scientific interest; and 3) levels of chance-corrected inter-referee agreement are rather low.

However, neither the study commentary nor the descriptions of the studies addressed the issue of truly random reviewer selection, and therefore the meaning of their conclusions is open to question. For example, what is the meaning of high reliability under these conditions? It could mean that the reviewers were able to identify and report accurately on the intrinsic quality of the manuscript or proposal, or it could mean that the reviewers were selected because of their extreme bias (positive or negative) toward the topic and the review manager did an outstanding job of selecting reviewers with similar biases.

One school of thought holds that chance-corrected inter-referee agreement should in fact be low, because the astute manager will pick reviewers who have sharply different viewpoints and expertise, so that they should be sensitive to different kinds of problems. From this perspective, too
much agreement may be a sign of weakness, that the system is not eliciting the full spectrum of opinion that the manager needs to make an informed decision.

A study of National Science Foundation (NSF) proposals [Cole, 1981], funded by NSF, using two sets of reviewers, showed a reversal rate (one group's decision would have been reversed by the other group) of about twenty-five percent. Since an entirely random process would have produced a reversal rate of fifty percent, it was concluded that the fate of a particular grant application is roughly half determined by the characteristics of the proposal and the principal investigator, and about half by apparently random elements. It was also concluded that the great bulk of reviewer disagreement observed is probably a result of real and legitimate differences of opinion among experts about what good science is or should be.

Similar reliability studies of research program reviews do not appear to be in the literature, probably because of the expense and effort of doing the replication involved in such studies, especially for panel reviews, and the question of whether the identical process is actually being replicated. The author's experience with reviews of existing and proposed research programs, a small fraction of which was documented and analyzed mathematically [Kostoff, 1992, 1997a], is that reliability is sufficient for practical purposes. As stated more fully in [Kostoff, 1996a], while a peer review can gain consensus on the proposed and existing research programs that are either outstanding or poor, there will be differences of opinion on the programs that cover the much wider middle range. For programs in this middle range, their fate is somewhat more sensitive to the reviewers selected. If a key purpose of a peer review is to insure that the outstanding programs are funded or continued, and the poor programs are either terminated or modified strongly, then the capabilities of the peer review instrument are well matched to its requirements.

The author's experience with the reliability of program peer reviews appears to be somewhat less negative than those above, or other similar studies reported in the literature. Why is this? It probably is due in large measure to how the peer review is conducted. In many proposal and manuscript reviews reported in the literature, there tends to be minimal feedback among the reviewers, and between the reviewers and authors or proposers. Probably at best there is one written rebuttal. This independence is undoubtedly valued, and is also less expensive than convening all the players to interact jointly.

The author's peer reviews involve extensive interaction among the reviewers and presenters. Many misunderstandings and differences in interpretation are clarified during the exchange of technical information before the scoring is performed. The initial scoring is performed independently by the reviewers. Then, differences in scores are discussed, and the reviewers are provided the opportunity to modify their scores. Usually, the final scores become closer. From the author's observations, this scoring variance reduction is not due to the dominance of more forceful or vociferous debaters, but rather is due to each reviewer's coming to a better understanding of the intrinsic nature of the material presented. Thus, rather than inter-reviewer agreement as the measure of reliability used for the journal manuscript analyses [Chicchetti, 1991], for research program peer review a better measure of reliability may be agreement of average panel scores after panels are conducted in the interactive mode suggested above.
EFFECTIVENESS AND PREDICTABILITY OF PEER REVIEW

Peer review predictability directly affects the credibility of technological forecasting. An organization peer reviewing research should consider relating the reviewers' scores to downstream impact on the organization's mission [Abrams, 1991; Van den Beemt, 1991, 1997]. A few studies have been done relating reviewers' scores on component evaluation criteria to proposal or project review outcomes (e.g., [DOE, 1982; Kostoff, 1992]). Some studies have been done in which reviewers' ratings of research papers have been compared to the numbers of citations received by these papers over time [Bornstein, 1991a; Bornstein, 1991b]. Correlations between reviewers' estimates of manuscript quality and impact and the number of citations received by the paper over time were relatively low. Bornstein concludes, after an extensive survey of peer review reliability and validity, that: "If one attempted to publish research involving an assessment tool whose reliability and validity data were as weak as that of the peer review process, there is no question that studies involving this psychometrically flawed instrument would be deemed unacceptable for publication." [Bornstein, 1991b].

The author is not aware of large-scale studies, singly or in tandem, that have related peer review scores and rankings of proposals to downstream impacts of the research on technology, systems, and operations, although some efforts toward this end have been initiated [Van den Beemt, 1991]. This type of study would require an elaborate data tracking system over lengthy time periods. No such tracking system currently exists. Thus, the value of peer review as a predictive tool for assessing the impact of research on an organization's mission (other than research for its own sake) rests on faith more than on hard, documented evidence.

GLOBAL DATA AWARENESS

In all of the decision aids, placement of the technology of interest in the larger context of technology development and availability world-wide is an absolute necessity. This tends to be a central deficiency of most management decision aids. Global data awareness is deficient because of the following factors (Kostoff, 2003).

1) Information Comprehensiveness is limited because there are many more disincentives than incentives for publishing S&T results. Except for academic researchers working on unclassified and non-proprietary projects, the remainder of S&T performers have little motivation for documenting their output.

a) For truly breakthrough research, from which the performer would be able to profit substantially, the incentives are to conceal rather than reveal. Proprietary research with these characteristics is especially difficult to document. As industrial sponsorship of, and participation in, academic research becomes more pervasive, and as many academic researchers also form small companies, there is decreasing incentive from this sector of academia to publish, as well.

b) For research that aims to uncover and correct product problems, there is little motivation (from
the vendor, sponsor, or developer) to advertise or amplify the mistakes made or the shortcuts taken.

c) For very focused S&T, the objective is to transition to a saleable product as quickly as possible; no rewards are forthcoming for documentation, and the time required for documentation reduces the time available for development.

d) For research of a classified or "grey" nature, especially in today's environment of fear of terrorism, there is no motive for documentation, at least in the open literature.

Therefore, only a very modest fraction of the S&T performed ever gets documented. This may sound surprising to people who have been bombarded with an "explosion" of technical documentation. However, much of this explosion may be due to a recent phenomenon known as "paper inflation." What would have been one substantive comprehensive technical paper three or four decades ago is now sub-divided into multiple papers, each covering a portion of the parameter range of interest. Additionally, very modest variants of a given paper are published in multiple forums.

Of the performed S&T that is documented, only a very modest fraction is included in the major databases. The contents of these knowledge repositories are determined by the database developers, not the S&T sponsors or the potential database users.

None of the research-sponsoring governments, including the United States, appear to have control over the contents of, or interfaces with, these large S&T databases. Basically, the Federal government is footing the bill for the research that makes these large databases useful tools, but the Federal government is at the mercy of the database developers in terms of addressing the government's needs for database contents and operational requirements. The present system is heavy on data generation and light on data dissemination.

Of the documented S&T in the major databases, only a very modest fraction is realistically accessible by the users because:

- the databases are expensive to access,
- not very many people know of their existence,
- the interface formats are not standardized, and
- many of the search engines are not user-friendly.

Insufficient documentation is not just an academic issue: in a variety of ways, it retards the progress of future S&T and results in duplication.

2) **Information Quality** is the product of amount of information provided and intrinsic quality of this information. Quality control is typically exerted through the peer review process, and the *pro bono* peer review process used today by the research journals has many well-known limitations. **Information Quality** content is limited because uniform guidelines do not exist for
contents of the major text fields in database records (Abstracts, Titles, Keywords, Descriptors), and because of logic, clarity, and stylistic writing differences. The medical community has some advantage over the non-medical technical community in this area, since many medical journals require the use of Abstracts that contain a threshold number of canonical categories – Structured Abstracts – while almost all non-medical technical journals do not.

Compatibility among the contents of all record text fields is not yet a requirement. As our studies have shown, this incompatibility can lead to different perspectives of a technical topic, depending on which record field is analyzed. This field consonance condition is frequently violated, because the Keyword, Title and Abstract fields are used by their creators for different purposes. This violation can lead to confusion and inconsistency among the readers.

3) **Information Retrieval** is limited because time, cost, technical expertise, and substantial detailed technical analyses are required to retrieve the full scope of related records in a comprehensive and high relevance fraction process. Of all the roadblocks addressed in this section, this is the one that attracts probably the most attention from the Information Technology (IT) community. Because much of the IT community's focus is on selling search engine software and automating the information retrieval process, they bypass the 'elbow grease' component required to get comprehensive and high signal-to-noise retrieval.

4) **Information Extraction** is limited because the automated phrase extraction algorithms required to convert the free text to phrases and frequencies of occurrence as a necessary first step in the text mining process leave much to be desired. This is especially true for S&T free text, which the computer views as essentially a foreign language due to the extensive use of technical jargon. Both a lexicon and technical experts from many diverse disciplines are required for credible information extraction.

Poor performance by the automated phrase extraction algorithms can result in:

- lost candidate query terms for semi-automated information retrieval;
- lost new concepts for literature-based discovery;
- generation of incomplete taxonomies for classifying the technical discipline of interest; and
- incorrect concept clustering.

For clustering in particular, the non-retrieval of critical technical phrases by the phrase extractor will result in artificial cluster fragmentation. Conversely, the retention of non-technical phrases by the phrase extractor will result in the generation of artificial mega-clusters.

Detailed labor-intensive manual cleanup is therefore crucial to success. Thousands of phrases must be examined and culled by technical experts to insure that the appropriate high technical content phrases are generated in usable form. This level of human effort required is not advertised by the software vendor community, and as a result, many users are disappointed by the results produced from the software alone.

5) Two types of **Technical Expertise** are required for a credible text mining study, text mining
technology expertise and technical (and related) domain expertise. Text mining technology
*Technical Expertise* is limited because the intrinsic complexity of text mining has not been
appreciated by the technical community, and resources have not been made available for the
development of text mining experts. In contrast, target domain and related technical expertise
exist, but their use in text mining studies has been limited both by tradition and by lack of
understanding of the role of technical domain experts in high quality text mining. Because much
information retrieval in the past and present has been performed by non-technical domain expert
library support staff, the need and cost for higher priced technical experts to participate in the text
mining studies is viewed as a non-essential expenditure. In addition, the developers of text
mining software promote the concept that intelligent agents and smart algorithms can substitute
for human experts.

An on-going text mining literature survey shows that there are in fact very few people actually
developing the true text mining processes globally and increasing the understanding of what text
mining can offer. For example, the only group actually publishing the results from the literature-
based discovery text mining application is Swanson and Smalheiser. Perhaps a couple of other
people, including the author, have written concept papers about literature-based discovery. The
literature-based discovery experience mirrors that of the other S&T text mining applications, as
well. The research impact road-mapping application is being addressed by only one group (the
author). There is a major mismatch between the potentially substantial benefits of these myriad
S&T text mining approaches and the number of researchers and developers who understand,
advance, and apply them.

**RELEVANCE OF EVALUATION CRITERIA TO FUTURE ACTION**

*Every S&T metric, and associated data, presented in a study or briefing should have a decision
focus; it should contribute to the answer of a question that in turn would be the basis of a
recommendation for future action.*

Almost every metrics briefing the author has attended failed this test. Metrics and associated data
that do not perform this function become an end in themselves. They offer no insight to the
central focus of the study or briefing, and contribute nothing to decision-making. Over time they
tend to devalue the worth of metrics in credible S&T evaluations. Because of:

- the political popularity and subsequent proliferation of S&T metrics,
- the widespread availability of data, and
- the ease with which this data can be electronically gathered, aggregated, and displayed,

most S&T metrics briefings and studies are immersed in data geared to impress rather than
inform. While metrics studies provide the most obvious examples, this conclusion can be easily
generalized to any of the evaluation methods.

**COSTS OF PERFORMING A PEER REVIEW**
Another problem with peer review is cost [ASTEC, 1991; Buechner, 1974; Hensley, 1980; Kostoff, 1995b, 1996a]. The true total costs of peer review, as will be shown, can be considerable but tend to be ignored or understated in most reported cases. Because there are many different types of peer review, it is very difficult to provide a total cost rule-of-thumb for generic peer review. Nevertheless, consider the following illustrative example for an order of magnitude estimate on total research program peer review costs [Kostoff, 1996a].

Assume that an interim peer review is desired of a $1M/yr program at a laboratory. The review mode of operation will be to bring a panel of experts to the laboratory site for two days, and hear presentations from the principal investigators. Assume that the panel consists of ten experts in research, technology, mission operations, etc., and that eight principal investigators will present their projects to the panel. The loaded cost (salary plus overhead) for each panel member is assumed to be $150,000 per year, and the loaded cost for each principal investigator is assumed to be $125,000 per year. Direct expenditures, such as panel per diem and travel costs, would be in the neighborhood of $6,000-8,000. Any honoraria would increase this cost.

Indirect expenditures, such as total reviewer, presenter, staff, and review audience time spent toward the review, would be in the range of $125,000 and would include at least the following:

1. Presenter time in preparing background material for reviewers to read before review, preparing the presentation, making dry runs for management, etc. [$40,000 estimate; 80 person-days];
2. Panel member time for reading background material (papers, reports, plans), traveling to review, spending time at meeting, writing report, etc. [$48,000-60,000 estimate; 80-100 person-days];
3. Agency staff time for identifying and soliciting reviewers, establishing review and coordinating with lab, writing reports, etc. [$10,000 estimate; 20 person-days];
4. Audience (lab management, other lab personnel, other agency representatives, etc.) time at review [$20,000 estimate; 40 person-days].

The main conclusion of this discussion is that for serious panel-type peer reviews, where sufficient expertise is represented on the panels, total real costs will dominate direct costs. This conclusion would also be true for mail-type peer reviews. While the total costs of mail-type peer reviews would be less than those of panel-type peer reviews due to the absence of travel costs, the ratio of total costs to direct costs for mail-type peer reviews would be very high. The major contributor to total costs for either type of review is the time of all the players involved in executing the review. With high quality performers and reviewers, time costs are high. 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.

**ETHICAL ISSUES IN PEER REVIEW**

The professional ethics of research must deal with, among other issues, scientific fraud, scientific misconduct, betraying confidential information, and unduly profiting from access to privileged information. There are both legal and unwritten, unspoken agreements and penalties that underlie
the maintenance of ethical standards in these areas. One subordinate objective of peer review, whether at the manuscript [Fox, 1994], proposal, or program level, is to maintain high ethical standards, especially as applied to fraud and misconduct. Since many of the fraud and misconduct violations have occurred in the written technical product, most of the reported applications of peer review in this area have emanated from journal peer review [Fielder, 1995; Goodstein, 1995; Gupta, 1996; Keown, 1996; Mokrasch, 1988; Moran, 1992; Southgate, 1992]. The maintenance of ethical standards in these areas tends to be through self-policing by the research community. The author has seen no program peer reviews in which fraud and misconduct were uncovered, and has not identified any such cases in the literature.

There is a fundamental ethical paradox that underlies any form of research peer review. For the review process to have credibility, experts must be employed, either for the right job function or the job right function. Contrary to popular opinion, it has been the author's experience (based on directed experiments and on personal observations during the conduct of reviews) that there are very few real experts in any specific research field. Armstrong [Armstrong, 1997] draws a similar conclusion relative to manuscript peer review, to the effect that the reviewers may work on similar areas but not the same specific problem, so that the reviewers have less experience on the total problem than do the authors. Thus, in order to obtain real experts for a panel, at least to evaluate the job right aspects of the research, a relatively small community must be accessed. Usually, the members of this community are acquainted with each other, and are either research collaborators or research competitors. They may compete for funds or awards or prestige or promotions, or other types of recognition. Thus, there is an inherent bias or conflict of interest in the process when real experts are desired as reviewers.

 Usually, in research program peer review, there are (or should be) documents that reviewers sign to protect the confidentiality of the research being reviewed, but pragmatically it is the adherence to the unwritten and unspoken ethical standards that restricts the unwarranted use of proprietary and sensitive information. There are also legal protections, and recently there have been court cases brought by those who felt their confidences and proprietary research had been violated through illegal expropriation of the results for personal reviewer gain.

No matter what documents reviewers sign, no matter how resolutely they wish to adhere to the highest ethical standards, they cannot help but be influenced by the privileged information to which they have access. The transfer of knowledge occurs through many pathways, and listening to detailed technical presentations or reading technical proposals are probably two of the more effective. Thus, the operative solution to the ethical dilemma posed by access to technical material is the principle of compromise rather than the compromise of principle. The ethical reviewer takes no conscious overt actions to reveal confidences or profit unduly from participation in the peer review, but rather accepts as his reward for participation the satisfaction of having aided the larger research enterprise and having improved his thought processes from exposure to different ideas. If the larger use of research program peer review becomes a reality, and if the outcomes are used to influence budgetary decisions, then more efforts need to be devoted to insure adherence to some of the ethical standards discussed here.
ALTERNATIVES TO PEER REVIEW

This report has identified a number of problems associated with the use of peer review. These problems conceptually transcend the different peer review applications of program, proposal, and manuscript evaluation, although the implementation severity of different problems is different for each of the applications. There have been a number of proposals for peer review modifications or complete alternatives [Forsdyke, 1991; Greene, 1991; Roy, 1981, 1984, 1985; Smith, 1988; Wick, 1996; Wood, 1997], in attempts to overcome the most egregious aspects of peer review. Most of these alternative concepts focus specifically on research proposal peer review, although some of their component ideas apply to the other applications of peer review as well. Two of the more widely known alternatives will now be presented and critiqued.

Bicameral Review
A modified form of peer review for project selection has been propounded in recent years by some Canadian scientists [Berezin, 1995; Forsdyke, 1991]. This methodology has been termed "Bicameral Review" by its originator, Dr. Forsdyke, and its essence is as follows.

The structure of Bicameral Review is founded on the assumption that the research funding system is highly error-prone due to the inherent uncertainty of predicting the outcome of basic research. If an evaluation system is highly error-prone, then that error-proneness has to be taken into account in system design. Two principles of decision-making in uncertain environments are: 1) place most weight on parameters most likely to be assessed with some degree of objectivity, and 2) hedge your bets.

In Bicameral Review, grant applications are divided into a major retrospective part (track record of proposers), and a minor prospective part (the work proposed), which are routed separately. The retrospective part only is subjected to peer review. The prospective part is subjected to in-house review by the agency, solely with respect to budget justification. The peers are required to assess not just productivity, but productivity per dollar received. Furthermore, they have to factor in the experience of the applicant. Young researchers are given more funding "rope" (the benefit of the doubt), until they have established a record. Funding is allocated on a sliding scale, replacing existing sharp fund-no fund cutoffs. Only those at the very top of the funding scale would get all the funds they needed to complete the work in a reasonable time. As the merit rating of the projects decreased down the funding scale, the fraction of requested funds would decrease as well.

Productivity-Based Formula Systems
A non-peer review alternative has been proposed [Roy, 1981, 1985], based on the principles that:

- past success is the best predictor of future performance,
- supporting small groups on a continuing basis for a reasonable time period increases probabilities of success and system efficiencies, and
- most innovative science is done with a minimum of micro-management.

This alternative proposes that researchers be funded essentially based on track record, and provides
an algorithm for allocating funds. In one algorithmic incarnation [Roy, 1985], the dollars awarded would be proportional to some weighted sum of numbers of publications, numbers of advanced degrees, dollar volume of research support from mission agencies, and dollar volume of research support from industry, and the award would be to a research unit (Departments, etc). Again, the underlying principle is that performance rather than promise will provide a much firmer basis for public accountability. New investigators added to a research unit would have extra shares added to the base formula allocation.

**Author's Commentary on Alternatives**

Ideally, a research proposal evaluation process should be able to allocate funds to the ideas with the greatest potential, independent of the source of these ideas. Such a process should be able to include ideas from established researchers with strong track records, established researchers with weak track records, and new researchers with no track records. It should be able to cover researchers from academia, government, and industry, ranging from one person operations to very large organizations, and cover classified and non-classified work with different venues and cultures for reporting research results. The allocation process should incorporate the best technical judgments in arriving at final decisions, recognizing the uncertainties involved in projecting the outcomes of fundamental research.

The two alternative approaches selected place heavy emphasis on awards to established researchers with strong track records. They differ in how the track records would be determined, with Bicameral using peers and productivity-based using a formula. Both minimize the use of true technical experts in the evaluation of the prospective portion of proposed research. In actual practice, these alternatives would not differ quite as significantly from existing peer review processes as might be imagined from first reading. As stated previously in this report, analyses have shown that Team Quality, a euphemism for performer track record, is the dominant factor in determining reviewer overall quality score for existing and proposed research. Thus, both the existing and alternative approaches de facto place heavy emphasis on track record. The real difference between the alternatives and the existing approaches, in the author's opinion, is the use of technical experts in evaluating the prospective portion of the proposal.

While both alternative approaches would reduce:

- the cost of submitting proposals to some degree,
- the impacts of reviewer bias,
- whatever pirating exists of novel ideas by competitors, and
- some unnecessary time expenditures in the review processes,

they have some drawbacks. Extremely heavy emphasis on track record to the exclusion of expert judgment on proposed concepts promulgates continuation of orthodox mainstream approaches by increasing the obstacles to new entrants into the research arena. Lack of technical expertise in the judgment of proposed research could lead to more non-technical factors predominating in the selection process, and the relative ascendance of form over substance in the evaluation.
In a zero-sum game, the Bicameral Review process appears to allocate some funds from the “best” proposals to the ‘worst' proposals because of the sliding scale and elimination of the sharp cutoff. It does, however, provide a 'safety-net' that allocates some funding to all, or almost all, researchers. The productivity based system has some analogies to the present GPRA approach addressed in the precursor Science article [Kostoff, 1997b], and suffers from many of the same drawbacks. Use of any metric or combination of metrics as a stand-alone approach for evaluating research is subject to error. The metrics chosen may or may not be a valid indicator of research quality; interpretation by peers is required to validate the credibility of the metrics. The formula based approach has the negative potential of driving researchers to achieve numerical output targets rather than fundamental understanding.

The productivity approach is similar to a recursive system of equations, and if the initial conditions are flawed, the final figure of merit would be flawed. For example, one of the formula terms is dollars received for research from mission agencies. Suppose a research team had received major grants that were 'earmarked' in legislation. This could lead to better numbers for at least two of the other formula terms as well, numbers of graduate students and papers produced, and then result in a high overall figure of merit that was not necessarily related to the intrinsic quality of the research program. This allocation based on flawed initial conditions would recur each year until it became a self-perpetuating system, even after the 'earmarking' was terminated. Thus, if any formula or combination of quantitative indicators is used, it must be accompanied by, and subordinate to, expert peer review, in order to avoid the occurrence of situations such as the one above.

These alternatives, and others of similar nature, are based on the premise that the peer review selection process does not yield the best research, and the tremendous expenditures of time and energy in generating proposals do not justify the continuance of such an inexact process. The validity of this basic premise can be challenged. While peer review has its imperfections and limitations, there is little evidence that the best researchers and ideas are going without funding, and far less evidence that the alternatives above would improve the situation.

**SCIENCE COURT**

A non-standard peer review approach for concept evaluations is the Science Court. As in a legal procedure, it has well defined advocates, critics, a jury, etc. It is a unique and potentially powerful technique, but like any tool, can be misused if not understood and applied properly. It was applied by the author to a review of alternate fusion concepts in the magnetic fusion office in 1977 [DOE, 1978].

The general format chosen for the evaluation was a panel review by selected evaluators with an adversary type of procedure. The main component groups in the process were a Steering committee, Evaluation Panel, Advocates, and Critics. These participants and their roles in the evaluation are described below.

The Steering committee consisted of fusion office representatives. The chief responsibilities of this committee were (1) to organize the evaluation, (2) to define the evaluation criteria, (3) to choose
members of the Evaluation panel, (4) to assist the Evaluation panel in the reviews, and (5) to receive the evaluators' conclusions and recommendations and draft a final report to the fusion office.

The Evaluation panel was composed of plasma physicists, fusion reactor systems experts, and a representative of the utility industry. The panel did not include active proponents of any of the concepts under consideration. In case of a remote conflict of interest, a panel member excuses himself from the deliberation on the particular concept involved. The panel was responsible for the technical evaluation of all concepts.

The Advocates of a concept were those scientists and engineers who were working on that particular concept. The Advocates were responsible for providing and defending scientific results and projections, as well as the technology and attractiveness of the reactor embodiment. A Chief Advocate was designated to coordinate the activities of the Advocates.

Critics were chosen for their special expertise in an area of physics or engineering that was important to a particular concept. The Critics' responsibility was to ferret out crucial physics and technology questions and to aid the Evaluation Panel in the review of experimental results and theoretical models. Proponents of one concept in some cases served as critics in the evaluation of another concept. One person was chosen as a Chief Critic and was given the responsibility of coordinating the activities of the Critics.

Any of the participants (Advocates, Critics, or the Evaluation Panel) were allowed to utilize outside experts as they deemed appropriate. This procedure probably had more debate and surfacing of crucial issues than any other concept evaluation seen by the author. However, it was time-consuming compared to a standard panel assessment.

**NETWORK-CENTRIC PEER REVIEW**

Network-centric peer review makes maximum use of information technology to eliminate many of the problems with traditional peer review. Appendix IV outlines the theory and proposed implementation of network-centric peer review.

**RECOMMENDATIONS FOR FURTHER RESEARCH IN PEER REVIEW**

The issues and concerns described above illuminate a number of gaps and deficiencies in the practice of research program peer review especially, and other forms of peer review as well. The overriding recommendation is that research be initiated in those aspects of research program peer review that have been analyzed for manuscript and proposal peer review. The literature is very sparse in studies of the practices and principles of program peer review. If program peer review undergoes an expansion to support GPRRA, then a much greater understanding of its strengths and weaknesses is required in order for it to become an effective and credible comparative diagnostic instrument.
One of the central problems in all types of peer review is lack of credibility in its predictive reliability. More studies are necessary to relate evaluations by peers of research proposals and existing research programs to future impacts of this research. Presently, the data to validate different predictive models does not exist. What is required is a database that allows tracking of the evolution of products of research in their various metamorphized stages. Having such a database would allow not only validation of peer review predictive models, but bibliometric predictive models and other quantitative predictive models as well. The database would allow predictive reliability to be determined for a number of different types of impact. These would include impact on the research area of interest, impact on allied research areas, impact on technology, impact on systems, impact on operations, etc.

Discussions of the validity and reliability of the peer review results can be found in Cicchetti [Cicchetti, 1991] and Daniel [Daniel, 1993], as well as in other commentary in the journal issue in which Cicchetti's article appears. To improve validity and reliability, research needs to be done on optimal numbers of reviewers utilized; ascertaining whether author anonymity impacts the results; and ascertaining whether training people to perform peer reviews would increase review quality as well as reliability and validity.

There are very few comparative studies of different types of peer groupings and the quality of the peer review product. Studies should be done varying mail versus panel review, the British model versus the standard non-British model (peer review using professionals instead of eminent persons), panel size, types of reviewer expertise, time expended by the reviewers and reviewees on the process, and correlating these variables with the quality of the product. Central to the result would be how the review’s cost impacts the quality of its product, and how this is affected by the different variables.

Normalization across many parameters (disciplines, panels, etc.) was identified previously as a major unknown. It is worth repeating again that research be performed on how to normalize across a variety of research program peer review parameters.

While the present report included a very approximate estimation of total peer review time and dollar costs for one peer review scenario, more accurate time and cost estimates would be required when comparing different types of peer review scenarios. Extensive data taking would be necessary, because of the many different types of peer reviews in existence. However, since total peer review costs can be substantial, and since cost reduction with consistent quality would be one of the goals of these different types of suggested studies, both the extensive data taking and development of improved peer review cost estimating procedures would be well justified from an economic viewpoint.

The application of expert systems and knowledge-based systems for proposal evaluation and program review could supplement peer review. Few studies have been done along these lines, but a 1993 dissertation [Odeyale, 1993] and follow-on studies [Odeyale, 1994a, 1994b] address this problem in detail. Much more work would be required to validate the application of these advanced technologies as useful supplements to peer review, but more research in this direction
could determine whether there is potential for real payoff.

One of the potential benefits resulting from a peer review is constructive feedback to the reviewees followed by an improvement in the reviewees’ conduct of research. Studies should be done to ascertain reviewees' perceptions of the peer review and the review's value in improving the conduct of research. An innovative study [Luukkonen, 1993] addresses peer review from the reviewee's perspective, but much more can be done to improve the information transfer from the reviewers to the reviewee, and to insure that the review's recommendations were translated into improved research.

V. PEER REVIEW PRACTICES

SELECTED PEER REVIEW PRACTICES: PROPOSED PROGRAMS

There are many approaches used by research sponsoring organizations to conduct peer reviews for selecting proposed research. This section focuses on selected peer review approaches that reflect the state of the art in the technical community and pays special emphasis to how research impact is incorporated into the peer review process. The four case studies presented include the National Science Foundation (NSF), the National Institutes of Health (NIH), the Office of Naval Research (ONR), and the Dutch Technology Foundation (STW). Grant proposals are also addressed by presenting the highlights of an excellent grant proposal study.

1) NSF

The two largest Federal sponsors of basic research are the National Institutes of Health (NIH) and the National Science Foundation (NSF) [NSF, 1996]. The NSF peer review process of research proposals illustrates how potential research impact influences selection of new research areas. In the NSF process, proposals received are assigned to program officers for review. The program officers select external peer reviewers and use mail and/or panel approaches to have the proposals assessed and rated. The program officers then perform their own assessment of the proposals and forward their recommendations to higher levels. These recommendations are rarely overturned [Frazier, 1987].

From the 1987 version of the NSF Brochure, Information for Reviewers, reviewers use four criteria to assess the proposals:

1. Research Performance Competence
2. Intrinsic Merit of the Research
3. Utility or Relevance of the Research
4. Effect of the Research on the Infrastructure of Science and Engineering

These criteria were adopted by the National Science Board in 1981 [NSF, 1997]. Research impacts are evaluated through the second, third, and fourth criteria. The second criterion, Intrinsic Merit, incorporates impact of the proposed research on other research fields in its
definition and is a measure of the nearer term impact of the proposed research. The third criterion, Utility, addresses potential contribution to an extrinsic goal such as a new technology. The fourth criterion, Infrastructure, incorporates impact on the nation's research/education/human resource base.

In 1996, the NSF merit review process was evaluated by a task force. The National Science Board recommended that the new review criteria proposed in the final task force report [NSF, 1997] be approved for implementation on October 1, 1997. The specific task force recommendations were that the following two criteria be adopted in place of the four criteria that were being used.

1. **What is the intellectual merit of the proposed activity?**
   The following are suggested questions to consider in assessing how well the proposal meets this criterion: How important is the proposed activity to advancing knowledge and understanding within its own field and across different fields? How well qualified is the proposer (individual or team) to conduct the project? (If appropriate, please 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 organized is the proposed activity? Is there sufficient access to resources?

2. **What are the broader impacts of the proposed activity?**
   The following are suggested questions to consider in assessing how well the proposal meets this criterion: 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 underrepresented groups (e.g., gender, ethnicity, geographic, etc.)? To what extent will it enhance the infrastructure for research and education, such as facilities, instrumentation, network, 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?

The task force further recommended that a cover sheet be attached to the proposal review form, which presents the context for using the criteria. The suggested language for this cover sheet was as follows:

**Important! Please Read Before Beginning Your Review!**

In evaluating this proposal, you are requested to provide detailed comments for each of the two NSF Merit Review Criteria described below. Following each criterion is a set of suggested questions to consider in assessing how well the proposal meets the criterion. Please respond with substantive comments addressing the proposal's strengths and weaknesses. In addition to the suggested questions, you may consider other relevant questions that address the NSF criteria (but you should make this explicit in your review). Further, you are asked to address only questions that you consider relevant to the proposal and that you feel qualified to make judgments on.
When assigning your summary rating, remember that the two criteria need to be weighted equally. Emphasis should depend upon either (1) additional guidance you have received from NSF or (2) your own judgment of the relative importance of the criteria to proposed work. Finally, you are requested to write a summary statement that explains the rating that you assigned to the proposal. This statement should address the relative importance of the criteria and the extent to which the proposal actually meets both criteria.

Regarding the 'ratings' issue, which was highlighted in the Discussion Report, the task force recommended that the NSF 'generic' proposal review form provide for the following:

1. Separate comments for each criterion
2. Single composite rating
3. A summary recommendation (narrative) that address both criteria

In the new process, research impacts are the focus of the second criterion. These include impacts on infrastructure, education, science, technology, and diversity. Thus, not only are technical impacts considered, but potential socio-political impacts are considered as well. Finally, it is unclear how other unwritten criteria, such as government vs industry appropriateness for funding, which may be important for a specific project/program, would impact the composite rating.


In July 2002, the NSF Director issued Important Notice 127 (Implementation of New Grant Proposal Guide Requirements Related to the Broader Impacts Criterion). This Important Notice reinforced the importance of addressing both criteria in the preparation and review of all proposals submitted to NSF. The Important Notice also indicated NSF’s intent to continue to strengthen its internal processes to ensure that both of the merit review criteria are addressed when making funding decisions. NSF also issued Important Notice No. 127 (http://www.inside.nsf.gov/pubs/2002/iin127/imptnot.pdf) to let the community know that effective October 1, 2002, NSF will return without review proposals that do not separately address both merit review criteria within the Project Summary. Also, NSF issued Important Notices 123 and 126 (http://www.inside.nsf.gov/pubs/2000/iin126/iin126.htm) to inform the community that merit review would be handled electronically.

2) NIH
In the NIH process, proposals are sent to initial peer review groups, composed mainly of active researchers at colleges and universities, where they are reviewed for scientific and technical merit. After receiving a priority rating from the peer reviewers, the proposals are then sent to a statutorily mandated advisory council, composed of scientists and public members, for a program relevance
review. After the council members recommend action to be taken on the proposals (usually concurrence with the peer group recommendations, but sometimes special action [Frazier, 1987]), the institute staff rank the proposals and initiate a funding strategy.

In response to a perceived need to refocus the review of grant applications on the quality of the science and the impact it might have on the field, rather than on details of technique and methodology, NIH has developed five new criteria for initial review of proposals for implementation in October 1997. Reviewers will be asked to apply the criteria in judging whether the proposed research is likely to have a substantial impact on advancing the goals of NIH-supported research: advancing understanding of biological systems, improving control of disease, and enhancing health. The new rating criteria are:

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

- **Approach:** Are the conceptual framework, design, methods, and analyses adequately developed, well-integrated, and appropriate to the aims of the project? Does the applicant acknowledge potential problem areas and consider alternative tactics?

- **Innovation:** Does the project employ novel concepts, approaches or method? Are the aims original and innovative? Does the project challenge existing paradigms or develop new methodologies or technologies?

- **Investigator:** Is the investigator appropriately trained and well suited to carry out this work? Is the work proposed appropriate to the experience level of the principal investigator and other researchers (if any)?

- **Environment:** Does the scientific environment in which the work will be done contribute to the probability of success? Do the proposed experiments take advantage of unique features of the scientific environment or employ useful collaborative arrangements? Is there evidence of institutional support?

In assigning a single global score for each application, the reviewers are to consider all criteria, weighting each criterion as appropriate for each application.

It appears that only the first criterion, Significance, relates to impact, and can include the relatively near term impact on allied research fields. Broader impact and relevance issues appear to be the purview of the advisory councils. The council members are asked to assess the fairness and appropriateness of the initial scientific review as well as the proposal's relevance to institute research program goals and broader societal health-related matters.

3) ONR
The ONR does not require formal peer review of individual research grants, but leaves the choice of
peer review to its scientific officers. Circa 1992, it required a competitive process among internal Navy organizations (claimants) with external reviewers for those accelerated program proposals that constituted about 30 per cent of the total ONR program [Kostoff, 1988, 1991, 1992]. The claimants that won the competition then went to the technical community (if their charter were extramural) and advertised their areas of interest for proposals, or, if their charter were intramural, performed the work in-house.

In a detailed description of the competition [Kostoff, 1988], all the accelerated programs proposed by the claimants (ARIs) were categorized into areas of similar science, and the proposals in each area were evaluated by a panel of experts external to ONR. The written portion of the evaluation required numbers and comments for factors related to research quality and Navy relevance. In this process, the factors on the scoresheet relating to potential research impact estimation were:

1. Research Merit (RM)
2. Potential Impact on Naval Needs (PINN)
3. Potential for Transition or Utility (PTU)

The Research Merit criterion incorporates the potential impact of the research, if successful, on allied research areas. The Potential Impact on Naval Needs criterion deals with downstream impact of the proposed research on naval systems and operations. The Potential for Transition or Utility criterion incorporates the potential nearer term impacts of the proposed research. Transition refers to the actual transfer of research programs to development and Utility refers to other mechanisms by which a program's results would be transmitted to, and used by, the technical community.

A key component of this process was the use of mixed levels of reviewers on the panels to evaluate the different potential impacts of research. The panels included bench-level researchers to address the impact of the proposed research on the field itself; broad research managers to address potential impact on allied research fields; technologists to address potential impact on technology and the potential of the research to transition to higher levels of development; systems specialists to address potential impact on systems and hardware; and operational naval officers to address the potential impact on naval operations. 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. The interplay among reviewers coming from different perspectives allowed each reviewer to incorporate elements of other perspectives into his decision-making process.

A multiple regression analysis showed RM to be the most important factor in determining the bottom line score [Kostoff, 1992]. PINN did not weigh as heavily in the reviewers' bottom line score as did PTU. The reviewers weighed nearer-term impact more heavily in their bottom line decisions, as evidenced by the higher correlations of PTU. Since the study also showed that the bulk of the proposed ARIs was viewed by the reviewers as basic research, and since the (possibly far) downstream naval impact of basic research may not be evident in many cases, it is not surprising that the more identifiable near-term impacts, such as transition to exploratory
development or utility of results by other researchers, would affect reviewers' bottom line decisions more than the long term impacts.

4) STW-NETHERLANDS
The Dutch Technology Foundation (STW) was founded in 1981. One of its main functions is to fund university research that is of high scientific quality and has the potential to lead to results that can be used by external bodies. In 1981, STW opted for a new system for the assessment and appraisal of research proposals from individual researchers (Van den Beemt, 1991, 1997). STW devised this new system in order to minimize the problems of selection by large committees, by colleagues, by a few peers only or by organizations belonging to the discipline concerned.

The system operates as follows: All applications belonging to the broad field of technology and engineering sciences are welcome. Every application is sent initially to six peers who are specialists in the topic covered by the proposal; some are university staff, others work in industry. STW asks peers, first by telephone and later by mail, to give comments based on two criteria: scientific quality and utilization potential.

These criteria incorporate the following sub-criteria:

- Scientific quality:
  - competence of a team,
  - originality of the proposal,
  - effectiveness of the proposed method,
  - the program itself,
  - time schedule,
  - available infrastructure and
  - estimated costs.

- Utilization potential:
  - applicability of the results,
  - commercial outcomes,
  - long-term contribution to technology,
  - influence on the competitive status of Dutch industry and
  - the importance of patents in the field.

From the comments received, the program officer at STW compiles a document in which the comments are sorted according to sub-criteria. This document is then sent to the principal investigator who is allowed to reply to each comment; the investigator's actual words are then typed in italics directly under each comment. The complete document, called a protocol, provides information for and against the proposal. When the protocols for 20 proposals (regardless of the topics concerned) are ready, a jury is formed consisting of 12 highly qualified persons coming from universities, government laboratories and industry. Their disciplines and backgrounds vary widely. No jury member knows who else is on the jury; names are not divulged. The work is done free of charge and each member of the jury is only allowed to participate once. The next 20 proposals are
handled by a new jury. The STW board gives a grant to at least the best 8 proposals. This minimum grant percentage of 40 per cent is never influenced by resource allocations. If STW resources were to become insufficient to operate this system, STW would stop accepting proposals for a while.

According to its proponents, this procedure has proved to be reproducible, and in the Netherlands it is widely accepted. Because the system is reproducible and objective, STW gets hardly any resubmissions. A proposal resubmitted to STW will be almost certain to receive the same assessment as the original proposal. A notable feature of the procedure is that it is very dynamic: for instance, there are no fixed groups of influential people within STW. Every year about 50 per cent of the peers are new. Jury members serve only once. The STW board does not set additional priorities once the priority rating has been established by the external assessors.

Opinions on the quality of the proposed research can differ considerably. STW has performed many studies to ascertain whether the STW process really works. They have checked the replicability of the jury judgment. They have also checked that their procedure does not discriminate with regard to age or budget. Their evaluation of the research results 10 years after the proposal was granted shows that there is a correlation between the outcomes and the jury's assessment of the utilization potential. Furthermore, their jury system ensures that original proposals receive grants, which would not be the case if STW had relied solely on bibliometric indicators [Van den Beemt & Van Raan, 1995].

After a proposal has been granted, STW immediately forms a users' committee for that particular research project. The committee meets twice a year at the university where the research is taking place. The research team gives an overview of their work, and discusses this with the 'users'. The “users” are mainly experts, but sometimes they are managers and/or, if appropriate, government representatives. STW regards this as an effective partnership. Most funding-agencies (after granting a project) neglect this aspect of the process, and ask only for annual reports on the granted research project, or they visit the groups once every two years. STW, on the other hand, constantly involves the potential users from society as the research progresses. They evaluate the projects one year and six years after the project has ended.

STW concludes that Peer Review can be relevant when it involves more than 5 peers and they are asked only for their comments. The comments of peers need to be assessed by a number of highly qualified people (non-peers). STW believes that the people involved in the peer and jury procedures must not meet and must work by mail. STW believes that it is not a good idea to work with fixed groups of peers and jury members. STW also believes that bibliometric indicators have nothing to do with scientific quality; they simply indicate numbers of publications and citations. They should not be used for the assessment of research proposals.

5) GRANT PROPOSALS

An excellent assessment of grant proposal peer review has recently been published. The highlights of this study are contained in Appendix VI-E-1.
SELECTED PEER REVIEW PRACTICES: EXISTING PROGRAMS

There are many approaches used by research sponsoring organizations to conduct periodic peer reviews to monitor the quality and potential impact of ongoing research [Salasin, 1980; Logsdon, 1985; DOE, 1993; Kostoff, 1995b; Ormala, 1989; Cozzens, 1987; Kerpelman, 1985; Luukkonen-Grunow, 1987; OTA, 1986]. This section focuses on selected peer review approaches that reflect the state of the art in the technical community and pays special emphasis to how research impact is incorporated into the peer review process. The first case study is the DOE review of its Office of Basic Energy Sciences (BES), and the evolution of that approach into present DOE practice. The second case study focuses on the ONR methods used to review extramural and intramural programs. The third and fourth case studies relate to the annual reviews of the National Institute of Standards and Technology (NIST) and the Army Research Laboratory (ARL) by the National Academy of Sciences (NAS), and the fifth case study addresses the annual review of the DOE national laboratories by the field offices. The final case study describes an approach used by the author to evaluate a program of small high-risk seed money projects.

1. DOE - BES
In 1981, the DOE performed an assessment of existing projects funded by its office of Basic Energy Sciences [DOE, 1982; Kostoff, 1988]. Out of approximately 1200 active projects supported by BES, a randomly selected sample of 129 projects was reviewed by panels of scientific peers. The projects were grouped by areas of similar science, and the reviews were conducted on 40 separate days by 40 separate expert panels, with an average of four members and three projects per panel. The reviewers were, for the most part, bench level scientists independent of the DOE.

The reviewers were asked to rate seven factors for each project:

1. Team Quality (TQ)
2. Scientific Merit (SM)
3. Scientific Approach (SA)
4. Productivity (P)
5. Importance to Mission (IM)
6. Energy Impact (EI)
7. Overall Project Quality (OPQ)

The three evaluation factors on the scoresheet that related to potential research impact were SM, IM, and EI. SM incorporated the potential impact of the research on allied research fields. IM covered the types of ways in which a research project could contribute to the Nation's energy needs. EI was the probable impact of the research project on energy development, conservation, or use.

After the scoring by the panels was completed, all possible linear regression models (ranging from six-factors to one-factor) were used to relate the OPQ rating factor (essentially the reviewers' bottom line score on each project) to the other rating factors for the 129 projects. The six-factor model produced a correlation coefficient of 0.89, which meant that the six-factors selected constituted the bulk of the considerations that the reviewers used to score the OPQ rating factor. In
fact, the best three-factor model derived to predict the OPQ rating factor score, consisting of TQ, SA, and IM, produced correlation coefficients within three percent of the complete six-factor model [DOE, 1982]. An updated version of the BES evaluation approach is used by the DOE Office of Program Analysis to conduct peer review assessments of DOE research and development [DOE, 1993]. Now, after a panel has completed the evaluation of all the projects assigned to it, the members are asked to identify research needs or opportunities available to the DOE research program. Since the panel members are very familiar with the program strengths and weaknesses at this point in the review, the opportunities and needs that they identify should be viewed as highly relevant and credible.

2. ONR
Each of ONR's review processes has a major peer evaluation component adapted to meet the particular needs of the organizational unit under review. The two reviews described here are those of ONR's two largest research claimants circa 1992, the Research Programs Department (RPD) and the Naval Research Laboratory (NRL).

The RPD sponsored extramural basic research mainly at universities, and consisted of 13 Divisions organized along science disciplines. Two separate groups contributed to the one day annual review of each Division. One group was the Division's Board of Visitors (BOV), which represented academia, industry, and non-ONR government. The majority of the BOV were members of the research community, but typically the BOV would include representatives from the technology development community and the operational Navy. The other group contributing to the review was the Research Advisory Board, the senior management of the RPD whose backgrounds spanned a wide range of scientific disciplines.

For the review, the Division Director overviewed the total Division, including programs, accomplishments, new opportunities, and management issues. The Division's program managers described their programs in detail, including the impact on science of their accomplishments, potential or ongoing transitions of their programs to development programs, some bibliometric measures such as publications, and potential impacts on the Navy if successful. The reviewers filled out comment sheets, focusing on Scientific Merit, Technical Approach, and Potential Naval Impact, and later discussed their findings with the RPD management.

Almost all of the NRL's programs are intramural, and it conducts full spectrum research in 60 task areas. On average, about 20 task areas will be reviewed per year, with 4 or 5 of these task areas reviewed using external reviewers, and the remainder reviewed by an internal NRL management group called the Research Advisory Committee (RAC). The external review group represents academia, industry, and non-NRL government. The RAC consists of NRL senior management whose backgrounds span a broad range of science disciplines.

The Coordinator of the task area reviewed by the external panel overviews the task area and investment strategy. Then, the principal investigators of the task area describe their work in detail, including the impact of their science accomplishments on the task area and allied science fields, transitions to more applied categories, bibliometric measures such as publications and
presentations, and potential impact of their research on the Navy. The reviewers fill out comment sheets, focusing on Scientific Merit, Technical Approach, and Potential Naval Impact, and afterward visit and review facilities. The reviewers draft a report and meet with ONR management and members of the RAC to present their preliminary findings. The remaining task areas are reviewed in detail by the RAC.

3. NIST
NIST is reviewed annually by two external groups, a general policy and management review, and a detailed technical review. The Visiting Committee on Advanced Technology reviews general policy, organization, budget, and programs of NIST. The Committee submits an annual report [NIST, 1991a] that includes reviews of progress in NIST's science, engineering and technology transfer programs.

The National Academy of Sciences' (NAS) Board on Assessment of NIST Programs performs a detailed technical review [NIST, 1991b]. Seventeen panels of reviewers (about ten people per panel) from industry and academia conduct program reviews based on 2 or 3-day site visits at NIST facilities. The panels address variants of research quality, and because of NIST's unique charter in supporting competitiveness, pay particular attention to technology transfer, industrial coupling, and emerging technologies. While quantitative indicators of research impact are not addressed in the panels' annual reports [NIST, 1991b], impacts of the research on technology and competitiveness are addressed extensively. Recommendations for improvement in these impact areas are provided.

4. ARL
In the mid-1990s, the ARL contracted with the NAS to establish a Technical Assessment Board (TAB) for the purposes of evaluating the quality of the ongoing research, assessing the state of the laboratory's facilities, and appraising the level of preparedness and functioning of the technical staff. The TAB has 15 members with expertise in fields aligned with ARL's six business areas (Vehicle Technologies, Weapons and Materials Research, Information Science and Technology, Sensors and Electronic Devices, Human Research and Engineering, Survivability and Lethality Analysis), and its members come mainly from Academia and Industry. The NAS established six review panels (one for each business area), each one consisting of about ten members including some TAB members. Each panel reviews one third of the program in its business unit area per year; each full business unit is therefore reviewed on a three-year cycle. Each review consisted of a two-day site visit by the panel. The review included:

• briefings on technical projects,
• touring the lab to assess the facilities and equipment,
• interacting personally with the research staff, and
• reviewing those portions of the ARL extended program being conducted with private sector partners under a Cooperative Agreement (Federated Laboratory; in essence, the addition of virtual lab divisions).

An annual report contains the review results [Brown, 1997].
5. DOE - NATIONAL LABS
The DOE has nine contractor-operated multiprogram laboratories. Each contractor’s laboratory management performance is evaluated annually by the DOE Field Office (FO) to which each laboratory is assigned [DOE, 1988]. The FO prepares an appraisal plan for the laboratory, which focuses on laboratory performance in four areas:

1. Institutional Management Performance, which includes different aspects of overall lab management
2. Programmatic Performance, which includes R&D achievements
3. Operations Support Performance, which includes technical functions that support mission objectives
4. Administrative Performance, which includes business management functions

In the programmatic performance areas, sources of input include DOE program officials, other agencies having substantial work at the laboratory, and FO program managers. For this annual review, DOE will utilize information from its own program advisory committees on the adequacy and impact of the laboratory’s R&D efforts in relation to the overall DOE program. Furthermore, DOE will use the reports of the scientific peer review committees established by the contractor, which provide an assessment of the quality of the laboratory's R&D programs.

There appears to be no formal requirement for using teams of external reviewers for the technical programs as in the ONR and NIST reviews. Instead, most input seems to come from the sponsors. Estimations of research impact appear to derive from the DOE program advisory committees and peer review assessments, which may be reflected in the annual appraisal.

In Europe, panel reviews have evolved where users of the research results together with scientific peers assess the impact of the research on scientific progress and industrial or social development. Another development line has been to commission evaluation experts either to support panels or to conduct independent assessments that may involve surveys, in-depth interviews, case studies, etc [Ormala, 1994]. A 1992 publication [Barker, 1992] describes how evaluation experts coming from two main communities (civil servants and academic policy researchers) interact in evaluation of R&D in the UK. The performance of evaluations, including the synthesis of evidence and the production of conclusions and recommendations, is done by professionals, as opposed to panels of eminent persons. No comparisons of reviews by the professionals with those of eminent persons are presented.

SEED MONEY REVIEW PROTOCOLS

Finally, many organizations have special programs that consist of small, high risk, finite duration projects. These programs have a variety of names, such as seed money or independent research. They may have a variety of purposes, such as attracting high level staff, maintaining staff technical competency, maintaining awareness of the cutting edge external R&D community, and identifying future investment areas for the organization. Because of these projects' small size and high risk nature, high intensity assessments during their lifetimes may be counterproductive. The remainder
of this section describes a protocol for evaluating these projects at the completion of their execution phase. The protocol combines the best of several different agencies' review practices of small projects, and recommends inclusion of some unique features. A process based on this protocol has been used by the author in the review of the Navy In-House Laboratory Independent Research program in the mid-1990s. This review process has produced excellent results, allowing very efficient review of all projects performed by the claimants.

For purposes of this discussion, it is assumed that the central evaluation mode is panel peer review. The underlying review philosophy is that it is neither cost-effective nor necessary for each project to be presented in its entirety before the panel, as would be the case with larger sized projects. If the main purpose of the program is to help the organization position itself for the future in cutting edge science and technology, then the project presentations need contain only that threshold amount of information that will describe the investment strategy that leads to the stated organizational goal. However, Lotka's Law states that only a small percentage of research projects will have substantial payoff, and assessment studies have shown that organizations need to have these few 'heavy-hitters' to maintain vigor and viability. Therefore, a few expanded presentations of the best projects will be required to determine whether the organization has its share of high payoff potential research projects.

For most of the projects presented, two or three vu-graphs of material would be sufficient. These viewgraphs should contain very short statements of the research objectives, the technical approach, the potential payoff to the organization (relevance to the organization's mission), results obtained, research products generated (paper and patent references, etc.), and coordination with other organizations (relation to complementary work in other organizations). Total presentation time for each of these projects should not exceed three or four minutes. The best of the projects would have presentation time expanded to about 15 minutes per project, would have more focus on results and transition possibilities, and would be subject to more detailed scrutiny by the review panel.

In order for this abbreviated presentation approach to be effective, the panel has to receive descriptive material about all the projects beforehand. These write-ups would be about two to five pages in length, and would contain the supporting details of the items summarized on the vu-graphs. Thus, the panel members would enter the review with some understanding about the technical details, and could focus on project linkages and investment strategy during the review.

Consider the following example. Assume a lab has a $3M per year program consisting of 60 seed money projects, and assume one third of the program is reviewed each year. Assume these projects can be aggregated equally into four technical disciplines, such as materials, acoustics, mechanics, and remote sensing. The review would consist of the following. The seed money program manager would spend about 30-45 minutes over-viewing the program. This would include the lab's mission, and how it relates to the corporate sponsor's mission. It would also include the seed money program's objectives, and how they relate to the lab's mission. It would describe selection and management criteria for the projects. Then, after the overview, an expert in each technical discipline would present the projects within that discipline. Four of the five projects within the discipline would require about 15 minutes total, and the fifth (best) project would require about 15
minutes by itself. Thus, each discipline would require about 30 minutes for presentation, and the total review, including overview, would be about three hours. By the end of the review, the panel would understand:

- the program's objectives,
- the strategy for choosing the projects,
- the importance of the projects to science and the organization,
- how the projects would help position the organization for the future, and
- whether some high quality results were obtained.

To close the loop, the reviewers' comments would be sent anonymously to the program manager. The manager would be required to respond in writing to the comments, including descriptions of actions to be taken as a result of the critiques. The manager's comments would be circulated to the reviewers to ascertain their satisfaction, and a final statement would be sent by the reviewers to the assessment manager.

VI. PEER REVIEW PROTOCOLS

The previous sections of this report have focused on concepts, principles, and issues related to research program peer review, as well as examples of selected federal agency peer review practices. The present section incorporates many of these ideas into a sample program peer review process. Sufficient detail is presented such that an organization could use this as a guide to developing a review process most appropriate to its needs. Most of the procedures and concepts described have been tested and found to produce very useful results.

Program Review Options

The guiding principle for review options is that evaluation should occur along the same structures and taxonomies by which the research is planned and executed. If the agency has a separate research unit, then the discipline should be evaluated as an integrated whole. In the nominal intra-agency review, quality and relevance could be evaluated concurrently or separately, as desired by the agency.

If research is vertically integrated with development, then the research could be evaluated as part of a total vertical structure R&D review [Kostoff, 1996a] or as part of the discipline, as desired by the agency. In the nominal intra-agency review, quality and relevance could be evaluated separately or concurrently. A key conclusion to be drawn from this paragraph is that research evaluation recommendations must take into account how research is structured, integrated, and managed within an agency.

Desirable characteristics of a high quality peer review were listed previously under the Objectives section. The generic protocol principles suggested for research program peer reviews are listed in Appendix II. The research programs should be reviewed on a trienniel cycle, based on the DOE BES evaluation results of 1982 [DOE, 1982], and on other agency practices.
The following considerations apply to a concurrent quality and relevance review. The reviewers should be external, have minimal conflicts with the program being reviewed, and should be selected with expertise in all facets of the research and potential impact areas. To evaluate the degree of horizontal coupling in the nominal intra-agency review, representatives of other Federal agencies should be considered as reviewers, or at least should be invited to participate as audience members. Thus, the review panel will be a heterogeneous mixture of research and relevance experts who can address the many facets of the science and areas of potential impact. Approaches for selecting a review panel are presented in Appendix I.

In the nominal concurrent quality and relevance review, quality and relevance should be the main review criteria. Research quality criteria should include research merit, research approach, productivity, and team quality. Relevance criteria should include short term impact (transitions and/or utility), long term potential impact, and some estimate of the probability of success of attaining each type of impact.

There should be an overview showing how the larger management unit (Division, Department, etc.) in which the programs are housed integrates into the total organization, and how the management unit's objectives relate to those of the larger organization. Then, the investment strategy of the larger management unit should be presented in detail. This would include the relative program priorities, the actual investment allocation to the different programs, and the rationale for the investment allocation. Finally, for each program presentation, the investment strategy for its thrust areas should be presented.

The investment strategy is perhaps the most crucial part of a program review, and deserves further discussion here. While investment is the allocation of resources among the program components, the investment strategy is the rationale for the prioritization and allocation of resources among the program components. The optimal investment strategy for a program, which should be a focal point of an assessment, is the allocation and rationale that will produce the most mission relevant high quality research for impacting the program's objectives. This will depend on the viewpoint of the assessor, and in particular how the assessor limits the role of the research within the national perspective.

The optimal investment strategy results from a timely confluence of research requirements (top-down driven) and promising research opportunities (bottom-up driven). Further, promising research opportunities result from a timely confluence of advances in theory, instrumentation, new experiments, new algorithms, and computers. Finally, research requirements result from a timely confluence of domestic and foreign, political and economic, strategic and tactical advances. All of the above factors should be included in a presentation of the investment strategy.

**Background Material**
While the emphasis is on peer review, bibliometric and other kinds of indicators should be used. In the protocol, it is recommended strongly that sufficient background material be supplied to the reviewers before the review. This would include organizational descriptive material, narrative descriptions of each program to be reviewed, and descriptive material of each work unit in the
program. It would also prove useful to include bibliometric output indicators for each program, with interpretive analytical material. This could include refereed papers, patents, awards and honors, presentations, etc. It would be useful to include narrative material on related programs in other agencies and industry. It would be useful to include Hindsight-type results of research that was funded years ago in the discipline under review and that recently came to fruition in a system or commercial technology.

In the following detailed guidance example, it is recommended that program managers include roadmaps with their technical presentations. It would be very valuable if the roadmaps were provided as background material as well. These roadmaps provide the global context in which the program is being performed. Their retrospective components show the program manager’s awareness of the breadth and depth of the intellectual heritage of the present program. The present roadmap components reflect the program manager’s awareness of the wide range of science and technology areas available to complement his program, and the degree of coordination and leveraging in which his program is involved. The prospective roadmap components indicate the program manager's vision and willingness to take risks, and his intrinsic understanding of how results from other science and technology programs could be exploited to enhance and expand the potential of his program. A certain amount of time and reflection is required to understand and fully appreciate the implications of a comprehensive roadmap, and the reviewers should receive these roadmaps well in advance of the actual review date. For the reader interested in obtaining more information about diverse aspects of roadmaps, a comprehensive document has been prepared replete with concepts, principles, and examples [Kostoff, 1997d, 2001].

Finally, although the following concept has never been tested to the author's knowledge, it would be valuable to incorporate the results of journal manuscript reviews in the research program peer review process. Appendix III outlines the benefits of such a proposal, and outlines how it could be accomplished.

Other Issues
A practical consideration concerns the length of the review. It is desirable to have the same group of reviewers present for the total review of the areas in which they have expertise. This allows normalization and continuity to occur. However, in the case of a program review, the larger the program, the more review time it will require. It becomes more difficult to retain high quality reviewers as the length of the review increases.

There are at least three approaches to circumvent this problem. First, the program could be broken into focused subprograms, and each subprogram could be reviewed separately with more focused experts. Second, the program could have its components aggregated, and the full program could be reviewed by the same panel at a lower level of detail. Third, the quality and relevance components could be divided for separate reviews.

The length of the review will be governed by the desired resolution detail of the technical area presentations as well as the breadth of coverage of the program. Two indicators are of value in the discussion of resolution detail. These are Spatial Presentation Intensity (SPI) and Temporal
Presentation Intensity (TPI). The SPI is the ratio of total dollar value of the program being reviewed to the number of reviewers, and the TPI is the ratio of total dollar value of the program being reviewed to total hours allotted to the review.

For the most detailed review, a review at the Principal Investigator (PI) level, the TPI should range from about $125K to $250K per hour (one to two projects per hour), and the SPI should range from about $100K to $250K per reviewer. These reviews could cover technical quality and agency relevance. For the second level detail of review, a program review that would cover both in-depth technical quality and agency relevance, both the SPI and TPI should range between $1M and $1.5M ($/reviewer, $/hour). The third level detail of review, a program review that would be a presentation aggregation of the second level of review and would cover agency relevance only, would have both the SPI and TPI range between $4M and $5M ($/reviewer, $/hour). The TPI estimates are based on review durations of one or more days, while the SPI estimates are based on one-day reviews. If the same reviewers are used for multi-day reviews, the SPI numbers increase sharply. Thus, if an agency wanted to do an in-depth technical quality and agency relevance review at the program level of a $50M program, then about 35-50 hours of presentation time would be required. If a different panel were used each day, then about 35-50 reviewers would be required, whereas if the same panel were used for the total review, then realistically about ten reviewers would be required.

Sample Peer Review Guidance

A) Overall Objectives
   1. Review 1/3 of organization's (Department, Division, Office, etc.) programs in depth each year; overview remainder of organization's programs; total organization program reviewed triennially.
   2. Review vertically integrated programs as a unit.
   3. Focus primarily on technical quality, but address relevance, integration, and investment strategy as well.
   4. Secure comments on the review from a Board of Visitors (BOV). Written comments provided independently to agency staffer, who produces report. The BOV consists of independent experts representing science, technology, customer, and other agencies.
   5. Invite customers, stakeholders, users, impactees, and other agency representatives.
   6. Deliver a summary report with responses to reviewers' comments and action items to agency senior management after review.

B) Sequence of Events
   1) Selection of Reviewers
      A science and technology taxonomy of the program to be reviewed in detail is generated, and brief descriptors of each taxonomy element are generated for reviewer selection purposes. The BOV is selected so that it can address in aggregate detailed science and technology quality, research and technology gaps and opportunities, broader technology and organizational issues, and mission relevance issues. Sources of reviewers could include Defense Sciences Board, NAS, NAE, AFSAB, NSB, AAC (NASA), and program manager recommendations. The names of proposed
reviewers are presented to the agency Director for approval before they are notified. All reviewers are required to sign non-conflict-of-interest statements.

2) Distribution of Background Material
To insure that review time is used most efficiently, reviewers and invited audience receive background material that will set the stage for the actual review. This background material includes the following administrative and technical canonical material:

a. Structural chart of agency, showing how organization fits into agency structure
b. Structural chart of organization, showing programs (including funding) and personnel associated with each program
c. Definitions of different generic types of programs that will be presented during review
d. Other administrative material (agenda, reimbursement, etc.)
e. Two page overview of each program being reviewed in detail (e.g. Weapons Technology), including program objective, program thrusts (e.g., Aerodynamics, Ordnance, G&C, etc.), and investment allocation among thrusts (three year trends)
f. Two page overview of each program thrust, including thrust objective and short descriptions of each technical sub-thrust (e.g., energetic propellants, combustion instability, propellant safety) pursued under the thrust as well as investment allocations among sub-thrusts.
Total program and thrust descriptive material should not exceed twenty pages.

3) Senior Management Introductory Presentation
To initiate the actual review, a senior agency manager provides a short introduction describing structure and mission of the agency, the role of the different corporate review processes in executing the mission, and a more detailed description of the purpose and goals of Department review. This person describes what is expected from BOV, and how BOV comments will be utilized.

4) Organization Head Presentation
The broader technical portion of the presentations is initiated by the Organization Head, and it includes:

a. Mission and objectives of organization
b. List of all programs in organization; describe objectives of each program, show funds and people associated with each program; note program to be reviewed in detail
c. Accomplishments and transitions of programs not being reviewed in detail; relation of accomplishments and transitions to organization's mission and potential national impact
d. Responses to actions from previous year's review

5) Program Manager Presentation
Each program manager then provides a more detailed overview of the program, including:

a. Objectives of program
b. Requirements to be met (for example, in the review of a military-oriented program: what is the present and evolving threat-identify documented sources, personal contact sources, etc.; what is the importance of the threat; what are the capabilities required to overcome threat)
c. Investment strategy
   c1. List of thrusts (e.g., propulsion, aerodynamics, G&C) and sub-thrusts (e.g., energetic propellants, combustion instability, propellant safety) selected to meet requirements
   c2. Objectives of each thrust
   c3. Thrust and sub-thrust funding and prioritization
   c4. Rationale for thrust and sub-thrust selection and prioritization (including bases for rationale and prioritization such as system studies, workshops, assessments, intuition, congressional and other mandates, etc.)
   c5. Integration of thrusts and sub-thrusts to form program
   c6. Coordination/ Roadmaps
      c6i. Roadmaps describe past, present, and future of program and linkage to other internal and external programs
      c6ii. Roadmaps contain at least the three dimensions of time, project title/ sponsor, and project funding
   d. Team quality (identify S&T performers)
   e. Summary of major accomplishments, transitions, milestones met

6) Technical Manager Presentation
The technical managers who support the program manager will present the following:

   a. Objectives of each sub-thrust
   b. Technical roadblocks to achieving the sub-thrust objectives
   c. Technical approach for overcoming the sub-thrust roadblocks
   d. Potential sub-thrust payoffs and capability enhancements
   e. Technical results achieved

7) Reviewers' Written Comments
The reviewers fill out an evaluation form, and provide it to the agency review manager at the end of the review. A sample short evaluation form follows.
PRESENTATION EVALUATION SHORT FORM

COMMENTS (PLEASE PROVIDE YOUR COMMENTS IN NARRATIVE FORM. WHERE APPLICABLE, INCLUDE YOUR ASSESSMENT OF RELEVANCE, GAPS AND OPPORTUNITIES, INVESTMENT STRATEGY, COORDINATION, TECHNICAL APPROACH, TEAM QUALITY, POTENTIAL PAYOFF, PRODUCTIVITY AND IMPACT. THESE EVALUATION CRITERIA HAVE BEEN DEFINED ON THE FIRST PAGE OF YOUR EVALUATION PACKAGE.)

Reviewers are invited to submit further written comments after they return home.

Other sample evaluation forms follow.

EVALUATION FORMS FOR EXISTING PROGRAMS - LONG FORM

PROGRAM EVALUATION FORM

TITLE OF PROGRAM.................................
REVIEWER NAME.................................

1A. RESEARCH MERIT (CIRCLE ONE NUMBER OR -)
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*LOW**    ***FAIR***    ****AVER****    **GOOD**    **HIGH**

1B. RESEARCH APPROACH/ PLAN/ FOCUS/ COORDINATION
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*LOW**    ***FAIR***    ****AVER****    **GOOD**    **HIGH**

1C. MATCH BETWEEN RESOURCES AND OBJECTIVES
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*LOW**    ***FAIR***    ****AVER****    **GOOD**    **HIGH**

1D. QUALITY OF RESEARCH PERFORMERS
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*LOW**    ***FAIR***    ****AVER****    **GOOD**    **HIGH**

1E. PROBABILITY OF ACHIEVING RESEARCH OBJECTIVES
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*LOW**    ***FAIR***    ****AVER****    **GOOD**    **HIGH**

1F. PROGRAM PRODUCTIVITY
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*LOW**    ***FAIR***    ****AVER****    **GOOD**    **HIGH**
### 2A. Potential Impact on Mission Needs (Research/Technology/Opers)

<table>
<thead>
<tr>
<th>Scale</th>
<th>LOW**</th>
<th>FAIR***</th>
<th>AVER****</th>
<th>GOOD**</th>
<th>HIGH**</th>
</tr>
</thead>
</table>

### 2B. Probability of Achieving Potential Impact on Mission Needs

<table>
<thead>
<tr>
<th>Scale</th>
<th>LOW**</th>
<th>FAIR***</th>
<th>AVER****</th>
<th>GOOD**</th>
<th>HIGH**</th>
</tr>
</thead>
</table>

### 2C. Potential for Transition or Utility

<table>
<thead>
<tr>
<th>Scale</th>
<th>LOW**</th>
<th>FAIR***</th>
<th>AVER****</th>
<th>GOOD**</th>
<th>HIGH**</th>
</tr>
</thead>
</table>

### 2D. Phase of R&D (DOD Terminology)

- **BASIC RES**
- **APPLIED RES**
- **EXPLORATORY DEV.**
- **ADV DEV**

### 3. Reviewer's Expertise in the Research Area of This Program

<table>
<thead>
<tr>
<th>Scale</th>
<th>LOW**</th>
<th>FAIR***</th>
<th>AVER****</th>
<th>GOOD**</th>
<th>HIGH**</th>
</tr>
</thead>
</table>

### 4. Overall Program Evaluation

<table>
<thead>
<tr>
<th>Scale</th>
<th>LOW**</th>
<th>FAIR***</th>
<th>AVER****</th>
<th>GOOD**</th>
<th>HIGH**</th>
</tr>
</thead>
</table>

---

*Page 60*
SCORING CRITERIA

The evaluation form contains factors generally related to research and naval relevance issues. The scoring bands for all criteria except 2D are identical, and are: 1-2 (LOW); 2.5-4 (FAIR); 4.5-6.5 (AVERAGE); 7-8.5 (GOOD); 9-10 (HIGH). Criterion 2D has its own scoring range defined.

DEFINITIONS OF CRITERIA ON PROGRAM EVALUATION FORM

1A. RESEARCH MERIT - Importance to the advancement of science of the question or problem addressed by the program. Consider the technical objectives, potential advancement of state-of-art, and uniqueness of contribution.

1B. RESEARCH APPROACH/ PLAN/ FOCUS/ COORDINATION - Quality of process employed to solve the research problem, including the quality and focus of the research plan, definition of research milestones, degree of innovation, understanding of field, balance between experiment and theory, and coordination with (or cognizance of) other related programs to minimize duplication or gaps.

1C. MATCH BETWEEN RESOURCES AND OBJECTIVES - Relationship between scientific objectives proposed and total resources requested. Also, adequacy of resources at performer level to ensure 'critical mass' for each performing unit.

1D. QUALITY OF RESEARCH PERFORMERS - Consider publications, honors, and awards, relevant experience, and other less tangible factors that contribute to team quality.

1E. PROBABILITY OF ACHIEVING RESEARCH OBJECTIVES - Probability that the program's research objectives will be achieved.

1F. PROGRAM PRODUCTIVITY - Volume and quality of work produced and relationship of this output to the resources available, costs incurred, and time elapsed since program initiation.

2A. POTENTIAL IMPACT ON MISSION NEEDS - Potential impact of this program on mission research/ technology/ operational needs if successful.

2B. PROBABILITY OF ACHIEVING POTENTIAL IMPACT ON MISSION NEEDS - Probability that the program will achieve its potential mission impact assuming that its research objectives have been met.

2C. POTENTIAL FOR TRANSITION OR UTILITY - Probability that results from this program will be transitioned to or utilized by technical community assuming that its research objectives have been met.

2D. PHASE OF R&D - Level of program development. Scale ranges from basic research (6.1) through exploratory development (6.2) to advanced development (6.3).

4. OVERALL PROGRAM EVALUATION - Single number description of overall program quality based on all relevant criteria. Provide detailed narrative of pros and cons and any recommendations under COMMENTS.
EVALUATION FORMS FOR PROPOSED PROGRAMS - LONG FORM

PROPOSED PROGRAM EVALUATION FORM

TITLE OF PROPOSED PROGRAM........................................
REVIEWER NAME......................................................

1A. RESEARCH MERIT (CIRCLE ONE NUMBER OR -)
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*LOW** ***FAIR*** ****AVER**** **GOOD** **HIGH**

1B. RESEARCH APPROACH/ PLAN/ FOCUS/ COORDINATION
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*LOW** ***FAIR*** ****AVER**** **GOOD** **HIGH**

1C. MATCH BETWEEN RESOURCES AND OBJECTIVES
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*LOW** ***FAIR*** ****AVER**** **GOOD** **HIGH**

1D. BALANCE BETWEEN EXPERIMENT AND THEORY
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*LOW** ***FAIR*** ****AVER**** **GOOD** **HIGH**

1E. PROBABILITY OF ACHIEVING RESEARCH OBJECTIVES
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*LOW** ***FAIR*** ****AVER**** **GOOD** **HIGH**

2A. MISSION NEED (PROBLEM OR NEED THAT THIS RESEARCH ADDRESSES)

2B. POTENTIAL IMPACT ON MISSION NEEDS (RESEARCH/TECHNOLOGY/OPERS)
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*LOW** ***FAIR*** ****AVER**** **GOOD** **HIGH**

2C. PROBABILITY OF ACHIEVING POTENTIAL IMPACT ON MISSION NEEDS
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*LOW** ***FAIR*** ****AVER**** **GOOD** **HIGH**

2D. POTENTIAL FOR TRANSITION OR UTILITY
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*LOW** ***FAIR*** ****AVER**** **GOOD** **HIGH**

2E. PHASE OF R&D (DOD TERMINOLOGY)
6.1-----------------------6.2------------------------6.3
3. **REVIEWER'S EXPERTISE IN THE RESEARCH AREA OF THIS PROGRAM**

<table>
<thead>
<tr>
<th></th>
<th>LOW**</th>
<th><em><strong>FAIR</strong></em></th>
<th>*<em><strong>AVER</strong></em></th>
<th><strong>GOOD</strong></th>
<th><strong>HIGH</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

4. **OVERALL PROGRAM EVALUATION**

<table>
<thead>
<tr>
<th></th>
<th>LOW**</th>
<th><em><strong>FAIR</strong></em></th>
<th>*<em><strong>AVER</strong></em></th>
<th><strong>GOOD</strong></th>
<th><strong>HIGH</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

-----------------------------------
SCORING CRITERIA

The evaluation form contains factors generally related to research and mission relevance issues. The scoring bands for all criteria except 2A and 2D are identical, and are: 1-2 (LOW); 2.5-4 (FAIR); 4.5-6.5 (AVERAGE); 7-8.5 (GOOD); 9-10 (HIGH). Criterion 2A has no scoring range, and criterion 2E has its own scoring range defined.

DEFINITIONS OF CRITERIA ON PROPOSED PROGRAM EVALUATION FORM

1A. RESEARCH MERIT - Importance to the advancement of science of the question or problem addressed by the program. Consider the technical objectives, potential advancement of state-of-art, and uniqueness of contribution.

1B. RESEARCH APPROACH/ PLAN/ FOCUS/ COORDINATION - Quality of process employed to solve the research problem, including the quality and focus of the research plan, definition of research milestones, degree of innovation, understanding of field, and coordination with (or cognizance of) other related programs to minimize duplication or gaps.

1C. MATCH BETWEEN RESOURCES AND OBJECTIVES - Relationship between scientific objectives proposed and total resources requested.

1D. BALANCE BETWEEN EXPERIMENT AND THEORY - Balance between experiment and theory proposed relative to optimum required to achieve performance targets.

1E. PROBABILITY OF ACHIEVING RESEARCH OBJECTIVES - Probability that the program's research objectives will be achieved.

2A. MISSION NEED - Identify the mission need or problem (operational, technological, research) to which this research relates.

2B. POTENTIAL IMPACT ON MISSION NEEDS - Potential impact of this program on mission research/ technology/ operational needs if successful.

2C. PROBABILITY OF ACHIEVING POTENTIAL IMPACT ON MISSION NEEDS - Probability that the program will achieve its potential mission impact assuming that its research objectives have been met.

2D. POTENTIAL FOR TRANSITION OR UTILITY - Probability that results from this program will be transitioned to or utilized by technical community assuming that its research objectives have been met.

2E. PHASE OF R&D - Level of program development. Scale ranges from basic research (6.1) through exploratory development (6.2) to advanced development (6.3).

4. OVERALL PROGRAM EVALUATION - Single number description of overall program quality based on all relevant criteria. Provide detailed narrative of pros and cons and any recommendations under COMMENTS.
VI-A. APPENDIX I - REVIEW PANEL SELECTION APPROACHES

A review panel should have at least the following characteristics:

1. Each member should be highly competent in the facet of the program for which he has been selected
2. The panel as a body should have sufficient competence to cover all major facets of the program being reviewed
3. Each member should be minimally conflicted with the program under review, and any conflicts or biases should be known to all the panel members before the review
4. Each member should agree to read all background material, attend all sessions, and protect any classified and proprietary information that arises during the review

Selection of an optimal review panel is more of an art than a science at present. It depends on:

- The selector's understanding of the program being reviewed,
- Her understanding of the experts available in the technical community, and
- Her ability to predict the interaction dynamics of a particular group of experts.

Presently, different Federal agencies’ approaches to panel selection range from assembling program manager recommendations to using an iterative co-nomination approach. Since the latter approach, properly done, is relatively objective with respect to the program being reviewed, the remainder of this attachment will focus on its description.

In essence, the iterative co-nomination approach is a multi-step process that starts with an input list of recommended experts and converges to a list of experts who have been multiply nominated by different experts. The first step is to define the specific technical areas to be reviewed, and the objectives and expected outputs of the review. Once the overall technical description of the program is generated, and technical descriptions of the sub-disciplines are provided, reviewer identification can be initiated.

Sources of candidate reviewers can include program manager recommendations, membership lists of prestigious organizations such as the National Academies, agency review boards, agency consultant pools, and other similar lists. (One of the real deficiencies in present day pools of reviewer candidates is the absence of a centralized updated pool of experts that spans the Federal agencies. With present computer capabilities, a centralized list that includes name, organization, biography, areas of expertise, previous panels and panel references for thousands of experts, and is easily accessible to assessment managers, would be simple to construct. It could be updated continuously with input from program managers as they become acquainted with new experts. Such a pool should be instituted immediately after multi-agency agreement.) Multiple names are chosen to cover each sub-discipline, the program as a whole, allied research disciplines, the technologies, systems, and operations that the program could potentially impact, and other elements of the customer, stakeholder, user, and impactee communities. This list of names is called level 1, or the initial list.
Each member of level 1 is asked to identify, or nominate, other experts in his particular area of expertise for the level 2 list. For example, assume that a Physics program is being assessed. Assume further that this program has three sub-disciplines: plasma physics, atomic physics, and molecular physics. The level 1 list may have two names for each of the sub-disciplines. To obtain the level 2 list for the plasma physics research area of expertise, each of the two plasma physics recommendees of level 1 would be asked to recommend two experts in plasma physics. If names appear more than once in the level 2 list, or between the level 1 and level 2 lists (multiply recommended individuals), then these people are assumed to be the leading experts in the fields to be assessed. If no multiple recommendations appear, then the experts in level 2 are asked to recommend two experts in plasma physics for level 3, and the co-nomination search is repeated. Convergence occurs when an adequate number of experts have been co-nominated. While this process may at first seem complex and open-ended, convergence is rapid because of the relatively small number of real experts in any well-defined technical discipline.

A primary and alternate list of co-nominees should be matrixed against selection requirements and criteria as shown below, where the matrix elements represent the reviewer's expertise in the different facets being examined. This matrix should be distributed to the program managers and performers who will be reviewed, and comments related to bias and conflict solicited. If strong objections can be supported, the list could be modified.

**REVIEWER/ CRITERIA MATRIX**

<table>
<thead>
<tr>
<th>SUB- REV</th>
<th>DIS1</th>
<th>DIS2</th>
<th>DIS3</th>
<th>PROG</th>
<th>DEP</th>
<th>EXPT</th>
<th>EXP</th>
<th>ALT</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAME.1. (OR1)</td>
<td>10..7</td>
<td>6...8</td>
<td>5...3</td>
<td>PRI.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAME.2. (OR2)</td>
<td>9...5</td>
<td>9...4</td>
<td>2...ALT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAME.3. (OR3)</td>
<td>6...7</td>
<td>10...7</td>
<td>5...PRI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAME.4. (OR4)</td>
<td>5...4</td>
<td>4...10</td>
<td>8...PRI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAME.5. (OR5)</td>
<td>2...3</td>
<td>3...8</td>
<td>10...PRI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAME.6. (OR6)</td>
<td>7...8</td>
<td>7...8</td>
<td>6...5...PRI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
VI-B. APPENDIX II - PROGRAM PEER REVIEW PROTOCOL

The best features of different organizations' peer review practices can be combined into a protocol for the conduct of successful peer review research program evaluations and impact assessments. The main aims of the protocol are to insure that the final assessment product has the highest intrinsic quality and that the assessment process and product are perceived as having the highest possible credibility. The protocol elements are:

1. The objectives of the assessment must be stated clearly and unambiguously at the initiation of the assessment by the highest levels of management, and the full support of top management must be given to the assessment. In turn, the objectives, importance, and urgency of the assessment must be articulated and communicated down the management hierarchy to the managers and performers whose research is to be assessed, and the cooperation of these reviewees must be enlisted at the earliest stages of the assessment;

2. The final assessment product, the audience for the product, and the use to be made of the product by the audience should be considered carefully in the design of the assessment;

3. One person should be assigned to manage the assessment at the earliest stage, and this person should be given full authority and responsibility for the assessment;

4. The assessment manager should report to the highest organizational level possible in order to insure maximum independence from the research units being assessed;

5. The reviewers should be selected to represent a wide variety of viewpoints, in order to address the many different facets of research and its impact [Kostoff, 1988]. These would include bench-level researchers to address the impact of the proposed research on the field itself; broad research managers to address potential impact on allied research fields; technologists to address potential impact on technology and the potential of the research to transition to higher levels of development; systems specialists to address potential impact on systems and hardware; and operational personnel to address the potential impact on downstream organizational operations. The reviewers should be independent of the research units being evaluated, and independent of the assessing organization where possible. The objectives of, and constraints on (if any), the assessment should be communicated to the reviewers at the initial contact;

6. Maximum background material describing the research to be assessed, related research and technology development sponsored by external organizations, the organization structure, and other factors pertinent to the assessment, should be provided to the reviewers as early as possible before the review. This will allow the reviewers and presenters to use their time most productively during the review;

7. Recommendations resulting from the assessment should be tracked to insure that they are considered and implemented, where appropriate. For research programs, planning, execution, and review are linked intimately. Feedback from the review outcomes to planning for the next cycle should be tracked to insure that the review/planning coupling is operable.

The following criteria and issues should be considered during the review as appropriate.

1. Quality and uniqueness of the work
2. Scientific and technological opportunities in areas of likely organization mission importance
3. Need to establish a balance between revolutionary and evolutionary work
4. Position of the work relative to the forefront of other efforts
5. Responsiveness to present and future organization mission requirements
6. Possibilities of follow-on programs in higher R&D categories
7. Appropriateness of the efforts for organization as opposed to other organizations
8. Coordination with related work in other organizations

In particular, when evaluating the investment strategy, adherence to the following investment principles should be assessed; i.e., actual program allocations in the following areas should be assessed against the desired target allocations:

1) Is the balance among technical thrust areas appropriate?
2) Is the balance among mission areas appropriate?
3) Is the balance among funding categories (6.1/ 6.2/ 6.3) appropriate?
4) Is the balance between discretionary and non-discretionary funding appropriate?
5) Is the balance between 'technology push' and 'requirements pull' appropriate?
6) Is the balance between revolutionary and evolutionary research appropriate?
7) Is the balance between technology advancement and demonstration appropriate?
8) Is the balance between high risk and low risk research appropriate?
9) Is the balance among short term, intermediate term, and long term research appropriate?
10) Is the balance between new projects and continuing projects appropriate?
11) Is the balance among performers (university/ government/ industry) appropriate?
12) Is the balance between individual research and joint projects (multi-department, multi-agency, multi-national, government-industry) appropriate?
13) Is the balance among single discipline, multiple discipline, and interdisciplinary research appropriate?
14) Is the balance between large and small projects appropriate?
15) Is the balance among research products (hardware, software, patents, presentations, reports, peer-reviewed journal papers) appropriate?
VI-C. APPENDIX III - USE OF PUBLISHED PAPERS IN RESEARCH PROGRAM EVALUATION

Research project or program peer reviews in many agencies appear designed more for the comfort of the participants rather than the efficient exchange of information. Especially in panel reviews, the presentation tends to focus on intricate technical details rather than investment strategy. The technical details address mainly the job right component of peer review, whereas the investment strategy has the focus of the right job component. Much of the detailed technical information could be supplied to the reviewers beforehand, and the valuable but usually quite limited presentation period could be devoted more to understanding the investment strategy rationale. However, the reviewers and presenters (and usually the audience) tend to be trained technically, are more comfortable in discussing technical details, and, because of their background expertise in the areas being reviewed, are usually willing to accept the right job aspects of the technical area as fundamentally important.

It is the author's firm contention that as much useful background information as possible should be supplied to the reviewers of a research program or project before the actual review occurs. In addition to the narratives suggested previously, there is another source of valuable information that has been almost completely neglected during any of the many different agency project and program reviews the author has attended. This information is the written peer reviews of the project's papers that were submitted, accepted, and/or published by refereed journals. The following discussion proposes that fuller use be made of these journal peer reviews in the research program peer review process.

A published paper is really not research, it is a documentation of research. However, while this observation mainly impacts the importance ascribed to bibliometric counts in assessing research productivity and quality, it says little about the intrinsic value of a published paper for use in research evaluation. Because of the effort generated by authors/editors/reviewers in the paper publication process, there is much information in the paper and the publication process that could be valuable in research program evaluation.

Under the present system of manuscript publishing, papers are submitted by a researcher(s) to a journal. The papers are then sent by the journal editor, or proxy, to one or more experts in the field for review (typically two or three experts). For a technical article, the author(s) tends to supply many details of the technical approach, as well as other useful information. During the manuscript review, typically the reviewers spend substantial time addressing the intricate details of the technical approach used in the research (as well as addressing other criteria). The paper may be accepted or rejected outright, or accepted pending approved revision. The reviewers' comments, and the submitter's rebuttal (if any) stay within the editor-submitter-reviewer group. Thus, if a researcher has one published paper during a year, and this is presented to a panel of experts as part of a project/program review, all the panel knows is that the paper passed the threshold requirements for a particular journal. The panel does not know how many journals rejected the article, what the comments of the rejecting peer reviewers were, what the rebuttal comments of the submitter were, or what the specific comments of the accepting journal peer reviewers were. This
information would be very useful to have during a project/program review, since it could reduce the need for the presentation of copious technical detail during the review, and allow more time for discussion of higher order issues such as investment strategy and relevance to organizational objectives.

Since the sponsoring agency pays for the research, it has every right to have full access to reviewers' comments on the products of the research. Otherwise, the agency is being excluded from external reviews of research that it has supported. The journal reviewers have typically expended much effort in the technical review process, and the valuable information contained in their comments is not being used for the fullest benefit to the rightful recipients of this information, the research sponsors. The following proposal addresses this deficiency.

For a paper that results from sponsored research, an agreement is required between the research sponsoring agencies or corporations and the research journals that the sponsor of the paper's research be identified when it is submitted for publication. Once the paper has been reviewed, a copy of the journal reviewers' comments would be sent to the sponsoring organization as well as to the article submitter. In return for the journal's efforts, the sponsoring organization would provide some financial compensation to the journal for the review and comments. Under this system, writers of low-to-average quality articles would be less motivated to submit randomly to different journals, since the peer reviews would be transmitted to their sponsoring organizations. This would have the positive effect of reducing the overwhelming volume of mediocre articles submitted to and published in the literature. Also, these journal reviews would be submitted to the sponsor's project evaluation panels as background material, and, as stated above, would reduce the need for detailed exposition of technical approach that presently consumes much of the presentation time of project reviews.

This approach would probably result in a positive Darwinian selection process. The good researchers who recognize that they are doing good research would be motivated to publish more, while the mediocre to average researchers who recognize that they are doing mid-level research would be motivated to publish less. The differences in numbers and quality of published papers between the good researchers and average researchers would be accentuated and would become more evident to the review panel, and the papers would then have more of an impact on the panel's evaluation of a project. The journals would be partially compensated for their efforts, and the journal reviewers could conceivably be partially compensated for their efforts. This could make journal reviewing a more attractive process to reviewers, and might improve some of the review quality issues described in the Quality section of this report.
VI-D APPENDIX IV  NETWORK-CENTRIC PEER REVIEW

INTRODUCTION

The objective of the proposed network-centric peer review is to evaluate a large ongoing S&T program, using a representative segment of the technical community, and employing whatever information technology is required to substantially enhance the quality of the review. Network-centric peer review uses the power of modern communication networks and information technology to expand greatly the number of people that can participate in real-time peer reviews, and expands greatly the access to data that can support all aspects of peer review. This technology allows diverse review operational modes such as the Science Court to be considered seriously, and allows the jury function of peer review to be independent from the higher conflict potential expert reviewer/ witness function. The operational architecture required for network-centric peer review may differ little from the architecture required for its parent network-centric strategic management. Since all strategic management components need to be integrated for optimal synergistic benefits, implementation of network-centric peer review should occur in parallel with implementation of the other components of network-centric strategic management.

This appendix addresses:

- Information technology advances and their potential impact on peer review
- An implementation procedure for a network-centric peer review process
- Research opportunities for network-centric peer review

INFORMATION TECHNOLOGY ADVANCES

In recent years, advances in computer hardware have resulted in much higher computational speed systems with massive amounts of rapidly-accessible storage space. In parallel with the hardware advances are software improvements that allow organization and ‘mining’ of the transmitted data, and architecture implementations that allow large networks of disparate data sources (whether sensors, humans, structured databases, or other types) to be linked. With such network architectures readily available, one person can communicate with many individuals at once, and the input from many individuals and data sources can be collected, integrated, and analyzed in real time. The implications for peer review in particular, and for strategic management in general, are enormous. One of the major (justified) criticisms of peer review (and of road-maps, metrics, data mining, information retrieval, S&T planning, S&T evaluation, S&T transitioning, and other strategic management decision support aids) has been that only a small fraction of the relevant communities and available data are being accessed when these decision aids are being exercised. Logistics costs and time delays have limited the magnitude of information and people available to contribute to these decision aids’ outputs, especially when time frames approximating real-time are required. Now, the hardware and software in combination with the network architectures, and especially supported by individuals who understand the relation between the information technology capabilities and the decision aid requirements, allow these logistics-based limitations to be removed.
POTENTIAL IMPACT OF INFORMATION TECHNOLOGY ADVANCES ON PEER REVIEW

First, the potential impact of information technology advances on the different temporal segments of peer review will be estimated. Then, the potential impact of information technology advances on the different quality principles will be discussed. In the following section, these concepts and estimates will be crystallized and integrated into a proposed network-centric review process.

Impact on Temporal Segments
This discussion will be based on the assumption that one component of a research program peer review will be a meeting that some, not necessarily all, of the participants will attend. Conduct of a meeting-based research program peer review can be categorized into three stages: a pre-meeting phase, the actual meeting, and a post-meeting phase.

Pre-Meeting Phase
The main goal of the pre-meeting phase is to inform and prepare all the participants sufficiently that little time is wasted during the actual meeting phase. Standard peer reviews today allow the various review participants to receive summary background material, to be read by the time of the meeting. An interdisciplinary workshop conducted by the author in December 1997 (Kostoff, 1999a) went one step further. Participants exchanged ideas by e-mail, and all participants were involved in each e-mail. By the time of the meeting, many of the issues had been greatly clarified. However, what could be envisioned in this pre-meeting phase if network-centric peer review were operable, utilizing much of the power of available information technology?

First, a substantially larger amount of data could be made accessible to each review participant, since the network could be structured to allow each node (participant) ready access to every other node (data source or participant). Second, a substantially larger number of participants could be involved in the review, limited only by the extent of the network architecture. Third, a real time iterative rating, learning, and subsequent presentation modification process could be established. New concepts could be discussed and improved. Presentations could be critiqued and given a preliminary rating, and then greatly modified for the meeting. Some types of reviews could be conducted entirely without physical presence, whereas those that required an actual meeting would have most of the time-delaying issues examined beforehand. In summary, this phase could accommodate substantially more data and participants than at present, could integrate and analyze this data in real-time, and could provide feedback in a continuous short-turnaround mode. It could also provide a period of reflection and gestation, as concepts became more integrated with the passage of time. How could this network-centric pre-meeting phase be envisioned to affect the next actual meeting phase?

Meeting Phase
First, the actual review panel could consist of hundreds or more of experts, some of whom are on-site and the remainder are off-site. All would be linked through the network architecture, and
the off-site participants may be video teleconferenced to the presentation material as well. These features allow the review process to be decentralized, either partially or fully, and provide much greater flexibility in time and location scheduling. They also allow a greater diversity of reviewers to be used, in technical areas ranging from closely aligned with the focused presentation themes to very disparate disciplines that could contribute innovative insights to the target themes and offer the possibility of real breakthroughs.

All data input would be mechanized, and instantly recorded. Statistical analyses could be performed on the data, at the level of each presentation and integrated over all presentations. This integrative analysis would show how each project’s ratings would influence overall rankings and overall parametric criteria, thus placing local decisions in their global context. All the background data, the reviewers’ ratings and comments, and other supportive data, would be available instantly to all participants. This latter feature would allow real-time Delphi processes, or modifications of comments and ratings, to be conducted at the end of the presentation period, or in dedicated Executive Sessions. The availability of large amounts of data of all types and large numbers of experts in diverse areas might allow the addition of extra evaluation criteria to be employed usefully, and offer additional perspectives on the S&T being reviewed. What impact could a network-centric meeting process have on the final post-meeting phase?

**Post-Meeting Phase**
The post-meeting phase would have some analogies to the pre-meeting phase, with more focus on integration of new concepts and identification of solutions/ modifications to problem areas identified, stimulated by the intense interactions from the highly efficient meeting phase. Final rankings, comments, and decisions would be obtained iteratively with the availability of the integrated comments and statistics, and a comprehensive integrated report could be assembled from the diverse reviewers effortlessly.

**Impact on Principles of High Quality**

**Need for Synergy and Integration**
In the preface to the high quality principles section, the main theme expounded was that peer review, and the complementary decision aids as well, needed to be an integral component of the overall strategic management process. If peer review, or any of these decision aids, are treated as add-ons or independent entities, the power of these techniques and value to the sponsoring organization are diminished substantially. These techniques are interlocking, their operation is symbiotic, and their benefits are synergistic. For network-centric peer review to achieve its full potential, it must be integrated fully into the network-centric strategic management process. Thus, the requirements for successful operation of network-centric peer review are more severe than for traditional peer review, because the operational targets and potential roadblocks are at a higher level.

For example, if data mining is not performed using all the global data sources available as well as the human and computer analytic and interpretive capabilities, then a gap will exist in the data available for comparing programs under review with the state-of-the-art. This in turn will affect
the use of metrics to gauge the comparisons, and road-maps to show project and technology linkages. The impact of data-deficient peer review on strategic planning will result in greater uncertainty in the planning process and products, and will be translated into greater uncertainty in the project selection, management, and transition processes and products.

Thus, a full-scale network-centric strategic management process must eventually be developed, of which the peer review component is one element. However, once the architecture has been established for a network that links the S&T performer, management, oversight, acquisition, operational, or vendor communities, then

- peer review can be accomplished readily in the network-centric mode,
- road-maps can be easily generated in the network-centric mode,
- planning can be performed efficiently in a network-centric mode,
- multi-discipline multi-category multi-performer multi-user programs can be coordinated and managed effectively in the network-centric mode,
- Integrated Product Teams can conduct planning and operations in a highly decentralized network-centric mode, and
- even marketing and sales can be conducted in a network-centric mode using all the resources of organizations, nations, and international communities.

The key point here is that it is the architectural structure, and the inherent logic that links the nodes of the network, that are central to the effective operation of all these seemingly diverse components of strategic management. Once the architecture has been constructed, and the data control established, successful operation of the strategic management tactical elements ceases to be a critical path item.

Impact on Specific Principles
The first three principles of high quality peer review listed in the Executive Summary focus on management commitment, incentives, motivation, and statement of objectives. These provide a context, or set the stage, for conducting a high quality peer review, but would not be impacted by the specific tools employed during the review.

The fourth principle, Evaluator Competency, could be impacted substantially by network-centric operation. Three of the critiques related to evaluator competency in peer reviews are:

- that not all technical areas are covered adequately by relatively small panels used in peer reviews,
- even in those covered areas, the sample of the community is too small to be representative, and
- there are many facets of related technical and non-technical areas that the panel does not cover as a body because of the narrow technical focus.

Network-centric operation would allow many representatives from any technical specialty of interest, representatives from all technical areas involved, and representatives from areas that go
beyond the purely technical (users of the technology, impactees, environmental, regulatory, etc.). Because time commitments of reviewers would be reduced due to less need for travel, and because high quality reviewers tend to be busy time-restricted people, more high quality reviewers would be available to participate in the review process, further raising the quality level of the review.

There is another potential benefit related to the Evaluator Competency criterion that deals with the evaluators’ operational mode. In the vast majority of traditional S&T peer reviews, the panel has a dual role or function. It serves as (hopefully) an impartial jury, and serves as an expert witness/reviewer body as well. This is intrinsically different from the legal system, where the jury is separate from the witnesses and experts, with separate responsibilities and separate individual requirements. Combining the jury with witnesses or experts has the potential to raise serious conflicts. The combination problem arises mainly due to the finite panel size, and the logistical inability to handle large numbers of witnesses and experts in parallel with panel operation.

There have been attempts to conduct peer reviews in which the jury function is executed by one group, and the expert or witness function by an entirely distinct group (DOE, 1978; Van den Beemt, 1997). The Science Court procedure used by the author to evaluate competing alternate magnetic fusion concepts is one example (DOE, 1978; Kostoff, 1997d). The author’s experience with the Science Court was that it was a very valuable process, but very time consuming and unwieldy. Network-centric operation would convert the Science Court into a much more manageable and powerful process.

Thus, network-centric operation offers potential benefits in either panel mode of operation. In the case where the panel operates as both the jury and expert/witness body, network-centric operation expands the number of participants to insure expertise coverage of all criteria. In the case where the jury and witness/expert body are separate, network-centric operation still insures expert coverage of all criteria, but allows the panel to function as a relatively independent conflict-free jury.

The next principle that could be affected by network-centric operation is Evaluation Criteria. With the expanded access to data allowed by network-centric operation, criteria could be added for which data could be obtained straight-forwardly. For example, suppose knowledge of specific types of impact was an important criterion, but the data by which impact would be evaluated were not readily available. Under traditional peer review, that criterion might not be used, but under network-centric operation, that criterion could be employed due to ready data availability on impact.

The criterion of Reliability would be impacted substantially by network-centric operation. With a large sample from the relevant communities, degree of representativeness is no longer an issue, and the repeatability of the results over different panels becomes a moot point. In addition, much more data becomes available for incorporation into the evaluation, and statistical representativeness effectively disappears as a data issue.
The Data Awareness criterion would obviously be affected to a large extent. Network-centric operation allows massive amounts of global data to be accessed, filtered, mined, interpreted, and evaluated. Bibliometric analysis capabilities will allow the performers, institutions, and countries that are sponsoring/performing S&T to be identified, thereby enhancing the potential for leveraging and exploitation, and minimizing the opportunities for excessive redundancy. Along with limited numbers of reviewers, limited access to data is a major deficiency of present day peer reviews that would be overcome by network-centric operation.

The Secrecy criterion could be impacted to some degree. Network-centric operation could allow people at remote sites to participate as reviewers/expert witnesses without their identity being revealed to other participants in the process. This enhanced anonymity would allow for greater open-ness and frank-ness, ultimately yielding a more useful product.

The Cost criterion would be impacted, due to the reduced travel requirement, and the reduced facilities requirement. Since time commitments would be reduced as well, high caliber typically busy people would be more likely to serve, and a higher quality product would also result concomitant with the lower cost.

**IMPLEMENTATION OF A NETWORK-CENTRIC REVIEW PROCESS**

**Background**

The author has conducted meetings and reviews that have made some use of network capabilities. These include the review of the Department of the Navy’s total Advanced Technology Development program (Kostoff, 2001), and an innovation workshop on Autonomous Flying Systems (Kostoff, 1999). The lessons learned from conducting these meetings/reviews will be integrated with the principles of high quality peer review in the Executive Summary and the network concepts of this appendix to outline an operational implementation for a high quality network-centric S&T program peer review.

The objective of the review is to evaluate a large ongoing S&T program, using a representative segment of the technical community, and employing whatever information technology is required to substantially enhance the quality of the review. For illustrative purposes only, the parameters of the Department of the Navy Advanced Technology Development program review will be used in the following discussion.

**Definition of Evaluation Criteria**

In the proposed network-centric review, after the objectives and goals have been specified, the first operational step would be to define the evaluation criteria. These are the metrics that would allow quantitative determination of progress toward the goals and objectives. For mission-oriented organizations, there tend to be two overarching evaluation criteria: mission-relevance and technical quality. For a variety of reasons, including the analysis of progress in achieving
sub-goals and objectives, additional supportive criteria tend to be employed in reviews. For the proposed review, assume the same criteria are used as were employed in the Department of the Navy illustrative example: Military Goal; Military Impact; Technical Approach/ Payoff; Program Executability; and Transitionability. In combination, these criteria will help answer the question: Will this program result in a high impact high-quality militarily relevant product with high probability of meeting cost, schedule, and performance targets?

Selection of Review Taxonomy

The second operational step is selection of a taxonomy for the review. A cardinal rule in assessment is that a program should be reviewed using the same taxonomy by which it was selected and managed. Otherwise, the program integration (linkages among the program’s sub-components) will appear fragmented, even though the sub-components may appear of high quality individually.

A taxonomy is analogous to a mathematical coordinate system, and the requirements for a high quality S&T taxonomy parallel those of a high quality coordinate system. These requirements/characteristics are:

Orthogonality - a good coordinate system has orthogonal axes, where the inner product between any two axes is zero. This avoids multiple counting and axis redundancy. Similarly, a good taxonomy should have categories as independent as possible.

Completeness - a good coordinate system has sufficient degrees of freedom to cover the full range of dimensionality of the physical problem. A 2-D coordinate system would be insufficient for representing a 3-D problem. Similarly, a good program taxonomy will have a sufficient range of categories to include the different technical disciplines that could occur.

Unit basis vectors - a good coordinate system has the unit vector for each dimension the same size. This avoids resolution mis-matches. In addition, the computational grid size should have adequate resolution to allow computational results to be compared to experimental results. Similarly, a good program taxonomy should include technical disciplines of relatively equal importance with relatively equal amounts of funding, with sufficient category resolution to allow equal levels of coherence about a central theme.

Alignment - a good coordinate system is aligned with the structure of the physical problem. This simplifies the solution by reducing the conversion/translation between the grid and the structure. A spherical coordinate system is more appropriate to representing a spherical body than a cartesian rectangular system. Similarly, a good program taxonomy should be impedance-matched to data availability.

Assume that these guidelines are followed in taxonomy selection for the proposed review, and a taxonomy of forty categories is defined to represent the total program.
Review Panel Selection

The third operational step is review panel selection. The availability of information technology capabilities will allow the following substantial panel enhancements relative to traditional peer review procedures.

Use of Group-Ware for entering data and computing summary rating statistics in real-time will allow a much larger and more representative segment of the technical community to actively participate in the process;

Having a larger panel will allow the expert witness function and the jury function to be decoupled, similar to the procedure of the Science Court (DOE, 1978);

Having a larger panel will also allow reviewers to be selected with expertise in a particular evaluation criterion for a specific technical area;

Use of data mining techniques in different literatures will allow a larger pool of experts to be identified as potential process participants.

For the proposed review, assume there is a central panel of perhaps fifteen individuals, and there are one hundred expert reviewers. The fifteen central panelists would not necessarily be expert in any of the areas reviewed, but would be high caliber individuals as free as possible of potential conflict with the programs under review. In the legal analogy, they would serve as the jury. The hundred expert reviewers would be divided equally among the five criteria, or twenty per evaluation criterion. In the legal analogy, they would serve as the expert witnesses. While complete independence from the programs reviewed would be preferable for the expert reviewers, it would not be the absolute requirement used for the fifteen central panelists.

The fifteen central panelists would be selected based on national reputation and absence of conflict. Their function would be to provide final ratings and comments on all the evaluation criteria for all forty programs under review. Their inputs would consist of background material provided by the program presenters, actual program presentations, and preliminary comments and ratings by the one hundred expert reviewers.

Expert reviewer selection would proceed as follows, using the Technical Approach/ Payoff criterion as an example. In parallel with recommendations for experts in the forty technical areas under review, the literature would be ‘mined’ using key phrases that describe the forty technical areas. A large number of reviewer candidates would be obtained. Bibliometrics would be employed to winnow this list through identification of those candidates with extensive publishing and citation records. Other reviewer selection criteria would be employed, to insure that bright younger people, who have not yet established a publication track record, would be included in the review process. All four of these selection approaches were used to nominate participants for the innovation workshop referred to previously, and have been used in part by the author for other types of reviews as well.
The twenty candidates selected as expert reviewers for the Technical Approach/ Payoff criterion would have two required output products. They would provide comments and preliminary ratings only on the single evaluation criterion for each of the forty programs. In order not to overwhelm the fifteen central panelists with comments and preliminary ratings from each of the twenty expert reviewers for each of the five criteria for each of the forty programs, one of the expert reviewers for each criterion for each program would be assigned the task of aggregating and summarizing the comments and preliminary ratings for the given criterion and program. To insure a balanced summary is presented from the expert reviewers to the central panelists, another of the expert reviewers for the criterion would have to approve the summary generated by the expert with primary authority. This expert with secondary authority would be selected based on maximum divergence with the viewpoints of the expert with primary authority, to the extent known beforehand. In the illustrative example, each expert reviewer would serve as the primary authority for Technical Approach/ Payoff for two programs, and would serve as the secondary authority for Technical Approach/ Payoff for two other programs.

Operational Review Process

Selection of the goals and objectives, evaluation criteria, review taxonomy, and reviewers, and definition of assignments and responsibilities, establish the structure of the review. The structure, in turn, provides the foundation for the operational review procedure that follows. The complete review process proposed here will consist of three phases: pre-presentation, presentation, post-presentation. The steps emphasized are those in which the use of information technology, especially in the network-centric mode, will enhance the efficiency and quality of the peer review process. Most of the procedures proposed have either been used or tested to some degree by the author, and their feasibility has been demonstrated.

Pre-Presentation Phase

The objectives of this phase are to provide as much information to all the review participants as is possible before the meeting occurs, and to clarify any outstanding questions and issues. This will allow the participants in the presentation phase to start on a much higher plane, and use the presentation period much more efficiently.

This pre-presentation phase has three distinct sub-phases. First is the distribution of background material. This sub-phase objective is to provide maximal information about the programs to be reviewed and about global efforts in the programs’ technical areas and allied disciplines. Since all reviewers are required to provide a preliminary rating on one criterion for every one of the forty programs, this sub-phase will provide the threshold level of understanding about each program necessary for casting an intelligent vote.

The second sub-phase consists of e-mail interaction among reviewers, where comments are exchanged about the program material and issues are clarified. At the end of this sub-phase, each reviewer has transmitted his or her comments on the assigned evaluation criterion for each of the forty programs to the individuals assigned primary and secondary responsibility for the specific
criterion for each program.

The third sub-phase consists of the primary and secondary principals responsible for each criterion for each program writing a brief summary based on the inputs of the other reviewers assigned to each criterion for each program. At the end of this sub-phase, these brief summaries will have been transmitted to the fifteen member central panel, along with the preliminary summary rating statistics for each criterion for each program.

Distribution of Background Material
This phase begins with the distribution of background material for the reviewers (and audience, if an audience is desired). In order for the background process to be most effective, the material should be distributed at least three months prior to the actual presentations. Two types of material are proposed.

First are narratives and vu-graphs describing in detail the material to be reviewed. The author distributes this type of background information routinely for S&T peer reviews. Requirements for this material have been detailed elsewhere (Kostoff, 1998). To maximize distribution efficiency, the material should be made available on the Internet, and the reviewers and audience informed of its location. If distribution of some of the material has to be restricted for proprietary or other reasons, then the Web site should be password-protected.

The second type of material is information related to the programs to be presented. This material is ‘data-mined’ from appropriate source S&T databases (e.g., Science Citation Index (basic research), Engineering Compendex (applied research and technology), NTIS Technical Reports (government-sponsored S&T reports), Medline (medical S&T), RADIUS (narratives of on-going government R&D programs). The author has distributed “data-mined” information to support reviews of technical areas of modest breadth. This information can be very valuable in identifying the scope of S&T performed globally in the specific technical area under review, in allied areas, and in disparate fields that have some thread of commonality with the specific area under review.

However, even for fields of moderate breadth, substantial effort is required to provide useful background information of this type. The query used has to be refined to satisfy two conditions: the coverage (records retrieved) should be comprehensive (large signal), and have minimal extraneous material (large signal-to-noise). Then, for most recipients, the records retrieved need to be summarized. The author has used the Database Tomography approach (Kostoff, 1999b) to develop queries with these properties, and to summarize the main pervasive technical themes in such retrieved record databases, and the relationships among these themes. While these computational linguistics and bibliometrics tools help substantially, they do not obviate the need for technical experts to spend substantial time and effort in developing this background material.
For the illustrative example used in this report, a forty sub-program Advanced Technology Development naval S&T program, the effort required for global data mining of the technical disciplines to be reviewed would be enormous. Nevertheless, if each reviewer’s rating is to be meaningful, then the reviewer needs to have some threshold level of understanding about each program reviewed. A substantial effort is necessary to provide such information, especially in summary form.

Individual Reviewer’s Comments
The discussion in this sub-section follows the experience of the innovation workshop in Autonomous Flying Systems mentioned previously. Even though the objectives of a workshop are different from those of a peer review, nevertheless, the principles learned from the workshop’s pre-presentation phase can be readily extrapolated to peer review application.

In the innovation workshop, each participant sent new concepts relating to the workshop theme to all the other participants by e-mail. An e-mail-based interactive discussion ensued among the participants to ‘flesh-out’ the concepts, and either clarify and/ or embellish them in preparation for the actual presentations. In order to stimulate this e-mail discussion, a facilitator was required to raise numerous questions. The discussion proved extremely successful in clarifying the concepts, but the need, and effort required, for facilitation of the discussion was appreciated only after the pre-presentation phase had begun.

In this phase of the peer review process, after the reviewers have received the background material, they would be expected to spend the next few weeks digesting the material and clarifying any outstanding or problematic issues. The primary and secondary principals for each criterion for each program would be expected to act as facilitators, to stimulate discussion on these issues. The total review group would not be involved in each e-mail discussion group; this would overwhelm the communication channels. Each e-mail discussion group, in the present example, would consist of the twenty experts for a given evaluation criterion for a given program, plus the individual who will be presenting the information. At the end of this phase, approximately two months before the presentations, each of the twenty experts would provide his/ her comments and preliminary ratings on the given evaluation criterion for the given program to the appropriate primary and secondary principals.

Summary Comments to Central Panel
After receiving the individual comments and preliminary ratings from each reviewer, the primary and secondary principals for each criterion for each program will generate a brief summary for each criterion for each program. If the two principals cannot agree on a specific summary, the secondary principal will contribute a dissenting addendum to the summary transmitted by the primary principal to the central panel. In any case, both the comment summary and a summary of the preliminary rating statistics are transmitted to each member of the central panel. In order for the central panel members to have time to absorb all the summary material, they would need to receive it no later than one month before the presentations.

In summary, the total pre-presentation time-line is as follows:
Distribution of background material to expert reviewers - three months before presentations

Transmission of comments and preliminary ratings to primary and secondary principals - two months before presentations

Transmission of summary comments and preliminary rating statistics to central panel members - one month before presentations.

Presentation Phase
In network-centric peer review, this phase is optional. There is no fundamental requirement for presentations. All of the review could be conducted through the network by e-mail, Internet, etc. However, there is a cultural aspect to peer review that rivals the information technology aspects in shaping the conduct of the review. Many cultures are not yet at the required comfort level with purely remote operation. In addition, there is value in real-time discourse with the presenters. Therefore, this presentation phase will be included in the present report.

For the scenario proposed in this report, presentations will be made to an on-site audience consisting of the fifteen member central panel and the one hundred member reviewer group. Presentations can also be made to a remote audience by video tele-conferencing. Under the present scenario, the role of the remote audience is observation.

All the members of the on-site audience will be linked by Group-Ware. During the presentations, the reviewers will enter final ratings and any additional comments they believe are important based on last-minute observations or insights. At the end of each presentation day, the remote transmission link will be closed, and the reviewers and central panel will meet in Executive Session. The Group-Ware algorithms will have computed each program’s statistics (panel averages for each evaluation criterion rating, etc) and any desired integrative statistics over multiple program groups as well. All these numerical results will be displayed graphically to all the on-site audience. The Group-Ware will have also aggregated the additional comments, and these comments will be displayed to all the participants. Both the ratings and the comments will be discussed for each evaluation criterion for each program presented. The central panel will then rate each evaluation criterion for each program presented, and these final program and integrative statistics will be displayed in real-time.

A note about Group-Ware. In the naval Advanced Technology Development review described in the text, Group-Ware was used in part. It had two components: computing summary and integrative statistics, and aggregating comments. Both these features operated in real-time. The immediate summary and integrative statistics feedback provides for high efficiency discussions, and its value increases as the number of programs reviewed and the number of experts used increase. The comment aggregation is valuable for documentation purposes. For an on-site panel, comment aggregation has little value, can serve to bias reviewers’ initial comments, and can be a distraction to some reviewers. For reviewers from remote locations, comment aggregation should prove to be of substantial value.
Post-Presentation Phase
This phase consists of writing the final review report. Depending on the contractual structure of
the review, either the staff of the organization sponsoring the review will write the report, or the
central panel will write the report. Because of the extensive pre-presentation preparation, the
involvement of a large segment of the community, and the extensive interactions that occurred
during all prior phases of the review, much of the available information will be ready for direct
insertion into the report.

RESEARCH OPPORTUNITIES IN NETWORK-CENTRIC PEER REVIEW

Opportunities for research into network-centric peer review abound. Issues to be addressed
include the following:

- How is peer review quality defined, especially in a network-centric mode? What are the
  metrics of quality; how can they be measured? What data is required to quantify these
  metrics, and how is this data obtained?

- What incentives and rewards have been employed to produce higher quality reviews, and
  what incentives and rewards should be tested for efficiency?

- *What types of network architectures should be developed for optimal review operation?
  How extensive should the networks be for successful operation? What are the
  implications of reviewer anonymity protection on the network architectures? What other
types of security and verification procedures are required to minimize review disruption
  and corruption problems? What levels of fault-tolerance need to be incorporated into the
  network? What are the hardware and software requirements for optimal large-scale
  operation?

- What are optimal reviewer selection processes, and what are the trade-offs among these
  processes?

- What are the cost-benefit considerations related to panel sizes, for different types of
  review objectives? What are the trade-offs of adding experts in a given technical area for
  statistical reliability and validity purposes verses broadening the expertise representation
  across many different fields? How far should the expertise diverge from the target S&T
  being evaluated, in order to access insights from other disciplines that could benefit the
  target discipline?

- What are the trade-offs involved in Science Court operation verses dual function jury-
  witness panel? What other panel operational modes are possible with network-centric
  operation? What has been the experience of these other operational modes; what is the
  potential of other operational modes, whether or not there has been some past history of
  operation?
• What credible processes exist, or could be devised, to normalize across panels and disciplines? How does network-centric operation complicate or simplify these diverse processes?

• How does the expanded capability of network-centric operation impact the selection of diverse evaluation criteria, and how does it impact the development of, and accession to, the data required to address these criteria?

• How are reliability and repeatability impacted by network-centric operation?

• How should the different types and sources of global data be accessed and integrated with the peer review process? What are the implications on the process operation and results on the availability of these different types of data? What data sources need to be developed and constructed to provide required information for peer reviews, and how does network-centric operation influence the composition and structure of these sources?

• What are the true costs and benefits of network-centric peer review, and what are the main parameters that affect cost-sensitivities? What steps could be instituted now to reduce potential high cost components of the network-centric peer review process?

• How should the larger network-centric strategic management process be constructed in order to maximize benefits from network-centric peer review, as well as optimize benefits organizationally and nationally from the strategic management process? What constraints do the other elements of the network-centric strategic management process place on efficient operation of the network-centric peer review component, and what enhanced capabilities for the peer review component do these other components offer? What are the common elements of all the components of the strategic management process, and what are the unique elements required for network-centric peer review? Are there benefits to constructing architectures that will encompass all the network-centric strategic management components, such that specific requirements for the peer review component will require a minimal additional requirement for resources?

SUMMARY AND CONCLUSIONS

Network-centric peer review uses the power of modern communication networks and information technology to expand greatly the number of people that can participate in real-time peer reviews, and expands greatly the access to data that can support all aspects of peer review. This technology allows diverse review operational modes such as the Science Court to be considered seriously, and allows the jury function of peer review to be independent from the higher conflict potential expert reviewer and witness function. The operational architecture required for network-centric peer review may differ little from the architecture required for its parent network-centric strategic management, and since all strategic management components need to be integrated for optimal synergistic benefits, implementation of network-centric peer review should occur in parallel with implementation of the other components of network-centric
strategic management.
VI-E. APPENDIX V – SIGNIFICANT RESOURCES

The following sub-appendices of Appendix V are particularly noteworthy resources for peer review information. VI-E-1 contains an evaluation of an excellent document on grants/proposals peer review by Wood and Wessely. This source document should be required reading for anyone interested in proposal peer review. VI-E-2 contains a broad outline to a DOE peer review guide presently under review. This comprehensive document should be of substantial help to any organization interested in the fundamentals and protocols of program peer review. VI-E-3 overviews the international congresses on biomedical peer review. These periodic congresses have covered a multitude of peer review topics, whose purview goes well beyond the biomedical community, and the proceedings of these congresses are required reading for anyone interested in improving the conduct of manuscript peer review.

VI-E-1. PROPOSAL PEER REVIEW

This appendix highlights the main issues addressed in a recent document that examines grant proposal reviews (Wood and Wessely, 2003). The author highly recommends reading the full document for anyone interested in grant proposal reviews.

SUMMARY OF DOCUMENT AUTHORS’ OVERVIEW

‘The document presents a systematic review of the empirical literature on peer review and grant applications. As a base for interpreting this review, brief historical and contextual information about research grant funding agencies and the peer review process is provided. The authors stress that peer review is only one means to an end – it is not the end itself.

There have been numerous criticisms of peer review in the context of grant-giving, chiefly centered on claims of bias, inefficiency, and suppression of innovation. The authors conclude that, with certain exceptions, peer review processes as operated by the major funding bodies are generally fair. The major tension exists in finding reviewers free from conflict of interest who are also true peers. The authors find little evidence to support a greater use of “blind” reviewing, or of replacing peer review by some form of citation analysis.

The document draws attention to the increased costs in both time and resources devoted to grant peer review, and suggests that some reforms are now necessary. The authors are unable to substantiate or refute the charge that peer review suppresses innovation in science – in general they conclude that peer review is an effective mechanism for preventing the wastage of resources on poor science – but whether it supports the truly innovative and inspirational science remains unanswerable. Finally, the document draws attention to the paucity of empirical research in an area of crucial importance to the health of science and recommends that ways for improving international understanding, debate and sharing of ‘best practice’ about grants peer review be investigated.’
The authors address a number of crucial issues related to proposal peer review. These include:

- Is peer review of grant applications fair? Do researchers think peer review of research proposals is fair, and are they satisfied with the peer review process? The authors’ conclusion is that applicants endorse the principle of peer review, but a substantial minority have practical criticisms. What is the evidence to support these criticisms?

- Are peer reviewers really peers? Applicants often complain that reviewers are not specialists in the relevant fields – in other words not true “peers”.

- Is there institutional bias? Is there a bias against lesser known individuals and/or institutions, either in the choice of reviewers or the decisions of grant committees? The authors conclude that in the grants literature, there is little evidence that the choice of reviewers reflects this bias.

- Do reviewers help their friends? A related issue is the perception of “cronyism”.

- Age and getting grants. Another frequent perception is that the system operates against younger researchers.

- Gender bias and grant peer review. Is peer review biased against women?

- Other biases. Many other biases have been claimed.

  - Reviewer responses were more likely to be favorable when dealing with their own discipline, just as reviewers are more likely to cite their own discipline within the context of general reviews, a possible interdisciplinary bias. On the other hand, there was a significant association between number of disciplines represented and success in obtaining grants from the UK National Health Service R&D program, suggesting a bias against uni-disciplinary research.

  - There is little evidence to suggest bias against clinical, as opposed to molecular research. Nonetheless, at government level, concern that patient-oriented research is adequately addressed by funding agencies is reflected in a number of initiatives.

  - Another claim, supported on the basis of personal observation by the authors, is that grants reviewed early in a session tend to be discussed more thoroughly and evaluated more critically than those reviewed later.
• Misuse of confidential information. The peer review system presumes a high level of objectivity, disinterestedness and honesty on the part of reviewers. However, this presumption has been challenged by a number of critics who believe that the system allows for “leakage” - a euphemism for theft of ideas by reviewers from the grants they review.

• Reliability of grant peer review. Are ratings reliable?

• Does peer review of grant applications serve the best interests of science? It has been frequently argued that peer review is inherently conservative and biased against speculative or innovative research. Those who write grant proposals agree, and may deliberately underplay the innovative parts of their proposals.

• Is peer review of grant applications cost effective? Many have observed with concern the amount of time spent in both writing and reviewing grants.

• Can peer review of grant applications be improved?
  - Blinding. There have been many suggestions of ways of improving the quality of peer review, albeit with few supported by empirical data. The question of blinding of referees to applicants and their institutions has already been considered under equity. Could it improve quality?

  - Signing. The other side of the coin is whether or not reviewers should sign their reports. This is currently the subject of controlled trials in the field of editorial peer review, and has been suggested for grant reviewing on several occasions.

  - Improving reliability. If reliability is a problem, can it be improved?

  - Tackling cronyism. Asking applicants to nominate referees is also often practiced, although we are unaware of any system where this is the only system. It is frequently thought that referees chosen in this manner will be more favorable than those selected by the grant-giving body. A comparison of scores carried out at the Medical Research Committee of the NHMRC found this was indeed the case, and discontinued the process.

  - Triage. The most popular way of improving efficiency has been to introduce some form of triage, in which not all grants receive the full process and deliberations of the full committee, but are rejected at an earlier stage. It has been used at the NIH, where a pilot study of reviewers suggested it was still fair, and a subsequent analysis verified that this did not introduce discrimination against ethnic minorities.
• Other suggestions. Other suggestions for which there is no empirical support or refutation include:

• adjusting individual reviewers’ scores according to their previous performance (akin to a golf handicap),

• paying referees, and

• restricting reviewers from receiving grants from the same source.

• Should peer review of grant applications be replaced? Many alternatives to peer review have been suggested.

• The most common replacement involves bibliometrics. This is the use of mathematical models of scientific productivity, since scientific work results in scientific publication. Various less orthodox suggestions to replace peer review have also been made.

• awarding of grants at random or after a lottery,

• cash prizes to stimulate research in key areas,

• random selection of reviewers from a pool, or

• a system of professional reviews, akin to theatre critics.

• The development of the chronometer is a historical precedent for funding by means of cash prizes, first pointed out by David Horrobin and subsequently the topic of a best-selling book.

• A recent, and apparently, successful approach to providing financial incentives for scientific innovations is the Web-based initiative ‘InnoCentive’ developed by the pharmaceutical company Eli Lilly. ‘InnoCentive’ posts a set of R&D ‘challenges’ to which scientists throughout the world can respond – the reward being both financial (the amount linked to the difficulty of the challenge – e.g. US$2000, or US$75000) and professional recognition. There have been some attempts to analyze the outcome of
research supported by different funding mechanisms, but there are no studies looking at the long term impact of different methods of peer review and its alternatives.

The chapter concludes the following (the author’s critiques of these conclusions are shown at appropriate points in the text below, in CAPS).

‘Peer review is a family of closely related procedures, differing not only between funding bodies, but also between programs within the same funding body. Results from one time period or one institution cannot necessarily be generalized to other settings. Given those caveats, what can be concluded about the many criticisms made of the process?

The most frequent criticism made by scientists about the day-to-day operation of peer review is that of institutional or gender bias. The authors suggest that this criticism is generally unfounded, with certain specific exceptions. However, even if biases can be established, the question of their adverse impact on research quality has not been systematically investigated. Indeed, claims regarding institutional and gender biases are usually couched in terms of issues to do with ‘equal shares of the funding pie’ which in themselves are not directly linked to research quality.

THE CENTRAL PURPOSES OF PEER REVIEW SHOULD BE TO IMPROVE THE QUALITY OF PAPERS/ PROPOSALS/ PROGRAMS, AND GENERALLY HELP ACCELERATE THE DEVELOPMENT OF S&T. TECHNICAL CORRECTNESS SHOULD BE THE PRIMARY OBJECTIVE, NOT POLITICAL CORRECTNESS. TO THE DEGREE THAT BIASES IMPACT THESE CENTRAL PURPOSES NEGATIVELY, THEY SHOULD BE ELIMINATED.

Lack of reliability has been found, but again may not be a fundamental weakness. Some is due to lack of reviewer expertise, which is potentially remediable, some due to reviewer age, but much results from the lack of consensus in areas on the frontiers of knowledge, which is where applications submitted to peer review are situated. In only one area is there clear consensus – the costs of peer review, both direct and indirect, are increasing.

The authors suggest there is no such thing as the perfect reviewer. Those too close to the subject may be influenced by jealousy or cronyism. More distant, and they may suffer from lack of expertise. Increasing the use of international reviewers is often suggested as a means of reducing conflict of interest and jealousy, but “off the record” observations from some grant officers are that these tend to produce more favorable, and hence less rigorous, evaluations. Perhaps a certain amount of competition is a spur to critical appraisal. There seems to be no substitute for grants officers who know the strengths and weaknesses of their reviewers.

MORE TO THE POINT, THERE IS NO SUBSTITUTE FOR GRANTS OFFICERS WHO UNDERSTAND THE SUBJECT MATTER THOROUGHLY, AND ARE ABLE TO DISCERN THE DIFFERENCES IN COMPETENCES OF REVIEWERS. SUCH GRANTS OFFICERS ARE MORE LIKELY TO UNDERSTAND THE SUBJECT MATTER IN DEPTH IF THEY HAVE BEEN, AND PREFERABLY STILL ARE, ACTIVE CONTRIBUTORS TO THE SCIENCE AND TECHNOLOGY DEVELOPMENT.
Until fairly recently publicly available information regarding the peer review process of research funding agencies was quite limited. However, demands for greater accountability have resulted in various efforts by funding agencies to improve the understanding of their operations and provide information on their peer review procedures. For example, changes to the peer review procedures of the US National Institutes of Health have been well-publicized and extensive consultation invited from a wide range of stakeholder groups. The UK Medical Research Council has also provided summary information on the recent assessment of its peer review system. The internet is clearly an important tool for achieving greater transparency about the operations of research funding councils and their peer review procedures. However, it is worthwhile noting the caveat of O’Neill that: ‘there is a downside to technologies that allow us to circulate and recirculate vast quantities of ‘information’ that is harder and harder to sort, let alone verify.’

Many of the questions addressed have not received definitive answers. As with journal review in the previous decade, there are now sufficient concerns with grant peer review to justify empirical research. Questions such as the role of blinding, feedback, and the balance of external and internal reviewers as well as gender and institutional bias require answers. Peer review of these questions would, as in other areas of scientific uncertainty, highlight the need for randomized controlled trials to address these issues. The paucity of trials in the area of scientific decision-making is therefore ironic.

Turning from the concerns of individual scientists about the fairness and reliability of the peer review system, the most important question to be asked about peer review is whether or not it assists scientists in making important discoveries that stand the test of time. The authors do not know. Furthermore, randomized trials will not address this most difficult, yet most important, question. This is a judgement which, by definition, can only be made with the passage of time. ‘ASSISTING IMPORTANT DISCOVERIES’ MAY BE AN OVERLY AMBITIOUS GOAL FOR PEER REVIEW. IMPROVING PROPOSAL QUALITY AND HELPING ACCELERATE S&T PROGRESS ARE SUFFICIENTLY CHALLENGING GOALS.

Likewise, does peer review impede innovation? It is desirable that resources are not wasted on poor science, but is this at the expense of the suppression of brilliance? This remains unproven, and possibly unprovable.

ONE CAN LEARN MUCH ABOUT THE ROLE OF PEER REVIEW IN IMPEDING OR ENABLING INNOVATION, AND PEER REVIEW’S VALUE IN PREDICTING S&T OUTPUTS AND OUTCOMES IN GENERAL, THROUGH CONTROLLED TRACKING OF PROPOSALS AND THEIR LONG-TERM FATE. UNFORTUNATELY, THE S&T SPONSORING COMMUNITY HAS DEVOTED SCANT RESOURCES TO LARGE SCALE COMPARATIVE STUDIES OF ALL PROPOSALS (WINNERS AND LOSERS) THAT WOULD DETERMINE THE FATE OF THESE PROPOSALS AND THEIR S&T IMPLEMENTATIONS, AND THEREBY DETERMINE THE CAPABILITY OF PEER REVIEW TO PREDICT THE WINNERS AND LOSERS. THE MAIN INTEREST IN STUDIES THAT TRACK PROPOSALS APPEARS TO BE THE HINDSIGHT-TRACES TYPE STUDIES, WHICH START WITH SUCCESSFULLY DEPLOYED ADVANCED TECHNOLOGY SYSTEMS, AND WORK BACKWARDS TO IDENTIFY CRITICAL
TECHNICAL BREAKTHROUGHS AND FAVORABLE ENVIRONMENTAL AND MANAGEMENT CONDITIONS (DOD, 1969).

The current interest shown by the scientific community in peer review has a pragmatic basis – the links between grants and the structures of scientific careers. Obtaining grants is increasingly an end in itself, rather than a means to an end.

PEER REVIEW IS NOT AN END IN ITSELF, BUT A MEANS TO THE END OF IMPROVING S&T QUALITY AND RESOURCE ALLOCATION (AND THEREBY ACCELERATING S&T DEVELOPMENT). FROM THE LARGER SOCIETAL PERSPECTIVE, OBTAINING ALLOCATED RESOURCES (GRANTS) SHOULD NOT BE AN END IN ITSELF, BUT A MEANS TO THE END OF PURSUING S&T FOR THE HIGHER PURPOSE OF ATTAINING SOCIETAL GOALS. UNFORTUNATELY, AS THE AUTHORS CONCLUDE, “OBTAINING GRANTS IS INCREASINGLY AN END IN ITSELF.” DEVELOPMENT OF RESEARCH EMPIRES FOR THE SOLE PURPOSE OF CONTROLLING EVER LARGER AMOUNTS OF FUNDING LEADS TO RESOURCE ALLOCATION DISTORTIONS, TO THE EVENTUAL DETRIMENT OF THE S&T ENTERPRISE.

Hence the fascination all scientists have in the process, and their willingness to express criticisms of it. Because obtaining grants is so important for scientists, it is proper to obtain further empirical data on questions such as equity and efficiency, but this should not blind us to the fact that such research can only answer short term questions rather than the real purpose of scientific endeavors.

Advances in medical research itself- for example in the area of stem cell research - have raised many new ethical and intellectual property issues for grants peer review. The overall accountability and regulatory environment for the conduct of research is also substantially different from that impacting on funding agencies several decades ago. And the scientific process itself has become increasingly internationalized, with greater stress on team based, collaborative research projects. The efficacy of peer review procedures in this new climate is clearly of great importance. In this regard, support for periodic independent reviews of the funding councils peer review processes have been strongly encouraged by governments in the UK and Canada with the former recommending the formal establishment of the research councils’ strategy group aimed at developing best practice in agency operations. Use of consultancy groups to provide independent assessments of agency peer review systems appears also to be on the increase (e.g. Segal Quince Wicksteed in the UK). In Europe, the heads of the national research councils of the European Union (Eurohorcs) meet twice a year primarily to discuss shared problems. Recently, the Swiss National Science Foundation celebrated its 50th anniversary in 2002 with a workshop on major challenges for research funding agencies at the beginning of the 21st Century. Representatives from twenty countries and the EU took part, identifying the issues and problems, and discussing ways of dealing with them. In 1999, the UK Economic and Social Research Council sponsored a global cyber-conference on peer review in the social sciences. However, despite so much activity taking place in various fora and domains, grants peer review (in contrast to editorial peer review or other topics regarding the conduct of science) seems to have attracted remarkable little
attention in the form of regular congresses/conferences intended to improve understanding and debate about its form and practice. This, the authors would argue, is an issue that warrants further investigation.

THE LATTER CONCLUSION COULD BE STATED EVEN MORE EMPHATICALLY FOR S&T PROGRAM REVIEW, ALTHOUGH SINCE THE PASSAGE OF THE GPRA BY THE US CONGRESS IN 1993, THERE HAS BEEN SUBSTANTIALLY MORE ATTENTION PLACED ON S&T PROGRAM REVIEW THAN IN PRIOR YEARS.
VI-E-2. PEER REVIEW GUIDE

For the past two decades, the U. S. Department of Energy has been a leader in advancing the use of peer review for evaluation of its programs. In 1982, a massive review was conducted of the Office of Basic Energy Sciences (DOE, 1983). The principles established from that review were used in many different sectors of DOE over the next decade, and a peer review guide was developed to formalize those principles (DOE, 1988). In 2003, a peer review task force in the Energy Efficiency and Renewable Energy (EERE) Office of DOE was assembled to develop a peer review guide for the evaluation of EERE programs. In January 2004, an external multi-agency group of peer review experts (Chaired by the author) met to review, and provide individual recommendations on, the peer review guide. As of this writing, the peer review guide is under review by EERE management, and details of its contents cannot be described until the document is finalized. However, the document’s contents can be generally summarized as the following.

The primary purpose of this guide is to provide managers and staff guidance in establishing formal in-progress peer review that provides intellectually fair expert evaluation of EERE RD³ and supporting business administration programs, both retrospective and prospective.

The guide focuses on activities that are planned, underway, or have recently been completed and does not directly cover merit review or readiness reviews, which are addressed in other EERE management procedures. In-progress peer review (or simply “peer review”) findings will be considered by DOE/EERE managers, staff, and researchers in setting priorities, conducting operations, and improving projects.

This guide provides information and examples useful for planning, conducting, and utilizing peer reviews based on best practices. Best practices are those that are (1) utilized with the most success by EERE’s own programs or by other institutions, or (2) identified as such by multiple widely recognized experts outside of EERE, including experts at the GAO and OMB.
VI-E-3. BIOMEDICAL PEER REVIEW CONGRESSES

The author would like to emphasize the papers published in the international congresses on peer review in biomedical publication. For anyone interested in biomedical peer review in particular, or journal peer review in general, there is no better starting point. There have been four congresses held since 1985, and the next one is scheduled for 2005. The congresses cover a wide swath of peer-review related topics including, but not limited to:

Mechanisms of peer review and editorial decision making
Evaluations of the quality, validity, and practicality of peer review and editorial decision making
Online and Web-based peer review and publication
Prepublication posting and release of information
Quality assurance for reviewers and editors
Authorship, contributorship, and responsibility for published material
Breakdowns, weaknesses, and biases
Conflicts of interest
Scientific misconduct
Peer review of grant proposals
Economics of peer review and scientific publication
Evaluations of the quality of print and online information
Methods for improving the quality, efficiency, and equitable distribution of biomedical information
Interactive digital systems and other new technologies that affect the dissemination of biomedical information
The future of scientific publication

Programs and Abstracts of the upcoming congress, as well as those of the previous congresses may be accessed at http://www.ama-assn.org/public/peer/peerhome.htm.
VI-F. LARGE AGENCY PEER REVIEW

This appendix has two sections. The first estimates the cost of a peer review for a large S&T funding agency. The second describes some of the ways that text mining could support such a review.

COST ESTIMATES OF TOTAL AGENCY REVIEW

This final appendix addresses the economics of peer review if it were implemented agency-wide in large organizations, and also describes the role that text mining could play in such a review.

Federal agencies conduct a variety of program reviews, at many different levels of detail, and at many different organizational levels. For those agencies that sponsor S&T programs, both technical and non-technical (business) reviews are conducted. Assume that it was desired to conduct technical program peer reviews of an agency with an annual S&T budget of $1B. What would be a reasonable approach to such a review, and what would be its cost estimate?

The first step would be to develop an agency review strategy. This would have two objectives: identify how each review integrates into the tactical and strategic management of the S&T, and consolidate reviews to eliminate overlaps and redundancies. The next step would to identify the scope of the technical review. In an S&T program review, three main questions are asked: 1) Is the S&T program doing the right job (adequacy of the existing S&T investment strategy and associated roadmaps); 2) Is the S&T program doing the job right (accuracy and efficiency of achieving the specified technical target. It evaluates the mechanics of the S&T development approach, and incorporates the cost, performance, schedule, and risk aspects of the mechanics); 3) Is the S&T program performing (is there associated productivity, impact, and progress)? From the author’s perspective, these three criteria would be evaluated at all levels of the organization, especially the first criterion at the highest levels.

Given this scope and these objectives, how would one conduct the review, and what would be its cost estimate? Many approaches exist; one will be presented here, based on the author’s recent experience, and costs will be extrapolated from those experiences.

From 1993 to 1998, the author conducted an annual review of his former Department. He used independent technical experts to participate in the review. The review process used is generalized, and described in the next appendix in some detail. From 1999 to 2003, the Naval Studies Board (NSB), an arm of the National Research Council of the National Academies, was contracted to conduct an annual review of the author’s former Department. The review was completely independent, with reviewers selected by the NSB. Each review constituted one three day meeting, consisting of two days of presentations by the S&T program managers, and one day of discussion and initial report drafting by the reviewers. A final report was issued about four-six months after the meeting. These reports are unclassified, and available from the National Academies Press (www.nap.edu). A list of titles is presented at the end of this appendix. From the author’s perspective, the reviews were conducted at the right level of detail for the objectives,
and provide a good model for scaling up to much larger reviews.

The Department was reviewed on a three year cycle, with one-third of the Department reviewed in detail annually. The funding of the programs reviewed varied from year-to-year, but averaged between $40-60M annually. Conservatively, about $20M worth of programs were presented per day of presentations. Today’s cost of such a review, including staff time, and reviewers’ travel and per diem, would be about $100K, or about $50K/$20M worth of presentations. For a $1B/Yr total agency program, about 50 days of presentations per year would be required, and an out-of-pocket cost of $2.5M per year would be incurred, or about a quarter of a percent of the total S&T program. Total costs, including preparation time for the presenters, and reviewer and audience time, would increase the cost substantially. Assume a factor of four multiplier, resulting in a total cost of $10M/Yr for the review, or one percent of the total program funding. For an agency with a $5B/Yr total S&T program funding, a simple scale-up would result in $12.5M/Yr in out-of-pocket costs, or $50M/Yr total costs, and 250 days worth of presentations. Thus, on average, there would be one presentation per day. If such a large agency review were conducted by one organization such as the NRC, a separate Board devoted to the review would have to be established, and economies of scale might be possible, decreasing the cost estimates somewhat. If multi-large agency reviews could be coordinated, then programs of similar generic themes could be combined in the review, and coordination issues could be observed and evaluated directly.

TEXT MINING TO SUPPORT LARGE AGENCY REVIEW

How can text mining help generate answers to each of the three major questions above?

Is the S&T program doing the right job?

In order of precedence, this is the first issue to address. It focuses on the adequacy of the existing S&T investment strategy and associated roadmaps. It starts with a vision, or description of the operational scenario. This is followed by an elucidation of the capabilities required in order for the vision to be implemented. The capabilities are quantified to provide the development targets. A roadmap of the S&T required to achieve the targets is generated, in parallel with the associated (ideal) investment strategy. This strategy consists of the investment allocation, and the rationale that supports the allocation.

The investment strategy can also be viewed as consisting of investment principles, investment allocations, and the investment rationale. Again, the actual can be compared against the ideal. What are some of these investment principles? Following are some of the investment principles used by the author in S&T investment strategy assessments:

- Is the balance among technical thrust areas appropriate?
- Is the balance among mission areas appropriate?
- Is the balance among funding categories (basic research, applied research, technology development) appropriate?
- Is the balance between discretionary and non-discretionary funding appropriate?
- Is the balance between 'technology push' and 'requirements pull' appropriate?
- Is the balance between revolutionary and evolutionary research appropriate?
- Is the balance between technology advancement and demonstration appropriate?
- Is the balance between high risk and low risk research appropriate?
- Is the balance among short term, intermediate term, and long term research appropriate?
- Is the balance between new projects and continuing projects appropriate?
- Is the balance among performers (university/ government/ industry) appropriate?
- Is the balance between individual research and joint projects (multi-department, multi-agency, multi-national, and government-industry) appropriate?
- Is the balance among single discipline, multiple discipline, and interdisciplinary research appropriate?
- Is the balance between large and small projects appropriate?
- Is the balance among research products (hardware, software, patents, presentations, reports, peer-reviewed journal papers) appropriate?

Obviously, additional investment principles are possible, depending on the specific review objectives, and review management interests. Once the desired S&T direction has been established, then the existing S&T program investment strategy is compared against the ideal investment strategy. Deviations of the existing from the ideal are noted, discussed, and corrective actions are taken, including personnel and budgetary.

Text mining could be used to support identification of the capabilities needed to implement the vision, development of the roadmap components, assessment of how well the investment principles are being followed, and evaluation of how the actual investment allocations compare with the desired investment allocations.

Support Identification of Capabilities
This would use the techniques employed, for example, in an Aircraft Investment Strategy study (Kostoff et al, 2000b). The evaluators would gather a number of different requirements documents, perform phrase frequency and proximity analyses, and identify technical capabilities to be pursued. The evaluators would then add planning documents, perform similar analyses, and identify enabling technologies for those capabilities.

Develop Roadmap Components
A roadmap is a network of technologies linked over space and time, aimed at achieving specific goals (Kostoff and Schaller, 2001c). A prospective roadmap is a network of science and technology areas to be developed in order to achieve the goals. Key issues in roadmap development center about whether all the blocks have been identified, all the linkages have been identified, and how accurate are the linkage strength quantifications.

Block identification comprehensiveness is a measure of how well the roadmap developers understand the mixture of technologies required to produce the higher level capabilities, and are aware of global S&T development. Advanced information retrieval, and associated clustering, can provide the mixture of technologies required to achieve the desired capabilities. Advanced information retrieval can certainly identify relevant S&T being developed globally.

Linkage identification is a measure of how well the roadmap developers understand the relationships among the roadmap technologies. Proximity and co-occurrence analyses,
performed on a database of technology narratives, should be able to provide the connections. Linkage strength quantifications measure how well the roadmap developers understand the strength of the relationships. Again, phrase proximity analyses, which provide the co-occurrence frequencies of specific phrases (number of times phrase pairs co-occur in the same linguistic domain- e.g., paper Abstract), should be able to estimate these relationship strengths.

Assess Adherence to Investment Principles
The combination of clustering and bibliometrics would address the relationship between the actual investment allocations, and the ideal. Clustering groups documents (or words/phrases) into categories, and if the core documents have associated attributes (funding, performers, institutions), then the weighted attributes in each category can be determined. Bibliometrics tend to count semi-structured data (authors, institutions, journals, countries) in each category.

Compare Actual Investment Allocations with Desired Investment Allocations
This is similar to what was done in the Aircraft investment strategy paper, although the far more powerful document clustering techniques developed recently (Kostoff et al, 2004d) could be used.

Is the S&T program doing the job right?
In order of precedence, this is the second issue to address. It focuses on the accuracy and efficiency of achieving the specified technical target. It evaluates the mechanics of the S&T development approach, and incorporates the cost, performance, schedule, and risk aspects of the mechanics. Most reviews concentrate on this component. Text mining could examine all the high frequency phrases, and all the cluster categories/themes. Then, judgments would be made as to balance (e.g., too much theory relative to experiment, insufficient North American contributions, etc). Examination of cluster themes and technical phrases has been done in almost every one of the author’s technology text mining studies, has been validated with world-class experts in those disciplines, and has been shown to be a remarkably accurate indicator of deficiencies in specific technologies.

Is the S&T program performing?
There are three components of performance: productivity, impact, and progress. Here, text mining can be very helpful, depending on the metrics selected. Bibliometrics can provide information relative to publications, patents, and citations, where the publications and patents are productivity metrics, and the citations are impact metrics. Citation Mining, a combination of text mining and citation analysis, can provide impacts and audience accessed for a research unit. Progress, in the present context, addresses how well a program is meeting its technology readiness levels, or milestone targets. The relation of text mining to progress assessment is untested and not clear at this point.

NSB DEPARTMENT REVIEW REPORTS
- 1999 Assessment of the Office of Naval Research’s Air and Weapons Technology Program
- 2000 Assessment of the Office of Naval Research’s Marine Corps Science and Technology Program
• 2001 Assessment of the Office of Naval Research’s Aircraft Technology Program
• 2002 Assessment of the Office of Naval Research’s Surface Weapons Technology Program
• 2003 Assessment of the Office of Naval Research’s Marine Corps Science and Technology Program
VI-G. DETAILED PEER REVIEW PROTOCOL

FIVE PHASES OF S&T PROGRAM PEER REVIEW

The S&T program peer review process can be divided chronologically into five somewhat independent phases. These are:

1. Initiation of the review
2. Establishing the foundations for the review
3. Preparing for the review
4. Conducting the review
5. Post-review actions

The following steps and considerations for each phase are recommended.

1. Initiation of the review
   A successful S&T program peer review requires full participation by the unit undergoing review. Recalcitrance by the reviewee(s) can result in unacceptable delays, lack of necessary background information, and poor presentations. These deficiencies will hamper the review process and affect the quality of review results.

   With few exceptions, no one likes or wants to be reviewed. How, then, can the unit undergoing peer review be motivated sufficiently to participate fully, and insure that the best review product will result? The author's experience from observing many different federal agencies' review processes is that motivation and participation derive from the actions of an organization's senior management at the initiation of the process. The management needs to communicate to the reviewees that they will be rewarded by appropriate participation and compliance in the review process, and penalized for non-compliance. Management needs to further communicate that critical judgments will be protected and handled with care. It is of the utmost importance that senior management send out an initial letter to all participants stating the following:

   • The purpose of the review and its importance to the organization.
   • The review's contribution to the larger agency GPRA response.
   • The goals, objectives, and scope of the review.
   • The identity and responsibilities of the review manager(s), the general responsibilities of the reviewees, and the responsibilities and reporting chain of the reviewers through all phases of the review process.
   • The reviewees' performance both during the review development process and in the actual review will be part of their performance evaluation.
   • The review manager will provide the input for the reviewees' performance during the review development process.

2. Establishing the foundations for the review:
   Once the responsibilities have been assigned by the senior management, the principles that
govern the review must be established. The review manager (ideally one person and not a committee) initiates this segment of the review by sending a letter to the senior management containing a detailed plan of how the total review process will be conducted. This letter is sent after extensive consultation on all review process aspects with the execution manager(s) of the unit(s) to be reviewed. Once this plan has been approved by senior management, the review manager sends a letter to the reviewees and all related support personnel, stating the following:

- The detailed objectives of the review.
- The process/approach to be followed in developing and conducting the review, including the evaluation criteria and the proposed disposition of the review report.
- A milestone schedule for completing all elements of the total review process, and
- assignment of personal responsibilities for completing each milestone.

The foundation elements to be discussed in detail in the plan, and in summary form in the reviewee letter, include the following items:

2.1. Identification of the boundaries of the program to be reviewed.
2.2. Establishment of a taxonomy that categorizes the program elements and defines the components by which the program will be reviewed.
2.3. Determination of the smallest unit (project, program) to be reviewed.
2.4. Identification of the evaluation criteria to be used.
2.5. Specification of the type of review group to be used (individual reviewer, fully independent panel).
2.6. Description of the different types of capabilities required by the review group (technical, managerial, application).
2.7. Identification of the types of attendees desired for the audience.

Considerations for each of these elements follow.

2.1. Identification of program boundaries
Identifying the scope of the program to be reviewed provides a framework for the remainder of the review. If the scope is defined too broadly (e.g., multiple partially-related projects/programs), then the review becomes very diffuse. This has consequences on the size and diversity of the panel required for a credible review. If the scope is defined too narrowly, the larger context and intrinsic integration and coordination with related projects may not be obvious. Unless there exist hard bureaucratic boundaries and requirements that automatically set the review's scope, the scope definition phase should be iterated to achieve a balance between dilute focus and incomplete context.

2.2. Establishment of program taxonomy
The guiding principle for review options is that evaluation should occur along the same structures and taxonomies by which the S&T is planned and executed. If the agency has a separate S&T unit, then the technical area should be evaluated as an integrated whole. If research is vertically integrated with development, with concurrent planning and execution, then
the research should be evaluated as part of a total vertical structure R&D review. A key conclusion to be drawn from this paragraph is that S&T evaluation recommendations must take into account how S&T is structured, integrated, and managed within an agency.

Establishing a taxonomy that represents the intrinsic nature of the program technically is analogous to selecting a mathematical coordinate system for solving a specific problem. Often, the ease of solving a particular technical problem, and sometimes the feasibility of solution, is highly dependent on selecting an appropriate coordinate system for the structure in question. This analogy holds for a program review as well. As in the mathematical system, the taxonomy selected should be orthogonal. This allows crisp presentations, each with a sharp focus, and minimal redundancy and overlap. Further, if the taxonomy contains too many categories, the review will be lengthened unnecessarily, and the program elements will appear to be discrete and fragmented. If the taxonomy has too few categories, it becomes very difficult to identify experts who can speak credibly for each component. Thus, a balance is required between selecting the appropriate number of review elements and ensuring that the review taxonomy remains aligned with the taxonomy used for program planning and execution. It has been the author's experience that time spent on the taxonomy definition phase results in time saved and problems eliminated downstream.

2.3. Determination of smallest review unit

Fiscally, an S&T, or research, program is a collection of funded S&T, or research, components. These elements could be subprograms, projects, or individual work units such as principal investigators (PIs). Conceptually, a program is greater than the sum of its components. A program includes the intelligence or inherent logic that links the components to each other and to the program's overall objectives. Thus, the intrinsic quality of an S&T or research program is not merely the sum of the qualities of the component projects, it depends on the quality of the structural relationships among and between the projects, as well as on the broader mission objectives.

Review of an S&T program can then be viewed as consisting of two elements:

2.3.1. "review of S&T projects," which examines the nature of the component projects, and is commonly referenced as an in-depth technical review; and
2.3.2. "review of an S&T program," which examines the nature of structural relationships among and between the projects and the mission objectives, and the relationships between the projects and the external environment.

This type of review is commonly referenced as a management review. These two elements, 2.3.1. and 2.3.2., can be merged operationally into a single review, or could be performed separately.

If review time were not a consideration, elements 2.3.1. and 2.3.2. would be recommended in total. This combination review would provide both depth and breadth necessary for a full understanding of program quality. In reality, review time is limited and it is desirable to have the
same group of reviewers present for the total review of the areas in which they have expertise. This allows normalization and continuity to occur during the review action. However, in the case of a program review, the larger the program, the more review time it will require. It becomes more difficult to retain high quality reviewers as the length of the review increases.

There are at least three approaches to circumvent this problem. First, the program could be broken into focused subprograms, and each subprogram could be reviewed separately with more focused experts. Second, the program could have its components aggregated, and the full program could be reviewed by the same panel at a lower level of detail. Third, the quality and relevance components could be divided for separate reviews. While all the above options are theoretically possible, some compromise in quantity and type of material presented is necessary to insure that the same group of reviewers is presented with, and can evaluate, the totality of program material.

The author's experience and recommendations for GPRA are that a hybrid of elements 2.3.1. and 2.3.2. be presented. Since a program is being evaluated, it is important that the reviewers understand the total program's objectives, both in isolation and in the context of the larger organizational unit's objectives. It is equally important that the reviewers understand

- how the component projects relate to each other and the mission objectives,
- how they are integrated within the program and within the larger organizational unit, and
- how they are coordinated with the external environment.

At the same time, the reviewers should have substantial evidence that high quality S&T is being performed within the program. Thus, the review would center around the structural relations emphasis of element 2.3.2, with copious examples of technical progress and output and impact woven in the presentations where applicable. Not all technical details are required. Nevertheless, enough examples of positive accomplishments are necessary to convince reviewers of the effectiveness of the program. Because of the output/outcome/impact emphasis of GPRA, program reviews performed to partially satisfy GPRA requirements should focus on the S&T products and their potential or actual consequences.

2.4. Identification of evaluation criteria
Identification and selection of evaluation criteria should be driven primarily by the mission and review objectives, as well as the nature of material being reviewed. In the specific case of selecting evaluation criteria for peer reviews performed to address GPRA requirements, additional consideration must be given to selecting criteria of interest to the review client, as well as to the eventual disposition and utilization of the criteria ratings. If promoting the highest quality S&T to the relative exclusion of other objectives is the main program objective, then the evaluation criteria should focus on S&T quality. If accelerating transitions from research to development to demonstration is the prime program consideration, with S&T quality a secondary program objective, then the evaluation criteria should include both transitions and S&T quality, with greater weight given to transitions. If other program objectives are the main focus, such as integrating disadvantaged groups into the sponsored programs, then the criteria should included
these goals and they should receive greater weight. In terms of the review mechanics, fewer
criteria should be specified whenever possible. While it may be easier to analyze reviewer
responses when many criteria are used, it forces the reviewers to fragment and channel their
thinking and writing. The author has found that some of the most useful and coherent inputs are
generated when the reviewers are allowed to provide comments in unstructured narrative form.

Reviews conducted by the author have allowed for a hybrid of both structured and unstructured
types of inputs. For a research program, the fundamental evaluation criteria are:

- research quality,
- research relevance, and
- overall program quality.

The evaluation criteria recommended for a basic research review are addressed in the Executive
Summary in the appendix. The criteria presented in the appendix resulted from separating
research quality into its major components:

- research merit,
- research approach, and
- team quality.

For some evaluations, as shown in the full paper (Kostoff, 1997c), the fundamental evaluation
criteria have been further subdivided into:

- research merit,
- research approach/plan/focus/coordination,
- match between resources and objectives,
- quality of research performers,
- probability of achieving research objectives,
- program productivity,
- potential impact on mission needs (research/technology/operations),
- probability of achieving potential impact on mission needs,
- potential for transition or utility, and
- overall program evaluation.

The full paper (Kostoff, 1997c) also presents sample evaluation criteria for more technology-
oriented programs. Along these lines, a 2001 paper describes the review of an advanced
technology development program in more detail (Kostoff et al, 2001b). If management or other
non-technical issues are to be evaluated as part of the program review, then the evaluation criteria
should be modified accordingly. Finally, the presenters should receive a copy of the evaluation
criteria at the earliest stages, so that they can begin to craft their presentations to focus on
addressing the criteria.

2.5. Review group type
Selection of the type of review group is a core issue, and should be addressed at the initiation of the review process. While many types of groups are possible, two will be discussed here. They are the independent panel (2.5.1) and the external reviewers group (2.5.2).

2.5.1. Independent panel.
The independent panel is a group of experts independent of the agency, and typically funded under a contract. The independent panel has a chairperson, attempts to reach consensus on issues, and generates a written report containing the results of the review and sometimes recommendations.

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

Each of the two review group approaches has value for specific applications. The group of external reviewers is less formal, and has fewer reviewer and audience restrictions. It is useful for internal reviews where structural program issues are paramount and need resolution or improvement, and where comparison with other programs is not the major focus. The independent panel is more formal. The independent reviewer panel has more specific reviewer, meeting, and audience selection constraints/requirements. If the panel is run under the auspices of one of the National Academy of Sciences boards, for example, there will be a more elaborate process used to select participants and review the final written product. From the agency's perspective, either group has very high utility for addressing the agency's program improvement needs. From a perspective external to the agency, the independent panel has higher credibility because of its independent nature. For GPRA application, the independent panel is more appropriate, because of its perceived independence.

However, operation of an independent panel under GPRA will be intrinsically different from past operation of this type of panel. If GPRA is viewed as a budgetary instrument with a potential for modifying resources (Brown, 1996), some additional factors must be considered in structuring and operating the two types of panels discussed. Since different types of panels may be used for different technical areas and different agencies, some means of normalizing review results across areas and agencies will be required. Also, because of the potential for errors or bias, some means of rebuttal or reclama must be provided for conclusions and recommendations produced by different panel types. Both these issues are summarized below.
2.5.3. Review report normalization
The author has not seen any fully satisfactory peer review normalization approaches due to the presence of many non-separable variables. However, one interesting normalization approach is used by the Dutch Technological Foundation for evaluating research proposals (Van den Beemt, 1991, 1997). Technical comments, but not quality ratings, are provided by technical peers. The comments and proposer responses for twenty different proposals are then provided to twelve people from a variety of disciplines. This "jury" of twelve provides the scores through an independent mail review. Essentially, the normalization is provided by having the twelve jurors common to all proposals.

The author has used two approaches to improve normalization across panels somewhat. First is the utilization of some individuals common to all panels. In a series of competitions for new accelerated research programs that was held in the late 1980s (Kostoff, 1988), the author served as de facto chairperson of all the different discipline panels. This resulted in some small measure of normalization among the different panels. Use of more individuals common to all panels would have provided an extra measure of normalization, and in this sense the presence of senior management during the reviews provided additional measures of normalization.

Obviously, the more closely the panels are related topically, the more valuable is the technical contribution of individuals common to the different panels. Secondly, in the above competitions, it was assumed that the difference in aggregated average scores for major disciplines (e.g., physical sciences and life sciences) was due to two factors: differences in intrinsic quality of the programs proposed and differences in the scoring severity of the reviewers. To normalize, a fraction of the differences in aggregated average scores for the major disciplines was removed. This was assumed to eliminate the scoring severity difference. Trial and error showed a fifty percent correction factor provided results that appeared reasonable to the audience members who had attended all the reviews. This normalization procedure had the added benefit of preserving and insuring representation from disciplines that had strategic value to the organization. This approach to normalization could have a second interpretation. If the research is viewed as having a strategic component and a quality component, with the reviewers' scores viewed as addressing the quality component only, the correction could be perceived as adjusting for the presence of the strategic component.

For example, assume a life sciences panel produced an average program score of five, and an engineering sciences panel produced an average score of ten. Assume further that each discipline had equal strategic value to the organization and that the strategic value (STRAT) was perceived by the organization to be of equal importance to the reviewers' scores (SCORE-assumed to be a total program quality score that includes mission relevance). Then the normalized total score (FOM) can be computed as $FOM = 0.5*STRAT + 0.5*SCORE$, and the difference between the two panels' scores would be reduced from five to 2.5. This correction factor can then be applied to the raw score of each program within the discipline to arrive at a final "normalized" score.

2.5.4. Rebuttal of review panel recommendations
In a 1997 paper (Armstrong, 1997), different studies of errors and superficial work by peer
reviewers of journal manuscripts are described. The conclusion one draws from these results is that the problem of manuscript reviewer error production is not insignificant. In most research program peer reviews, commission of technical errors by reviewers due to the relaxed standards resulting from anonymity and lack of financial incentives is probably not nearly as serious as in manuscript reviews. In the author's experience, panel members tend to suppress overt expressions of biases, and they typically make statements they are able to defend. Studies of the extent of errors, or bias, committed by research program peer reviewers remain to be done. If these panels eventually have substantial input to the budgetary process under GPRA, an appeals system for program reviews may have to be established to resolve errors or perceived biases.

2.6. Specification of review group capabilities required
Even with the strongest support from an organization's top management, and the direction of an unbiased and competent review leader, the quality of a review will never go beyond the competence of the reviewers. Two dimensions of competence that should be considered for a program peer review are the individual reviewer's technical competence for the subject area, and the competence of the review group as a body to cover the different facets of S&T issues (research impacts, technology and mission considerations and impacts, infrastructure, political and social impacts). The quality of a review is limited by the biases and conflicts of the reviewers. The biases and conflicts of the reviewers selected should be known as well as possible to the leader and among the reviewers themselves.

One common error in panel selection is limiting the choice of S&T experts to those who have specific expertise in the subdisciplines of the existing program. This provides an answer to the question of whether the job is being done right, but not whether the right job is being done. The former question relates to detailed technical quality, while the latter relates more to investment strategy in the broadest sense (investment strategy is the rationale for the prioritization and allocation of resources among the program components.). To answer the latter question, people with broad expertise in the area covered by the overall program's highest level objectives should also be selected. They will be able to address the investment strategy more objectively, and determine whether the mix of subdisciplines and the allocation of resources among the subdisciplines is appropriate. The review group, then, would be able to address the central question of whether the right job is being done right.

One of the major criticisms of peer review, whether manuscript, proposal, or program, is that it tends to perpetuate orthodox and conservative paradigms, and tends to reject new paradigms that threaten the structure of the status quo. If one of the objectives of an S&T program peer review is in fact to ensure that innovation is recognized, that truly revolutionary research with attendant new paradigms will be promoted and rewarded, then the selection of reviewers to address the right job issue in parallel with reviewers to address the job right issue becomes of paramount importance.

In summary, a review panel should have at least the following characteristics:

- Each member should be highly competent in the facet of the program for which he/she has
been selected; this assures the presence of sufficient depth on the panel.
- The panel as a body should have sufficient competence to cover all major facets of the program being reviewed; this assures the presence of sufficient breadth on the panel.
- Each member should be minimally conflicted with the program under review, and any conflicts or biases should be known to all the panel members before the review; this assures the presence of independence and objectivity on the panel.
- Each member should agree to read all background material, attend all sessions, and protect any classified and proprietary information that surfaces during the review; this assures the presence of preparedness and security on the panel.

2.7. Identification of audience types
A program review provides an excellent forum for disseminating program information and results to a wide audience. In addition, a program review is a useful mechanism for providing coordination with intra- and inter-organization related programs. Care should be taken to insure that the review audience includes:

- actual and potential customers,
- stakeholders and other oversight groups,
- co-sponsors,
- users, and
- other agency representatives.

Judicious use of the many databases that are now accessible, and algorithms that expand the identification of potentially related technical areas and their contact points (Kostoff, 1997e, 1999b, 2000a, 2001c, 2001d, 2003a, 2003b, 2003c, 2003d) can help develop a broadly-based audience for maximum impact.

3. Preparing for the review
The schedule and milestones originally submitted to senior management to obtain approval for initiating the review should be further detailed. A tracking system for schedule progress should be initiated and periodic status reports sent to senior management. The author has found weekly status reports to be adequate.

3.1. Developing the agenda
Once the taxonomy has been developed, the structural elements of the agenda can be easily identified. The main elements include:

- an introduction by the review manager to identify the goals of the review, set the stage for the remainder of the review, and handle any administrative issues;
- an overview by the program manager of:
  - the role of the program in its larger context,
  - the vision of the operational scenario to which the program will contribute,
  - the requirements necessary for the vision to be achieved,
• the technical capabilities defined by the requirements and the S&T necessary to produce the capabilities,
• promising S&T opportunities that could result in capabilities not yet defined by requirements,
• the overall investment strategy that links the above components to each other and to the external environment and will allow the capabilities to be obtained, and
• the detailed technical presentations to follow.

• detailed technical presentations and, if these are held at a laboratory, tours could be included in this segment;
• question and answer time allocated to each presentation;
• written evaluation periods after each presentation;
• an executive discussion period at the end of each day; and
• administrative break periods (coffee, lunch, etc.).

3.2. Developing the presentations

3.2.1. Assignment of responsibilities
The presentation development phase begins by assigning the responsibility for the presentations to the program manager. The program manager is sent a letter detailing these responsibilities, identifying:

• overall time available on the agenda for presentations,
• fraction of presentation time reserved for questions and answers,
• taxonomy to be used for evaluating the program, and
• criteria by which the program will be evaluated.

The program manager then has to decide:

• the amount of time to be devoted to addressing each taxonomy category,
• how to address the category, and
• who should make the presentations for each category.

There is a wide range of combinations of potential presenters for the total program being reviewed. At one extreme, the total program presentation could be made by the program manager alone. At the other extreme, each taxonomy category could be presented by selected PIs (the performers). The level of presenter selected depends on the objectives, type, and location of the review. For a GPRA-type program review conducted at a sponsor's headquarters, the author's preference would be to have as few different presenters as is feasible. Each presenter should be as high in the program management chain as possible while still having an acceptable grasp of the technical material. This allows the program integration message to be communicated to the audience most effectively. For a smaller program review conducted at a laboratory, in which tours of the working environment may be incorporated, PI-level presentations could be included.
3.2.2. Reducing presentation problems
The reasoning behind recommending that presenters be relatively high in the program management chain is the following. For the large federal S&T sponsoring agencies with which the author is familiar, technical competence of the performers is not a major issue or problem. The number of proposals to these agencies far exceeds the funding available, and with the use of in-house and external experts to provide advice in proposal selection, typically only the 'cream-of-the-crop' is selected. Reviews in which the author has participated that focus mainly on technical quality at the PI level invariably arrive at the conclusion that the technical work is of high quality. This conclusion appears almost invariant of the agency or type of panel or reviewer selection process employed. If a problem is surfaced, it tends to focus on the following issues of integration and coordination:

- Are the different projects coordinated with each other and with other agency projects?
- Do they form a cohesive program or are they a collection of isolated and fragmented efforts?
- Are the projects coordinated/jointly planned/jointly managed with external organizations and is the total program coordinated in this way with the external community?

The actual S&T performers tend to focus on the technical details, and the coordination and integration issues are best addressed by those somewhat removed from the actual performance of the tasks.

Another presentation problem that appears to emerge in every agency presentation the author has attended overlaps somewhat with the technical detail/coordination issue described above. The problem stems from the training and characteristics of many S&T performers. Technical personnel are trained to pay careful attention to details, and very good technical people seem to have an innate interest and predilection for details. While some technical presentation skills are included in technical training, they typically constitute a small portion of that training. Consequently, many program level presentations remain immersed in technical details and tend to be far too long. While this level of presentation is most comfortable for the technical specialist making the presentation, it acts to the detriment of presenting the program in its larger context. In addition, because of the concentration on details, the main message tends to become diluted and diffuse and overwhelmed by material extraneous to the main message. It is very important that the main message to be delivered be kept in focus at all times when structuring the presentations. More specifically, the presentations should be kept short and the number of view graphs should be few. Every line (and word) on each view graph should contribute to the central message that the presenter wants to communicate. If it does not, it should be removed. The producers of TV commercials have learned this lesson well. Unfortunately, these fundamental communication principles and techniques have not found their way to many technical program presenters.

3.2.3. Presentation content

3.2.3.1. Outline of presentations
In alignment with the agenda outline, the detailed contents of the specific presentations should
incorporate the following. There should be an overview showing how the larger management unit (division, department, etc.) in which the programs are housed integrates into the total organization, and how the management unit's objectives relate to those of the larger organization. Then, the investment strategy of the larger management unit should be presented in detail. The investment strategy presentation should include the:

- relative program priorities,
- actual investment allocation to the different programs, and
- rationale for the investment allocation.

Finally, for each program presentation, the investment strategy for its thrust areas should be presented. The investment strategy is perhaps the most crucial part of a program review, and deserves further discussion here.

Investment is the allocation of resources among the program components. Investment strategy is the rationale for the prioritization and allocation of resources among the program components. The optimal investment strategy for a program is the specific allocation and rationale that will produce the most mission relevant high quality S&T for impacting the program's objectives. This will depend on the viewpoint of the assessor and, in particular, how the assessor limits the role of the S&T within the national perspective.

The optimal investment strategy should be a focal point of an assessment. The optimal investment strategy results from a timely confluence of:

- S&T requirements (top-down driven) and
- promising S&T opportunities (bottom-up driven).

Further, promising S&T opportunities result from a timely confluence of advances in:

- theory,
- instrumentation,
- new experiments,
- new algorithms, and
- computers.

Finally, S&T requirements result from a timely confluence of:

- domestic and foreign,
- political and economic, and
- strategic and tactical advances.

All of the above factors should be included in a presentation of the investment strategy.

3.2.3.2. Specific presentation content
The senior management presentation.
To initiate the actual review, a senior agency manager provides a short introduction describing structure and mission of the agency, and a more detailed description of the purpose and goals of the program review. Senior management describes what is expected from the reviewers, and how their comments have been, and will be, utilized.

The review manager presentation
The review manager provides the details of the organization's structure, the types of reviews within the agency, and the integration of the present review with the other reviews and with the total organization's management processes. The review manager also describes the steps of the specific evaluation process, including the meeting agenda, and presents all the administrative details and procedures to be followed.

Organizational unit head presentation
The broader technical portion of the presentations is initiated by the head of the organizational unit in which the program resides, and it includes the following informational material:

- The mission and objectives of organizational unit,
- a list of all programs in organizational unit,
- a description of objectives of each program,
- the funds and people associated with each program and with the program to be reviewed,
- an overview of the accomplishments and transitions of programs not being reviewed, and their relation to the accomplishments and transitions of the organizational unit's mission and potential national impact, and
- responses to actions taken as a result of the previous year's reviews of the organizational unit's programs

Program manager presentation
The program manager(s) then provides a more detailed overview of the program under review, including:

- objectives of program under review.
- requirements to be met and derived target capabilities for the S&T initiative (For example, in the review of a military-oriented program, what is the present and evolving threat? Identify documented sources, personal contact sources, etc.? What is the importance of the threat and what are the capabilities required to overcome the threat?).
- investment strategy.
- list of targeted thrust areas selected to meet program requirements (e.g., propulsion, aerodynamics, G&C) and sub-thrusts (e.g., energetic propellants, combustion instability, propellant safety).
- objectives of each thrust that will include:
  - thrust and sub-thrust funding and prioritization,
● rationale for thrust and sub-thrust selection and prioritization (including the bases for rationale and prioritization such as system studies, workshops, assessments, intuition, Congressional and other mandates, etc.),

● integration of thrusts and sub-thrusts to form overall program coordination/roadmaps (Road maps are graphical displays of the inter-connectivity among diverse S&T projects and potential applications. They describe the past, present, and future of the program, and its linkage to other internal and external programs, as well as linkage to institutional capabilities and requirements. They offer a convenient focal point for discussing complementary and related programs sponsored by other external organizations.),

● team quality (identify S&T performers), and

● a summary of major accomplishments, transitions, milestones met.

The technical manager presentation.
The technical managers who support the program manager will present the following:

● Objectives of each sub-thrust
● Technical roadblocks to achieving the sub-thrust objectives
● Technical approach for overcoming the sub-thrust roadblocks
● Potential sub-thrust payoffs and capability enhancements
● Technical results achieved

3.2.4. Dry runs
After the presentations have been developed and reviewed within the performer organizations, there should be at least two series of "dry runs" before the review manager. If possible, senior management should be in attendance as well. The dry run presentations should be polished from the presenter viewpoint, and the main purpose is to assure that all the separate taxonomy category presentations appear cohesive and integrated. The dry runs are not forums in which diplomacy and tact, and the preservation of fragile egos, are paramount. One key objective is that all questions and issues and weak points that could arise in the final presentations are surfaced and discussed in the dry runs. The earlier such issues are resolved, or at least recognized, the better for all participants.

3.3. Selecting and inviting the reviewers
Selection of an optimal review panel is more of an art than a science, and depends on:

● the selector's understanding of the many facets of the program being reviewed,
● his/her understanding of the experts available in the technical community, and
● his/her ability to predict the interaction dynamics of a particular group of experts.

Presently, different federal agency approaches in panel selection range from assembling program manager recommendations as potential reviewers to using an iterative co-nomination approach for reviewer identification and selection. Since the latter approach, properly done, is relatively objective to the program being reviewed, it will be the focus of this discussion.
In essence, the iterative co-nomination approach is a multi-step process that starts with an input list of recommended experts and results in a list of experts who have been multiply nominated by different experts. Once the overall technical description of the program is generated, and technical descriptions of the taxonomy categories (technical sub-areas) are provided, reviewer identification can be initiated. Sources of candidate reviewers can include:

- program manager recommendations,
- membership lists of prestigious organizations such as the National Academies of Science and Engineering and the Institute of Medicine,
- agency review boards,
- agency consultant pools,
- contributors to technical databases (such as journal article authors or technical report authors), and
- other similar lists.

Multiple names are chosen to cover:

- each sub-discipline,
- the program as a whole,
- allied research disciplines,
- the technologies, systems, and operations that the program does or could potentially impact,
- and
- other elements of the customer, stakeholder, user, and impactee communities.

This list of names is called level 1, or the initial list. Each member of level 1 is asked to identify, or nominate, other experts in his/her particular area of expertise to generate the level 2 list. For example, assume that a physics program is being assessed. Assume further that this program has three subdisciplines: plasma physics, atomic physics, and molecular physics. The level 1 list may have two names for each one of the subdisciplines. To obtain the level 2 list for the plasma physics research area of expertise, each of the two plasma physics recommendees of level 1 would be asked to recommend two experts in plasma physics. If names appear more than once in the level 2 list, or between the level 1 and level 2 lists (multiply recommended individuals), then these individuals are assumed to be the leading experts in the fields to be assessed. If no multiple recommendations appear, then the experts in level 2 are asked to recommend two experts in plasma physics for level 3, and the co-nomination search is repeated. Convergence occurs when an adequate number of experts have been co-nominated. While this process may at first seem complex and open-ended, convergence is rapid because of the relatively small number of real experts in any well-defined technical discipline.

A primary and alternate list of co-nominees should be matrixed against selection requirements and criteria, where the matrix elements represent the reviewer's expertise in the different facets being examined. This matrix should be distributed to the program managers and performers who will be reviewed, and comments related to bias and conflict solicited. If strong objections can be supported against one or more nominees, the list could be modified. Some additional constraints
should be placed on the list of reviewer candidates. Because the iterative co-nomination approach focuses on identifying recognized experts in a field, there is always the danger of excluding younger reviewers of high caliber with fresh perspectives on the topical area. Therefore, the co-nomination approach has to be tempered with other selection processes that allow for the recognition of lesser known experts of high quality.

In practice, the author uses a hybrid combination of reviewer sources and selection approaches to insure that a diversified portfolio of appropriate experts is represented on the review team. There needs to be a balance of continuity and turnover among reviewers. The ratio between these two considerations will be heavily dependent on review frequency. For three year period reviews, the author has tended to use about 25-33% continuity. Total number of reviewers is another important consideration. As the number of reviewers on the panel increases, more coverage of depth and breadth is possible, and the diversity of opinion on a given topic area is increased. At the same time, the cost of conducting the review increases, and the logistics of controlling the panel increases. The author has found that a range of panel sizes from about eight to fourteen is desirable, with the actual size depending on the range of material covered by the review. Once the list has been finalized incorporating the above considerations and constraints, potential candidates are contacted by phone. If there are no conflicts-of-interest, invitations are then extended, preferably at least three months in advance of the review date.

3.4. Selecting and inviting the audience
As stated earlier, care should be taken to insure that the review audience includes actual and potential customers, stakeholders and other oversight groups, co-sponsors, users, impactees, and other agency representatives. The invitation may come from the program manager(s). Databases, however, can help in the identification of other participants. Depending on how the GPRA reviews are conducted, especially who is conducting them and where they are being conducted, announcements to the general public may be advertised. While a large audience in a review room may serve to restrict discussion, with the present-day ease of establishing video transmissions, separate rooms can be reserved for general public audiences remote from the review room. Once the desired audience has been identified, invitations should be sent at least three months in advance of the review. This substantial advance notice will insure that the busy schedules of high caliber attendees can accommodate the review. The invitation package should include many of the elements sent to the reviewers, including the background material.

3.5. Selecting and distributing background material
It is strongly recommended that a variety of background material be supplied to the reviewers (and the invited audience) before the review. This should include:

- material focused strictly on the internal program under review,
- material focused on related external programs, and
- material that shows how the totality of these internal and external programs are inter-related and coordinated.

The internal program material should include:
• organizational descriptive material,
• narrative descriptions of each program to be reviewed, and
• descriptive material of each work unit in the program.

It would also prove useful to include bibliometric output indicators for each program, with interpretive analytical material. This could include refereed papers, patents, awards and honors, presentations, etc.

Specifically, internal program background material should include the following administrative and technical information:

• Structural chart of the agency showing how the organization under review fits into agency structure.
• Structural chart of organization, showing programs (including funding) and personnel (including background and expertise) associated with each program.
• Definitions of different generic types of programs that will be presented during the review.
• Administrative material (agenda, reimbursement, conflict-of-interest forms, proprietary protection forms, etc.).
• Two page overview of each program being reviewed in detail (e.g., weapons technology), including:
  • program objective,
  • program thrusts (e.g., aerodynamics, ordnance, guidance and control, etc.),
  • investment allocation among thrusts (three year trends),
  • milestones where appropriate, and
  • progress made toward achieving these milestones.

• Two page overview of each program thrust, including:
  • thrust objective,
  • short descriptions of each technical sub-thrust (e.g., energetic propellants, combustion instability, propellant safety) pursued under the thrust, as well as
  • investment allocations among sub-thrusts.

Total program and thrust descriptive material should not exceed twenty pages. It would be useful to include narrative material on related external programs in other agencies and industry, including descriptions of papers and other output material from these programs, as well as narrative descriptions of ongoing programs. Choice of material sent to reviewers should be very selective, since an excessive amount will go unread. However, it would be useful to include hindsight-type results of research that was funded years ago in the technical area under review, and which recently have come to fruition in a system or commercial technology.
It would also be valuable if roadmaps (Kostoff, 1997d, 2001a) were provided as background material (i.e., visual depictions of the structural relationships among the program components and the mission objectives). These roadmaps provide the global context in which the program is being performed. Retrospective roadmap components depict the program manager's awareness of the breadth and depth of the intellectual heritage of the program being reviewed. Present roadmap components reflect the program manager's awareness of the wide range of S&T areas available to complement his/her program, and the degree of coordination and leveraging in which the program is involved. Prospective roadmap components provide indication of the program manager's vision and willingness to take risks, and his/her intrinsic understanding of how results from other S&T programs could be exploited to enhance and expand the potential of the program. A certain amount of time and reflection is required on the part of the reviewer to understand and to fully appreciate the implications of a well-prepared, comprehensive roadmap. As a result, roadmaps should be sent to reviewers well in advance of the actual review date.

4. Conducting the review

Once the reviewers are assembled, they should be provided with a document containing hard copies of the viewgraphs to be presented, as well as documented evidence of program accomplishments. These accomplishments should include bibliometric information (papers and reports published, conference proceedings, books, awards, etc.), and write-ups of significant accomplishments. Each accomplishment write-up should describe:

- the actual scientific or technological accomplishment,
- what impact it has had, or will have, on
  - other science or technology initiatives,
  - the agency and its national mission, and
  - the performer and performing organization.

The presentations should then occur in the sequence described in section 3.2.3.2. Briefly, a senior agency representative should welcome the reviewers and audience, and describe the purpose of the review from the agency's perspective. The review manager then provides the details of the organization's structure, the types of reviews within the agency, and the integration of the present review with the other reviews and with the total organization's management processes. The review manager also describes the detailed steps of the evaluation process, including the meeting agenda, and presents all the administrative details and procedures to be followed. The head of the organizational unit describes the mission and programs of the unit, and how the program to be reviewed integrates with the remainder of the unit. These presentations constitute the introductory material for the total audience. The program manager then describes the larger context in which the program operates, the structure and contents of the program, and the investment strategy that guides the specific program element allocations. Approximately 1/3 of the presentation period should be devoted to questions and answers.

After the program manager's presentation, time is allotted for written evaluation before proceeding to the next presenter. There is a school of thought that written evaluations should
only be performed after a group of presentations rather than after each presentation. This would allow for each presentation to be evaluated in the context of the other presentations, both relative to individual presentations and to the larger collective body of presentations. However, the author has found that an element of spontaneity and freshness is lost by not performing evaluations directly after each presentation. The integrative aspect can be incorporated into the review by allowing for some reflective time, after the day's presentations have been completed, for modifying the written comments, if desired. The executive session at day's end allows for further integration through discussion.

Each of the technical managers then describes his/her S&T sub-category within the program. Again, approximately 1/3 of the presentation time is devoted to questions and answers (Q&A). After each of these presentations, time is allotted for written evaluation before proceeding to the next presenter.

At the end of each presentation day, about one to two hours should be devoted to an executive session, in which the reviewers and review manager meet to discuss each presentation. At the end of the executive session of the final presentation day, all the written evaluation forms are collected. The importance of the verbal (and written) comments made by the discussants depends not only on their intrinsic merit, but on the context in which they are made. It is extremely valuable to have a separate technically knowledgeable observer present throughout the review, who can discuss any contextual issue with the review manager or chairman after the discussions have concluded. This allows key issues to be framed within their proper context in the final report, and allows the credibility of the report to be raised substantially among the sophisticated readers.

5. Post-review actions
After the actual review meetings have been completed, all the information must be assembled, analyzed, and reported. Then actions following the report recommendations must be taken, and the responses to those actions tracked and analyzed. The detailed steps follow.

5.1. Integrating additional comments
Any additional comments about the review, either from the reviewers, the external audience, or senior management should be considered and integrated into the review report, where appropriate. For the reviewers in particular, they have had a chance to integrate all aspects of the review and can provide a cohesive narrative of their views on the program. Either review type, independent panel or individual external reviewer, should insure that this avenue for additional information remains open, not to be arbitrarily closed for some artificial expediency.

5.2. Writing a final report
There should be two forms of the final report, a long version and a short version. The long version should include all the written material that was generated during the course of the review. It provides an archival record of exactly what was done during the review. This report version would include:
• the initial review charter,
• invitation letters,
• background material,
• completed evaluation forms with reviewer identification deleted,
• other reviewer/audience input, and
• the final report write-up.

The short version would summarize the process details, and would focus on reviewer comments and other significant inputs, conclusions, and recommendations. The final report should include the viewpoints of all the reviewers, with appropriate weightings given for judgment and expertise of specific contributors. Dissenting viewpoints should be identified. Based on the diverse inputs, the report author should specify conclusions on the health of the program, and recommendations for action in modifying the program, if required.

5.3. Assigning action items
Under GPRA, there will be at least two clients for the report, internal management, and the Federal government oversight organization. If internal management accepts the conclusions and recommendations of the report, action items should be assigned to the appropriate personnel for responding to problems identified in the report. There are many types of responses possible (e.g., a corrective action, or a rebuttal disagreeing with the conclusion and recommendations). Maximum flexibility and leeway should be given to the program manager for the initial response.

5.4. Evaluating response to action items
Each action item should have a deadline for response. After the deadline, the response should be evaluated, and appropriate follow-up action taken. These action items, responses, and follow-up actions should be presented at the introduction of the next annual review. This provides evidence to the reviewers that their input has impact on the program, and will motivate them to participate in the review process further.

BIBLIOGRAPHY FOR APPENDIX VI-G


On Management Of Technology, Miami, FL (February 17-21). (Larger Text Available From Author.)


VII. BIBLIOGRAPHY AND RELATED REFERENCES

The author expresses his appreciation to Drs. Fiona Wood, Scott Armstrong, and Drummond Rennie for their contributions to this bibliography.

Following are important peer review references, for anyone interested in obtaining more detailed information from the peer review literature. There are many thousands of articles that address the peer review process, and provide examples of peer reviews. The focus of the following references is articles that address the peer review process, although some that describe actual peer reviews may contain advances in the peer review process as well. The references are followed by a listing of the most cited references.

Listing of References for Suggested Further Reading

Abbott A. Peer review cuts power of Italy's 'barons'. Nature. 1998 Apr 9;392(6676):531.
Agnew, B. New NIH procedures to shield clinicians from grants bias SCIENTIST. 1998 MAY 11 12 10 1 +
Agres-T, "NSF Pushed to Open Up Peer-Review", SCIENTIST, 1988, Vol 2, Iss 1, pp 6-6
Aguillo, I. A new generation of tools for search, recovery and quality evaluation of World Wide Web medical resources ONLINE INFORMATION REVIEW. 2000 24 2 138 143
Ahmed, HS. Does peer review process exist in developing-country journals?: The Bangladesh experience JOURNAL OF SCHOLARLY PUBLISHING. 2003 JUL 34 4 236 248
Ahmed, HS. The current status of science journals published in Bangladesh JOURNAL OF SCHOLARLY PUBLISHING. 1998 OCT 30 1 38 46
Aisen ML. Judging the judges: keeping objectivity in peer
Aisen, ML. Judging the judges: Keeping objectivity in peer review - Guest editorial JOURNAL OF REHABILITATION RESEARCH AND DEVELOPMENT. 2002 JAN-FEB 39 1 VII VIII
Albert T. Peer review: thinking the unthinkable BMJ. 1999 Sep 25;319(7213):861.
Aldenkamp, AP, Vermeulen, J. Effects of antiepileptics drugs on cognition REVISTA DE NEUROLOGIA. 2002 MAY 1 34 9 851 856
Allan B, Morlet N, Wormald R. Journals and the Internet. Use of the Internet for on line peer review must be explored further. BMJ. 1997 May 3;314(7090):1352.
Allen, E. M. 1960 Why are research grant applications disapproved?Science. 132, 1532-1534.
Altman, E. Silencing scientists and scholars in other fields: Power, paradigm controls, peer review, and scholarly communication. JOURNAL OF ACADEMIC LIBRARIANSHIP. 1998 NOV 24 6 496 497


Amato, I. 1992 Rustum Roy: PR is a better system than peer review. Science. 258, 736.


Ament, L. Quality management activities in the obstetric triage setting JOURNAL OF NURSE-MIDWIFERY. 1999 NOV-DEC 44 6 592 599


American Chemical Society (1985). Ethical Guidelines to Publication of Chemical Research. Columbus, Ohio


Anderson, J. Silencing scientists and scholars in other fields: Power, paradigm controls, peer review, and scholarly communication JOURNAL OF INFORMATION ETHICS. 2002 SPR 11 1 83 86


ARC. 1990 Advice and instructions to applicants for 1991 research grants.


Armstrong, J. Scott, "Editorial Policies for the Publication of
Armstrong, J. Scott,, "Cheating in Management Science" (with commentary), Interfaces, 13 (August 1983), 20-29.
Armstrong, JS. Discovery and communication of important marketing findings - Evidence and proposals JOURNAL OF BUSINESS RESEARCH. 2003 JAN 56 1 69 84
Arrowsmith, J, Sisson, K, Marginson, P. What can 'benchmarking' offer the open method of co-ordination? JOURNAL OF EUROPEAN PUBLIC POLICY. 2004 APR 11 2 311 328
Asch-S Goldzweig-CL Lee-P, "Do We Understand the Effects of Managed Care in Ophthalmology - A Review and Analysis", ARCHIVES OF OPHTHALMOLOGY, 1997, Vol 115, Iss 4, pp 531-536
Aske, JR. Evidence: No medical peer review privilege in discrimination actions JOURNAL OF LAW MEDICINE & ETHICS. 2001 FAL-WIN 29 3-4 411 413
Association, A. D. E. National Institute for Dental and Craniofacial Research.
ASTEC. 1990 Setting directions for Australian research. Canberra: AGPS.
Atkinson-M, "Regulation of Science by Peer-Review", STUDIES IN HISTORY AND PHILOSOPHY OF SCIENCE, 1994, Vol 25, Iss 2, pp 147-158
AU Brown, C. The role of electronic preprints in chemical communication: Analysis of citation, usage, and acceptance in the journal literature JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY. 2003 MAR 54 5 362 371
AU van Raan, AFJ. The pandora's box of citation analysis: Measuring scientific excellence - The last evil? WEB OF KNOWLEDGE - A FESTSCHRIFT IN HONOR OF EUGENE GARFIELD
Averch, H. A. 1987 Measuring the cost-efficiency of basic
Bachand, RG, Sawallis, PP. Accuracy in the identification of scholarly and peer-reviewed journals and the peer-review process across disciplines SERIALS LIBRARIAN. 2003 45 2 39 59
Bader-GE Bloom-AE, "How to Do Peer-Review", TRAINING & DEVELOPMENT, 1992, Vol 46, Iss 6, pp 61-
Baez, B. Confidentiality and peer review: The paradox of secrecy in academe REVIEW OF HIGHER EDUCATION. 2002 WIN 25 2 163 +
Bailey, SJ, Deen, MY. Development of a web-based evaluation system: A tool for measuring life skills in youth and family programs FAMILY RELATIONS. 2002 APR 51 2 138 147
Baker, D. The peer review process in science education journals RESEARCH IN SCIENCE EDUCATION. 2002 32 2 171 180

Page 133


Baliga, S, Sjostrom, T. Optimal design of peer review and self-assessment schemes RAND JOURNAL OF ECONOMICS. 2001 SPR 32 1 27 51


Ball, WJ. Political science - The state of the art in on-line publishing and (slightly) beyond SOCIAL SCIENCE COMPUTER REVIEW. 1997 SPR 15 1 13 26


Barnett, SR, dosReis, S, Riddle, MA. Improving the management of acute aggression in state residential and inpatient psychiatric facilities for youths JOURNAL OF THE AMERICAN ACADEMY OF CHILD
AND ADOLESCENT PSYCHIATRY. 2002 AUG 41 8 897 905
Barry, J, Chandler, J, Clark, H. Between the Ivory Tower and the academic assembly line JOURNAL OF MANAGEMENT STUDIES. 2001 JAN 38 1 87 101
Barton, AC, Johnson, V. Truncating agency: Peer review and participatory research RESEARCH IN SCIENCE EDUCATION. 2002 32 2 191 214
Battin, DB, Ceci, SJ. Editorial peer review: Its strengths and weaknesses CONTEMPORARY PSYCHOLOGY-APA REVIEW OF BOOKS. 2003 FEB 48 1 41 43
Bazeley, P. Peer review and panel decisions in the assessment of Australian Research Council project grant applicants: what counts in a highly competitive context? HIGHER EDUCATION. 1998 JUN 35 4 435 452


Bedian, AG. The manuscript review process - The proper roles of authors, referees, and editors JOURNAL OF MANAGEMENT INQUIRY. 2003 DEC 12 4 331 338


Bence, V, Oppenheim, C. The role of academic journal publications in the UK Research Assessment Exercise LEARNED PUBLISHING. 2004 JAN 17 1 53 68


Benjamin, M. Third international congress on peer review in biomedical publication LEARNED PUBLISHING. 1998 JAN 11 1 69 70


Berezin, AA. From peer review to peer recommendation SCIENTIST. 2000 APR 17 14 8 6 6
Berezin, AA. Peer review: Open, not anonymous SCIENTIST. 2001 JUL 23 15 15 6 6
Berg, LD. Masculinism, emplacement, and positionality in peer review PROFESSIONAL GEOGRAPHER. 2001 NOV 53 4 511 521
Berger, G. Do we need a European Research Council?
Bergquist, M, Ljungberg, J. The power of gifts: organizing social relationships in open source communities INFORMATION SYSTEMS JOURNAL. 2001 OCT 11 4 305 320
Blanpied, W. A. & Borg, A. F. 1979 Peer Review of Science Education proposals at the National Science Foundation. Science Education. 63,417-421.


Bloch, S, Walter, G. The Impact Factor: Time for change. AUSTRALIAN AND NEW ZEALAND JOURNAL OF PSYCHIATRY. 2001 OCT 35 5 563 568


Board of Danish Research Councils Newsletter No.2 Dec 2001, in. 2001.


Boden, M. 1990 Peer review. A report to the Advisory Board for the Research Councils from the Working Group on Peer Review. ABRC.
Bornstein, RF. Peer-review in neuropsychology: Can we increase effectiveness without sacrificing rigor? CORTEX. 2002 JUN 38 3 403 405
Manuscript Evaluation Procedures. American Psychologist, 27, 221-225
Boyd, EA. Bureaucratic authority in the "company of equals": The interactional management of medical peer AMERICAN SOCIOLOGICAL REVIEW. 1998 APR 63 2 200 224
Braun, D. The role of funding agencies in the cognitive development of science RESEARCH POLICY. 1998 DEC 27 8 807 821

Brennan, J, Shah, T. Quality assessment and institutional change: Experiences from 14 countries HIGHER EDUCATION. 2000 OCT 40 3 331 349


Brennan, M. 'Big Seven' debate must focus on community benefit - not institutions or researchers, in, Campus Review, Vol. 3. 1993.


Brezsnyak, M. Presenting psychology research on the Internet: Obstacles and opportunities BEHAVIOR RESEARCH METHODS INSTRUMENTS & COMPUTERS. 1999 MAY 31 2 281 286

Brickley, P. Proposed peer-review rule calls academicians biased scientist. 2003 DEC 1 17 23 52 52


Brinn, T, Jones, MJ, Pendlebury, M. Measuring research quality: peer review 1, citation indices 0 OMEGA-INTERNATIONAL JOURNAL OF MANAGEMENT SCIENCE. 2000 APR 28 2 237 239


Brown HM. Peer review should not be anonymous. BMJ. 2003 Apr 12;326(7393):824. .
Brown LC, Stanton WC, Paye W. Facing the limits on uses of medical and peer review information: are high technology and confidentiality on a collision course? Whittier Law Rev. 1997 Fall;19(1):97-118. .
Brown, C. The changing face of scientific discourse: Analysis of genomic and proteomic database usage and acceptance JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY. 2003 AUG 54 10 926 938
Brown, C. The Matthew effect of the Annual Reviews series and the flow of scientific communication through the World Wide Web SCIENTOMETRICS. 2004 60 1 25 36
Brown, G, Manogue, M. AMEE Medical Education Guide No. 22: Refreshing lecturing: a guide for lecturers MEDICAL TEACHER. 2001 MAY 3 23 3 231 244
Brown, JS, Duguid, P. Balancing act: How to capture knowledge without killing it HARVARD BUSINESS REVIEW. 2000 MAY-JUN 78 3 73 +
Brown, KS. A winning strategy for grant applications: Focus on impact SCIENTIST. 1997 APR 28 11 9 13 14

Brunsma, DL, Rockquemore, KA. Statistics, sound bites, and school uniforms: A reply to bodine JOURNAL OF EDUCATIONAL RESEARCH. 2003 NOV-DEC 97 2 72 77


Buddeberg, C, Buchi, S, Buddeberg-Fischer, B. What is an original contribution in the fields of psychosomatic medicine and psychotherapy ZEITSCHRIFT FUR PSYCHOSOMATISCHE MEDIZIN UND PSYCHOANALYSE. 1998 44 1 88 98


Burnard, P, Hannigan, B. Reviewing the review process: towards good practice in the peer review of manuscripts submitted to nursing journals NURSE EDUCATION TODAY. 2001 APR 21 3 238 242


Burns, CW. Colonizing the academy: Assessment, accountability, and quality EDUCATIONAL POLICY. 1998 JUL 12 4 419 431

Burns, K, Welk, D. American Board of Nursing Specialties: Past, present, and future NURSING OUTLOOK. 1997 MAY-JUN 45 3 114 117


Cahill, LS. Social ethics of embryo and stem cell research WOMENS HEALTH ISSUES. 2000 MAY-JUN 10 3 131 135
Calvert, P. Editorial peer review: Its strengths and weaknesses ONLINE INFORMATION REVIEW. 2001 25 6 409 409
Campanario J. Compensation for the scientist - sometimes it is hard to publish papers that are later highly cited. Soc Stud Sci 1993;23:342-63.
Campanario, JM. Peer review for journals as it stands today - Part 1 SCIENCE COMMUNICATION. 1998 MAR 19 3 181 211
Campanario, JM. Peer review for journals as it stands today -


Carlson, LM, Thompson, NJ. Methodologies for informing beneficiaries of their rights: Evaluating the important message from Medicare PATIENT EDUCATION AND COUNSELING. 1997 JUL 31 3 205 213


Carroll, BC, Cotter, GA. A new generation of grey literature: The impact of advanced information technologies PUBLISHING RESEARCH QUARTERLY. 1997 SUM 13 2 5 14

Carter C. Peer review, citations, and biomedical research policy: NIH grants to medical school faculty. Rand Institute Monograph DHEW R-1583, 1974.


Casparie, AF. Quality management of occupational health services: The necessity of a powerful medical profession OCCUPATIONAL MEDICINE-OXFORD. 1998 APR 48 3 203 206

Cassidy MA. Peer review immunity: history, operation, and recent decisions--has HCQIA accomplished its goals? Health Care Law Mon. 2002 May;:3-9.

Cate, MJ. Physician peer review - Serving the patient or the physician? JOURNAL OF LEGAL MEDICINE. 1999 DEC 20 4 479 506


Center for Scientific Review Expanding Opportunities. Center for Scientific Review
Center for Scientific Review Recommendations for Change at the NIH's Centre for Scientific Review
Chait, J. Peer review - William F. Buckley steals a colleague's work. NEW REPUBLIC. 2001 JUL 9 225 2-3 18 19
Chalk, R. Silencing scientists and scholars in other fields: Power, paradigm controls, peer review, and scholarly communication. ISIS. 2003 SEP 94 3 552 553
Champion, H. R. and Mabee, M. S. The Coalition for Trauma Care
Chan, ZCY. Peer review gives voices to novice researchers NURSING RESEARCH. 2003 JAN-FEB 52 1 67-68
Chapman, M. Peer review and intellectual honesty JOURNAL OF INFORMATION SCIENCE. 2000 26 2 129 129
Christiansen D (1978) The perils of publishing. IEEE Spectrum 15,
No. 5, 27.
Chubin, DE. Much ado about peer review, part 2 - Commentary on "Peer review and innovation" (Spier) SCIENCE AND ENGINEERING ETHICS. 2002 JAN 8 1 109 112
Cicchetti, D. V. & Showalter, D. 1988 A computer program for determining the reliability of dimensionally scaled data when the numbers and specific sets of examiners may vary at each assessment. Educational and Psychological Measurement. 48, 717-720.
Statistics Section), 22, 596-600
Cicchetti, DV. Guardians of science: Fairness and reliability of peer review. JOURNAL OF CLINICAL AND EXPERIMENTAL NEUROPSYCHOLOGY. 1999 JUN 21 3 412 421
Cicchetti-DV, "The Reliability of Peer-Review for Manuscript and Grant Submissions - Its Like Deja-Vu All over Again - Authors Response", BEHAVIORAL AND BRAIN SCIENCES, 1993, Vol 16, Iss 2, pp 401-403
Clark, A, Singleton-Jackson, J, Newsom, R. Journal editing: Managing the peer review process for timely publication of articles PUBLISHING RESEARCH QUARTERLY. 2000 FAL 16 3 62 71
Clarke, A, Parsons, E, Williams, A. Outcomes and process in genetic counselling CLINICAL GENETICS. 1996 DEC 50 6 462 469
Clarke, AL. Assessing the carrying capacity of the Florida Keys POPULATION AND ENVIRONMENT. 2002 MAR 23 4 405 418

Page 150
Coelho, RJ, La Forge, J. Manuscript characteristics affecting reviewers' decisions for rehabilitation counseling related journals JOURNAL OF REHABILITATION. 2000 APR-JUN 66 2 4 8
Coimbra CE Jr. Challenges for the evaluation of scientific...
Cole, ER, Stewart, AJ. Invidious comparisons: Imagining a psychology of race and gender beyond differences POLITICAL PSYCHOLOGY. 2001 JUN 22 2 293 308
Cole, S. Guardians of science: Fairness and reliability of peer review MINERVA. 1998 SUM 36 2 179 189
Colliver JA. Commentary on Cicchetti's "reliability of peer review". Teach Learn Med. 2002 Summer;14(3):142-3.
Colman-AM, "Unreliable Peer-Review - Causes and Cures of Human
Colvin, AJS.  Institutional pressures, human resource strategies, and the rise of nonunion dispute resolution procedures.  INDUSTRIAL & LABOR RELATIONS REVIEW. 2003 APR 56 3 375 392


Committee on Research and Peer Review in EPA, Board on Environmental Studies and Toxicology. 1995. Interim Report of the Committee on Research and Peer Review in EPA. National Academy Press, Washington, D.C.


Conn, VS, Valentine, JC, Cooper, HM, Rantz, MJ. Grey literature in meta-analyses NURSING RESEARCH. 2003 JUL-AUG 52 4 256 261


Cordes, C. 1987 Competition for federal money is said to have damaging effects on scientists. Chronicle of Higher Education. 33, 6,7,11.


Cosser, M. Towards the design of a system of peer review of teaching for the advancement of the individual within the university HIGHER EDUCATION. 1998 MAR 35 2 143 162
Cox, JE. Publishers, publishing and the Internet: How journal publishing will survive and prosper in the electronic age ELECTRONIC LIBRARY. 1997 APR 15 2 125 131
Cozzens, S. E. 1986 Editor's introduction in "Funding and Knowledge Growth" - Theme Section. Social Studies of Science. 16, 9-22.
Crigger, NJ. What we owe the author: Rethinking editorial peer review NURSING ETHICS. 1998 SEP 5 5 451 458
Cronin, B. Bibliometrics and beyond: Some thoughts on web-based citation analysis JOURNAL OF INFORMATION SCIENCE. 2001 27 1 1 7
Cronin, B. Peer review and the stuff of scholarship LIBRARY JOURNAL. 2001 SEP 15 126 15 57 57
Cronin, B. Silencing scientists and scholars in other fields: power; paradigm controls, peer, review and scholarly communications. JOURNAL OF DOCUMENTATION. 1999 JUN 55 3 359 360
Culver, DM, Gilbert, WD, Trudel, P. A decade of qualitative research in sport psychology journals: 1990-1999 SPORT PSYCHOLOGIST. 2003 MAR 17 1 1 15
Cunliffe, B. The benefits of assessment, and some risks., in, Antiquity. 1996.
Cunningham, CM, Helms, JV. Sociology of science as a means to a more authentic, inclusive science education JOURNAL OF RESEARCH IN SCIENCE TEACHING. 1998 MAY 35 5 483 499


Daucourt, V, Michel, P. Results of the first 100 accreditation procedures in France INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE. 2003 DEC 15 6 463 471


Davies, R, Berrow, T. An evaluation of the use of computer supported peer review for developing higher-level skills. COMPUTERS & EDUCATION. 1998 JAN-FEB 30 1-2 111 115
Davis RM. Peer review-the "who" and the "how". Tob Control. 1998 Summer;7(2):109-11.
der Castro, R. Reviewing peer review SCIENTIST. 2003 MAR 24 17 6 18 19
de Rond, M. Reviewer 198, the hedgehog, and the fox - Next generation theories in strategy JOURNAL OF MANAGEMENT INQUIRY. 2002 MAR 11 1 35 45
de Silva, MAT, Yapa, G, de Silva, ED. Interactive peer review as a productive evaluation process RESEARCH EVALUATION. 2002 DEC 11 3 119 128
De Waal, A. What's new in the 'New Partnership for Africa's Development'? INTERNATIONAL AFFAIRS. 2002 JUL 78 3 463 +
Decker, O, Brahtler, E. Quality control and bibliometric features - publications in psychosomatic medicine examplified by the Zeitschrift fur Psychosomatische Medizin und Psychoanalyse ZEITSCHRIFT FUR PSYCHOSOMATISCHE MEDIZIN UND PSYCHOTHERAPIE. 1998 44 4 370 384
Del Carmen Davo M, Vives C, Alvarez-Dardet C. Why are women underused in the JECH peer review process? J Epidemiol Community
Dickson-D, "Peer-Review - NIH Urged to Streamline Bids ... as Britain Seeks to Reassure Doubters over Policy Changes", NATURE, 1994, Vol 370, Iss 6486, pp 170-171
Vol 151, Iss 1, pp 93-96
Dirk, L. A measure of originality: The elements of science SOCIAL STUDIES OF SCIENCE. 1999 OCT 29 5 765 776
Dirk-L, "From Laboratory to Scientific Literature - The Life and Death of Biomedical-Research Results", SCIENCE COMMUNICATION, 1996, Vol 18, Iss 1, pp 3-28
Druce HM. Peer review and manuscript triage. Am J Respir Crit Care Med. 2003 Aug 1;168(3):397; author reply 397; discussion 397-8.
Drummond, MF. A reappraisal of economic evaluation of pharmaceuticals - Science or marketing? PHARMACOECONOMICS. 1998 JUL 14 1 1 9
Drummond-MF Jefferson-TO, "Guidelines for Authors and Peer Reviewers of Economic Submissions to the Bmj", BRITISH MEDICAL JOURNAL, 1996, Vol 313, Iss 7052, pp 275-283
Eamon, W. 1985 From the secrets of nature to public knowledge: the origins of the concept of openness in science. Minerva. 23, 321-347.
Eckmann, B. (1977). Qualitätsskriterien wissenschaftlicher


Eisenhart, M.  The paradox of peer review: Admitting too much or allowing too little?  RESEARCH IN SCIENCE EDUCATION.  2002  32  2  241  255

Eisenman, R.  Peer review of scholarly and scientific work: The good, the bad, and the ugly  JOURNAL OF INFORMATION ETHICS.  1998  FAL  7  2  6  10

Eldredge, JD.  Identifying peer-reviewed journals in clinical medicine  BULLETIN OF THE MEDICAL LIBRARY ASSOCIATION.  1997  OCT  85  4  418  422


Ellerbeck, EF, Kresowik, TF, Hemann, RA, Mason, P, Wiblin, RT, Marciniak, TA. Impact of quality improvement activities on care for acute myocardial infarction INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE. 2000 AUG 12 4 305 310
Ellis-PM Mellsop-GW Peace-KA Wilson-JM, "Peer-Review As an Aid to Improving the Completeness of Psychiatric Case Notes", MEDICAL EDUCATION, 1987, Vol 21, Iss 6, pp 493-497
Elvik, R. Are road safety evaluation studies published in peer reviewed journals more valid than similar studies not published in peer reviewed journals? ACCIDENT ANALYSIS AND PREVENTION. 1998 JAN 30 1 101 118

ERC's funding instruments and research areas, in, Centre National de la Recherche Scientifique (CNRS) France, pp. 15. 2002.


Espin, AV, Pina, JAL. Methodology used in the evaluation of universities PSICOTHEMA. 2000 12


Evans J T, Nadjari H I & Burchell S A (1990) Quotational and
Fallon WF Jr, Barnoski AL, Mancuso CL, Tinnell CA, Malangoni MA.
Benchmarking the quality-monitoring process: a comparison of outcomes analysis by trauma and injury severity score (TRISS) methodology with the peer-review process. J Trauma. 1997 May;42(5):810-5; discussion 815-7.
Federal Research: Peer Review Practices at Federal Science


Fine, P. Canadian network is denied funding, in, The Times Higher Education Supplement. 2001.
Finn, R. Researchers get ready for NIH reforms SCIENTIST. 1997 AUG 18 11 16 1 &
Finnegan, GA. Silencing scientists and scholars in other fields: Power, paradigm controls, peer review, and scholarly communication. COLLEGE & RESEARCH LIBRARIES. 1999 NOV 60 6 604 605
Fitzgerald, ME, Molinari, GF, Bausell, RB. The empowering potential of quality improvement data. EVALUATION & THE HEALTH PROFESSIONS. 1998 DEC 21 4 419 428
Foss A, Westcott M. Major journals should peer review trials at protocol stage. BMJ. 1997 Jun 7;314(7095):1691-2.
Frank-E, "Editors Requests of Peer Reviewers - A Study and a Proposal", PREVENTIVE MEDICINE, 1996, Vol 25, Iss 2, pp 102-104
Franklin, J. J. 1988 Selectivity in funding: evaluation of research in Australia. Prometheus. 6, 34-60.
Franklin, R. Problems with peer review - Comment JOURNAL OF INFORMATION ETHICS. 1998 FAL 7 2 4 4
Fraser-RC Mckinley-RK Mulholland-H, "Consultation Competence in General-Practice - Testing the Reliability of the Leicester
Freeman J. Peer review is critical to quality health care. Iowa Med. 1997 Nov-Dec;87(9):357.
Fuhrmann, J, Hubner, M. Peer review methods in diagnosis-related groups GESUNDHEITSWESEN. 2001 AUG-SEP 63 8-9 A14 A14
Fuller, D. R. Dr. Rick Fuller, Director of Science Funding of the Wellcome Trust, outlines ways for ERS members to apply for grants and funding., in, ERS Newsletter. 2002.
Furness-P, "Promoting Research into Peer-Review - Referees Should
Provide References", BRITISH MEDICAL JOURNAL, 1994, Vol 309, Iss 6953, pp 539-539


Garfield-E, "Training Our Cites on the Stars - ABT, Helmut Observations on Astrosociology .1. Age and Productivity, Peer-Review, and Big Science Little Science - An Introduction to the
General Accounting Office 1987b University Funding: Information on the role of peer review at NSF and NIH. Fact sheet for the Ranking Minority Member, Committee on Appropriations, United States Senate. Washington, DC.
GeneWatch GeneWatch UK Challenges the MRC, Wellcome Trust and Department of Health to allow an independent scientific peer review of Biobank UK. GeneWatch,
Georghiou, L, Roessner, D. Evaluating technology programs: tools and methods RESEARCH POLICY. 2000 APR 29 4-5 657 678

Gerard, K, Smoker, I, Seymour, J. Raising the quality of cost-utility analyses: lessons learnt and still to learn HEALTH POLICY. 1999 MAR 46 3 217 238

Gerlach, FM, Beyer, M, Romer, A. Quality circles in ambulatory care: state of development and future perspective in Germany INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE. 1998 FEB 10 1 35 42

GERMANY, "On Our Own Behalf - Safeguarding Quality of Medical-Journals Through Peer-Review", SCHWEIZERISCHE MEDIZINISCHE WOCHENSCHRIFT, 1988, Vol 118, Iss 8, pp 255-256


Giangrego E. Mediation and peer review. CDS Rev. 1999 Sep;:24-7.


Gibbs, G. Improving teaching, learning and assessment JOURNAL OF GEOGRAPHY IN HIGHER EDUCATION. 1999 JUL 23 2 147 155


Gillett-R Harrow-J, "Prescriptions for Medical-Research .2. Is Medical- Research Well-Served by Peer-Review", BRITISH MEDICAL
Gillig, PM, Barr, A. A model for multidisciplinary peer review and supervision of behavioral health clinicians COMMUNITY MENTAL HEALTH JOURNAL. 1999 AUG 35 4 361 365
Glantz SA, Bero LA. Inappropriate and appropriate selection of "peers" in grant review. JAMA 1994;272:114-16.
Goebel, LJ. A peer review feedback method of promoting compliance with preventive care guidelines in a resident ambulatory care clinic JOINT COMMISSION JOURNAL ON QUALITY IMPROVEMENT. 1997 APR 23 4 196 202
Goering, M, Wilson, W. Implementing preterm labor guidelines: A collaborative care improvement process JOURNAL OF PERINATAL & NEONATAL NURSING. 2002 JUN 16 1 47 57
Goldbaum, M. Peer review of scientific papers REVISTA DE SAUDE PUBLICA. 1999 AUG 33 4 327 328
Goldbeck-Wood S. Secrecy and openness in peer review--time for a change of culture? Ital J Gastroenterol Hepatol. 1999
Goldman, RL, Ciesco, E. Improving peer review: Alternatives to unstructured judgements by a single reviewer. JOINT COMMISSION JOURNAL ON QUALITY IMPROVEMENT. 1996 NOV 22 11 762 769
Gordon, M. (1978). A study of the evaluation of research papers by primary journals in the UK. University of Leicester: Primary Communications Research Centre
GPRA, Government Performance and Results Act of 1993 (PL 103-62), 1993


Gray, DO. Government-sponsored industry-university cooperative research: an analysis of cooperative research center evaluation approaches RESEARCH EVALUATION. 2000 APR 9 1 57 67


Greenbaum, D, Lim, J, Gerstein, M. An analysis of the present system of scientific publishing: what's wrong and where to go from here INTERDISCIPLINARY SCIENCE REVIEWS. 2003 DEC 28 4 293


Greenberg, D. S. Peer review: and the winner is..... The Lancet 354(9195), 2092. 1999.


Greene, JP, Peterson, PE. If the peer review attack fails, attack something else PS-POLITICAL SCIENCE & POLITICS. 2000 JUN 33 229 231


Grimmett, MR, Sulmasy, DP. The call of the sirens: Ethically navigating the sea of nonvalidated therapies JOURNAL OF REFRACTIVE SURGERY. 1998 SEP-OCT 14 5 559 566


Guidelines for Ensuring and Maximizing the Quality, Objectivity,
Utility and Integrity of Information Disseminated by Federal Agencies, 2002.
Guilford, WH. Teaching peer review and the process of scientific writing ADVANCES IN PHYSIOLOGY EDUCATION. 2001 SEP 25 3 167 175
Gunn IP. Evidence-based practice, research, peer review, and publication. CRNA. 1998 Nov;9(4):177-82. Review.
Glitz, P. (1990). This manuscript must either be drastically reduced or fully oxidized - Die "Angewandte Chemie" im Spannungsfeld der Interessen. Unverffentlichtes Manuskript eines Vortrags, gehalten im November 1990 beim Ortsverband Frankfurt am Main der Gesellschaft Deutscher Chemiker
Hackerman-N, "Peerless Science - Peer-Review and United-States Science Policy, by D.E. Chubin, E.J. Hackett", JOURNAL OF HIGHER


Hackett, EJ. Four observations about "Six Domains of Research Ethics" SCIENCE AND ENGINEERING ETHICS. 2002 APR 8 2 211 214

Hadjistavropoulos, T, Bieling, PJ. When reviews attack: Ethics, free speech, and the peer review process CANADIAN PSYCHOLOGY-PSYCHOLOGIE CANADIENNE. 2000 AUG 41 3 152 159

Hafferty, FW. What medical students know about professionalism MOUNT SINAI JOURNAL OF MEDICINE. 2002 NOV 69 6 385 397


Hallman, G. Peer review SCIENTIST. 2001 NOV 26 15 23 6 6


Halpin, G. (et al). 1986 Reliability of ratings of research proposals submitted for funding. In Annual Meeting of the Mid-
Hamilton, NW. Peer review - The Linchpin of academic freedom and tenure ACADEME-BULLETIN OF THE AAUP. 1997 MAY-JUN 83 3 14 19
Hammond, M. Communication within on-line forums: the opportunities, the constraints and the value of a communicative approach COMPUTERS & EDUCATION. 2000 DEC 35 4 251 262
Handysides S, Goalen V, Soltanpoor N. CDR review's editorial


Harley, S, Lee, FS. Research selectivity, managerialism, and the academic labor process: The future of nonmainstream economics in UK universities HUMAN RELATIONS. 1997 NOV 50 11 1427 1460


Harman, G. Supporting quality research in institutions of higher education AUSTRALIAN JOURNAL OF EDUCATION. 1998 NOV 42 3 285 302


Harnad, S. Learned inquiry and the Net: the role of peer review, peer commentary and copyright LEARNED PUBLISHING. 1998 OCT 11 4 283 292


Harris S, Buchinski B, Grzybowski S, Janssen P, Mitchell GW, Parquharson D. Induction of labour: a continuous quality...


Harrison, J. Going for gold LIBRI. 2002 DEC 52 4 199 203


Main: Campus
Hartmann-P Bott-U Grusser-M Kronsteiner-P Jorgens-V, "Effects of Peer-Review Groups on Documentation of Diabetes-Related Patient Data in Physicians Practice - Results of a Prospective Controlled-Study", DIABETOLOGIA, 1994, Vol 37, Iss S1, pp A176-A176
Hauptman, R. Problems with peer review JOURNAL OF INFORMATION ETHICS. 1998 FAL 7 2 3 3
Hauser, JR. Research, development, and engineering metrics MANAGEMENT SCIENCE. 1998 DEC 44 12
Hawkins JW. Peer review: mark of a profession or "who's in and who's out?". Clin Excell Nurse Pract. 1999 Sep;3(5):253..
Hayhoe, GF. Something so right: Why i'm proud of our review process TECHNICAL COMMUNICATION. 2001 AUG 48 3 261 262
Healy, P. 2 Key lawmakers seek further study of federal agencies' peer-review policies. The Chronicle of Higher Education 45(34), A32. 1999.
Heidemann, EG. Moving to global standards for accreditation processes: the ExPeRT Project in a larger context INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE. 2000 JUN 12 3 227 230
Heijl A. Developing the peer review system of Acta


Henderson, A. Undermining peer review SOCIETY. 2001 JAN-FEB 38 2 47 54


Hernon, PP, Schwartz, CC. Peer review LIBRARY & INFORMATION SCIENCE RESEARCH. 2001 23 1 1 3


Higgins-AC, "Peerless Science - Peer-Review and United-States
Hirschhorn, L, Kulish, D, Blaylock, E. Work design in R&D: Time for a change? RESEARCH-TECHNOLOGY MANAGEMENT. 2004 JAN-FEB 47 1 44 48
Hirshon, A. Jam tomorrow, jam yesterday, but never jam today: Some modest proposals for venturing through the looking-glass of scholarly communication SERIALS LIBRARIAN. 1998 34 1-2 65 87
Hodson, D, Maher, I. The open method as a new mode of governance: The case of soft economic policy co-ordination JOURNAL OF COMMON MARKET STUDIES. 2001 NOV 39 4 719 746
Hofer, TP, Bernstein, SJ, DeMonner, S, Hayward, RA. Discussion between reviewers does not improve reliability of peer review of hospital quality MEDICAL CARE. 2000 FEB 38 2 152 161
Hojat M, Gonnella JS, Caelleigh AS. Impartial judgment by the
Hoke-F, "Science in the Courtroom - What Evidence Is Admissible - And Who Decides - Some Scientists Say a Supreme-Court Decision to Deemphasize Peer-Review Has Led to Better Court Science .1.", SCIENTIST, 1994, Vol 8, Iss 12, pp 1+
Holbrook, A. Evaluation of research sponsored by federal granting councils in Canada: the social contract RESEARCH EVALUATION. 2000 APR 9 1 47 56
Holloway, M. 1991 The reward of ideas that are wrong. Sci. Am. 265, 16-17.
Homan, FM. Whither peer review: Prague '97 BULLETIN OF THE MEDICAL LIBRARY ASSOCIATION. 1998 JUL 86 3 421 422
Honer, WG, Linseman, MA. The physician-scientist in Canadian psychiatry JOURNAL OF PSYCHIATRY & NEUROSCIENCE. 2004 JAN 29 1 49 56
Hornbostel, S. Third party funding of German universities. An indicator of research activity? SCIENTOMETRICS. 2001 MAR-APR 50 3 523 537
Horowitz, S, Van Eyck, S, Albanese, M. Successful peer review of courses: A case study ACADEMIC MEDICINE. 1998 MAR 73 3 266 271
Horrobin DF. Peer review of statistics in medical research. Rationale for requiring power calculations is needed. BMJ. 2002 Aug 31;325(7362):491; author reply 491.
Horrobin DF. The philosophical basis of peer review and the suppression of innovation. JAMA 1990;263:1438–41.
Howard, L, Wilkinson, G. Peer review and editorial decision-making BRITISH JOURNAL OF PSYCHIATRY. 1998 AUG 173 110 113
http://www.mrc.ac.uk/index/funding/funding-specific_schemes/funding-evaluation_of_schemes/funding-peer_review_study.htm
Huang, P, Hsu, YHE, Kai-Yuan, T, Hsueh, YS. Can European external peer review techniques be introduced and adopted into Taiwan's hospital accreditation System? INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE. 2000 JUN 12 3 251 254


Humphreys, LG. Problems in individual differences research with peer review, some peer reviewers, and suggestions for reform. MULTIVARIATE BEHAVIORAL RESEARCH. 2002 37 2 283 295


Huston-P Elmslie-T, "Program Descriptions - Information for Authors and Peer Reviewers", CANADIAN MEDICAL ASSOCIATION JOURNAL, 1996, Vol 155, Iss 8, pp 1069-1071

Huston-P Squires-BP, "Case-Reports - Information for Authors and Peer Reviewers", CANADIAN MEDICAL ASSOCIATION JOURNAL, 1996, Vol 154, Iss 1, pp 43-44


Huston-P, "Information for Peer Reviewers", CANADIAN MEDICAL ASSOCIATION JOURNAL, 1994, Vol 150, Iss 8, pp 1211-1216


Huston-P, "Reporting on Surveys - Information for Authors and Peer Reviewers", CANADIAN MEDICAL ASSOCIATION JOURNAL, 1996, Vol 154, Iss 11, pp 1695-1698

Hwang, NH. Defaming a physician's career - The double edged sword of peer review privilege and immunity. JOURNAL OF LEGAL MEDICINE. 2004 MAR 25 1 95 109


in Environmental Technology Development Programs, National Academy Press, 1998. http://www.nap.edu/openbook/0309063388/html/27.html> (Type in BM_1 in the Go to Page field at the bottom of the page.)
Isetts, BJ, Brown, LM, Schondelmeyer, SW, Lenarz, LA. Quality assessment of a collaborative approach for decreasing drug-related morbidity and achieving therapeutic goals ARCHIVES OF INTERNAL MEDICINE. 2003 AUG 11 163 15 1813 1820
JAMA, The Effects of Blinding on the Quality of Peer Review. Journal of the American Medical Association, 263 (10), 1371-1376
Jans, MP, Schellevis, FG, Le Coq, EM, Bezemer, PD, van Eljk, JTM. Health outcomes of asthma and COPD patients: the evaluation of a project to implement guidelines in general practice INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE. 2001 FEB 13 1 17 25
Jans, MP, Schellevis, FG, Van Hensbergen, W, Van Emden, TD, Van
Eijk, JTM. Management of asthma and COPD patients: feasibility of the application of guidelines in general practice. INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE. 1998 FEB 10 1 27 34


Jayasinghe, UW, Marsh, HW, Bond, N. A multilevel cross-classified modelling approach to peer review of grant proposals: the effects of assessor and researcher attributes on assessor ratings. JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES A-STATISTICS IN SOCIETY. 2003 166

Jayasinghe, UW, Marsh, HW, Bond, N. Peer review in the funding of research in higher education: The Australian experience. EDUCATIONAL EVALUATION AND POLICY ANALYSIS. 2001 WIN 23 4 343 364


Jefferson, T, Demicheli, V. Methodological quality of economic modelling studies - A case study with hepatitis B vaccines

JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION. 1998 JUL 15 280 3 275 277


Johnson, KB, Miller, RA. The JAMIA student editorial board: Peer review education in biomedical Informatics JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION. 2004 JAN-FEB 11 1 87 88


211-212
Johnson-DM  Roen-DH, "Complimenting and Involvement in Peer Reviews - Gender Variation", LANGUAGE IN SOCIETY, 1992, Vol 21, Iss 1, pp 27-57
Johnson-Greene, D, Bechtold, KT. Ethical considerations for peer review in forensic neuropsychology CLINICAL NEUROPSYCHOLOGIST. 2002 16 1 97 104
Jones, AW. The impact of Alcohol and Alcoholism among substance abuse journals ALCOHOL AND ALCOHOLISM. 1999 JAN-FEB 34 1 25 34
Jones, B. 1985 Government funding of scientific research. Vestes. 28, 4-10.
Jones, MJ. Critically evaluating an applications vs theory framework for research quality OMEGA-INTERNATIONAL JOURNAL OF MANAGEMENT SCIENCE. 1999 JUN 27 3 397 401
Joyce J, Rabe-Hesketh S, Wessely S. Reviewing the reviews: the
Judson-HF, "Structural Transformations of the Sciences and the End
of Peer-Review", JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION,
1994, Vol 272, Iss 2, pp 92-94
Juhasz, S., Calvert, E., Jackson, T., Kronick, D. A. & Shipman, 
Transactions on Professional Communication, 18 (3), 177-185 
Jukes-TH, "Peer-Review of Scientific Manuscripts", FASEB JOURNAL, 
1995, Vol 9, Iss 13, pp 1381-1381 
Justesen-DR, "An Exercise in Open Peer-Review", 
BIOELECTROMAGNETICS, 1995, Vol 16, Iss 1, pp 1-1 
author identity improve peer review quality? A randomized 
controlled trial. PEER Investigators. JAMA. 1998 Jul 
Justice, A. C., Cho, M. K., Winker, M. A., Berlin, J. A. and 
Rennie, D. Does Masking Author identity Improve Peer Review 
1998. 
Justice, A. C., Cho, M. K., Winker, M. A., Berlin, J. A., 
Waeckerle, J., Callahan, M. and Rennie, D. Masking Author identity 
in Peer Review. The Journal of the American Medical Association 
Justice-AC Berlin-JA Fletcher-SW Fletcher-RH , Goodman-SN, 
"Do Readers and Peer Reviewers Agree on Manuscript Quality", JAMA- 
JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION, 1994, Vol 272, Iss 2, 
pp 117-119 
Juzych-MS Shin-DH Coffey-JB Parrow-KA Tsai-CS , Briggs-KS, 
"Pattern of Publication of Ophthalmic Abstracts in Peer-Reviewed 
Kaa, D. van de 1993 Picking the winners by consensus: grant- 
giving Practice in the Netherlands. In Research Grants Management 
and Funding (ed.F. Q. Wood & V. Meek), pp. 63-82. Canberra: 
Anutech. 
Kadzielski M, Marino SP. Civil rights claims and physician peer 
Kaewsarn, P, Moyle, W, Creedy, D. Thai nurses' beliefs about 
breastfeeding and postpartum practices JOURNAL OF CLINICAL 
NURSING. 2003 JUL 12 4 467 475 
Kaiser J. Biomedical politics. Sex studies denounced, NIH's 
. 
Kaiser J. Regulatory policy. White House softens disputed peer-
Kaiser J. Science and regulation. A White House mandate for more 
Kaiser, H. Time To Close R&D Funding Gap. AGB Reports 27(4), 20-
Kalberer J. Treatment of young investigators in the National 
Karlawish, JHT, Hougham, GW, Stocking, CB, Sachs, GA. What is the quality of the reporting of research ethics in publications of nursing home research? JOURNAL OF THE AMERICAN GERIATRICS SOCIETY. 1999 JAN 47 1 76 81
Karmel, P. 1987 Management of research funding. Higher Education Research and Development. 6, 69-76.
Kawski, S, Koch, U. Quality management in the field of psychosomatic rehabilitation PSYCHOTHERAPIE PSYCHOSOMATIK MEDIZINISCHE PSYCHOLOGIE. 1999 SEP-OCT 49 9-10 316 325


Kellogg, ND, Lamb, JL, Lukefahr, JL. The use of telemedicine in child sexual abuse evaluations. CHILD ABUSE & NEGLECT. 2000 DEC 24 12 1601 1612

Kelly, JA. Scientific meeting abstracts: Significance, access, and trends. BULLETIN OF THE MEDICAL LIBRARY ASSOCIATION. 1998 JAN 86 1 68 76


Kendall, D. Doing a good deed or confounding the problem? Peer review and sociology textbooks. TEACHING SOCIOLOGY. 1999 JAN 27 1 17 30


Kennedy, RH, Gubbins, PO, Luer, M, Reddy, IK, Light, KE. Developing and sustaining a culture of scholarship: AMERICAN JOURNAL OF PHARMACEUTICAL EDUCATION. 2003 67 3


Kiernan, V. Ingelfinger, embargoes, and other controls on the dissemination of science news SCIENCE COMMUNICATION. 1997 JUN 18 4 297 319


Kim, MJ. A comparative study of citations from papers by Korean scientists and their journal attributes JOURNAL OF INFORMATION SCIENCE. 1998 24 2 113 121

Kim, YM, Putjuk, F, Basuki, E, Kols, A. Self-assessment and peer review: Improving Indonesian service providers' communication with clients INTERNATIONAL FAMILY PLANNING PERSPECTIVES. 2000 MAR 26 1 4 12


King-J, "Ruling Could Inhibit Peer-Review Candor - Reviewers Will Have to Think Twice, Now That Tenure Evaluation Letters Can Be Used As Evidence in Bias Cases", SCIENTIST, 1990, Vol 4, Iss 13, pp 1-


Kirby, W. S. 1990 Toward a model policy for federally supported


Klanica, K. Managed care: Immunity for peer review under HCQIA JOURNAL OF LAW MEDICINE & ETHICS. 2003 SPR 31 1 160 161


Kobayashi, S. Applying audition systems from the performing arts to R&D funding mechanisms: quality control in collaboration among
the academic, public, and private sectors in Japan RESEARCH POLICY. 2000 FEB 29 2 181 192
Kohane, IS, Altman, RB. The new peer review JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION. 2000
Kostoff, R. N., "Science and Technology Roadmaps",
Kostoff, RN. Bypassing peer review SCIENTIST. 2000 MAR 20 14 6 4 4
Kostoff, RN. Peer review SCIENTIST. 1997 FEB 3 11 3 12 12
Kostoff, RN. Peer review: The appropriate GPRA metric for research SCIENCE. 1997 AUG 1 277 5326 651 652


Kostoff, R. N. and Stanford, L. B. “Science and Technology Asset Management:
Kraemer, HC, Gardner, C, Brooks, JO, Yesavage, JA. Advantages of excluding underpowered studies in meta-analysis: Inclusionist versus exclusionist viewpoints PSYCHOLOGICAL METHODS. 1998 MAR 3 1 23 31

Kreber, C. The scholarship of teaching: A comparison of conceptions held by experts and regular academic staff HIGHER EDUCATION. 2003 JUL 46 1 93 121


Krinisky G. How to avoid "unblinding" the peer review process. AJR Am J Roentgenol. 1999 Jun;172(6):1474.


Kritchevsky-D, "Senior Scientists Could Play a Key Role in Resolving Big Problems in Peer-Review", SCIENTIST, 1990, Vol 4, Iss 2, pp 12-12


Krogh, MC. A skill, process, and person-oriented graduate seminar on teaching TEACHING SOCIOLOGY. 2000 OCT 28 4 333 345


La Manna, M, Young, J. The Electronic Society for Social Scientists: from journals as documents to journals as knowledge exchanges INTERLENDING & DOCUMENT SUPPLY. 2002 30 4 178 182


Laflin, MT, Horowitz, SM, Nims, JK. A researcher's guide to health education journals AMERICAN JOURNAL OF HEALTH BEHAVIOR. 1999 MAY-JUN 23 3 210 232
Langfeldt, L. The decision-making constraints and processes of grant peer review, and their effects on the review outcome SOCIAL STUDIES OF SCIENCE. 2001 DEC 31 6 820 841
Lanier, DC, Roland, M, Burstin, H, Knottnerus, JA. Doctor performance and public accountability LANCE. 2003 OCT 25 362 9393 1404 1408
Larochelle, M, Desautels, J. On peers, those 'particular friends' RESEARCH IN SCIENCE EDUCATION. 2002 32 2 181 189
Lazarus A. Equitable peer review. Psychiatr Serv. 1998 Dec;49(12):1535-6, 1539.
Lebowitz, B. D. 1981 Funding agencies and the research community.Gerontologist. 21, 382-387.
Lederberg, J. 1989 Does scientific progress come from projects or people?Current Contents. 20, 4-12.
Lederman, L. M. 1991 Science: The End of the Frontier? Science:
Page 210


Lee, D, Lopez, L. An invitational workshop on collaboration between quality improvement organizations and business coalitions. JOINT COMMISSION JOURNAL ON QUALITY IMPROVEMENT. 1997 Jun 23 6 334 341


Lehrl, S. Evaluating scientific performances by impact factors - the right for equal chances. STRAHLENTHERAPIE UND ONKOLOGIE. 1999 Apr 175 4 141 +

Lehrman, NS, Sharav, VH. Ethical problems in psychiatric research. JOURNAL OF MENTAL HEALTH ADMINISTRATION. 1997 SPR 24 2 227 250

Leipzig, RM, Adams, MA, Brand, J. Peer review for online geriatric educational products. JOURNAL OF THE AMERICAN GERIATRICS SOCIETY. 2004 Apr 52 4


Leopold-AC, "The Peer-Review System - Pique and Critique", SCIENTIST, 1988, Vol 2, Iss 13, pp 11-


Lepkowski No gender bias in its grant awards, says NSF. Chemical & Engineering News 75, 10. 1997.

Lepkowski, W. 1991a Bromley at midterm is confident, in charge in science adviser job. Chemical & Engineering News. 69, 21-22.
Lepkowski, W. 1992b NSF gears up for stricter oversight of academic research support. Chemical and Engineering News. 70, 9-14.
Lepkowski, W. 1992d Science-Technology policy: Clinton victory will mean vast changes. Chemical and Engineering News. 70, 4-6.
Lepkowski, W. 1993a UK research structure. Chemical and Engineering News. 71, 4-5.
Leta, J, Lewison, G. The contribution of women in Brazilian science: A case study in astronomy, immunology and oceanography SCIeNtOMeTRiCS. 2003 57 3 339 353
Leutner, D. Constancy within change? Diagnostica as a German psychology journal on its way into the third millennium DIAGNOSTICA. 2004 50 1 18 21
Lewis, JN, Rising, W. Improving appropriate selection of radical prostatectomy for prostate cancer EVALUATION & THE HEALTH PROFESSIONS. 1998 DEC 21 4 429 441
Hopkins University Press
Lewison, G, Lipworth, S, Rippon, I, Roe, P, Cottrell, R. Geographical equity between outputs of biomedical research grants and research capability as an indicator of the peer-review process for grant applications RESEARCH EVALUATION. 2003 DEC 12 3 225 230
Leysen, JM, Black, WK. Peer review in Carnegie Research Libraries COLLEGE & RESEARCH LIBRARIES. 1998 NOV 59 6 512 522
Lievrouw, LA. Silencing scientists and scholars in other fields: Power, paradigm controls, peer review, and scholarly communication. LIBRARY QUARTERLY. 2000 JAN 70 1 163 166
Light, CM, Chappell, PH, Kyberd, PJ. Establishing a standardized clinical assessment tool of pathologic and prosthetic hand function: Normative data, reliability, and validity ARCHIVES OF PHYSICAL MEDICINE AND REHABILITATION. 2002 JUN 83 6 776 783
Lim, TO, Soraya, A, Ding, LM, Morad, Z. Assessing doctors' competence: application of CUSUM technique in monitoring doctors' performance INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE. 2002 JUN 14 3 251 258
Lin, SSJ, Liu, EZF, Yuan, SM. Web-based peer assessment: feedback for students with various thinking-styles JOURNAL OF COMPUTER
ASSISTED LEARNING. 2001 DEC 17 4 420 432
LindgensStrache, U.  Peer review - A proven method for assuring and improving the quality of audits in Germany? Some comments BETRIEBSWIRTSCHAFTLICHE FORSCHUNG UND PRAXIS. 1997 MAY-JUN 49 3 266 291
Lindsey, D. Ensuring standards in social work research RESEARCH ON SOCIAL WORK PRACTICE. 1999 JAN 9 1 115 120
Liu, EZF, Lin, SSJ, Chiu, CH, Yuan, SM.  Web-based peer review: The learner as both adapter and reviewer IEEE TRANSACTIONS ON EDUCATION. 2001 AUG 44 3 246 251
Liu, JG, Pysarchik, DT, Taylor, WW. Peer review in the classroom BIOSCIENCE. 2002 SEP 52 9 824 829
Liversidge-A, "Pnas Publication of AIDS Article Spurs Debate over Peer- Review - Is the National Academy to Be Praised or Chastened for Disseminating the Views of Controversial Scientist Duesberg, Peter", SCIENTIST, 1989, Vol 3, Iss 7, pp 1-
Livingston, EH, Harwell, JD. Peer review AMERICAN JOURNAL OF SURGERY. 2001 AUG 182 2 103 109
Lloyd M E (1990) Gender factors in reviewer recommendations for manuscript publication. Journal of Applied Behavior Analysis 23, 539-543. EX


Lombarts, MJMH, Klazinga, NS. A policy analysis of the introduction and dissemination of external peer review (visitatie) as a means of professional self-regulation amongst medical specialists in The Netherlands in the period 1985-2000 HEALTH POLICY. 2001 DEC 58 3 191 213

Lombarts, MJMH, Klazinga, NS. Supporting Dutch medical specialists with the implementation of visitatie recommendations: a descriptive evaluation of a 2-year project INTERNATIONAL
JOURNAL FOR QUALITY IN HEALTH CARE. 2003 APR 15 2 119 129
Long RM. From the National Institute of General Medical Sciences. Learning the system: changes pending in peer review and around the NIH. Pharm Res. 1999 Jun;16(6):783-4.
Luce-BR Lyles-CA Rentz-AM, "The View from Managed Care Pharmacy", HEALTH AFFAIRS, 1996, Vol 15, Iss 4, pp 168-176
Luce-RD, "Reliability Is Neither to Be Expected Nor Desired in Peer-Review", BEHAVIORAL AND BRAIN SCIENCES, 1993, Vol 16, Iss 2, pp 399-400
Lumsden, DB. Jump-starting a journal's paper flow: Fourteen tested, effective methods JOURNAL OF SCHOLARLY PUBLISHING. 2000 JAN 31 2 87 95
Lundberg, GD. The ethics and practice of authorship and peer review ACTA PHARMACOLOGICA SINICA. 1998 NOV 19
Luukkonen, T. Research evaluation in Europe: state of the art
RESEARCH EVALUATION. 2002 AUG 11 2 81 84
Luzi, D. E-print archives: a new communication pattern for grey literature INTERLENDING & DOCUMENT SUPPLY. 1998 26 3 130 +
MacDonald, R. J. 1989 Research - basic or applied? Australian Physicist.26, 83.


Madden, AD. When did peer review become anonymous? Comment ASLIB PROCEEDINGS. 2000 SEP 52 8 273 276


Magnus, D. Politics and peer review AMERICAN JOURNAL OF BIOETHICS. 2004 WIN 4 1 VII VIII


Mahoney M (1977) Publication prejudices: An experimental study of confirmatory bias in the peer review system. Cognitive Therapy and Research 1, 161-175. EXR


Maisonneuve, H, Dutson, E, Marescaux, J. The virtual university MINIMALLY INVASIVE THERAPY & ALLIED TECHNOLOGIES. 2002 MAR 1 11 2 61 65
Malcolm, L, Wright, L, Seers, M, Guthrie, J. An evaluation of pharmaceutical management and budget holding in Pegasus Medical Group NEW ZEALAND MEDICAL JOURNAL. 1999 MAY 14 112 1087 162 164 
Manwell, C. 1979 Peer review: a case history from the Australian Research Grants Committee. Search. 10, 81-86. 
Maricic, S, Sorokin, B, Papes, Z. Croatian journals at the end of the 20 century - A bibliometric evaluation DRUSTVENA ISTRAZIVANJA. 2000 9 1 1 17 
Marin, BV, Diaz, RM. Collaborative HIV Prevention Research in Minority Communities Program: A model for developing investigators of color PUBLIC HEALTH REPORTS. 2002 MAY-JUN 117 3 218 230 
Marsh, HW, Bazeley, P. Multiple evaluations of grant proposals by independent assessors: Confirmatory factor analysis evaluations of reliability, validity, and structure MULTIVARIATE BEHAVIORAL RESEARCH. 1999 34 1 1 30
Martin, F. The economic impact of Canadian university R & D. RESEARCH POLICY. 1998 NOV 27 7 677 687
Martin, N. Referee! Two critical reports question the standing of peer review. PSYCHOLOGIST. 2000 DEC 13 12 627 627
Martin, PA. Research peer review: A committee when none is required. APPLIED NURSING RESEARCH. 1998 MAY 11 2 90 92
Martin-BR, "The Use of Multiple Indicators in the Assessment of Basic Research", SCIENTOMETRICS, 1996, Vol 36, Iss 3, pp 343-362
Marton, J. (1983). Causes of Low and High Citation Potentials in Science: Citation Analysis of Biochemistry and Plant Physiology Journals. Journal of the American Society for Information Science, 34 (4), 244-246
Marts, SA. Interdisciplinary research is key to understanding sex differences: Report from the society for women's health research meeting on understanding the biology of sex differences JOURNAL OF WOMENS HEALTH & GENDER-BASED MEDICINE. 2002 JUL-AUG 11 6 501 509

Martsolf, DS, Dieckman, BC, Cartechine, KA, Starr, PJ, Wolf, LE, Anaya, ER. Peer review of teaching: Instituting a program in a College of Nursing JOURNAL OF NURSING EDUCATION. 1999 OCT 38 7 326 332

Martsolf, DS, Dieckman, BC, Heiss, MA. Cultural factors related to the peer review of teaching NURSING CONNECTIONS. 1998 FAL 11 3 41 54


Masood-E, "Peer-Review Deemed Essential for Blue Skies Award Scheme", NATURE, 1996, Vol 382, Iss 6591, pp 483-483


Maxwell, S, Christiansen, K. 'Negotiation as simultaneous equation': building a new partnership with Africa INTERNATIONAL AFFAIRS. 2002 JUL 78 3 477 +


Mccahy-P, "Usefulness of Comments of Previous Peer Reviewers", BRITISH MEDICAL JOURNAL, 1994, Vol 309, Iss 6958, pp 880-880


McCarthy-P, "Critics Sharpen Assault on Peer-Review - Some Even Recommend Abolishing the Traditional Process Used by Science
McCarty, R. Science, politics, and peer review - An editor's dilemma. SCIENTIST, 1994, Vol 8, Iss 11, pp 1+


Mccaul, KD, Johnson, RJ, Rothman, AJ. The effects of framing and action instructions on whether older adults obtain flu shots. HEALTH PSYCHOLOGY. 2002 NOV 21 6 624 628


McDowell, SA. Internal conflict and the peer review process - A case study. JOURNAL OF INFORMATION ETHICS. 1998 FAL 7 2 68 77


McGourty, J. Using multisource feedback in the classroom: A computer-based approach IEEE TRANSACTIONS ON EDUCATION. 2000 MAY 43 2 120 124
McKay, S. Quantifying quality: Can quantitative data ("metrics") explain the 2001 RAE ratings for social policy and administration? SOCIAL POLICY & ADMINISTRATION. 2003 OCT 37 5 444 467
McLellan F. Peer-review meeting participants urge greater accountability. Lancet. 2001 Sep 22;358(9286):991.
Medical Research Council UK Assessing the Medical Research Council's Peer Review System Medical Research Council UK MRC Bibliometric Analyses Pilot Study, 1999 http://www.mrc.ac.uk/index/funding/funding-specific_schemes/funding-evaluation_of_schemes/funding-bibliometric_analyses_pilot_study.htm
Meehan, TP, Weingarten, SR, Holmboe, ES, Mathur, D, Wang, Y,
Petrillo, MK, Tu, GS, Fine, JM. A statewide initiative to improve the care of hospitalized pneumonia patients: The Connecticut pneumonia pathway project AMERICAN JOURNAL OF MEDICINE. 2001 AUG 15 111 3 203 210
Meisel, A, Snyder, L, Quill, T. Seven legal barriers to end-of-life care - Myths, realities, and grains of truth JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION. 2000 NOV 15 284 19 2495 2501
Melia, RP, Pledger, C, Wilson, R. Disability and rehabilitation research - Opportunities for participation, collaboration, and extramural funding for psychologists AMERICAN PSYCHOLOGIST. 2003 APR 58 4 285 288
Menz, HB. Publication patterns and perceptions of the Australian podiatric medical faculty JOURNAL OF THE AMERICAN PODIATRIC MEDICAL ASSOCIATION. 2001 APR 91 4 210 218
and academic science in the emergence and legitimation of a new global model of sequence stratigraphy SOCIOLICAL QUARTERLY. 2002 SPR 43 2 307 334
Milbank, D. 1990 Scientists have to beat the bushes for money to stay inbusiness. The Wall Street Journal. 116, A1, A8.
Milburn, LAS, Brown, RD, Mulley, SJ, Hilts, SG. Assessing academic contributions in landscape architecture LANDSCAPE AND URBAN PLANNING. 2003 JUL 15 64 3 119 129
Mills, ME. Authors of the world unite: Liberating academic content from publishers' restrictions - Book review of Tenopir & King on Electronic Journals PSYCOLOQUY. 2001 12 21 1 3

Page 227
Moos, MK, Bartholomew, NE, Lohr, KN. Counseling in the clinical setting to prevent unintended pregnancy: an evidence-based research agenda CONTRACEPTION. 2003 FEB 67 2 115 132
Moran, G. Peer review and academic paradigms - Referees and information ethics JOURNAL OF INFORMATION ETHICS. 1998 FAL 7 2 19 29
Morin, KH, Ashton, KC. A replication study of experienced graduate nurse faculty orientation offerings and needs. JOURNAL OF NURSING EDUCATION. 1998 OCT 37 7 295 301
Morse, JM. Considering the "peer" in peer review. QUALITATIVE HEALTH RESEARCH. 2002 MAY 12 5 579 580
Mota, M. Reviewing peer review. SCIENTIST. 2003 MAR 24 17 6 19 19
Mruck, K, Mey, G. Between printed past and digital future. RESEARCH IN SCIENCE EDUCATION. 2002 32 2 257 268
Munk-Jorgensen, P. Peer review in health sciences. ACTA PSYCHIATRICA SCANDINAVICA. 2001 OCT 104 4 319 319
Murray, FB, Raths, J. Factors in the peer review of reviews. REVIEW OF EDUCATIONAL RESEARCH. 1996 WIN 66 4 417 421
Musacchio-JM, "American Science in Crisis - The Need to Revise the NIH Funding Policy", FASEB JOURNAL, 1994, Vol 8, Iss 10, pp 679-683
Nabitz, U, Klazinga, N, Walburg, J. The EFQM excellence model: European and Dutch experiences with the EFQM approach in health care INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE. 2000 JUN 12 3 191 201
Naiburg, S. Mentors at the gate: Editors talk about clinical writing for journal publication CLINICAL SOCIAL WORK JOURNAL. 2003 FAL 31 3 295 313
Naleppa, MJ. Families and the institutionalized elderly: A review JOURNAL OF GERONTOLOGICAL SOCIAL WORK. 1996 27 1-2 87 111
National Science Foundation 1989 Annual Report.
National Science Foundation Grant Policy Manual.
National Science Foundation Grant Proposal Guide - Chapter on "Review Criteria", in, National Science Foundation. 2002.
Needleman H L (1992) Salem comes to the National Institutes of Health: Notes from inside the crucible of scientific integrity. Pediatrics 90, No. 6, 977-981.


Nelson, ML. A model for scholarship in nursing: The case of a private liberal arts college NURSING OUTLOOK. 2001 SEP-OCT 49 5 217 222

Nelson, S. Engineering and technology student perceptions of collaborative writing practices IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION. 2003 DEC 46 4 265 276


Neuhauser, D. Peer review and the research commons - A problem of success MEDICAL CARE. 1997 APR 35 4 301 302


Newcombe, NS. Five commandments for APA AMERICAN PSYCHOLOGIST. 2002 MAR 57 3 202 205


NHMRC Guidelines fit the conduct of grant review panels 2002. 17.
2002.
Nijm, LM. Pitfalls of peer review - The limited protections of state and federal peer review law for physicians JOURNAL OF LEGAL MEDICINE. 2003 DEC 24 4 541-556
NIST, "Annual Report, 1990," Visiting Committee on Advanced
Noah, L. Sanctifying scientific peer review: Publication as a proxy for regulatory decision making. UNIVERSITY OF PITTSBURGH LAW REVIEW. 1998 SPR 59 3 677 717
Noyons, E. Bibliometric mapping of science in a science policy context SCIENTOMETRICS. 2001 JAN 50 1 83 98
O'Brien P, Wakley G. Peer review at the british journal of
Okamura, K, Bernstein, J, Fidler, AT. Assessing the quality of infertility resources on the world wide web: Tools to guide clients through the maze of fact and fiction JOURNAL OF MIDWIFERY & WOMENS HEALTH. 2002 JUL-AUG 47 4 264 268
Olsson, G. Women in the university - guests at the men's table. NORDISK PSYKOLOGI. 1999 MAR 51 1 59 76
O'Neill, P. Analyzing the impact of childhood sexual encounters - Comment on the article by Forouzan and Van Gijseghem CANADIAN PSYCHOLOGY-PSYCHOLOGIE CANADIENNE. 2004 FEB 45 1 81 82
Opthof T, Coronel R, Janse MJ. The significance of the peer review process against the background of bias: priority ratings of reviewers and editors and the prediction of citation, the role of geographical bias. Cardiovasc Res. 2002 Dec;56(3):339-46. Review.
Ormerod, RJ. Is content analysis either practical or desirable for research evaluation? OMEGA-INTERNATIONAL JOURNAL OF MANAGEMENT SCIENCE. 2000 APR 28 2 241 245
Orntoft TF, Nielsen OS. [Comment to the article "Nepotism in peer review of applications for research grants"] Ugeskr Laeger. 2003 Nov 24;165(48):4638-40. Danish.
Osborne, MD, Brady, DJ. The room is long and narrow RESEARCH IN SCIENCE EDUCATION. 2002 32 2 163 169
the Reports of Apas Peer-Review Panel in the Congressional Review
of the USS Iowa Incident", JOURNAL OF PERSONALITY ASSESSMENT,
1993, Vol 61, Iss 3, pp 425-442
Oullier, O, Michel-Kerjan, E. Does scientific publication need a
peer consensus? Commentary on Stodolsky on consensus-journals
PSYCOLOQUY. 2001 12 22 1 4
Overberger, C. G. Universities and the Federal government: a
marriage that has survived. Chemical & Engineering News 56(43),
Overmann-RJ, "Peerless Science - Peer-Review and United-States
Science Policy, by D.E. Chubin, E.J. Hackett", ISIS, 1991, Vol 82,
Iss 314, pp 711-712
Paar, GH. The quality assurance programme of the social security
insurance authorities. Perspective of a user in a Psychosomatic
hospital PSYCHOTHERAPEUT. 1997 MAY 42 3 156 162
Pabalan-R.T., "Report on the Peer Review of the Sorption Modeling
for High-Level Waste Performance Assessment Research Project."
Southwest Research Inst., San Antonio, TX. Center for Nuclear
Pace F. Secrecy and openness in peer review. Dig Liver Dis. 2000
Jan-Feb;32(1):71. .
Paces-V Viklicky-V, "Peer-Review in the Czech Republic", SCIENCE,
1993, Vol 262, Iss 5137, pp 1193-1193
Paddock-R.A.; Lerman-A.; Ditmars-J.D.; Macdonald-D.D.; Peerenboom-
J.P., "Radioactive Waste Isolation in Salt: Peer Review of the
Office of Nuclear Waste Isolation's Draft Report on a Multifactor
Test Design to Investigate Uniform Corrosion of Low-Carbon Steel.
85p.
Page RL, Harrison BD. Interdepartmental peer review. BMJ. 1997
Mar 15;314(7083):765-6. .
Page-RL Harrison-BDW, "Setting Up Interdepartmental Peer-Review -
The British- Thoracic-Society Scheme", JOURNAL OF THE ROYAL
COLLEGE OF PHYSICIANS OF LONDON, 1995, Vol 29, Iss 4, pp 319-324
Paice E. How to write a peer review. Hosp Med. 2001
Palca, J. 1990a NSF: hard times amid plenty. Science. 248, 541-
543.
Palca, J. 1990b Researchers declare crisis, seek funding
Palca, J. 1990c Young investigators at risk. Science. 249, 351-
353.
Science.251, 153-154.
257, 1620.
Palli-D Confortini-M Biggeri-A Russo-A Caragiagi-P , Carozzi-
F Minuti-PA, "A Quality-Control System Involving Peer-Review of
Abnormal Cervical Smears", CYTOPATHOLOGY, 1993, Vol 4, Iss 1, pp
Page 237
Program Planning. 16,11-23.
Paulus, K. How to survive peer review LEARNED PUBLISHING. 2002 OCT 15 4 312 312
Payne, AA. Measuring the effect of federal research funding on
private donations at research universities: Is federal research funding more than a substitute for private donations?

INTERNATIONAL TAX AND PUBLIC FINANCE. 2001 8 5-6 731 751


Pellizzoni, L. Reflexive modernization and beyond - Knowledge and value in the politics of environment and technology THEORY CULTURE & SOCIETY. 1999 AUG 16 4 99 +

Pemberton, M, Honey, M. Military peer review NATION. 1997 MAY 26 264 20 24 24


Peterson, MW, Galvin, JR, Dayton, C, D'Alessandro, MP. Realizing the promise - Delivering pulmonary continuing medical education over the Internet CHEST. 1999 MAY 115 5 1429 1436

Petruzzi, J. M. (1985). Peer Review in ANALYTICAL CHEMISTRY (Editor's Column). Analytical Chemistry, 57 (8), 868 A - 870 A

Petty, RE, Fleming, MA. The review process at PSPB: Correlates of interreviewer agreement and manuscript acceptance PERSONALITY AND SOCIAL PSYCHOLOGY BULLETIN. 1999 FEB 25 2 188 203

Peuskens-J, "Peer-Review of Risperidone in the Treatment of


Pierce, SJ. Silencing Scientists and Scholars in other fields: Power, paradigm controls, peer review and scholarly communication. LIBRARY & INFORMATION SCIENCE RESEARCH. 1999 21 3 415 417

Pilous B. What is a peer review organization anyway? What nurses need to know. Ohio Nurses Rev. 2003 Apr;78(4):5-7, 10-1. Review. .


Pitney, WA, Ilsley, P, Rintala, J. The professional socialization of certified athletic trainers in the National Collegiate Athletic Association Division I context JOURNAL OF ATHLETIC TRAINING. 2002 JAN-MAR 37 1 63 70


Plebani, M. Role of inspectors in external review mechanisms: criteria for selection, training and appraisal CLINICA CHIMICA ACTA. 2001 JUL 20 309 2

Plerou, V, Amaral, LAN, Gopikrishnan, P, Meyer, M, Stanley, HE. Similarities between the growth dynamics of university research and of competitive economic activities NATURE. 1999 JUL 29 400 6743 433 437

Plummer WP. Screening for depression in primary care. Scientific and statistical errors should have been picked up in peer review. BMJ. 2003 May 3;326(7396):982; author reply 982. .
Pollex, M. Silencing scientists and scholars in other fields: Power, paradigm controls, peer review, and scholarly communication JOURNAL OF SCHOLARLY PUBLISHING. 2000 JAN 31 2 106 110
Pomerantz, AM. Could peer review inadvertently lead to "educational malpractice"? Lessons from clinical psychology TEACHING OF PSYCHOLOGY. 2000 SPR 27 2 137 140
Poorterman, JHG, Van Weert, CM, Eijkman, MAJ. Quality assurance in dentistry: the Dutch approach INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE. 1998 AUG 10 4 345 350
Poschl, U. Interactive journal concept for improved scientific
publishing and quality assurance LEARNED PUBLISHING. 2004 APR
17 2 105 113
Posey LM. Changes in JAPhA's peer review system. J Am Pharm
Posner-KL Caplan-RA Cheney-FW, "Variation in Expert Opinion in
Medical Malpractice Review", ANESTHESIOLOGY, 1996, Vol 85, Iss 5,
pp 1049-1054
Pouris A. Peer review in scientifically small countries. R&D
Powell, TW, Muller, N, Ball, MJ. Electronic publishing:
opportunities and challenges for clinical linguistics and
phonetics CLINICAL LINGUISTICS & PHONETICS. 2003 JUN-AUG 17
4-5 421 426
Power-E, "Always a Need for Peer-Review", IRISH VETERINARY
Powers-T.J., "Filter Blank Contamination in Asbestos Abatement
Monitoring Procedures: Proceedings of a Peer Review Workshop Held
at Cincinnati, Ohio on April 24-25, 1986. ", Environmental
Protection Agency, Cincinnati, OH. Water Engineering Research
PPARC Astronomy Research Grants: Description of the Peer Review
Prager, D. Trends in Governmental and Academic Institutions
Prathap-G, "A Modest Proposal for Glasnost in the Peer-Review
Pravin Kumar E. Peer review and appeal: flawed but trusted?
Search. 22,101-103.
Pressman S (1994) Simultaneous multiple journal submissions: The
case against. American Journal of Economics and Sociology 53,
316-333.
Preston, JA, Grady, JN, Schulz, AF, Petrillo, MK, Scinto, JD. The
impact of a physician intervention program on older women's
mammography use EVALUATION & THE HEALTH PROFESSIONS. 1998 DEC
21 4 502 513
(pp 469-470). New York: Americana Corporation
Price, D. J. d. S. 1978 Endless frontier or bureaucratic morass?
Daedalus.107, 75-92.
Procedures for Peer Review Assessments. Office of Energy
Research, DOE/ER-0357.
Prout-DM, "Checks and Balances in Peer-Review - Advice from the
Patrick Case", ANNALS OF INTERNAL MEDICINE, 1988, Vol 109, Iss 9,
pp 689-690.
Pruthi S, Jain A, Wahid A, Mehra K, Nabi S. Scientific community
and peer review system - a case study of a central government
funding scheme in India. J Scientific Industrial Res 1997;56:398-
Quinlan, KM, Akerlind, GS. Factors affecting departmental peer collaboration for faculty development: Two cases in context HIGHER EDUCATION. 2000 JUL 40 1 23 52
Quinlan, KM. Inside the peer review process: how academics review a colleague's teaching portfolio TEACHING AND TEACHER EDUCATION. 2002 NOV 18 8 1035 1049
Radaelli, CM. The code of conduct against harmful tax competition: Open method of coordination in disguise? PUBLIC ADMINISTRATION. 2003 81 3 513 531
Ramasarma-T, "The Peer Reviewer Should Be Fair, Secret or Otherwise", FASEB JOURNAL, 1995, Vol 9, Iss 14, pp 1500-1500
Rangel, S. J., Efron, B. and Moss, R. L. Recent Trends in National Institutes of Health funding of surgical research. Lippincott Williams & Wilkins online.

Reed, M, Mitchell, B. Using information technologies for collaborative learning in geography: a case study from Canada. JOURNAL OF GEOGRAPHY IN HIGHER EDUCATION. 2001 NOV 25 3 321 339


Reinhardt, UE. Making economic evaluations respectable. SOCIAL SCIENCE & MEDICINE. 1997 AUG 45 4 555 562

Reinking, D, Alvermann, DE. The RRQ peer-review process. READING RESEARCH QUARTERLY. 2003 APR-JUN 38 2 168 171


Relyea, HC. Silencing scientists and scholars in other fields: Power, paradigm controls, peer review, and scholarly communication. GOVERNMENT INFORMATION QUARTERLY. 1999 16 2 193 195


Rennie D. Fourth International Congress on Peer Review in


Rennie, D. The Development and Rationale of Peer Review. Peer Review in Health Sciences,


Report of the Managing Editor of the 'American Political Science Review', 1986-87. PS, 20, 1006-1016

Restuccia, JD, Shwartz, M, Kreger, BE, Payne, SMC, Ash, AS, Iezzoni, LI, Heineke, J, Selker, HP, Gomes, T, Labonte, A, Butterly, JR. Does more "appropriateness" explain higher rates of cardiac procedures among patients hospitalized with coronary heart disease? MEDICAL CARE. 2002 JUN 40 6 500 509


Reynolds, CF, Martin, C, Brent, D, Ryan, N, Dahl, RE, Pilkonis, P
Marcus, MD, Kupfer, DJ. Postdoctoral clinical-research training in psychiatry - A model for teaching grant writing and other research survival skills and for increasing clarity of mentoring expectations ACADEMIC PSYCHIATRY. 1998 FAL 22 3 190-196


Ribbons, B, Vance, S. Using e-mail to facilitate nursing scholarship COMPUTERS IN NURSING. 2001 MAY-JUN 19 3 105-110


Richardson, JV. The peer review process: Acceptances, revisions, and outright rejections LIBRARY QUARTERLY. 2002 JAN 72 1 V XI


Rinia, EJ, van Leeuwen, TN, van Vuren, HG, van Raan, AFJ. Comparative analysis of a set of bibliometric indicators and central peer review criteria - Evaluation of condensed matter physics in the Netherlands RESEARCH POLICY. 1998 MAY 27 1 95 107

Rinia, EJ, van Leeuwen, TN, van Vuren, HG, van Raan, AFJ. Influence of interdisciplinarity on peer-review and bibliometric evaluations in physics research RESEARCH POLICY. 2001 MAR 30 3 357 361

Ritzler, B, Erard, R, Pettigrew, G. A final reply to Grove and Barden - The relevance of the Rorschach comprehensive system for expert testimony PSYCHOLOGY PUBLIC POLICY AND LAW. 2002 JUN 8 2 235 246
Roberts, LW, Geppert, C, Connor, R, Nguyen, K, Warner, TD. An invitation for medical educators to focus on ethical and policy issues in research and scholarly practice ACADEMIC MEDICINE. 2001 SEP 76 9 876 885
Roberts, LW. On the centrality of peer review ACADEMIC PSYCHIATRY. 2002 WIN 26 4 221 222
Rodrigues, R. Care in the community - Continuous quality improvement within community care services: A peer review project AUSTRALASIAN JOURNAL ON AGEING. 2000 FEB 19 1
Roediger, H L (1987) The role of journal editors in the scientific...


Rojo, JM. Peer review - Experiences at National and European level SCIENTOMETRICS. 1999 JUL-AUG 45 3 497 500


Romano PE. Peer review; cursing presbyopia; DUI and dealing with the law; Cadillac II: Robocop. Binocul Vis Strabismus Q. 2000 Winter;15(4):372-3


Rosen, GM. Litigation and reported rates of posttraumatic stress disorder PERSONALITY AND INDIVIDUAL DIFFERENCES. 2004 APR 36 6 1291 1294

Rosen, RC. Assessment of female sexual dysfunction: review of validated methods FERTILITY AND STERILITY. 2002 APR 77 4


Rosenfeld, MR. Contracts: Third Circuit upholds limited immunity of peer review actions under HCQIA JOURNAL OF LAW MEDICINE & ETHICS. 1999 SUM 27 2 200 201


Rosenzeig, R. 1992a Issues in American Science Policy. In Research and Higher Education. The United Kingdom and the United
Ross, CF, Wall, CE. Mammalian feeding and primate evolution: An overview. AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY. 2000 AUG 112 4 449 453
Rossiter, J.R. Qualifying the importance of findings. JOURNAL OF BUSINESS RESEARCH. 2003 JAN 56 1 85 88
Roth, WM, Tobin, K. Peer review in science education: An introduction. RESEARCH IN SCIENCE EDUCATION. 2002 32 2 127 134
Roth, WM. Editorial power/authorial suffering. RESEARCH IN SCIENCE EDUCATION. 2002 32 2 215 240
Rothblum, ED. Contradictions and confounds in coverage of obesity: Psychology journals, textbooks, and the media. JOURNAL OF SOCIAL ISSUES. 1999 SUM 55 2 355 369
Rothwell, PM, Martyn, CN. Reproducibility of peer review in clinical neuroscience - Is agreement between reviewers any greater than would be expected by chance alone? BRAIN. 2000 SEP 123
Rowland, F. The peer-review process. LEARNED PUBLISHING. 2002 OCT 15 4 247 258
Roworth, WW. Pandora's dilemma: Some reflections on peer review. ACADEME-BULLETIN OF THE AAUP. 1997 MAY-JUN 83 3 35 38
Roy, R. 1981 An Alternative Funding Mechanism. Science 211, 1377
Roy, S, Nagpaul, PS, Mohapatra, PK. Developing a model to measure the effectiveness of research units INTERNATIONAL JOURNAL OF OPERATIONS & PRODUCTION MANAGEMENT. 2003 23 11-12 1514 1531
Rudy, EB, Kerr, ME. Auditing research studies NURSING RESEARCH. 2000 MAR-APR 49 2 117 120
Rugg-DL Levinson-R Diclemente-R Fishbein-M, "Centers-for-Disease-Control and Prevention Partnerships with External Behavioral and Social-Scientists - Roles, Extramural Funding, and Employment", AMERICAN PSYCHOLOGIST, 1997, Vol 52, Iss 2, pp 147-


Russo, E. Bypassing peer review SCIENTIST. 2000 MAR 6 14 5 1 +

Russo, E. NIH to revise the peer review process SCIENTIST. 1999 AUG 30 13 17 5 5


Salinas, AM, Coria, I, Reyes, H, Zambrana, M. Effect of quality of care on preventable perinatal mortality INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE. 1997 APR 9 2 93 99


Samet, JM, Burke, TA. Turning science into junk: The tobacco industry and passive smoking AMERICAN JOURNAL OF PUBLIC HEALTH. 2001 NOV 91 11 1742 1744

Sangalli, A. 1995 A challenge to the cake-slicers. New
Scientist. 146, 50-51.
Sattin, RW. The prevention research initiative and the peer review process at CDC PUBLIC HEALTH REPORTS. 2001 116
Scantlebury, K. A snake in the nest or in a snake's nest: What counts as peer review for a female science educator in a chemistry department? RESEARCH IN SCIENCE EDUCATION. 2002 32 2 157 162
SCEET. 1994 Senate Standing Committee on Employment, Education and Training. Inquiry into the organisation and funding of research in higher education. Canberra: AGPS.
Schauffler, HH, Mordavsky, JK. Consumer reports in health care: Do they make a difference? ANNUAL REVIEW OF PUBLIC HEALTH. 2001 22 69 89
Scheutzow, SO. State medical peer review: High cost but no benefit - Is it time for a change? AMERICAN JOURNAL OF LAW & MEDICINE. 1999 25 1 7 60
Schilling, J, Cranovsky, R, Straub, R. Quality programmes, accreditation and certification in Switzerland INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE. 2001 APR 13 2 157 161
Dec;126(6):1683-5. 
Sen, BK, Pandalai, TA, Karanjai, A. Ranking of scientists - A new approach JOURNAL OF DOCUMENTATION. 1998 DEC 54 5 622 628
Shabot, MM, Polaschek, JX, Duncan, RG, Langberg, ML, Jones, DT. A novel governance system for enterprise information services JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION. 1999
Sharma, G. & Sanyal, B. 1990 Funding mechanism of thrust areas of higher education in India. Paris:
Shashok, K. Author's editors: facilitators of science information transfer LEARNED PUBLISHING. 2001 APR 14 2 113 121
Shaw, CD. External quality mechanisms for health care: summary of the ExPeRT project on visitatie, accreditation, EFQM and ISO assessment in European Union countries INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE. 2000 JUN 12 3 169 175
The Chronicle of Higher Education 42 (August 16), A12 & A17.
Shea, B, Moher, D, Graham, I, Pham, B, Tugwell, P. A comparison of
the quality of Cochrane reviews and systematic reviews
published in paper-based journals EVALUATION & THE HEALTH
PROFESSIONS. 2002 MAR 25 1 116 129
Sheahan SL, Simpson C, Rayens MK. Nurse practitioner peer
Mar;13(3):140-5.

ASSOCIATION, 1990, Vol 264, Iss 24, pp 3142-3142
Shields-WD, "Peer-Review of Expert Medical-Legal Testimony - A
Proposal for Child Neurology", JOURNAL OF CHILD NEUROLOGY, 1992,
Vol 7, Iss 3, pp 237-239
Shields-WD, "Peer-Review of Medical Expert Witnesses - Reply",
JOURNAL OF CHILD NEUROLOGY, 1994, Vol 9, Iss 2, pp 217-217
Shiflett, L. Editorial peer review: Its strengths and weaknesses.
COLLEGE & RESEARCH LIBRARIES. 2002 MAR 63 2 205 206
Confidentiality of Assessment by Peers Guarantee the Quality of
policy and national goals. Cambridge, Mass.: The MIT Press.
Shimp CP. Ambiguity, logic, simplicity, and dynamics:
Wittgensteinian evaluative criteria in peer review of quantitative
research on categorization. Behav Processes. 2004 Jun
Ship-JA Duffy-V Jones-JA Langmore-S, "Geriatric Oral Health and
Its Impact on Eating", JOURNAL OF THE AMERICAN GERIATRICS SOCIETY,
1996, Vol 44, Iss 4, pp 456-464
Short-TG Oregan-A Lew-J Oh-TE, "Critical Incident Reporting in
an Anesthetic Department Quality Assurance Program", ANAESTHESIA,
1993, Vol 48, Iss 1, pp 3-7
Shum, SB, Sumner, T. JIME: an interactive journal for interactive
media (Reprinted from First Monday, vol 6, 2001) LEARNED
PUBLISHING. 2001 OCT 14 4 273 285
SI Sp. Iss. SI 147 154
Review (Editor's Page). Radiology, 178, 637-642
Silagyi-CA Jewell-D Mant-D, "An Analysis of Randomized Controlled
Trials Published in the Us Family Medicine Literature, 1987-1991",
Silver MP, Antonow JA. Reducing medication errors in hospitals:
a peer review organization collaboration. Jt Comm J Qual Improv.
Silver-K, "The Yellowed Archives of Yellowcake", PUBLIC HEALTH
REPORTS, 1996, Vol 111, Iss 2, pp 116-127
Simmons, M. Review of Clinical Research. Center for Scientific
Singer, JD. Micromanagement and peer review PS-POLITICAL SCIENCE & POLITICS. 2001 MAR 34 1 1 1
Siow, A. Tenure and other unusual personnel practices in academia JOURNAL OF LAW ECONOMICS & ORGANIZATION. 1998 APR 14 1 152 173
Skootsky, SA, Slavin, S, Wilkes, MS. Attitudes toward managed care and cost containment among primary care trainees at 3 training sites AMERICAN JOURNAL OF MANAGED CARE. 1999 NOV 5 11 1397 1404
Sluijs, EM, Dekker, J. Diffusion of a quality improvement programme among allied health professionals INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE. 1999 AUG 11 4 337 344
Smith BR. Integrity of the peer review process. Can J Cardiol.


So, CYK. Citation ranking versus expert judgment in evaluating communication scholars: Effects of research specialty size and individual prominence SCIENTOMETRICS. 1998 MAR-APR 41 3 325 333
Sollenberger, JF. Editorial peer review: Its strengths and weaknesses. JOURNAL OF THE MEDICAL LIBRARY ASSOCIATION. 2002 JAN 90 1 114 116
Somit, A, Peterson, SA. Journal response time: A case for multiple submission - Commentary BEHAVIORAL AND BRAIN SCIENCES. 1996 SEP 19 3 533 534
Speck, W. B., "Publication Peer Review", Westport, CT.
Spier, RE.  Peer review and innovation SCIENCE AND ENGINEERING ETHICS. 2002 JAN 8 1 99 108
Spinney, L.  UK autism fracas fuels calls for peer review reform NATURE MEDICINE. 2004 APR 10 4 321 321
Squires-BP, Elmslie-TJ, "Reports of Case-Control Studies - What Editors Want from Authors and Peer Reviewers", CANADIAN MEDICAL ASSOCIATION JOURNAL, 1990, Vol 143, Iss 1, pp 17-18
Squires-BP, Elmslie-TJ, "Reports of Randomized Controlled Trials - What Editors Want from Authors and Peer Reviewers", CANADIAN MEDICAL ASSOCIATION JOURNAL, 1990, Vol 143, Iss 5, pp 381-382
Squires-BP, "Biomedical Review Articles - What Editors Want from Authors and Peer Reviewers", CANADIAN MEDICAL ASSOCIATION JOURNAL, 1989, Vol 141, Iss 3, pp 195-197
Squires-BP, "Descriptive Studies - What Editors Want from Authors and Peer Reviewers", CANADIAN MEDICAL ASSOCIATION JOURNAL, 1989, Vol 141, Iss 9, pp 879-880
Squires-BP, "Editorials and Platform Articles - What Editors Want from Authors and Peer Reviewers", CANADIAN MEDICAL ASSOCIATION JOURNAL, 1989, Vol 141, Iss 7, pp 666-667
Stanton, WR, Willis, M, Balanda, KP. Development of an instrument for monitoring adolescent health issues HEALTH EDUCATION RESEARCH. 2000 APR 15 2 181 190
Steve son, L. Peer review and peer support: Implementing quality improvement in home and community care (HACC) services AUSTRALIAN JOURNAL ON AGEING. 1998 NOV 17 4 207 207
Stine-M.D., "Position paper - peer review and design verification of selected activities.", Westinghouse Hanford Co., Richland, WA., Report Number WHCSDW236ATI014, Sep 94. 12p.
Stodolsky, D. Scientific publication needs a peer consensus - Reply to Oullier and Michel-Kerjan on Stodolsky on consensus-journals PSYCHOLOGY. 2002 13 2 1 3
Stricker, RB, Goldberg, B. Time for peer review reform SCIENTIST. 2000 APR 17 14 8 6 6
Stromholm, S. Peer review - Experience at National and European level SCIENTOMETRICS. 1999 JUL-AUG 45 3 491 495
Sugden-D, "Funding UK Research - Comment on the Report to the Abrc
Swan, A, Brown, S. Authors and electronic publishing: what authors want from the new technology LEARNED PUBLISHING. 2003 JAN 16 1 28 33
Swierc, SF, Routh, DK. Introduction to the Special Issue on International Clinical Psychology JOURNAL OF CLINICAL PSYCHOLOGY. 2003 JUN 59 6 631 634
Sze-PY, "Problems with Peer-Review", TECHNOLOGY REVIEW, 1992, Vol 95, Iss 2, pp 8-8
Talbott-SP, "Peer-Review Drives Compensation at Johnsonville", PERSONNEL JOURNAL, 1994, Vol 73, Iss 10, pp 126+
Tamarelli-AW, "Daubert V Merrell-Dow-Pharmaceuticals - Pushing the Limits of Scientific Reliability - The Questionable Wisdom of Abandoning the Peer-Review Standard for Admitting Expert Testimony", VANDERBILT LAW REVIEW, 1994, Vol 47, Iss 4, pp 1175- 
Page 264
1203
Tavernor-D, "Peer-Review", JOURNAL OF SMALL ANIMAL PRACTICE, 1992, Vol 33, Iss 9, pp 413-413
TAX., "Disclosure Allowed for Peer-Review, Illness, or Death", JOURNAL OF TAXATION, 1991, Vol 74, Iss 3, pp 164-
Taylor, JA, Dana, TM. Secondary school physics teachers' conceptions of scientific evidence: An exploratory case study JOURNAL OF RESEARCH IN SCIENCE TEACHING. 2003 OCT 40 8 721 736


The Wellcome Trust - Types of decision that may be taken by an advisory committee. The Wellcome Trust, http://www.wellcome.ac.uk/en/1/ 2002.

Theoharakis, V, Hirst, A. Perceptual differences of marketing journals: A worldwide perspective MARKETING LETTERS. 2002 NOV 13 4 389 402


Thomas, JW, Guire, KE, Horvat, GG. Is patient length of stay related to quality of care? HOSPITAL & HEALTH SERVICES ADMINISTRATION. 1997 WIN 42 4 489 507

Thomas, PR, Watkins, DS. Institutional research rankings via
bibliometric analysis and direct peer review: A comparative case study with policy implications SCIENTOMETRICS. 1998 MAR-APR 41 3 335 355


Thorpe, D. Medical library online, in, The Australian. 2002.


Tilden, V. Peer review: Evidence-based or sacred cow? NURSING RESEARCH. 2002 SEP-OCT 51 5 275 275


Tobin, K, Roth, WM. The contradictions in science education peer review and possibilities for change RESEARCH IN SCIENCE EDUCATION. 2002 32 2 269 280

Tobin, K. The multiple faces of peer review in science education RESEARCH IN SCIENCE EDUCATION. 2002 32 2 135 156


Tomlins, CL. Just one more 'zine? Maintaining and improving the scholarly journal in the electronic present: a view from the humanities LEARNED PUBLISHING. 2001 JAN 14 1 33 40

Tonks-A, "Reviewers Chosen by Authors", BRITISH MEDICAL JOURNAL, 1995, Vol 311, Iss 6999, pp 210-210Keywords Plus: PEER REVIEWERS


Toulmin, S. 1964 The complexity of scientific choice: a stock
taking. Minerva. 2, 343-359.
Townsend DW. Hospital peer review is a kangaroo court. Med Econ. 2000 Feb 7;77(3):133-6, 141.
Traversa G. [Which peer review is appropriate for studies carried out during a judicial inquiry?] Epidemiol Prev. 2001 Jan-Feb;25(1):27. Italian.
Tuckett, D. Evaluating psychoanalytic papers - Towards the development of common editorial standards INTERNATIONAL JOURNAL OF PSYCHO-ANALYSIS. 1998 JUN 79
Turner, R. An end to great publishing myths - Book review of Tenopir and King on electronic-journals PSYCOLOQUY. 2001 12 12 1 4
Tyrer P. Chairman’s action; the importance of executive decisions in peer review. Behav Brain Sci 1991;14:164-5.
UIPC, "Louisiana Department of Natural Resources, Office of
Valdez-Martinez, E, Garduno-Espinosa, J, Martinez-Garcia, MDC, Lavielle-Sotomayor, P, Amato-Martinez, JD, Blanco-Favela, F, Munoz-Hernandez, O. Satisfaction related factors of attendants to a medical research meeting REVISTA DE INVESTIGACION CLINICA. 2000 MAY-JUN 52 3 261 265
Van den Beemt, F.C.H.D, "The Right Mix: Review by Peers as well as
Van Lange, PAM. Why authors believe that reviewers stress limiting aspects of manuscripts: The SLAM effect in peer review JOURNAL OF APPLIED SOCIAL PSYCHOLOGY. 1999 DEC 29 12 2550 2566
Van Raan, AFJ. The pandora's box of citation analysis: Measuring scientific excellence - The last evil? ASIST MONOGRAPH SERIES. 2000 301 319
van Rooyen, S. A critical examination of the peer review process LEARNED PUBLISHING. 1998 JUL 11 3 185 191
van Rooyen, S. The evaluation of peer-review quality LEARNED PUBLISHING. 2001 APR 14 2 8591
Van Weert, C. Developments in professional quality assurance towards quality improvement: some examples of peer review in the Netherlands and the United Kingdom INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE. 2000 JUN 12 3 239 242
Vanhemel-OJS Elferinkstinkens-PM Brand-R, "How to Compare and
Report Department Specific Mortality- Rates for Peer-Review Using the Perinatal Database of the Netherlands", EUROPEAN JOURNAL OF OBSTETRICS GYNECOLOGY AND , REPRODUCTIVE BIOLOGY, 1994, Vol 56, Iss 1, pp 1-7
Vaughn, TE, McCoy, KD, BootsMiller, BJ, Woolson, RF, Sorofman, B, Tripp-Reimer, T, Perlin, J, Doebbeling, BN. Organizational predictors of adherence to ambulatory care screening guidelines MEDICAL CARE. 2002 DEC 40 12 1172 1185
Venables, A, Summit, R. Enhancing scientific essay writing using peer assessment INNOVATIONS IN EDUCATION AND TEACHING INTERNATIONAL. 2003 AUG 40 3 281 290
Verhoeven, BH, Verwijnen, GM, Muijtjens, AMM, Scherpber, AUJA, van der Vleuten, CPM. Panel expertise for an Angoff standard setting procedure in progress testing: item writers compared to recently graduated students MEDICAL EDUCATION. 2002 SEP 36 9 860 867
Victoroff MS. Peer review of the inexpert witness, or ... do you trust chickens to guard the coop? Manag Care. 2002 Sep;11(9):14-6.
Viner, N, Powell, P, Green, R. Institutionalized biases in the award of research grants: a preliminary analysis revisiting the principle of accumulative advantage RESEARCH POLICY. 2004 APR 33 3 443 454
Vinkler, P. General performance indexes calculated for research institutes of the Hungarian Academy of Sciences based on scientometric indicators SCIENTOMETRICS. 1998 JAN-FEB 41 1-2 185 200
Visser HK. [The significance of publication in Dutch medical
Wager, E, Jefferson, T. Shortcomings of peer review in biomedical journals LEARNED PUBLISHING. 2001 OCT 14 4 257 263
Wagner, AK, Boninger, ML, Levy, C, Chan, L, Gater, D, Kirby, RL. Issues in physical medicine and rehabilitation AMERICAN JOURNAL OF PHYSICAL MEDICINE & REHABILITATION. 2003 OCT 82 10 790 802
Wagner, J. That was then and this is now: An economist's wish list for the LLRW siting paradigm NATURAL RESOURCES JOURNAL. 1998 FAL 38 4 635 649

Walsh, E, Rooney, M, Appleby, L, Wilkinson, G. Open peer review: a randomised controlled trial BRITISH JOURNAL OF PSYCHIATRY. 2000 JAN 176 47 51
Warner, J. A critical review of the application of citation studies to the Research Assessment Exercises JOURNAL OF INFORMATION SCIENCE. 2000 26 6 453 459
Washington, MJ. The practice of peer review in the international nuclear safety regime NEW YORK UNIVERSITY LAW REVIEW. 1997 MAY 72 2 430 469
Watson, LC, Pignone, MP. Screening accuracy for late-life depression in primary care: A systematic review JOURNAL OF FAMILY PRACTICE. 2003 DEC 52 12 956 964


Wegmann-R, "Continuing the Peer Reviewing - What Does It Mean - Foreword", CELLULAR AND MOLECULAR BIOLOGY, 1994, Vol 40, Iss 6, pp R3-R4


Weinberg, A. M. 1978 The obligations of citizenship in the republic of science. Minerva. 16, 1-3.


Weingart, SN, Iezzoni, LI, Davis, RB, Palmer, RH, Cahalane, M, Hamel, MB, Mukamal, K, Phillips, RS, Davies, DT, Banks, NJ. Use of administrative data to find substandard care - Validation of the complications screening program MEDICAL CARE. 2000 AUG 38 8 796 806


Weinmann, C. Quality improvement in health care - A brief history of the Medicare Peer Review Organization (PRO) initiative EVALUATION & THE HEALTH PROFESSIONS. 1998 DEC 21 4 413 418


Weller AC. Peer review is essential for both print and
Weller, A. Editorial peer review, its strengths and weaknesses (vol 23, pg 37, 2001) LIBRARY & INFORMATION SCIENCE RESEARCH. 2002 24 3 305 305
Weller, AC. Qualitative and quantitative measures of indexed health sciences electronic journals JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION. 2002 JUN 5 287 21 2865 2866
Wenneras, C, Wold, A. Nepotism and sexism in peer-review NATURE. 1997 MAY 22 387 6631 341 343
Wessely-S, "What Do We Know About Peer-Review", PSYCHOLOGICAL MEDICINE, 1996, Vol 26, Iss 5, pp 883-886
West, W. 'Defective' liaison barrier to industry, in, Australian. 1990.
Wets, K, Weedon, D, Velterop, J. Post-publication filtering and evaluation: Faculty of 1000 LEARNED PUBLISHING. 2003 OCT 16 4 249 258
Whalley, WB, Rea, BR. Two examples of the use of 'electronic posters' JOURNAL OF GEOGRAPHY IN HIGHER EDUCATION. 1998 NOV 22 3 413 417
Whatling P. Having non-medical readers of papers on internet will enhance peer review. BMJ. 1999 Apr 24;318(7191):1144-5.
White, MD. Editorial peer review, its strengths and weaknesses LIBRARY & INFORMATION SCIENCE RESEARCH. 2001 23 4 371 372
Whitworth J, Pettigrew A. Peer review in the Medical Research Council (MRC)
Wiemelt, J. Toward an activity-based conception of writing and school writing contexts WRITTEN COMMUNICATION. 2001 APR 18 2 107 179
Wilkes, MS. The public dissemination of medical research: Problems and solutions JOURNAL OF HEALTH COMMUNICATION. 1997 JAN-MAR 2 1 3 15
Williams, B. 1979  Education, Training and Employment. Canberra: AGPS.
Williamson, A. What will happen to peer review? LEARNED PUBLISHING. 2003 JAN 16 1 15 20
Williamson, E, Kent, J, Goodenough, T, Ashcroft, R. Social science gets the ethics treatment SOCIOCOLOGICAL RESEARCH ONLINE. 2002 NOV 30 7 4
Willmott, H. Commercialising higher education in the UK: the state, industry and peer review STUDIES IN HIGHER EDUCATION. 2003 28 2 129 141
Wilson, JR. Responsible authorship and peer review SCIENCE AND ENGINEERING ETHICS. 2002 APR 8 2 155 174
Winterhager (Hrsg.), Indikatoren der Wissenschaft und Technik: Theorie, Methoden, Anwendungen (S. 91-111). Frankfurt/Main: Campus


Wood, D, Hurst, P. Online peer review: perceptions in the biological sciences LEARNED PUBLISHING. 2000 APR 13 2 95 100

Wood, D. Electronic submission and peer review - An update on the ESPERE project LEARNED PUBLISHING. 1997 APR 10 2 157 159

Wood, D. Online peer review: current options LEARNED PUBLISHING. 2001 APR 14 2 151 158


Wood, DJ. Peer review and the Web: The implications of electronic peer review for biomedical authors, referees and learned society publishers JOURNAL OF DOCUMENTATION. 1998 MAR 54 2 173 197


Wood, F. Q., Meek, V. L. & Harman, G. H. 1992 The research grant application process. Learning from failure? Higher Education. 24,
Wood, M, Roberts, M, Howell, B.  The reliability of peer reviews of papers on information systems JOURNAL OF INFORMATION SCIENCE. 2004 30 1 2 11
Working Group on Review of Bioengineering and Technology and instrumentation development research, 7.
Wouters, P. 1994  Dutch science observed. EASST Newsletter. 13, 7-10.
Wouters-P, "Citation Cycles and Peer-Review Cycles", SCIENTOMETRICS, 1997, Vol 38, Iss 1, pp 39-55
Young, SN.  Peer review of manuscripts: theory and practice JOURNAL OF PSYCHIATRY & NEUROSCIENCE. 2003 SEP 28 5 327 330
Youngs, BL, Green, A. A successful peer writing assistant program FOREIGN LANGUAGE ANNALS. 2001 NOV-DEC 34 6 550 558

Zembal-Saul, C, Munford, D, Crawford, B, Friedrichsen, P, Land, S. Scaffolding preservice science teachers' evidence-based arguments during an investigation of natural selection RESEARCH IN SCIENCE EDUCATION. 2002 32 4 437 463


Zhang, YH, Yuan, YC, Jiang, YF. An international peer-review system for a Chinese scientific journal LEARNED PUBLISHING. 2003 APR 16 2 91 94


Zinn C. Australia proposes peer review of alleged negligence cases. BMJ. 2002 Sep 14;325(7364):565.


Zuckerman, H. 1987 Foreword. In Scientific Excellence (ed. J. P.
Zurer-PS, "NIH Peer Reviewers to Watch for High-Risk, High-Pay off Proposals", CHEMICAL & ENGINEERING
Zwarenstein M. Peer review of statistics in medical research. Journal reviewers are even more baffled by sample size issues than grant proposal reviewers. BMJ. 2002 Aug 31;325(7362):491; author reply 491.

[Anon]. The mysteries of the peer review process JOURNAL OF LITERACY RESEARCH. 2003 FAL 35 3 XI XIII
[Anon]. The mysteries of the peer review process JOURNAL OF LITERACY RESEARCH. 2003 FAL 35 3 XI XIII
[No authors listed] Peer review--reject, accept, or major
[No authors listed] Peer review--reject, accept, or major revision? Lancet Neurol. 2003 Sep;2(9):517. 
AAMC National Clinical Research Summit Unites Stakeholders, Sparks Legislative Agenda.

Listing of Most Cited References (First Author, Year Published, Volume, Page)
(Most cited by recent peer review articles retrieved from the Science Citation Index)

CICCHETTI DV, 1991, BEHAV BRAIN SCI, V14, P119
RUBIN HR, 1992, JAMA-J AM MED ASSOC, V267, P2349
GOLDMAN RL, 1992, JAMA-J AM MED ASSOC, V267, P958
HORROBIN DF, 1990, JAMA-J AM MED ASSOC, V263, P1438
PETERS DP, 1982, BEHAV BRAIN SCI, V5, P187
ZUCKERMAN H, 1971, MINERVA, V9, P66
WENNERAS C, 1997, NATURE, V387, P341
BOYER EL, 1990, SCHOLARSHIP RECONSID
DANIEL HD, 1993, GUARDIANS SCI FAIRNE
HAYWARD RA, 1993, ANN INTERN MED, V118, P550
MCNUTT RA, 1990, JAMA-J AM MED ASSOC, V263, P1371
VANROOYEN S, 1999, BRIT MED J, V318, P23
SEGLEN PO, 1997, BRIT MED J, V314, P498
FISHER M, 1994, JAMA-J AM MED ASSOC, V272, P143
CHUBIN DE, 1990, PEERLESS SCI PEER RE
LOCK S, 1985, DIFFICULT BALANCE ED
HARNAD S, 1996, SCHOLARLY PUBL, P103
KASSIRER JP, 1994, JAMA-J AM MED ASSOC, V272, P96
SMITH R, 1999, BRIT MED J, V318, P4
GARFIELD E, 1986, CURRENT CONTENTS, V3, P3
GOODMAN SN, 1994, ANN INTERN MED, V121, P11
LANDIS JR, 1977, BIOMETRICS, V33, P159
MORAN G, 1998, SILENCING SCI SCHOLA
JENC KS SF, 1992, JAMA-J AM MED ASSOC, V268, P900
MAHONEY MJ, 1977, COGNITIVE THERAPY RE, V1, P161
SILBERG WM, 1997, JAMA-J AM MED ASSOC, V277, P1244
VANRAAN AFJ, 1996, SCIENTOMETRICS, V36, P397
GILBERT JR, 1994, JAMA-J AM MED ASSOC, V272, P139
FISKE DW, 1990, AM PSYCHOL, V45, P591
COLE S, 1981, SCIENCE, V214, P881
BAILAR JC, 1985, NEW ENGL J MED, V312, P654
BRENNAN TA, 1991, NEW ENGL J MED, V324, P370
BLACK N, 1998, JAMA-J AM MED ASSOC, V280, P231
LINDSEY D, 1978, SCI PUBLICATION SYST
LOCALIO AR, 1996, ANN INTERN MED, V125, P457
MARSH HW, 1989, J EXP EDUC, V57, P151
HARNAD S, 1990, PSYCHOL SCI, V1, P342
KRONICK DA, 1990, JAMA-J AM MED ASSOC, V263, P1321
NYLENNA M, 1994, JAMA-J AM MED ASSOC, V272, P149
HORROBIN DF, 1996, LANCET, V348, P1293
GOTTFREDSON SD, 1978, AM PSYCHOL, V33, P920
WILKES MS, 1995, J GEN INTERN MED, V10, P443
SMITH MA, 1997, JAMA-J AM MED ASSOC, V278, P1573
TRAVIS GDL, 1991, SCI TECHNOL, V16, P322
SCHU TZW SO, 1999, AM J LAW MED, V25, P7
STEPHENSON J, 1997, JAMA-J AM MED ASSOC, V278, P1389
WELLER AC, 2001, EDITORIAL PEER REVIE
WELLER AC, 1990, JAMA-J AM MED ASSOC, V263, P1344
GARDNER MJ, 1990, JAMA-J AM MED ASSOC, V263, P1355
ALT MAN LK, 1996, LANCET, V347, P1382
BRENNAN TA, 1989, MED CARE, V27, P1148
ARMSTRONG JS, 1997, SCI ENG ETHICS, V3, P63
GLANTZ SA, 1994, JAMA-J AM MED ASSOC, V272, P114
GARFIELD E, 1972, SCIENCE, V178, P471
DRUMMOND MF, 1996, BRIT MED J, V313, P275
BLANK RM, 1991, AM ECON REV, V81, P1041
GOLDBECKWOOD S, 1999, BRIT MED J, V318, P44
BEYER JM, 1978, SOCIOLOGICAL Q, V19, P68
BERWICK DM, 1989, NEW ENGL J MED, V320, P53
DONOVAN B, 1998, LEARN PUBL, V11, P179
*AM PSYCH ASS, 1994, DIAGN STAT MAN MENT
ALTMAN LK, 1996, LANCET, V347, P1459
RENNIE D, 1993, JAMA-J AM MED ASSOC, V270, P2856
PIERIE JPEN, 1996, LANCET, V348, P1480
ROTHWELL PM, 2000, BRAIN 9, V123, P1964
MERTON RK, 1968, SCIENCE, V159, P56
MOED HF, 1995, SCIENTOMETRICS, V33, P381
MARSH HW, 1981, J EDUC PSYCHOL, V73, P872
KUHN TS, 1970, STRUCTURE SCI REVOLU
SCOTT WA, 1974, AM PSYCHOL, V29, P698
SHARP DW, 1990, JAMA-J AM MED ASSOC, V263, P1390
SCHARSCHMIDT BF, 1994, J CLIN INVEST, V93, P1877
SCHUETZOW SO, 1992, J LAW HLTH, V7, P169
STRAYHORN J, 1993, AM J PSYCHIAT, V150, P947
TOPPING K, 1998, REV EDUC RES, V68, P249
WESSELY S, 1998, LANCET, V352, P301
GLASSICK CE, 1997, SCHOLARSHIP ASSESSED
BRADLEY JV, 1981, B PSYCHONOMIC SOC, V18, P31
GOLDMAN RL, 1994, EVAL HEALTH PROF, V17, P3
GINSPARG P, 1994, COMPUTATION PHYSICS, V8, P390
BURNHAM JC, 1990, JAMA-J AM MED ASSOC, V263, P1323
CHUBIN D, 1990, PEERLESS SCI PEER RE
FELDMAN KA, 1987, RES HIGH EDUC, V26, P227
FLEISS JL, 1981, STAT METHODS RATES P
COLE S, 1977, SCI AM, V237, P34
COLAIANNI LA, 1994, JAMA-J AM MED ASSOC, V272, P156
CREECH CD, 1988, NC L REV, V67, P179
ERNST E, 1994, J LAB CLIN MED, V124, P178
GARFUNKEL JM, 1994, JAMA-J AM MED ASSOC, V272, P137
CICCHETTI DV, 1980, AM PSYCHOL, V35, P300
CAMPANARIO JM, 1998, SCI COMMUN, V19, P181
BAKANIC V, 1987, AM SOCIOL REV, V52, P631
GARFIELD E, 1979, CITATION INDEXING
GARFUNKEL JM, 1990, JAMA-J AM MED ASSOC, V263, P1369
ABRAMS PA, 1991, SOC STUD SCI, V21, P111
GERARD K, 1992, HEALTH POLICY, V21, P249
ADAMS ME, 1992, MED CARE, V30, P231
CAMPANARIO JM, 1998, SCI COMMUN, V19, P277
CICCHETTI DV, 1991, BEHAV BRAIN SCI, V14, P167
BROWN C, 2001, J AM SOC INF SCI TEC, V52, P187
DICKERSIN K, 1990, JAMA-J AM MED ASSOC, V263, P1385
RENNIE D, 1997, JAMA-J AM MED ASSOC, V278, P579
RENNIE D, 1994, JAMA-J AM MED ASSOC, V272, P91
OPTHOF T, 1997, CARDIOVASC RES, V33, P1
KERR S, 1977, ACAD MANAGE J, V20, P132
HUTCHINGS P, 1996, MAKING TEACHING COMM
LINDSEY D, 1976, AM PSYCHOL, V31, P799
OXMAN AD, 1991, J CLIN EPIDEMIOL, V44, P91
RENNIE D, 1990, JAMA-J AM MED ASSOC, V263, P1317
RIND B, 1998, PSYCHOL BULL, V124, P22
HARGENS LL, 1990, SOC SCI RES, V19, P1
HARGENS LL, 1988, AM SOCIOL REV, V53, P139
MARTIN BR, 1983, RES POLICY, V12, P61
GRAMSCI A, 1971, SELECTIONS PRISON NO
KASSIRER JP, 1994, NEW ENGL J MED, V331, P669
OXMAN AD, 1995, CAN MED ASSOC J, V153, P1423
KLAZINGA N, 2000, INT J QUAL HEALTH C, V12, P183
LABAND DN, 1994, JAMA-J AM MED ASSOC, V272, P147
LATOUR B, 1986, LAB LIFE CONSTRUCTION
HULKA BS, 1979, MED CARE S, V17, P1
HARTER SP, 1998, J AM SOC INFORM SCI, V49, P507
JEFFERSON T, 1998, JAMA-J AM MED ASSOC, V280, P275
JEFFERSON T, 2002, JAMA-J AM MED ASSOC, V287, P2784
YANKAUER A, 1991, AM J PUBLIC HEALTH, V81, P843
SEWELL WH, 1992, AM J SOCIOL, V98, P1
SHAW CD, 2000, INT J QUAL HEALTH C, V12, P169
VANROOYEN S, 1999, J CLIN EPIDEMIOL, V52, P625
VANWEERT C, 2000, INT J QUAL HEALTH C, V12, P239
WESSELY S, 1996, PSYCHOL MED, V26, P883
SEGLEN PO, 1992, J AM SOC INFORM SCI, V43, P628
SEGLEN PO, 1997, ALLERGY, V52, P1050
SHULMAN L, 1993, CHANGE, V25, P6
SEGLEN PO, 1994, J AM SOC INFORM SCI, V45, P1
SMALL H, 1974, SCI STUD, V4, P17
UDVARHELYI S, 1992, ANN INTERN MED, V116, P238
STEWART WW, 1987, NATURE, V325, P207
STOSSEL TP, 1985, NEW ENGL J MED, V312, P658
SCHULMAN K, 1994, JAMA-J AM MED ASSOC, V272, P154
TAUBES G, 1993, SCIENCE, V259, P1246
WOOD DJ, 1998, J DOC, V54, P173
STAKE RE, 1995, ART CASE STUDY RES
ZIMAN J, 1994, PROMETHEUS BOUND SCI
TENOPIR C, 2000, ELECT J REALITIES SC
ZIMAN JM, 1968, PUBLIC KNOWLEDGE ESS
WILSON JD, 1978, J CLIN INVEST, V61, P1697
STRAUSS AL, 1990, BASICS QUALITATIVE R
WEINMANN C, 1998, EVAL HEALTH PROF, V21, P413
SCOTT C, 1991, MD L REV, V50, P316
WILLIS CL, 1990, KNOWLEDGE, V11, P363
WENSING M, 1994, INT J QUAL HEALTH C, V6, P115
WHITE M, 1997, J AM COLL CARDIOL A, V29, P181
SEGLEN PO, 1989, TRENDS BIOCHEM SCI, V14, P326
SCHMADER K, 1994, J AM GERIATR SOC, V42, P1241
WINER BJ, 1971, STAT PRINCIPLES EXPT
WESTEN D, 1997, AM J PSYCHIAT, V154, P895
SCRIVENS E, 1998, INT J QUAL HEALTH C, V10, P1
SCHERER RW, 1994, JAMA-J AM MED ASSOC, V272, P158
SCHAUFLER HH, 1999, HEALTH AFFAIR, V18, P134
WIDNALL SE, 1988, SCIENCE, V241, P1740
SCHNEIDER EC, 1998, JAMA-J AM MED ASSOC, V279, P1638
TIERNEY WG, 1996, PROMOTION TENURE COM
WALSH E, 2000, BRIT J PSYCHIAT, V176, P47
SMITH R, 1994, BRIT MED J, V309, P143
VANLEEUVEN TN, 1996, 9609 CWTS CTR SCI TE
SOKAL A, 1996, SOCIAL TEXT, V46, P217
VANDERWAART MAC, 1992, HUISARTS WET, V35, P437
VANDERVELUTEN CPM, 1996, MED TEACH, V18, P103
SONNERT G, 1995, SOC STD SCI, V25, P35
SONNERT G, 1996, AM SCI, V84, P63
VANRAAN AFJ, 2000, PRACTICING INTERDISC
STEEL K, 1981, NEW ENGL J MED, V304, P638
SMITH K, 1995, ACCOUNTING PERSPECTI, V1, P21
THOMSON R, 1995, BRIT MED J, V311, P237
STEPHAN PE, 1996, J ECON LIT, V34, P1199
STERLING TD, 1959, J AM STAT ASSOC, V54, P30
STERLING TD, 1995, AM STAT, V49, P108
TENOPIR C, 2000, ELECT J REALITIES SC, P91
TENOPIR C, 2000, ELECT J
TAUBES G, 1996, SCIENCE, V271, P767
TAUBES G, 1996, SCIENCE, V271, P764
TABACHNICK BG, 1989, USING MULTIVARIATE S
SWEITZER BJ, 1994, JAMA-J AM MED ASSOC, V272, P152
TUCKETT D, 1994, INT J PSYCHOANAL 5-6, V75, P865
WALLACE DP, 1986, J AM SOC INFORM SCI, V37, P136
WEINBERG RA, 1993, RESPONSIBLE SCI ENSU, V2, P66
SHASHOK K, 1992, EUROPEAN SCI EDITING, V45, P5
WEBER EJ, 2002, JAMA-J AM MED ASSOC, V287, P2790
COTTON P, 1993, JAMA-J AM MED ASSOC, V270, P2775
BAILEY CW, 2002, SCHOLARLY ELECT PUBL
BOND N, 1998, REPORT READER TRIAL
GOODSTEIN D, 1995, AM SCI, V83, P401
CHASSIN MR, 1998, JAMA-J AM MED ASSOC, V280, P1000
BAILAR JC, 1991, BEHAV BRAIN SCI, V14, P137
FEUERER ID, 1994, JAMA-J AM MED ASSOC, V272, P98
GARCIA J, 1981, AM PSYCHOL, V36, P149
2000, ULRICHS INT PERIODIC
CHASE JM, 1970, AM SOCIOL, V5, P262
2001, NATURE, V413, P93
GOODSTEIN LD, 1970, PSYCHOL REP, V27, P835
GALLAGHER EJ, 1998, ANN EMERG MED, V31, P83
BOHIGAS L, 1996, INT J QUAL HEALTH C, V8, P583
CHAN EJ, 1995, NEW YORK U LAW REV, V70, P100
COLE S, 1981, SCIENCE, V214, P885
FLEISS JL, 1981, STAT METHODS RATES P, P212
BOHIGAS L, 2000, INT J QUAL HEALTH C, V12, P231
GALLIN JI, 1997, JAMA-J AM MED ASSOC, V277, P651
*HIGH ED FUND COUN, 1992, IMP 1992 RES ASS EX
FEREE MM, 1996, AM SOCIOL REV, V61, P929
CENTRA JA, 1993, REFLECTIVE FACULTY E
BORNSTEIN RF, 1990, J SOC BEHAV PERS, V5, P71
DAVIS DA, 1997, CAN MED ASSOC J, V157, P408
ABBY M, 1994, JAMA-J AM MED ASSOC, V272, P105
CARSON JG, 1996, J 2 LANGUAGE WRITING, V5, P1
BOTTEMA BJA, 1992, HUISARTS WET, V35, P430
DAVIS DA, 1992, JAMA-J AM MED ASSOC, V268, P1111
EDGERTON R, 1991, TEACHING PORTFOLIO C
GARVEY WD, 1979, COMMUNICATION ESSENC, P202
*AM ASS ADV SCI, 1993, BENCHM SCI LIT
CARMINES EG, 1979, RELIABILITY VALIDITY
CARMICHAEL HL, 1988, J POLITICAL EC, V96, P453
ANGELO T, 1993, CLASSROOM ASSESSMENT
ACS ZJ, 1992, AM ECON REV, V82, P363
COOPER H, 1994, HDB RES SYNTHESIS
ABRAMI PC, 1982, REV EDUC RES, V52, P446
GARFIELD E, 1986, ANN INTERN MED, V105, P313
GARFIELD E, 1990, CURR CONTENTS, V22, P5
ELLERBECK EF, 1995, JAMA-J AM MED ASSOC, V273, P1509
CENTRA JA, 1979, DETERMINING FACULTY
BLUMENTHAL D, 1993, JAMA-J AM MED ASSOC, V269, P2775
BEREZIN AA, 1998, KNOWLEDGE TECHNOLOGY, V11, P5
BEGG CB, 1988, J ROY STAT SOC A STA, V151, P419
FINZEN A, 1996, PSYCHIAT PRAX, V23, P1
ANDERSON J, 1996, BRIT MED J, V312, P762
DAVIDSON F, 1997, JAMA-J AM MED ASSOC, V277, P1808
COLLOPY BT, 2000, INT J QUAL HEALTH C, V12, P211
EICHORN DH, 1985, AM PSYCHOL, V40, P1309
DAVIDSON RA, 1986, J GEN INTERN MED, V1, P155
DAVIES R, 1998, COMPUT EDUC, V30, P111
FINE MA, 1996, AM PSYCHOL, V51, P1190
CECI SJ, 1984, AM PSYCHOL, V39, P1491
BINGHAM C, 1996, MED J AUSTRALIA, V164, P8
BLUMENTHAL D, 1996, NEW ENGL J MED, V335, P891
FOLTZ FA, 2000, B SCI TECHNOL SOC, V20, P427
FOGG L, 1993, AM PSYCHOL, V48, P293
BIGGS M, 1990, LIBR TRENDS, V39, P145
BIGLAN A, 1973, J APPL PSYCHOL, V57, P195
CRONIN B, 1995, J AM SOC INFORM SCI, V46, P700
COZZENS SE, 1995, HDB SCI TECHNOLOGY S, P533
EVANS JT, 1990, JAMA-J AM MED ASSOC, V263, P1353
BAZELEY P, 1996, 50 NAT BOARD EMPL ED
FRAZER M, 1989, WAYNE L REV, V35, P1181
BARRETT E, 1999, J COMPUT ASSIST LEAR, V15, P48
FREEMAN M, 1995, ASSESSMENT EVALUATIO, V20, P289
EPSTEIN WM, 1990, SCI TECHNOL, V15, P9
CUCIC S, 2000, INT J QUAL HEALTH C, V12, P217
CRONIN B, 1996, J DOC, V52, P163
CRANDALL R, 1982, BEHAV BRAIN SCI, V5, P207
COHEN J, 1988, STAT POWER ANAL BEHA
COHEN J, 1960, EDUC PSYCHOL MEAS, V20, P37
EXTEJT MM, 1990, J MANAGE, V16, P539
*NAT RES COUNC, 1996, NAT SCI ED STAND
EXNER JE, 1993, RORSCHACH COMPREHENS, V1
*NAT RES COUNC COM, 1986, ENV TOB SMOK MEAS EX
BIERIG JR, 1988, ST LOUIS U LJ, V32, P977
FOUCAULT M, 1975, SURVEILLER PUNIR NAI
FREIDSON E, 1970, PROFESSION MED STUDY
FOUCAULT M, 1980, POWER KNOWLEDGE SELE
BETANCOURT H, 1993, AM PSYCHOL, V48, P629
BEYER JM, 1995, ACAD MANAGE J, V38, P1219
CRANDALL R, 1978, AM PSYCHOL, V33, P623
FABOS B, 1999, REV EDUC RES, V69, P217
BATES DW, 1995, JAMA-J AM MED ASSOC, V274, P29
HARNAD S, 1996, SCHOLARLY PUBLISHING
REISMAN A, 1994, OPER RES, V42, P577
REISMAN A, 1995, OPER RES, V43, P731
HURD JM, 1996, PRINT ELECT TRANSFOR, P9
MCNAY I, 1997, IMPACT 1992 RAE I IN
MOBILY PR, 1991, J NURS EDUC, V30, P73
HUBBARD R, 1994, INT J RES MARK, V11, P233
RELMAN AS, 1981, NEW ENGL J MED, V305, P824
HARNAD S, 1992, SERIALS REV, V18, P58
HARNAD S, 1991, PUBLIC ACCESS COMPUT, V2, P39
LEAPE LL, 1995, JAMA-J AM MED ASSOC, V274, P35
HARNAD S, 1986, NATURE, V322, P24
HARNAD S, 1985, SCI TECHNOL, V10, P55
LEAPE LL, 1991, NEW ENGL J MED, V324, P377
HARTZ AJ, 1989, NEW ENGL J MED, V321, P1720
MOHER D, 1999, LANCET, V354, P1896
PARKER M, 1995, ORGANIZATION, V2, P319
KLAHR D, 1985, AM PSYCHOL, V40, P148
MCCAIN KW, 1995, SCI COMMUN, V16, P403
LAWRENCE M, 1993, MED AUDIT PRIMARY HL
LATOUR B, 1987, SCI ACTION FOLLOW SC
REIMHERR FW, 1998, AM J PSYCHIAT, V155, P1247
KING J, 1987, J INFORM SCI, V13, P261
HUBBARD R, 1992, MARKET LETT, V3, P127
KING DW, 1981, SCI J US THEIR PRODU
KING CR, 1997, IMAGE J NURSING SCHOL, V29, P163
KLAZINGA N, 1996, QUALITY MANAGEMENT M
HARNAD S, 2001, NATURE, V410, P1024
LAWTHERS AG, 2000, MED CARE, V38, P785
LATOUR B, 1979, LAB LIFE CONSTRUCTIO
HARNAD S, 1982, PEER COMMENTARY PEER
PAE JB, 1997, FORDHAM LAW REV, V66, P975
MCDONALD CJ, 1984, ANN INTERN MED, V100, P130
RELMAN AS, 1988, NEW ENGL J MED, V318, P918
HOWARD L, 1997, BRIT J PSYCHIAT, V170, P109
MULLINS NC, 1973, THEORIES THEORY GROU
MCCUTCHEN CW, 1991, TECHNOL REV, V94, P28
HOWARD GS, 1987, AM PSYCHOL, V42, P975
MARSHALL MN, 2000, JAMA-J AM MED ASSOC, V283, P1866
MARTIN BR, 1996, SCIENTOMETRICS, V36, P343
KERR EA, 1995, ANN INTERN MED, V123, P500
RENNIE D, 2000, JAMA-J AM MED ASSOC, V284, P89
RICE RE, 1991, METROPOLITAN U, V1, P7
MARSH HW, 1999, MULTIVAR BEHAV RES, V34, P1
MCDOWELL L, 1995, INNOV EDUC TRAIN INT, V32, P302
IEZZONI LI, 1992, QUALITY REV B, V18, P361
IEZZONI LI, 1994, INQUIRY-J HEALTH CAR, V31, P40
HAYWARD RA, 1993, MED CARE, V31, P394
HAMMEL E, 1980, REPORT TASK FORCE FA
RABINOVICH BA, 1996, AM PSYCHOL, V51, P1190
PATTON MQ, 1990, QUALITATIVE EVALUATI
RICHARDSON FM, 1972, MED CARE, V10, P29
HUTCHING P, 1994, AAHE B, V47, P3
RELMAN AS, 1989, NEW ENGL J MED, V321, P827
KIESLER CA, 1991, BEHAV BRAIN SCI, V14, P151
PAGE G, 1997, J PUBLISHING
HARLEY S, 2000, CRIT PERSPECT, V11, P549
OVER R, 1982, BEHAV BRAIN SCI, V5, P229
RENIK O, 1994, INT J PSYCHOANAL 5-6, V75, P1245
HATTIE J, 1996, REV EDUC RES, V66, P507
MITROFF II, 1979, SOC STUD SCI, V9, P199
LAPORTE RE, 1995, BRIT MED J, V310, P1387
OWENS A, 1984, MED EC 0820, P167
RENNIE D, 1998, JAMA-J AM MED ASSOC, V280, P214
MORAVCSIK MJ, 1975, SOC STUD SCI, V5, P86
OVER R, 1995, AUSTR U REV, V38, P32
MOOSSY J, 1985, J NEUROPATH EXP NEUR, V44, P225
NIEVA VF, 1980, ACAD MANAGE REV, V5, P267
HOFER TP, 1999, JAMA-J AM MED ASSOC, V281, P2098
GROL R, 1995, QUALITY IMPROVEMENT
JEFFERSON T, 1995, HEALTH ECON, V4, P383
JEFFERSON T, 1994, HEALTH ECON, V3, P25
NORD E, 1993, INT J TECHNOL ASSESS, V9, P37
PORTER A, 1985, SCI TECHNOL HUM VALU, V10, P33
GRANT J, 1997, NATURE, V390, P438
LLOYD ME, 1990, J APPL BEHAV ANAL, V23, P539
ROTH WM, 2001, TEACH TEACH EDUC, V17, P741
KOSTOFF RN, 1997, SCIENCE, V277, P651
LINK AM, 1998, JAMA-J AM MED ASSOC, V280, P246
KOHUT H, 1971, ANAL SELF
KATES N, 1999, CAN FAM PHYSICIAN, V45, P2143
OBRIEN BJ, 1997, JAMA-J AM MED ASSOC, V277, P1802
JAUCH LR, 1989, ACAD MANAGE J, V32, P157
OBRIEN RM, 1991, SOC SCI RES, V20, P319
MACROBERTS MH, 1989, J AM SOC INFORM SCI, V40, P342
JUHASZ S, 1975, IEEE T PROFESSION PC, V18, P177
PETERS DP, 1982, BEHAV BRAIN SCI, V5, P246
NOYONS ECM, 1999, IN PRESS RES POLICY
GROUSE LD, 1981, JAMA-J AM MED ASSOC, V245, P375
LIJIMER JG, 1999, JAMA-J AM MED ASSOC, V282, P1061
MACKENZIE KR, 1999, CAN J PSYCHIAT S1, V44, S4
NEDERHOF AJ, 1987, SCIENTOMETRICS, V11, P333
GRECO PJ, 1993, NEW ENGL J MED, V329, P1271
KAHN KL, 1989, STRUCTURED IMPLICIT
KAHN KL, 1989, N3016HCFA RAND
LINDNER UK, 1997, ANAESTHESIST, V46, P1
KOLLOCK P, 1999, COMMUNITIES CYBERSPA
PHILLIMORE AJ, 1989, RES POLICY, V18, P255
KOLB D, 1984, EXPT LEARNING EXPERI
ROTH WM, IN PRESS MIND CULTUR
METOYERDURAN C, 1993, COLL RES LIBR, V54, P517
LJUNGBERG J, 2000, EUROPEAN J INFORMATI, V9
LYOTARD JF, 1984, POSTMODERN CONDITION
NELSON PG, 1982, TRENDS NEUROSCI, V5, P229
GRIMSHAW JM, 1993, LANCET, V342, P1317
KRITCHEVSKY SB, 1991, JAMA-J AM MED ASSOC, V266, P1817
JADAD AR, 1998, JAMA-J AM MED ASSOC, V279, P611
POLAK JF, 1995, AM J ROENTGENOL, V165, P685
NUNNALLY JC, 1978, THEORY MEASUREMENT E, P190
LINCOLN YS, 1985, NATURALISTIC INQUIRY
GROENEVELD L, 1975, SOC STUD SCI, V5, P343
KOSTOFF RN, 1997, HDB RES IMPACT ASSES
GROL R, 1988, PEER REV GEN PRACTIC
GROL R, 1992, QUAL HEALTH CARE, V1, P184
ROSENTHAL R, 1991, BEHAV BRAIN SCI, V14, P160
KOREN G, 1986, NEW ENGL J MED, V315, P1298
MEYER GJ, 2000, J PERS ASSESS, V75, P46
NEWTON G, 2001, ALA L REV, V52, P723
LYKETSOS CG, 1997, J NEUROPSYCH CLIN N, V9, P556
GUSTON DH, 1996, SCI PUBL POLICY, V23, P229
PETTY RE, 1986, ADV EXPT SOCIAL PSYC, V19, P123
LUUKKONEN T, 1991, RES EVALUAT, V1, P21
MERRIAM SB, 1988, CASE STUDY RES ED QU
PRESSMAN S, 1994, AM J ECON SOCIOI, V53, P315
ROSSITER JR, 2001, MARK THEORY, V1, P9
PINCUS HA, 1995, AM J PSYCHIAT, V152, P596
HENDRICK C, 1976, PERSONALITY SOCIAL P, V2, P207
ROTH WM, 1999, J CURRICULUM STUD, V31, P501
NEUMANN PJ, 1997, MED DECIS MAKING, V17, P402
KEALEY T, 1996, EC LAWS SCI RES
NARIN F, 1997, RES POLICY, V26, P317
HERSHEY N, 1992, QUAL ASSUR UTIL REV, V7, P23
ROTEMBERG JJ, 1995, RAND J ECON, V26, P630
LEWISON G, 2002, AUSTRIAN BIOMEDICAL
ROTH WM, 1998, J RES SCI TEACH, V35, P399
MULSANT BH, 1999, J CLIN PSYCHIAT S20, V60, P9
PFIZER F, 1993, ACAD MANAGE REV, V18, P599
PIKE I, 1984, ADULT EDUC QUART, V34, P71
ROGOFF B, 1991, PERSPECTIVES SOCIAL, P349
JUDD LL, 1998, ARCH GEN PSYCHIAT, V55, P694
NARIN F, 1976, EVALUATIVE BIBLIOMET
GUEDON JC, 2001, OLDENBURGS LONG SHAD
MEADOWS AJ, 1974, COMMUNICATION SCI
JADAD AR, 1998, JAMA-J AM MED ASSOC, V280, P278
LOCK S, 1993, BRIT MED J, V307, P382
LOCK S, 1994, ANN INTERN MED, V121, P60
KEELER EB, 1992, JAMA-J AM MED ASSOC, V268, P1709
LONG JS, 1992, SOC FORCES, V71, P159
MAHONEY MJ, 1987, J SOC BEHAV PERS, V2, P165
KURTZ H, 1991, WASHINGTON POST 0616, A10
KAWANO T, 1993, J SCHOOL PSYCHOL, V31, P407
RUBENSTEIN LV, 1991, N3033HCFA RAND
JONES MJ, 1999, OMEGA-INT J MANAGE S, V27, P397
LOCK SP, 1991, BEHAV BRAIN SCI, V14, P156
RUBIN HR, 1990, N3066HCFA RAND
MILGROM P, 1980, J AM DENT ASSOC, V101, P258
MAHONEY MJ, 1979, SOC STUD SCI, V9, P349
LEWISON G, 1999, SCIENTOMETRICS, V46, P529
RUBIN HR, 1993, J GEN INTERN MED, V8, P255
ROSSI PH, 1993, EVALUATION SYSTEMATI